Using a Network of Single Site-Specific Cysteine Mutations Coupled with Crosslinking Mass Spectrometry (CX-MS) to Refine the Structure and Dynamics of the Human Alpha 1 Glycine Receptor (GlyR)

Kayce Tomcho

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation

This One-year Embargo is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection.
USING A NETWORK OF SINGLE SITE-SPECIFIC CYSTEINE MUTATIONS
COUPLED WITH CROSSLINKING MASS SPECTROMETRY (CX-MS) TO REFINE
THE STRUCTURE AND DYNAMICS OF THE HUMAN ALPHA 1 GLYCINE
RECEPTOR (GLYR)

A Dissertation
Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for
the degree of Doctor of Philosophy

By
Kayce Alexandra Tomcho
May 2021
USING A NETWORK OF SINGLE SITE-SPECIFIC CYSTEINE MUTATIONS COUPLED WITH CROSSSLINKING MASS SPECTROMETRY (CX-MS) TO REFINE THE STRUCTURE AND DYNAMICS OF THE HUMAN ALPHA 1 GLYCINE RECEPTOR (GLYR)

By

Kayce Alexandra Tomcho

Approved December 10, 2020
ABSTRACT

USING A NETWORK OF SINGLE SITE-SPECIFIC CYSTEINE MUTATIONS COUPLED WITH CROSSLINKING MASS SPECTROMETRY (CX-MS) TO REFINE THE STRUCTURE AND DYNAMICS OF THE HUMAN ALPHA 1 GLYCINE RECEPTOR (GLYR)

By
Kayce Alexandra Tomcho

May 2021

Dissertation supervised by Dr. Michael Cascio

A network of site-specific single Cys-mutations coupled with CX-MS can be used to elucidate a more refined structure of GlyR and obtain a more definitive understanding of pentameric ligand-gated ion channel (pLGIC) allostery. Each Cys-mutant is introduced into an α1 homomeric Cys null background (C41S/C290A/C345S), or in the same background with F207G/A288G mutation that allows non-desensitizing GlyR activation by ivermectin (IVM). State-dependent crosslinking with methanethiosulfonate benzophenone (MTS-bzp) to a single thiol of purified, vesicle reconstituted GlyR are conducted after enriching the receptor in different allosteric states: resting (no ligand), open (F207G/A288G + IVM), or desensitized (excess glycine). Digested peptides are analyzed via liquid chromatography mass spectrometry to identify sites of intra- and intermolecular crosslinking. Tandem MS of mass-shifted precursor ions further refine
these distance constraints. Independent comparative studies targeting different single Cys GlyR (M287C, K116C, K206C) provides evidence of allosteric changes between the three states, as well as direct topological information of regions that are unresolved in other high-resolution structures of pentameric ligand gated ion channels (pLGICs), most notably the M3-M4 loop. The unresolved loop is shown to interact with the rest of the protein primarily in non-conducting states, possibly highlighting its role in structure in those states. Regions of receptor mobility have also been identified and are both novel and congruent with what is reported in literature. Reported mobile regions include the M2-M3 loop, the C-loop, M3-M4 loop and very near the N-terminus. These findings can be applied to further refinement of GlyR structure, and provide information on GlyR molecular mechanism. Additionally, methods reported in this work can also be applied to the understanding of other members of the pLGIC family, and other transmembrane proteins that are waiting further structural information.
DEDICATION

To my sweet Kit Kat. I love you more than you know, the soul to my soul.
ACKNOWLEDGEMENT

To My PI:

I would like to thank my PI, Dr. Mike Cascio, first and foremost, for your unwavering support of me throughout the years. I know you will disagree, but I know I can be a lot to handle at times. Thank you for dealing with my “out-loud worrying” in your office, all the times you “talked me off the ledge” and all of the tears. So. Many. Tears. And yet, you were always willing to continually reassure me that everything would be okay. You’ve taught me more than I can say about structural biochemistry, but also music, books, movies, politics, and so much more – you know, the more important stuff. Thank you for allowing me to decorate the lab for Christmas – in October. Thank you for always encouraging me and standing in my corner. Thank you for respecting and trusting my opinions, and always thinking of me as a professional. I’ve learned, through you, that I am more capable than I realized. But most importantly, thank you for believing in me when I didn’t believe in myself. Which, as you very well know, was most of the time.

To My Dissertation Committee:

I would also like to thank my dissertation committee: Dr. Rita Mihailescu, Dr. Mike Van Stipdonk and Dr. Ben Kolber. I might be biased, but I am truly lucky to have such a supporting group of individuals as my committee members. You all have contributed so much to my professional development, and I am very grateful.

Rita, I’ve always been in awe of you. You’re an incredible person and a wonderful role model. You’ve inspired me as a person, as a woman, as a scientist and as a teacher. I hope that I can be half as good of a teacher that you are someday. Thank you for teaching
me that’s it’s okay to say ‘no’. Thank you for keeping an eye on how much I over-extend myself, and for telling me when it was too much. And thank you for being there for me whenever I needed you, and for always supporting me throughout my time here.

MVS, thank you for making me adopted member of your lab family, even though I’m not at all qualified to work there. Thank you for seeing my ability early on, my “poise and professionalism” as you put it. Thank you for letting me confide in you, and for helping me plan that surprise birthday party. And, thank for at least trying to teach me how to be diplomatic. We both know it didn’t really stick.

Ben, thank you for bringing your intensity to every one of our meetings. You always inspire me to do better and try harder. I have learned so much from you, from the very first days during orientation, from taking your class, seeing you present your tenure talk, during our update meetings, and the times I sought you out for advice. I especially thank you for providing an example of good presentation skills, and how best to communicate with an audience.

To Members of the Cascio Lab:

Many of you have become like family to me throughout the years, and I am grateful for the time we had together. Thank you firstly, to Nick Ferraro for being an incredible mentor, and for putting up with me for all these years. I am most thankful for you sticking up for me through the difficult times, and for treating me with respect. You are honestly like a brother to me and I’m so thankful it was you that taught me what I needed to know to survive – and thrive – in graduate school. Though we both know without you, there wouldn’t have been anyone else.
Thank you to Emily Cooper. You and I have so much in common, and honestly, the most wonderful thing about my time at Duquesne is meeting you. I am so grateful to you for so many things, but about the lab specifically, that we were able to make it run as a smooth, organized ship, regardless of what was going on. Thank you for having my back and for always supporting me and believing in me. Thank you for your friendship. I could go on and on about what you mean to me, but I’ll leave it at this: we both know what it’s like to have so many brothers, and finally, we both have a sister. I love you so much.

Thank you to Hannah Gering. You have contributed so much to this work, and I’m grateful to your intelligence and dedication to this scientific endeavor. But more, importantly, you more than any student helped me grow as a mentor and teacher. I’m still a hover-mother, and always will be, but it’s because of your independence that I’ve learned when to let you soar. Furthermore, I’m impressed by your bravery – when something wasn’t working for you, when you weren’t getting what you needed, you spoke up. And we made it right.

Thank you to Amanda Pellegrino. What would I have done without you? You have been more than instrumental to this project, and I am so, so glad you decided to spend some of your valuable time in our lab, working hard as you always do for the benefit of my project. I have known you since you were a freshman, and it has been my pleasure to watch you mature and accomplish many impressive things. You are a wonderful person, and I can’t wait to see what else you conquer as you continue in science.

Thank you to Adam Gargano. We made a great team, and I’m so glad that you were responsible, mature, dedicated and passionate enough to step up and go beyond your duty as an undergraduate student to become my co-captain. I am so thankful for the time you
put in. Furthermore, it was a pleasure working with you in general, as you are intelligent, warm and generous. You’ve always impressed and inspired me, and I’m very proud of you.

Thank you to Andrew DeMarco. You’re such a good person, and if not for you I very well might have gone insane much sooner in my career. We always had good talks; where we would do our best to save the world. Even though you come off as a sourpuss, you really are a sweetheart. I’m most grateful for your honesty, your kindness and your friendship. Thank you for your support, and your belief in me, always.

Thank you especially to the other members of the GlyR team: Lauren Cocozza, Britton Study, Regan Harmon, Adriana Duncan; you guys made coming to the lab everyday a much better place. I’m proud of all of you, and can’t wait to see what else you do in your futures.

Thank you to Warren Lowther. I’m glad you decided to join the lab; we’ve become close in the short time we had together, and I know you’ll continue to do good work as you continue in your career. Thank you for being smart, capable and competent, as it was much easier to train you because of it. Thank you for your friendship and your kindness. We will always share a dual connection since both of our alma maters are identical, so you can’t get rid of me even if you tried!

Thank you to Sara Lutty and Rathna Veeramachaneni for welcoming me into the lab and teaching me as much as you could in the short time before you both graduated. I am indebted to your kindness and your patience.

And lastly, thank you to the many, many other undergraduate students I had the pleasure of mentoring throughout the years. Whether you were on my project, or not, I gained valuable experience from every single one of you.
To My Support System Outside of the Lab:

It takes a village for this kind of work, and oftentimes, the support and advice that I received outside of my dissertation committee was able to help me grow as a professional and was then beneficial to my productivity in the research lab and my knowledge as a scientist. I knew early on in my career that teaching was something that was important to me and that I was passionate about – because of you all, I’ve also become good at it.

To those that helped me grow as a teacher: Dr. Skip Kingston, Dr. Tim Evans, Dr. Mary Sherman and Dr. Alicia Paterno. You all recognized my abilities as a teacher and helped me reach my potential. In different ways, you have all supported me in this endeavor, whether it was acclimating to life after graduate school, preparing for answering difficult questions about education, recognizing my strengths, giving me opportunities that I wouldn’t have otherwise had, or lending a supporting ear when I was going through a difficult time. I appreciate all of you, and thank you.

Skip, you’ve gone the extra mile to support me in developing my career, and helping me determine what I’m meant to do. I’m especially grateful for your belief in me, your unyielding support and all of your kindness. And Tim, you are such a dedicated educator, and I’m so lucky you were willing to share your experiences with me. You more than most understand my journey here, as we have a lot of common ground, and I’m so grateful you were always willing to support me. Thank you for caring so much about the work you do, and for always going the extra mile for all of us graduate students, and undergraduate students. I respect you and your guidance so much.
'Thank you to Aimee Ciesielski, Lisa Zander, Dr. Nithya Vaidyanathan, Scott Boesch and Heather Costello. Thank you for always being willing to answer all of my stupid questions. And thank you for your support.

Thank you to my friends who have also supported me throughout my journey, especially to Cassandra Hanley, Lauren Stubbert, Tell Lovelace, Joshua Imperatore, Sarah Hejnosz, Jeremiah Jamrom, Jen Armen, Luke Metzler, Amy Carlson, Jason Heming and others. You all know what your support has meant to me. And you were always there for me through the good, the bad and the ugly. I’m grateful to you all. Thank you.

To My Family:

Lastly, thank you to my family: Dad, Mom, Andy, Mike and Jamie. Thank you for instilling in me a sense of integrity, a good work ethic and the notion to “leave every place better than when you found it”.

xii
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xv</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xvii</td>
</tr>
<tr>
<td>Chapter 1: Crosslinking Mass Spectrometry</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2: Pentameric Ligand Gated Ion Channels</td>
<td>33</td>
</tr>
<tr>
<td>Chapter 3: Materials and Methods</td>
<td>68</td>
</tr>
<tr>
<td>Chapter 4: Protein-Protein Interactions in the Resting (apo-) State Network</td>
<td>80</td>
</tr>
<tr>
<td>Chapter 5: Protein-Protein Interactions in the Open and Desensitized State Networks</td>
<td>103</td>
</tr>
<tr>
<td>Chapter 6: Scientific Impact and Future Directions</td>
<td>146</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Forward Oligo Sequences ................................................................. 69
Table 2. M287C Resting State Crosslinks ..................................................... 89
Table 3. K116C Resting State Crosslinks ...................................................... 94
Table 4. K206C Resting State Crosslinks ...................................................... 97
Table 5. M287C Open State Crosslinks ....................................................... 110
Table 6. M287C Desensitized State Crosslinks ........................................... 111
Table 7. K116C Open State Crosslinks ....................................................... 117
Table 8. K116C Desensitized State Crosslinks ........................................... 118
Table 9. K206C Open State Crosslinks ....................................................... 123
Table 10. K206C Desensitized State Crosslinks .......................................... 124
Table 11. Distance Measurements ............................................................... 129
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>General Workflow of CX-MS</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>The Synaptic Cleft</td>
<td>33</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Architectural Structure of a pLGIC</td>
<td>38</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>pLGIC Conformational States</td>
<td>39</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>Extracellular Domain of pLGICs</td>
<td>42</td>
</tr>
<tr>
<td>Figure 6.</td>
<td>ECD-TMD Interface</td>
<td>47</td>
</tr>
<tr>
<td>Figure 7.</td>
<td>Methanethiolsulfonate (MTS)-Benzophenone Crosslinker</td>
<td>83</td>
</tr>
<tr>
<td>Figure 8.</td>
<td>Network of Mutants</td>
<td>84</td>
</tr>
<tr>
<td>Figure 9.</td>
<td>MS and MSMS Representative Spectra</td>
<td>87</td>
</tr>
<tr>
<td>Figure 10.</td>
<td>M287C Resting State Crosslinks</td>
<td>91</td>
</tr>
<tr>
<td>Figure 11.</td>
<td>K116C Resting State Crosslinks</td>
<td>95</td>
</tr>
<tr>
<td>Figure 12.</td>
<td>K206C Resting State Crosslinks</td>
<td>98</td>
</tr>
<tr>
<td>Figure 13.</td>
<td>M287C Resting, Open and Desensitized State Crosslinks</td>
<td>113</td>
</tr>
<tr>
<td>Figure 14.</td>
<td>K116C Resting, Open and Desensitized State Crosslinks</td>
<td>119</td>
</tr>
<tr>
<td>Figure 15.</td>
<td>C-loop Crosslinking Comparison in K116C</td>
<td>120</td>
</tr>
<tr>
<td>Figure 16.</td>
<td>K206C Resting, Open and Desensitized State Crosslinks</td>
<td>125</td>
</tr>
<tr>
<td>Figure 17.</td>
<td>Network Comparison of Open State Crosslinking</td>
<td>131</td>
</tr>
<tr>
<td>Figure 18.</td>
<td>Network Comparison of Desensitized State Crosslinking</td>
<td>136</td>
</tr>
<tr>
<td>Figure 19.</td>
<td>Intracellular M3-M4 Loop of α1-GlyRs</td>
<td>138</td>
</tr>
<tr>
<td>Figure 20.</td>
<td>Total Intracellular M3-M4 Loop Crosslinking</td>
<td>141</td>
</tr>
<tr>
<td>Figure 21.</td>
<td>Proposed Mobile Regions of α1-GlyRs</td>
<td>144</td>
</tr>
</tbody>
</table>
Figure 22. MTS-Diazirine Crosslinker .................................................................149
LIST OF ABBREVIATIONS

5-HT$_3$R – Serotonin Receptor
Ambic – ammonium bicarbonate
ChEP – Chromatin
ChIP – Chromatin Immunoprecipitation
CID – Collision Induced Dissociation
CNS – Central Nervous System
COSEY – Correlation Spectroscopy
CPP – cell-penetrating peptide
Cryo-EM – Cryo-Electron Microscopy
CX-MS – Crosslinking Mass Spectrometry
DHSO - dihydrazide sulfoxide
DMDSSO - dimethyl disuccinimidyl sulfoxide
DNA – deoxyribonucleic acid
DSBU - disuccinimidal dibutyric urea
DSSO – disuccinimidyl sulfoxide
DTT – dithiolthreitol
ECD – Extracellular Domain
E. coli – Escherichia coli
EDC – 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
EDTA – ethylenediamine tetraacetic acid
EGTA – ethylene glycol-bis(β-aminoethylether)-N,N,N’,N’-tetraacetic acid
ELIC – *Erwinia chrysanthemi*

ESI – Electrospray Ionization

GABA<sub>A</sub>R – Gamma-aminobutyric Acid Type A Receptor

GLIC – *Gloeobacter violaceous*

GlyR – Glycine Receptor

GPCR – G Protein Coupled Receptor

HDX – Hydrogen Deuterium Exchange

HSQC – Heteronuclear Single Quantum Coherence

ICD – Intracellular Domain

IMS – Ion Mobility Mass Spectrometry

LBD – Ligand Binding Domain

LC – Liquid Chromatography

MALDI – Matrix-assisted laser desorption ionization

MS – Mass Spectrometry

MS/MS – Tandem Mass Spectrometry

MTS-bzp – Methanethiolsulfonate benzophenone

nAChR – Nicotinic Acetylcholine Receptor

nESI – nanoESI

NHS – N-hydroxysuccinimide esters

NLS – nuclear localization signal

nMS – native MS

NMR – Nuclear Magnetic Resonance

NOESY – Nuclear Overhauser Effect Spectroscopy
PA – phosphatidic acid

*P. aeruginosa* – *Pseudomonas aeruginosa*

PC – phosphatidylcholine

PCR – Polymerase Chain Reaction

PDB – Protein Data Bank

PE – phosphatidyl ethanolamine

PG – phosphatidylglycerol

PI – phosphatidylinositol

PIP₂ – phosphatidylinositol 4,5-bisphosphate

PKA/C – Protein Kinase A/C

pLGIC – pentameric Ligand Gated Ion Channels

PTM – Post-translational modification

QTOF – Quadrupole Time of Flight

RNA – ribonucleic acid

SILAC – stable isotope labeling by amino acids in cell culture

SID – surface induced dissociation

SDS-PAGE – sodium dodecyl sulfate – polyacrylamide gel electrophoresis

*Sf*9 – *Spodoptera frugiperda*

TMD – Transmembrane Domain

TOCSY – Total Correlated Spectroscopy
CHAPTER 1. CROSSLINKING MASS SPECTROMETRY (CX-MS)

1.1 Introduction

The 1980s saw a surge of developments in mass spectrometry that paved the way for mass spectrometry proteomics. The Nobel prize in chemistry was awarded to John Fenn, Koichi Tanaka and Burt Wuthrich for the development and utilization of a “soft” ionization source, electrospray ionization (ESI) which allowed for the study of “molecular elephants”, or proteins.\textsuperscript{1} ESI and also MALDI (matrix-assisted laser desorption/ionization) allow protein samples to become ionized while preserving both covalent and non-covalent interactions that are crucial to the folding and assembly of proteins. These ionization sources are usually coupled with time-of-flight (TOF) or quadrupole-time-of-flight (Q-TOF) mass analyzers to separate the ions to provide accurate mass measurements.\textsuperscript{2}

A protein sample can be introduced to the mass spectrometer and is generally analyzed in two ways, as a whole protein or as peptide fragments. This refers to either the top-down or bottom-up approach, respectively. Both of these methods can provide information on the sequence, structure, higher order assembly, binding properties and kinetics of proteins and thus, are very useful.\textsuperscript{3}

In bottom-up studies, the protein is enzymatically digested fully before being separated by liquid chromatography, ionized by either ESI or MALDI and then shuttled to the mass analyzer. Tandem MS (or MS/MS) allows selected peptide ions (in this case) to be selected by m/z and be further fragmented before undergoing a second mass spectrometric analysis. These precursor ions, the peptides chosen, are most commonly fragmented using collision induced dissociation (CID) which typically breaks peptides at their peptide bond, or at the alpha carbon, producing what is referred to as b, y and a ions.
(see Figure 1 for representations of these fragments) depending on which side of the peptide (N or C-terminus) the charge resides. These ions are referred to as product ions, and are used to match to predicted fragmentation patterns in databases. This can be used for a variety of approaches such as de novo sequencing, protein identification, quantitation of proteins, and most relevant to this work – protein structure determination.3,4

Over the last two decades, the ability to study protein structure using MS has further evolved into several different techniques. Of particular interest here is crosslinking mass spectrometry (CX-MS). CX-MS has emerged as a useful tool in determining protein structure and dynamics,5–10 uncovering protein constituents in large protein assemblies,6,11–13 identifying protein-lipid interactions,14–16 protein-DNA interactions useful in epigenetics research,17 studying the interactome18,19 and analyzing proteins in situ.20

CX-MS still makes use of the soft ionization techniques spoken of previously, and uses MS/MS with CID to produce fragmentation spectra in order to do this. A few of the other mass spectrometric approaches to studying protein structure will be addressed in Section 1.2.4. Continued advancements to a growing database of crosslinked peptides,7,21,22 and a better matching system that also identifies the confidence of a match,7 as well as methods to better quantitate crosslinking6,7 will aid in the further popularity and usefulness of CX-MS.

The overall principle of this technique relies on generating a covalent bond between residues that are close in space. After crosslinking occurs, the sample is proteolytically digested to allow for better ionization of these smaller peptide fragments and then analyzed via mass spectrometry and various software databases. A series of generated crosslinks give rise to distance constraints and create a topological map of the
complex/protein/interactome in question. The topological map can be used to aid in the development or refinement of current structural models and can provide necessary details into the molecular mechanisms of the protein in various conformations and its interactions with its surroundings.

The crosslinker used to generate these distance constraints consists of at least two parts; reactive groups that are either homo- or hetero-functional and a spacer connecting the two that determines the distance between the proximal residues. The spacer can be thought of as a ‘molecular ruler’ as its length can be variable, from zero-length to longer. Using multiple sites of initial attachment of the crosslinker within the protein of interest can provide a network of interactions that aid in model building and structural refinement.

Typical chemistries used based on limited reactive groups of 20 natural amino acids are amine, carboxylate, thiol and photoreactive. Some crosslinkers also include additional components such as moieties to allow for enrichment of crosslinked species.

CX-MS is a high-throughput and versatile technique. Though it has low-resolution, it benefits from having complementarity to other structural techniques.\textsuperscript{10,19,23} CX-MS is also highly sensitive, as only femtomolar amounts of total protein are required to generate many detectable crosslinks using modern MS platforms.\textsuperscript{5,8} Sample purity, homogeneity and solvent requirements for MS are more inclusive than x-ray crystallography or NMR,\textsuperscript{10,24} and thus can be applied to larger proteins,\textsuperscript{13,23} membrane proteins\textsuperscript{5,25,26} and protein assemblies\textsuperscript{19,20} in their native environments. Perhaps most important, conformational changes and protein dynamics can be analyzed as CX-MS allows investigators to examine functional complexes under wide ranges of physiological conditions. In other high-resolution structures, the protein is often truncated, bound with
stabilizers, solubilized or mutated for stability, often resulting in a non-functional protein, leaving the determined structure a static image that does not truly represent a dynamic full-length protein in its native environment.\textsuperscript{27,28}

1.2. Methodologies for Determining Structure - Why Choose CX-MS?

As previously mentioned, CX-MS is very versatile and lends itself to analyzing proteins not only in native conformations and environments but also in the context of larger systems. Since this will be a discussion about the utility of MS methodologies, particularly CX-MS, as a useful complement to other methods for studying the structure of proteins this section will first discuss some of the methods with a particular focus on membrane proteins.

Membrane proteins represent approximately 25 to 30\% of the entire proteome, but a disproportionate amount of them are targets for therapeutics.\textsuperscript{29–31} However, the characterization of integral membrane proteins, such as pentameric ligand-gated ion channels (pLGICs, see Chapter 2), is particularly challenging as they are typically expressed in low abundance, and because they are very hydrophobic, can only be purified by solubilization in detergents.\textsuperscript{25,32,33} They are often hard to crystallize due to the need for solubilizing detergents.\textsuperscript{33} The toolbox used by structural biochemists has a variety of methods to deal with these complications, though there are limitations to each of them.

1.2.1. NMR

NMR (nuclear magnetic resonance) spectroscopy studies the quantum mechanical properties of the nucleus of an atom by analyzing their behavior in a magnetic field and quantizing their nuclear spin. These properties depend on the local environment in which the atom is situated, and in the case of protein NMR, this information can be used to build
a map of how the atoms in the protein are linked together chemically, how close they are in space, and how they move with respect to one another.\textsuperscript{34} For proteins, the atom that is analyzed is usually the unpaired spin states $^1$H, $^{13}$C, $^{15}$N, or a combination of them.\textsuperscript{34} There are several different types of 2D-NMR techniques that are used in structural biochemical studies, including; COSY (correlation spectroscopy), NOESY (nuclear Overhauser effect spectroscopy), TOCSY (total correlation spectroscopy), HSQC (heteronuclear single-quantum correlation spectroscopy), among others, which are reviewed quite thoroughly in these reviews.\textsuperscript{35–41} All of them require similar steps to gain structural information, however; including sample preparation, which requires highly purified, stable and concentrated protein,\textsuperscript{19,42,43} data acquisition, spectra processing which matches NMR resonances to each atom in the protein and structural analysis which builds the structure using either distance geographic methods or molecular dynamics.\textsuperscript{35,42,44}

The major advantage to this technique is that protein structure can be determined for soluble proteins in their natural state. Because of this, NMR structures are able to represent protein dynamics.\textsuperscript{34,36,44} It is also a high-resolution technique, that provides good details and insights into protein structure. The disadvantages, however; are that the sample must be highly pure, in high concentration and stable, and be no larger than approximately 50 kDa.\textsuperscript{45–49} Also, membrane proteins, because of their large hydrophobic regions are difficult to study due to large micelle size of solubilized proteins, whose slow tumbling results in peak broadening.\textsuperscript{10,25,50} Solid-state techniques have been utilized to study some membrane proteins, but NMR studies have limited success in studying full-length integral membrane proteins, and instead has been used to examine smaller truncated soluble domains of membrane proteins.
1.2.2. X-ray Diffraction

X-ray diffraction uses x-rays to determine the arrangement and position of atoms in a crystal. By exposing a crystalline sample with x-ray beams, the generated angles and intensity of the diffracted x-rays as a function of sample orientation can be measured, and deconvoluted to create a three-dimensional image of the electron density of the crystallized molecule. This information can be used in fitting atom positions, chemical bonds, and structural details of a protein in question.\textsuperscript{42,51,52}

Major advantages to this methodology are the model building, high resolution and broad molecular weight range. Furthermore, x-ray diffraction is well studied and has been used for years.\textsuperscript{42} Though it can be used for a wide variety of protein types and for proteins in complex with DNA or RNA, the preparation, diffraction and crystallization for the technique is laborious and difficult.\textsuperscript{19,52} The samples used to generate structures must be able to be crystallized and must have high purity.\textsuperscript{52} There is also difficulty in studying membrane proteins, highly flexible proteins and large protein complexes.\textsuperscript{5,19,24,53} The biggest disadvantage though to this technique, and thus, the drive to use CX-MS, is that the structures generated do not represent proteins in their native environments and are typically liganded, to reduce the energy state of the complex to promote crystallization. X-ray structures are static images, and do not directly account for protein dynamics and conformational changes.\textsuperscript{5} Furthermore, mutations, truncations and bound-stabilizers can detract from native functionality and dynamics or render a protein non-functional.\textsuperscript{27,28} It is imperative to apply CX-MS to these incomplete protein structures to gain insight into protein dynamics and to refine structure in areas that are not well resolved.

1.2.3. Cryo-Electron Microscopy (cryo-EM)
Similar to X-ray diffraction, cryo-EM works by measuring diffraction using an electron beam, rather than x-rays, targeted at a cryopreserved sample in a vacuum. The beam passes through a thin layer of amorphous ice and scatters, creating an image that is magnified and detected. However, this high energy electron beam is highly destructive, causing damage to proteins and necessitating low dose imaging. Technical and computational limitations have only recently been surmounted. This information is used to generate a three-dimensional image of the structure of the protein. The versatility of this technique is immense, cryo-EM allows for small sample size and can be used on very large molecules and molecular complexes, such as virions and ribosomes. And unlike its predecessor, cryo-EM has the ability to study membrane proteins. The invention of polymer-based or lipid-based systems, such as amphipols, styrene-maleic acid (SMA) and lipid nanodiscs have made the analysis of membrane proteins easier.

Commonly, detergents are used to solubilize membrane proteins, but when used in cryo-EM structure determinations, the detergents can cause significant issues, such as decreasing the contrast of resulting images, disrupting the hydrophobic interactions between protein-protein and protein-lipid interactions, and not providing a native environment of lipids for the proteins. Amphipols, like detergents, have different physical properties than a native lipid bilayer, and differ in their electrostatic potential and dielectric constant as well. This can result in non-natively folded proteins. A study by Efremov et al. showed that the rabbit ryanodine receptor (RyR1), when solubilized in a lipid environment, compared to detergent environments such as Tween-20 or CHAPS has a similar structure but represents a different conformation. This is an important
point, as membrane proteins rely on direct interactions with certain lipids in their environment to influence their function.\textsuperscript{55,66,67}

Lipid nanodiscs provide a more optimal way to provide membrane protein structures. Briefly, lipid nanodiscs are composed of a section of lipid bilayer surrounded by a membrane scaffold protein (MSP), which is a derivative of apolipoprotein A-1, and consists of short amphipathic helices.\textsuperscript{55,68} MSP molecules are arranged in either a parallel or anti-parallel manner surrounding the piece of lipid bilayer which can be of variable size.\textsuperscript{55} A membrane protein can be inserted into the nanodisc and then can be imaged using cryo-EM. Though this tool has paved the way towards building the structural database of membrane proteins, which currently only represent a small fraction in the PDB, it does have some limitations. There are different biophysical properties between lipids used in nanodiscs, and the lipids that are contacting the MSP show perturbations.\textsuperscript{55,58,69,70} Furthermore, a study analyzing an ABC transporter, MsbA, showed an increase of activity in nanodiscs that contain a higher number of lipids.\textsuperscript{71} Another study using EPR showed that lipids in nanodiscs are also more highly ordered and may resemble a lipid bilayer environment in the presence of cholesterol.\textsuperscript{72}

These drawbacks may be managed in time, as this technique is still fairly new. However, the major disadvantages to this technique cannot be as easily addressed. Cryo-EM, despite its versatility, is very costly, and thus, not available for many researchers.\textsuperscript{51,52} Like cryo-EM, CX-MS is not hindered by the added detergents and membrane proteins can be easily reconstituted into various lipid bilayer environments to allow for native or near-native environments, but at a much more reasonable cost.
The other main disadvantage is that cryo-EM does not provide a good context in which to study protein dynamics, similar to x-ray diffraction, the resulting structures are static. The protein must be first trapped in a conformational state before rapid freezing, which makes studying dynamics involved in functional cycles, ligand or lipid binding, and infrequently populated states harder to discern. 29

1.2.4. Other Types of Mass Spectrometry

The structure of a protein can be analyzed through a bottom-up approach, meaning the protein is subjected to proteolysis and its fragments are analyzed to obtain information about the protein as a whole, or top-down analysis, where an intact protein is studied. CX-MS, as well as hydrogen/deuterium exchange (HDX), and oxidative labeling are examples of bottom-up proteomics, but they can also be performed in a top-down manner. Native (or non-covalent) MS, often used in conjunction with ion mobility spectrometry (IMS) is a good example of top-down proteomics.

Native MS (nMS) relies on keeping the non-covalent interactions intact when subjected to electrospray ionization (ESI), which allows for tertiary and/or quaternary structures to remain intact. 73–75 This is particularly useful for the study of protein subunit stoichiometry. 29 The Robinson lab has been quite successful in obtaining structural information for large membrane protein complexes using nMS methods, including stochiometric information for V-type ATPases from *T. thermophiles* and *E. hirae*, 76 an ATP binding cassette (ABC) transporter, 77–79 among others. 12,80–83 When nMS is coupled with IMS, protein complex connectivity can be determined, which has also been shown by the Robinson lab, 80,84–88 and others. 89–92 Besides determining architecture and stoichiometry, nMS with and without IMS has also helped developed the literature on the
importance of lipids in membrane protein structure and function. Specific lipids provide 
stability, alter function and can promote dimerization. Some challenges exist with 
this technique, in regards to studying unstructured loop regions and intrinsically disordered 
proteins, but these can be managed by using bottom-up techniques such as CX-MS in tandem with nMS. There are also similar issues involving the use of detergents, 
amphipols and lipid nanodiscs as seen in cryo-EM, reviewed by Calabrese and Radford.

Using a bottom-up approach with HDX-MS, CX-MS or various other labeling 
techniques such as oxidative labeling, one may study protein structure by following a 
generic workflow; labeling the protein/protein complex, using proteolytic digestion, and 
finally MS analysis. Usually, the MS analysis includes soft ionization techniques such as 
ESI and either collision induced dissociation (CID) or surface induced dissociation (SID) 
to create analyzable fragments that include the site of modification. These can then be used 
to build a map to infer structural information, binding interfaces, and/or conformational 
dynamics (See Section 1.4 for more details about the workflow of CX-MS).

1.2.4.1. HDX-MS

HDX-MS involves diluting a protein sample in deuterated buffer, allowing for 
hydrogen atoms that are accessible to the solvent to be exchanged with deuterium. The 
exchange rate between side chain amines and carboxylic acids is too fast to be measurable, 
so these experiments focus on main chain amide hydrogens. The exchange between 
hydrogen and deuterium is allowed to occur for defined periods of time and then the 
reaction is quenched by lowering the pH of the solution to approximately 2.5, where the 
exchange rate is negligible. As with any bottom-up approach, the sample is then digested, 
separated by liquid chromatography (LC) and analyzed via MS.
Commonly, HDX-MS is used to study protein conformation, dynamics, ligand binding sites and allosteric effects. Several studies have been conducted to study conformational changes of membrane proteins including a study by Mehmood et al. of a bacterial ATP-binding cassette (ABC) transporter in both an inward facing and outward facing conformation. Their studies showed different flexibilities of regions between the two states. Another study, focusing on the Na+/H+ antiporter from *E. coli* analyzed lithium binding at physiological pH and provided a mechanism for ion translocation. Similarly, Adhikary et al. analyzed LeuT a bacterial homolog of the neurotransmitter sodium symporter family to discern a conformational mechanism for transport. Protein complexes and their configurations are also a common target for HDX-MS studies. The β2-adrenergic receptor has been studied in complex with a G protein coupled receptor (GPCR), GRK5, a GPCR kinase, and β-arrestin.

HDX-MS is also a useful technique for understanding protein folding and discovering motifs within proteins. An example of this is a study that analyzed human green opsin, a GPCR, where researchers found a conserved Pro-Pro motif in an extracellular loop. Loops that are unstructured are often difficult to crystallize and thus, using methods like HDX-MS can help provide structural details not found in high-resolution methods.

1.2.4.2. Oxidative Labeling

Other labeling techniques, such as oxidative labeling, are often coupled with MS to provide similar results to HDX-MS and CX-MS. Labeling with hydroxyl radicals is generally not very specific, and thus, multiple residues can be non-specifically targeted. Like HDX-MS, the stipulation to labeling is that the residue must be solvent accessible,
but it must also have intrinsic reactivity as well. Residues most likely to be labeled are those with sulfur atoms, Met and Cys, followed by aromatic residues, Trp, Tyr and Phe. Additionally, His, Arg, Lys, Gln, Glu, Leu, Ile, Val and Pro can also be modified.\textsuperscript{110}

Oxidative labeling is most commonly used in much the same ways as both HDX-MS and CX-MS, in that is provides constraints useful in modeling, specifically for proteins that are not amenable to other methods,\textsuperscript{110–112} and maps protein-protein interfaces to show connectivity and stoichiometry of protein complexes.\textsuperscript{113–118} Various studies have analyzed the interface of monomeric subunits of galectin-1,\textsuperscript{112} unfolding properties of apomyoglobin,\textsuperscript{117} and have characterized the tertiary structure of CD4 bound to an envelope protein component of HIV, gp120.\textsuperscript{116} Of particular importance to this work is the analyses of integral membrane proteins using this technique. Two studies by the Konermann lab\textsuperscript{119,120} investigated structural characterization of bacteriorhodopsin in a natural lipid environment. These studies focused primarily on labeling Met residues as they are found in abundance in bacteriorhodopsin. They were interested in discerning which Met residues were solvent accessible and comparing their results to native structures of bacteriorhodopsin. They also analyzed the protein under semi-denatured conditions and found that noncovalent contacts between bacteriorhodopsin and the bilayer provide stability to certain helices within the structure.

There are some drawbacks to this technique, such as secondary oxidation reactions, oxidation of non-solvent accessible residues, incomplete conversion and specificity.\textsuperscript{121} But, overall, this technique provides complimentary to high resolution structures as does the other bottom-up approaches and is thus, useful towards understanding the complex fields of protein dynamics and protein-protein/lipid/drug interactions.
Though not a complete discussion of all the methods used to study protein structure and dynamics, this section has provided a glance into the toolbox a structural biochemist uses. The next section will look deeper into the wide array of uses for CX-MS and why it is the specific tool of choice for this work.

1.3. Applications of CX-MS

1.3.1. Protein Structure and Modeling

Determinations of tertiary protein structure are most commonly solved using NMR, x-ray crystallography and now, cryo-EM. As of 2018, x-ray crystallographic and NMR structures totaled 98% of all structures found in the Protein Data Bank (PDB).\textsuperscript{19} Most of the structures in the PDB are of soluble proteins, however; this is changing, and more membrane proteins are being included due to recent advancements in cryo-EM. Though these methodologies provide good structural details, they do have disadvantages as was previously discussed. One common way to utilize CX-MS was to aid in structural refinements as CX-MS can help mitigate some of the issues seen in other methodologies and provide structural refinements important towards understanding conformational dynamics in native-like environments.

Recent successes of the technique, towards structure determination are derived from crosslinking data, resulting from computational or experimental studies. There have also been publications that focused on using crosslink data from the literature with software tools to assess how to best analyze and understand the crosslinking results in terms of protein structure.\textsuperscript{122-125} Kahraman \textit{et al.}\textsuperscript{126} and Hoffmann \textit{et al.}\textsuperscript{127} used simulated crosslinks with various analytical software tools to provide information on visualization and validation and scoring solvent-accessible crosslinking, respectively.\textsuperscript{128}
Analysis of the bovine basic fibroblast growth factor with intramolecular CX-MS identified the fold of the protein using sequence threading and Lys-Lys crosslinking with bis-sulfosuccinimidyl suberate\textsuperscript{129}. A study by Brodie \textit{et al.}\textsuperscript{130} analyzed myoglobin and FK506 binding protein using CX-MS and discrete molecular dynamic simulations to create distance constraints. They were particularly interested in comparing the ability of CX-MS to identify proteins made up of mostly α-helices versus β-sheets\textsuperscript{130}. Belsom \textit{et al.}\textsuperscript{131} used high-density CX-MS to analyze human serum albumin and compared their results to available crystal structures. The study used a photoactivatable crosslinker, sulfo-SDA (sulfo-NHS-diazirine or sulfosuccinimidyl 4,4’– azipentanoate), which benefits from being able to react with any amino acid on the diazirine end, thus providing more crosslinking data. A particular focus of this study was to see how this methodology performed in terms of identifying unknown structures\textsuperscript{131}. This study, as well as a study by Fischer \textit{et al.}\textsuperscript{132}, compared the success of analyzing α-helices versus β-sheets. According to these studies, CX-MS has greater success solving protein structures that were mostly α-helical in nature\textsuperscript{131,132}. This may be due to crosslinker length, which has a relative distance from Cα to Cα of approximately 20 to 25 Å, whereas; the distance between β-strands in a β-sheet is only 5 Å\textsuperscript{10,131}. However, a good way to combat this would be through the use of different length crosslinkers, to create a network of constraints to hone in on minute structural details.

Structural refinement, through the means of protein-protein interactions within a singular protein will be the focus of this work. A heterobifunctional MTS-benzophenone crosslinker will be used (details on common crosslinkers used can be found in Section 1.5).
and the target will be a transmembrane protein, which as previously covered, have unique complications involved when studying their structure.

1.3.2. Protein-Protein, Protein-Lipid and Protein-DNA Interactions

Outside of using protein-protein interactions to determine structure and refine models, protein-protein interactions, as well as proteins crosslinked to other macromolecules can provide context into the protein’s native environment.

Protein-protein interactions can be studied between multiple proteins in multi subunit complexes, or between binding partners. Two of the most well-known examples are the elucidation of the nuclear pore complex\textsuperscript{24,133,134} and RNA polymerase II and its binding partners.\textsuperscript{13,135–137} The architecture of the entire pore complex was determined by using several techniques, including a low-resolution cryo-EM density map as a starting point, affinity purification and pull-down assays, MALDI MS, immunoblotting, constraint mapping and computational studies.\textsuperscript{24} Predicted locations of proteins in the pore complex as well as protein folds typical of coat proteins that were close to the nuclear membrane surface, such as clathrin were assigned.\textsuperscript{133,134} Pore spokes were arranged in vertical columns and it was hypothesized to be due to gene duplication for these types of structures.\textsuperscript{24} RNA polymerase II was shown to interact with TFIIF through over 250 crosslinks between the two proteins,\textsuperscript{13} which visualized their orientation towards each other. Before this study, the understanding of transcription initiation was provided via docking studies, which can be inaccurate when used alone.\textsuperscript{24} This work has led to the investigation via CX-MS of other systems, including the eIF1-eIF3 translation initiation complex,\textsuperscript{138} transcription and DNA repair factor TFIIH,\textsuperscript{139} and a modulator of the TORC1 pathway, the SEA complex.\textsuperscript{140}
Protein-lipid interactions can be helpful in elucidating protein structure and also provides information, specifically when studying membrane proteins, regarding local membrane environment. According to a review by Yeagle\textsuperscript{67} in 2014, 70 membrane proteins have been crystallized in the presence of lipids, including cholesterol, diphosphatidylglycerol (cardiolipin), phosphatidic acid (PA), phosphatidylglycerol (PG), phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI) and phosphatidylinositol bisphosphate (PIP\textsubscript{2}). Furthermore, studies have shown that different lipids have an effect on both protein structure and function.\textsuperscript{93,95,141–144} Studies using CX-MS and analyzing protein and lipid interactions, like those conducted by Ferraro \textit{et al.}\textsuperscript{14} have shown that cholesterol concentration results in different apo-states of the glycine receptor and has elucidated specific binding sites for cholesterol in a state dependent manner.\textsuperscript{14} DeMarco \textit{et al.}\textsuperscript{145} showed cholesterol binding to the N and C-terminus of the serotonin transporter, regions absent in all crystal structures of monoamine transporters. A large-scale study by Wang \textit{et al.}\textsuperscript{15} looked at protein-lipid interactions using an alkyne-containing choline head group and a diazirine modified fatty acid to crosslink \textit{in situ}, with phospholipid interacting proteins identified by MS. The field of protein-lipid interactions, unlike protein-protein interactions is still unexplored and needs further studies to develop the lipidomics side of CX-MS.

Interactions between protein and DNA can show the dynamics of binding, structural details related to these conformations and the bigger picture regarding the local environment. Because the interactions between nucleic acids (DNA or RNA) are noncovalent interactions, and thus weak, trapping them via crosslinking can facilitate analysis. Formaldehyde is most commonly used,\textsuperscript{146,147} but photoactivatable crosslinkers
that produce pyrimidinyl radical are also used.\textsuperscript{148,149} Both of these are zero-length crosslinks.\textsuperscript{146,149} This protein and nucleic acid crosslinking was later applied to MS analyses,\textsuperscript{150–153} including a study where a fibroblast growth factor was crosslinked to uracil.\textsuperscript{153} Later advancements in protein pull-down assays and chromatin immunoprecipitation (ChIP) led to more large-scale studies.

Qiu and Wang\textsuperscript{154} searched for unknown DNA repair and regulation proteins using formaldehyde crosslinking and LC/MS-MS. A study analyzing histone proteins in intact nuclei used protein-protein and protein-DNA interactions to help aid in understanding the total landscape of the nuclei.\textsuperscript{155} ChIP, with added MS-based analysis and Chromatin enrichment proteomics (ChEP) both use a crosslinking step with formaldehyde that crosslinks DNA to proteins.\textsuperscript{17,156,157} These methods are gaining popularity as they provide a way to better understand the role of PTMs, particularly in chromatin, in affecting DNA binding, which is useful for the growing field of epigenetics.

1.3.3. The Interactome and Studying Proteins \textit{in Situ} – The Future of CX-MS

More recently, CX-MS has been used to study larger interaction networks and proteins \textit{in situ}, within their crowded physiological context. This is ideal for studying protein-protein and protein-macromolecule interactions, as it provides the most authentic environment.

A study by Ryl \textit{et al.}\textsuperscript{20} in 2019, analyzed the human mitochondria for protein network interactions. They used a non-cleavable Lys-Lys crosslinker, disuccinimidyl suberate (DSS) and identified over 5000 distance constraints, and had crosslinking between over 700 proteins. Studies similar to this have also been conducted on murine mitochondria with cleavable crosslinkers\textsuperscript{18,158} and several other large systems already mentioned,
including the 26S proteasome\textsuperscript{135} components of RNA polymerase I, II, III\textsuperscript{13,136,137} and the nuclear pore complex\textsuperscript{133,134} have been elucidated structurally by CX-MS. Other large-scale protein-protein interaction studies focus on intact cells, including \textit{P. aeruginosa} cells\textsuperscript{159}, \textit{E.coli} cells\textsuperscript{160}, MR-1 cells\textsuperscript{161} and HeLa cells\textsuperscript{162}. There are also studies that have used viruses\textsuperscript{163}. These studies highlight the future territory of CX-MS and its role in discovering the molecular machinery of large-scale systems and how proteins interact \textit{in situ}.

Fürsch \textit{et al.}\textsuperscript{164} conducted a proteome-wide study to analyze the propensity for high-abundant protein crosslinking. They used an \textit{in vitro} mimic of a crowded cell environment and compared it to eukaryotic cell lysates as they hypothesized that by optimizing analysis parameters towards a first order kinetic model would allow for more low-abundant protein crosslinking sites to be determined. The importance of their work is to show limitations in current CX-MS analysis techniques (see Section 1.5) and to provide better methodologies to improve these techniques.

The use of crosslinkers that are cleavable inside the mass spectrometer also offers new insights into understanding proteins in their natural environment. These include incorporation on unnatural amino acids in proteins using photoactivatable amino acids, such as photo-leucine and photo-methionine\textsuperscript{165,166}, oxidative crosslinkers\textsuperscript{167} and photoactivatable crosslinkers\textsuperscript{168,169}, such as disuccinimydal dibutyric urea (DSBU) and disuccinimidyl sulfoxide (DSSO).\textsuperscript{5,10} A study by Arlt \textit{et al.}\textsuperscript{170} analyzed the full-length tumor suppressor protein, p53, used DSBU and mapped distance constraints applying it to p53 crystal structures bound with DNA.\textsuperscript{10,171–173} It is noted that the length between Ca – Ca using DSBU is approximately 27Å, and the group found that the majority of both intra- and intermolecular crosslinking matched what was found in the crystal structure.\textsuperscript{10,170,173}
Of particular importance, p53 is an intrinsically disordered protein\textsuperscript{174} and thus has a very flexible protein structure, so using CX-MS is ideal for studying it in a native context.

Regardless of the protein being analyzed, its size, its environment and what it is being crosslinked to, CX-MS has the ability to aid in structural refinement, modeling, and understanding protein dynamics. Of particular interest to this work is the availability of CX-MS to analyze membrane proteins, which have proven difficult to study by other methods. CX-MS is a versatile, cost-efficient, and high-throughput method that is a valuable tool in fully developing the structure of a protein. It is becoming increasingly important in understanding proteins in a larger setting, something that was unforeseen a mere decade or two ago. Because of its affordability and accessibility, it will continue to be a well-sought-after tool in the structural biochemist’s toolbox.

1.4. Workflow of CX-MS

As previously discussed, CX-MS can be used to study a protein/protein complex either by a top-down or a bottom-up proteomic approach. Because the focus of this work uses the bottom-up approach, this section will discuss only this one at length. The general workflow of a bottom-up CX-MS experiment (Figure 1) consists of primarily three parts; (1) crosslinking, (2) proteolytic digestion and (3) MS/MS analysis. The overall goal of this type of experiment is to map protein interfaces or determine low-resolution structures of proteins. This section will focus on the more specific workflow involved in using CX-MS to determine and refine protein structure.

Before beginning this type of experiment, it is essential that the sequence of the
protein(s) that are being analyzed is completely known, allowing them to be matched using
databases. It is also ideal to optimize all conditions such as crosslinker concentrations,
protein concentrations, buffer pH and reaction times to maximize sequence coverage, and

Figure 1. General Workflow of CX-MS. The general workflow of CX-Ms is split into several parts: 1) crosslinking, which in this example, shows intrasubunit crosslinking in yellow and intersubunit
crosslinking in green, 2) proteolytic digestion, which maintains the covalent crosslinking, 3) MS analysis
via ESI-Q-TOF, which identifies precursor ions, 4) MS/MS analysis which allows mass-selected
precursor ions to be targeted for fragmentation with CID, creating product ions which are then matched
with predicted fragmentation patterns to build crosslinking maps and refine structure.

more fully characterize the protein(s) in question. Control samples that have not been
crosslinked should be used as controls to rule out any possible protein aggregate formation
occurring that would be confused for crosslinking.

The chosen crosslinker will determine the specificity of what residues are being
crosslinked as well as the graphic scale of the crosslinking map. This information is a
-crucial step towards refining structure (see Section 1.5 for common crosslinkers used in
CX-MS experiments). For this work, methanethiol sulfonate benzophenone (MTS-bzp) will
be used as the molecular ruler, because the benzophenone group does not show specificity
towards any amino acid, and because the MTS functional group can be targeted towards
recombinantly introduced single cysteines as a point of initial attachment. In general, a
crosslinking experiment occurs in solution, after enrichment of a protein in a desired conformational state. Depending on the protein or protein complex and the crosslinker used, the end result may give both intra- and intermolecular/subunit/protein crosslinks.

In a bottom-up approach, the next step is proteolytic digestion. In order to do this, first the crosslinked protein must be separated from the rest of the reaction mixture which can be done by SDS-PAGE or by size-exclusion chromatography. If SDS-PAGE is used, the piece of the gel containing the crosslinked protein can be excised and be subjected to \textit{in gel} proteolytic digestion. If size-exclusion chromatography is used, the digestion happens in solution. In the case of multi-subunit or protein complexes, there are different bands or fractions, consisting of a single protein (or subunit) with only crosslinks within itself, and those containing crosslinking between one protein and another protein or one subunit to another subunit. Different proteases can be used for digestion, depending on the sequence and/or buffer conditions. Sometimes, multiple proteases are used.

The crosslinked peptides can now be analyzed via MS. This can be done using MALDI or ESI in a bottom-up approach. This section will focus only on the latter. Usually the peptide mixture is introduced into the mass spectrometer by first allowing LC separation. Then, in ESI, the liquid sample is sprayed into the instrument while being subjected to an electric field which produces small, highly charged droplets. These ionized samples are then introduced to the mass analyzer. Identifying the crosslinked residues from the produced data happens in a two-step process; (1) identifying the peptides that were crosslinked and (2) refining the area to a single residue (or close to single) of that peptide where the crosslinker bound. In the first step, a series of spectra are obtained and precursor ions are selected. These precursor ions are the peptides that have been mass-shifted by the
mass of the crosslinker. ESI-MS/MS can then be used to further refine these precursor peptides. The precursor ions are targeted for dissociation, in the case of these studies CID was used, to produce product ions of reduced mass that may be subsequently separated by their respective m/z by further MS. Identifying fragments with or without the covalently attached crosslinker allow for the refinement of crosslinking sites to an exact or approximate site within the precursor mass ion. Successful refinement of the site of crosslinking is dependent on the quality and dynamic range of the spectra produced.

Peptide fragmentation by CID generally occurs at the peptide bond, produces two main fragment types, N-terminal $b$-ions and C-terminal $y$-ions, depending on which species retains the charge. An alternative to CID is electron transfer dissociation which generally gives N-terminal $c$-ions and C-terminal $z$-ions. Because peptides and proteins fragment in these expected ways, this methodology is very useful towards identification of peptides.

The extracted data from the fragmentation spectra must be matched at a certain level of confidence with a theoretical crosslinked peptide match. This leads to unambiguous matches that can refine structure. A large component of using CX-MS is computer software data analysis programs. This aspect of the workflow, more than any other, is laborious and time consuming due to the complexity and sheer size of the data produced. This is overall, the biggest limitation to CX-MS. There are many different computer algorithms that can be used to match spectra with potential crosslinked peptides, and ultimately, that is the main goal. Many of the publicly available software for crosslink identification, visualization, modeling and quantitation can be found in the review by Yu and Huang. More complete searchable databases and a generally
acceptable method for CX-MS experiments, in addition to new, better equipped software programs will increase the usability and popularity of this technique. The final aspect of this technique is mapping the distance constraint data onto available crystal structures or designing structures based on the distance constraints. Structural modeling techniques can assist with this, once a suitably sized network of crosslinks is obtained. This information can be used to guide complex protein arrangements and stoichiometry, tertiary structural modeling, and conformational changes.

1.5. Crosslinkers used in CX-MS

The location of a crosslinker imposes a distance restraint of known length, which after enzymatic digestion, crosslinked peptides can be analyzed with mass spectrometry. This information can be used to help construct a three-dimensional structural image of the protein. These distance restraints can also be applied to help narrow down the likeliness of a preexisting model, generated from crystallography or by computational studies. Some important considerations that can be used to reduce the number of predicted structures include the accuracy of the probe to limit interatomic distances, the nature of the approximations used in making the models, and the similarities and dissimilarities in the examined models.

Another important consideration is what crosslinker is chosen. Several factors contribute when choosing a crosslinker; its reactivity, its functionality and its length. Common reactivities used are sulfhydryl, amine and photoreactive. Sulfhydryl reactive compounds include maleimides, haloacetylts and pyridyl disulfides. Their reaction mechanism includes the formation of either a thioether bond or a disulfide bond to Cys residues in proteins. Because of their specific reactivity, using this type of crosslinker
can limit the structural details obtained, and the need to reduce disulfide bonds could also lead to a distortion of structure.\textsuperscript{187} The abundance of Cys residues in proteins is quite small, only about \(2.3 - 3.3\%\),\textsuperscript{188,189} so structural information using just the Cys available in the protein minimizes the amount of information that can be obtained.

Amine reactive crosslinkers most commonly used are N-hydroxysuccinimide (NHS) esters, imidoesters, and carbodiimides.\textsuperscript{186} All of these react with primary amines, which can be found at the N-terminal end of a protein and at the side chain of Lys residues.\textsuperscript{190,191} Lys-Lys crosslinking is commonly used with this type of crosslinker.\textsuperscript{192–197} An example of a commonly used carbodiimide is EDC (1-(3-dimethylaminopropyl)-3ethyl-carbodiimide hydrochloride), which is a zero-length crosslinker.\textsuperscript{198} The abundance of Lys residues in proteins, is higher than Cys, at \(7.2\%\),\textsuperscript{199–202} so there are more available; however, they are usually only solvent exposed. The main disadvantage, though, to using these types of crosslinkers is that tryptic cleavage sites are removed.\textsuperscript{5,8}

Photoreactive crosslinkers are those that are induced to react when exposed to UV light. Common ones are diazirines, aryl azides and benzophenones. What makes these molecules ideal is their high reactivity, low selectivity, stability, and reactivity within a range that is not destructive to biological samples.\textsuperscript{5,198} Because of their non-selectivity or low-selectivity, they are not limited to binding specific amino acid residues and thus are useful for unbiased structural determinations.\textsuperscript{198} Aryl azides are the most commonly used crosslinker of this type, and work by forming short-lived nitrenes after photoactivation that can insert into target molecules through active hydrogen bonds at C-H and N-H sites as well as through addition reactions at double bonds.\textsuperscript{191,203,204} Diazirines, when induced by UV light, generate a carbene that inserts into C-H bonds or heteroatom-H bonds.\textsuperscript{5} This type
of crosslinker is used often to analyze protein-DNA/RNA interactions\textsuperscript{205–207} as well as for structural studies.\textsuperscript{208–210} Finally, benzophenones differ from the previous types because they are produced via a non-photo-dissociative mechanism and thus reversible. After being activated via light, a biradical is formed. The oxygen radical abstracts a hydrogen radical from a bond of the reaction partner. The created alkyl radicals form a new C-C bond between the photoactive molecule and the protein target.\textsuperscript{5,211,212} Several structure determination studies have been conducted using benzophenones as the crosslinker of choice.\textsuperscript{213–217}

An additional affinity tag can also be applied to crosslinkers, such as biotin,\textsuperscript{218} or an alkyne or azido group which can be used to enrich crosslinks by covalently attaching them to a solid phase support.\textsuperscript{198} There are also anti-crosslinker antibody based enrichment tags.\textsuperscript{198} The incorporation of these tags can help increase the signal-to-noise ratio by providing for enrichment of crosslinked peptides before MS.\textsuperscript{218}

Crosslinkers of all reactivities can also be under the umbrella of cleavable crosslinkers. Both photo-\textsuperscript{219} and chemical-reactivities\textsuperscript{220} can be cleaved by various means depending on their unique properties before MS analysis. But there are also MS-inducible cleavages that result from fragmentations that happen inside the mass spectrometer, including via CID and ETD.\textsuperscript{19} The advantages to using crosslinkers that are cleavable in the MS are two-fold: one, there is effective correlation between separated crosslink constituents and the parent ion, as each fragment carries a piece of the fragmented crosslinker, and two, this results in simplification of data processing and analysis.\textsuperscript{19} An important consideration for these crosslinkers is that their cleavable bonds are more labile than peptide bonds and that their fragmentation should be independent of peptide charge.
and/or sequence. Several important ones to note, designed in the Huang lab are DSSO (disuccinimidyl sulfoxide), DMDSSO (dimethyl disuccinimidyl sulfoxide), and DHSO (dihydrazide sulfoxide). All of these are sulfoxide containing crosslinkers which are commonly utilized reactive groups in other heterobifunctional crosslinkers.

Functionality is another factor used when choosing a crosslinker. These fall under two general categories; homo- or heterofunctional. Homofunctional crosslinkers have identical reactivity groups on each side of the spacer arm. This allows for the crosslinking between identical functional groups, such as Lys-Lys or Cys-Cys crosslinking. Though these crosslinkers are very common and used often, there are some issues associated with using them. First, amino acids may be distributed unevenly throughout a protein, and thus, if crosslinking to only a specific type, it may lead to lack of global information. Also, there can be a tendency to form protein aggregates through intermolecular crosslinking. This can be mitigated by using less concentrated protein samples or by performing a two-step reaction where a single protein is crosslinked first before adding a second protein of interest. This is only useful though if the study is targeting a protein interface between multiple proteins, instead of looking at a singular protein.

Heterobifunctional crosslinkers will have at least two different reactive groups, but can also have more than that. These are useful as they target more than one functional group and thus are benefitted naturally by a multi-step process that decreases the chance of forming higher-order aggregates. Photoreactive crosslinkers are especially useful in that they are stable until induced with UV light. Methanethiosulfonate-benzophenone (MTS-bzp) is a good example of a heterobifunctional crosslinker that targets both Cys residues
one reactive end, and has no specificity on the benzophenone site, allowing it to target any amino acid.\textsuperscript{211,212} Of particular note, this work will use MTS-bzp.

Often, crosslinkers will have more than one reactivity, which can help to provide less selectivity and thus a better chance of reacting with more of the protein providing more distance constraint information. For example, it is common that photoreactive groups are also paired with amine- or sulfhydryl groups, making them multi-functional.\textsuperscript{5}

Lastly, the length of the crosslinker provides another level of classification. As stated previously, there are zero length crosslinkers like formaldehyde (commonly used to fix cells and in DNA-protein interactions studies). This class of crosslinkers require both functional groups being crosslinked to be in close enough proximity to form a covalent bond. MTS-bzp is approximately 25Å from Cα to Cα. Regardless of the crosslinker chosen, it is important to consider the range of flexibility of the crosslinker as well as protein dynamics. This defines the proximity between the two points, but this distance can vary over time, which is useful for structural mapping. Studies by Ryl \textit{et al.},\textsuperscript{20} Merkley \textit{et al.}\textsuperscript{224} and Kahraman \textit{et al.}\textsuperscript{123} showed that many crosslinks formed exceed the length of the crosslinker used. It is also useful to use multiple length crosslinkers to expand or shrink the grid scale of the map, to provide additional details related to structure from a particular point of reference.

1.6. Limitations and Challenges of CX-MS

CX-MS is a valuable tool for the structural biochemist, however; there is room for improvement. The limitations and challenges of this technique can be split into two categories: technique-based and data analysis-based, specifically for the purpose of modeling.
1.6.1. Technique-based

Proteolytic digestion is most commonly performed with trypsin, and because it cannot cleave a modified C-terminal Lys or Arg residue, this results in missed cleavages. Also, there are many crosslinkers that employ the modification of a primary amine, and so trypsin cannot be as effective here, either. Many data analysis algorithms account for missed cleavages, and this will mitigate missing peptides that have modified Lys residues. However, missed cleavages also result in larger peptides, which is another issue altogether.

The larger the peptide piece is, the greater the likelihood of it being harder to detect with the mass spectrometer. Large peptides may also result from crosslinked products with low charge states. ESI can create these peptides from a loss of positive charge after modification of Lys residues. There is also an issue with an increase in the number of peptides that have the same mass. This is increased as the number of amino acids in the peptide increases.

Using in-gel digestion of crosslinked peptides via excised gel pieces can result in a loss of sample. This could result from proteins not being well-resolved and thus not excising them properly. Size exclusion chromatography could be used to mitigate this. However, because of the high sensitivity of the mass spectrometer, not much sample is needed.

1.6.2. Data Analysis – How is this information best used for modeling?

Crosslinking data can provide a wealth of distance constraints, that can then be applied to model building or model validation. While CX-MS is not the only tool that creates distance constraints, it does create a multitude of them that can be used for modeling.
purposes. This means that some consideration and care should be taken when applying this information towards building models.

The distance between two crosslinked residues should be limited by the distance of the crosslinker between them. However, this does not take into account that both the protein and the crosslinker will have inherent flexibility. The crosslinker will have rotations of bonds as well as vibrations that cause a variability in its length. And the protein, as a dynamic molecule, will sample energy states that may not be representative of its lowest energy state and differ from available high-resolution structures. One way to provide additional information refining the available conformational states is by using multiple crosslinkers of variable lengths.

Matching fragmentation spectra with expected fragments of peptides in databases yield unambiguous data. There is also evidence that supports very little experimental error in regards to misassigned crosslinks, as data is only interpreted from fits with a very high confidence match. Since CX-MS is able to look at protein dynamics, whereas crystal structures do not, it makes sense that the variability in conformation and flexibility will be larger. This is also seen in peptides that cannot be resolved to a single residue; it may be due to flexibility of the region, or alternative conformations, instead of lack of fragmentation spectra due to hard to ionize species. Thus, there is value to allowing lower-confidence matches being taken into account. This is also important in terms of showing multiple structures as a result of lower confidence matches, as these might be indicative of multiple conformational states in a singular sample.

Depending on the crosslinker used, as well as the accessibility of the crosslinker to various regions in the protein, there may be missed data. If a crosslinker is only targeting
Lys residues, there may be an uneven distribution of them in the protein; the same can be said for any crosslinker with certain specificities. Photoactivatable crosslinkers, that do not have residue specificity are a good alternative. However, depending on the hydrophobicity or hydrophilicity, there may be issues penetrating the transmembrane regions of the protein or the solvent accessible regions,\textsuperscript{225–228} this is particularly important when studying membrane proteins. Similar to the distance issue, this can be mitigated by choosing a crosslinker that best suits the protein of interest being analyzed, or by using multiple types of crosslinkers to obtain global information.

As previously mentioned, larger peptides are harder to detect in the mass spectrometer. Also, if these larger peptides are picked up, they may be hidden by background noise due to their lack of ionizability.\textsuperscript{177} There are also issues of hydrophobic peptides being lost through sample preparation.\textsuperscript{7} This means that not all peptides that are crosslinked are represented in the data. The interpretation of this can be two-fold; as the absence of crosslinked peptides either means that it is not in close proximity to the initial point, or that the two pieces would crosslink, but it is not represented in the sample. These difficulties are reduced by having good mass coverage in control studies. A developed network of various points can also help mitigate this. This way, not only is new information valuable, but also reciprocity between regions. If A crosslinks to B and B crosslinks to C, it is important to see if A also crosslinks to C.

Another setback of CX-MS data analysis is quantitation. Mass spectra peaks represent the intensity of an ionizable peptide, not its abundance. MS quantitation has been typically done using a labeling technique, by using stable isotopes, either introducing them to the protein sample through the growth media (SILAC – stable isotope labeling by/with
amino acids in cell culture)\textsuperscript{229–231} or the crosslinker.\textsuperscript{232} But comparing the yield of crosslinked peptides between samples is difficult, and because their ‘flyability’ the intensity in which a given peptide is observed, differs between experiments, individual peptides cannot be compared between samples.\textsuperscript{7} Another issue with quantitation is low signal intensity.\textsuperscript{233} This means if peaks are even visible to be detected, it is much harder to produce reliable fragmentation spectra. Often an Orbitrap mass spectrometer is used to trap a precursor ion and produce fragmentation spectra through tandem MS.\textsuperscript{179,233} The biggest issue with analyzing quantitation data is the software is currently not equipped to handle such complex data. In 2016, Chavez \textit{et al.}\textsuperscript{234} proposed using an open source software package Skyline to analyze quantitative CX-MS data. It is available to researchers who have LC-MS capabilities, and works with the already available quantitative methods, like SILAC. There are other packages like this one that have been proposed to help with the vast amount of data produced.\textsuperscript{179,233} In 2019, Muller \textit{et al.}\textsuperscript{179} discussed a data-independent acquisition of analyzing quantitation CX-MS data.

1.7. Summary.

Crosslinking-mass spectrometry (CX-MS) is a sensitive and useful tool to characterize spatial relationships of proteins. It is a preferred technique to study membrane proteins as it is not hindered by their mass or their hydrophobicity. By using CX-MS, a network of distance constraints will be obtained, and these can be used to build more highly resolved and accurate molecular models. This will provide complementary information of a full-length, functional protein in a near-native environment, which cannot be achieved using high resolution structural techniques. Ultimately, as protein structure is intimately
linked to protein function, these studies will aid in future modulation of this protein and
the development of targeted therapeutics.
CHAPTER 2. PENTAMERIC LIGAND GATED ION CHANNELS

2.1 Introduction

Pentameric ligand gated ion channel (pLGICs) proteins are transmembrane channel proteins that allow cells to respond quickly to external changes in their environment, particularly by mediating fast neurotransmission in the nervous system. When a presynaptic nerve is excited or stimulated, it releases a neurotransmitter that then binds its target protein on the post synaptic nerve. As a result of neurotransmitter, or ligand, binding, a conformational change results which opens a pore in the pLGIC and allows either cations (Na\(^+\), K\(^+\), Ca\(^{2+}\)) or anions (Cl\(^-\)) to enter the cell. The resulting ion flux is driven by the electrochemical gradient and is thus passive. The result of this communication between neurons via electrochemical response is an action potential, which is when the membrane potential rapidly rises and falls from its normal ‘resting’ state, often referred to as ‘firing’.

Figure 2. The Synaptic Cleft. The glycine receptor is located at the post-synaptic cleft and responds to glycine binding. Its role is to allow the entry of chloride ions into the cell to hyperpolarize the cell, which decreases the likelihood of an action potential. The image below shows the release of glycine from synaptic vesicles into the cleft, where it can find its target receptor and bind. Chloride ions can then move into the cell.
The path along which action potentials occur starts approximately at -60 millivolts (mV), slightly more negative than its threshold potential, and as voltage-gated sodium channels open, and positive ions pass into the cell, the potential rises to the threshold potential, where at that point, an action potential will either happen or not. If it does pass the threshold, there is a surge of more sodium ions entering the cell, and the potential increases to approximately 40 mV. This is called depolarization and is an excitatory response. As the sodium channels begin to close and potassium channels open, repolarization occurs. Instead of settling around the potential in which it started, it becomes more negative, as a result of efflux of potassium and influx of chloride ions. This hyperpolarizes the cell, and is known as the refractory period, which prevents another action potential from happening too quickly; this is the inhibitory response. Both the excitatory and inhibitory responses are controlled by pLGICs.\textsuperscript{238,239}

There are several members that belong to this family, that also goes by the name of Cys-loop receptors, named as such due to a defining loop of residues formed by a disulfide bond in the extracellular domain (ECD) of these receptors,\textsuperscript{235,240–242} which will be detailed later during the discussions about structure. In mammals, there are the excitatory cation-permeable: nicotinic acetylcholine receptors (nAChRs) and serotonin type 3 receptors (5-HT\textsubscript{3}Rs), and the inhibitory anion-permeable γ-aminobutyric acid type A receptors (GABA\textsubscript{A}Rs) and glycine receptors (GlyRs), as well as zinc-activated ion channels.\textsuperscript{238,243} There are also pLGICs found in bacteria and archaea,\textsuperscript{242,244} including a proton-gated ion channel found in cyanobacteria, *Gloeobacter violaceous* (GLIC)\textsuperscript{245,246} and a GABA/cysteamine-gated channel isolated from *Erwinia chrysanthemi* (ELIC).\textsuperscript{247–249}
though these do not have the characteristic disulfide bridge found in their eukaryotic relatives.

pLGICs are distributed throughout many parts of the body and by mediating fast neurotransmission, and thus, neural communication, they play an important role in sensory, motion, metabolism, memory and attention, cognition, sleep and wakefulness, pain, and mood.\textsuperscript{243,250,251} nAChR’s are located in the central and peripheral nervous system, in the muscles, and other tissues,\textsuperscript{252,253} and are the most well-understood in the family. Their primary agonists are acetylcholine and nicotine,\textsuperscript{252} and they are primarily responsible for the depolarization of muscle fiber by Na\textsuperscript{+}, K\textsuperscript{+} and Ca\textsuperscript{2+} ions,\textsuperscript{254,255} mediating post-synaptic neurotransmission in the autonomic ganglia, and triggering intracellular metabolism,\textsuperscript{252} due to their high permeability to calcium.\textsuperscript{256} Medically important, nAChRs are targets of pharmaceuticals used to treat several diseases, including, schizophrenia,\textsuperscript{257} ADHD,\textsuperscript{258} Alzheimer’s disease,\textsuperscript{259} Parkinson’s disease\textsuperscript{260} and Crohn’s disease,\textsuperscript{261} as well as others, and they are also targets of drugs treating smoking cessation.\textsuperscript{262,263}

5-HT\textsubscript{3}Rs, the only ionotropic member of 5-HTRs subtypes (the others are G-protein coupled receptors\textsuperscript{264}) are found distributed throughout the CNS, particularly in the hippocampal formation, amygdala, and cerebral cortex,\textsuperscript{265,266} but also are found in the gut.\textsuperscript{267,268} In the CNS, they mediate the release of many neurotransmitters such as dopamine, acetylcholine, GABA and glutamate,\textsuperscript{243,269,270} and in the gut they play an important role in gastrointestinal motility and survival and development of dopaminergic neurons.\textsuperscript{267,268} Due to their role in the gut, these receptors are drug targets for irritable bowel syndrome\textsuperscript{271,272} as well as treating emesis, particularly in cancer patients.\textsuperscript{273,274}
GABA\(_{A}\)Rs are located throughout the CNS, in many parts of the brain, but also are found in the retina, liver cells and immune cells.\(^{275,276}\) They are the predominant source of fast inhibitory neurotransmission,\(^{243,277}\) and contribute to the subsequent integration of neuronal excitation.\(^{276}\) Defects to the receptor can cause epilepsy.\(^{278}\) They also respond to a wide variety of drugs, including benzodiazepines and other anesthetics.\(^{275}\)

GlyRs, to contrast, are more localized and are found in specific regions, such as the brainstem, spinal cord, and substantia nigra, as well as the retina.\(^{236}\) They have also been found in non-neuronal tissues, such as immune cells, endothelial cells, hepatocytes, and renal cells, though the significance of non-neuronal expression is not yet understood, but could potentially play a role in disease.\(^{279}\) In CNS, GlyRs also play a role in inhibitory neurotransmission,\(^{236,237}\) and like GABA\(_{A}\)Rs are hot spots for many pharmaceutical targets, including anesthetics and cannabinoids,\(^{280,281}\) and thus, play a role in pain management,\(^{282-284}\) a topic that will be discussed in more detail in Section 2.3.3.

The stoichiometry of subunits in pLGICs varies widely depending on the protein type, the region where it is found, the developmental stage of the organism, the specific job intended for the pLGIC and influences how many ligands bind to the receptor to open the channel. A total of 17 subunits have been identified for nAChRs, \(\alpha1-10, \beta1-4, \gamma, \delta, \text{ and } \varepsilon,\) and with the exception of \(\alpha8\) which is found in avian receptors, all are found in mammals.\(^{235,243,285}\) In adult somatic neuromuscular junctions, the stoichiometry is \((\alpha1)_2\beta1\gamma\varepsilon,\) whereas, in embryonic and denervated skeletal muscles the stoichiometry is \((\alpha1)_2\beta1\gamma\delta.\(^{235}\) Also, some of the subunits do not form functional pentamers, \(\alpha5\) and \(\beta3,\) as homomers or as pairs,\(^{235,285}\) whereas; \(\alpha7\) and \(\alpha9\)-nAChRs can form functional homopentamers.\(^{243}\)
For 5-HT₃Rs there are five different subunits, A-E.²³⁵,²⁴³,²⁸⁶–²⁸⁸ Heteromeric subunits containing 5-HT₃A and B subunits are well-characterized, and it is known that the receptors containing A subunits form homomers,²⁸⁹ and that heteromers containing A and B differ in their allosteric regulation of several anesthetics and alcohols.²⁸⁹–²⁹¹ 5-HT₃C, D and E subunits, however; cannot form functional homomers, and must be combined with 5-HT₃A to function; heteromers containing these subunits are expressed in the submucosal plexus of the large intestine, the colon, and in the cell bodies of myenteric neurons and are important in gut mobility.²³⁵,²⁸⁶–²⁸⁸

GABAₐ receptors consist of many different subunits, α₁-6, β₁-3, γ₁-3, ρ₁-3, δ, ε, π and θ.²⁴³,²⁹²–²⁹⁴ Many GABAₐRs contain α, β, and γ subunits, with the α₁β₂γ₂ pentamer contributing to the largest percentage in the CNS, followed by α₂β₃γ₂ and α₃β₃γ₂.²³⁵ Functional homopentamers can be formed with ρ₁-3 subunits.²³⁵ In addition to these homopentamers, approximately ten other GABAₐRs have been characterized and well-studied. There are a multitude of specific examples of positive allosteric modulators for GABAₐRs, but some noteworthy ones are diazepam and flunitrazepam which are not selective on α₄ or α₆ subunits.²³⁵ Ligands that bind in the benzodiazepine binding site can distinguish α subunits from γ subunits.²⁹²

GlyR can be expressed as either a homopentamer of α subunits of which there are four varieties (α₁-4, though α₄ is not detected in humans), or as a combination of α and β subunits.²³⁵,²⁴³ The most commonly found stoichiometry in adult mammals is 2α subunits and 3β,²⁹⁵ and the homopentameric α₁ and α₃ are both well-characterized and have been crystallized.²⁹⁶,²⁹⁷ α₂ can also form homopentamers in neonates.²³⁵,²⁴³ The β subunit is
known to anchor the receptor to gephyrin through an amphipathic sequence found in the intracellular loop, and it also reduces channel conductance and alters pharmacology.\textsuperscript{298–300}

2.2. The Structure of pLGICs

Figure 3. Architectural Structure of a pLGIC. A. Represented by the schematic is the structural components in a single subunit of a pLGIC and its orientation in the membrane. A large N-terminal domain preceding M1, four TM domains labeled M1-M4 that have α-helical secondary structure. M2 is in green as it, forms a side of the pore. Also shown is a large intracelluar loop connecting M3 and M4, as well as other loops connecting the TM regions. B. Top down view of the TM domain. Four α-helices per subunit for a total of twenty. M2, in green, one from each of the five subunits, lines the pore of the channel. C. A typical stoichiometric assembly of GlyR, with two α-subunits, and three β-subunits.

In recent years, due to advances in crystallography and cyro-EM techniques, many structures of pLGICs have been solved. The first structures of a pLGIC was from x-ray crystallography experiments in 2001, by Brejc \textit{et al.},\textsuperscript{301} of a soluble acetylcholine binding protein (AChBP), and electron microscopy experiments by Cartaud \textit{et al.} and Miyazawa \textit{et al.},\textsuperscript{302,303} using nAChR found in the electric organ of \textit{Torpedo marmorata}. Nigel Unwin solved the full-length nAChR structure in 2005, also using the electric organ of the eel.\textsuperscript{304} The structure solved by Unwin and the structure of AChBP represent a closed structure and a ligand-bound structure, respectively.\textsuperscript{303,304} Later, in 2016, a human nAChR structure was determined via cyro-EM.\textsuperscript{305}
Recent successes in studying eukaryotic members of the pLGIC family were preceded and aided by crystallographic determinations of simpler prokaryotic members, ELIC\textsuperscript{243,306,307} and GLIC,\textsuperscript{308,309} which, aided by the assessment of nAChRs in different conformations, gave broader information on the family as a whole. The overall layout of these proteins is described as urn-like in appearance,\textsuperscript{296} with a wider ECD region composed of a mainly β-sheets, a narrower α-helical TMD region and a partially resolved intracellular region.\textsuperscript{238,304} Each pLGIC consists of a ring of five subunits of various types, that pack together to form a central pore, lined by M2 in each of the five subunits in which to allow ion flow in response to channel opening via neurotransmitter binding. The ECD also contains the ligand-binding domain (LBD) in which two to five ligands bind between subunits, depending upon receptor type.\textsuperscript{238,304,310}

Figure 4. pLGIC Conformational States. Typical conformational dynamics of a pLGIC involves three allosteric states: resting (blue) where ligand is not bound and the channel is closed. The ECD (blue ovals) is a “bloomed” conformation. The open, or active state, (green) is short-lived and responds to ligand binding by allowing the passage of ions through a now open pore. The desensitized state (pink), is non-conducting and longer lived. The receptor has higher affinity for its ligand in this state.
Successively, other eukaryotic pLGIC structures were solved, including: the 5-HT₃Rs from *Mus musculus*; a glutamate gated chloride channel (GluCl) from invertebrate *Caenorhabditis elegans*; human homomeric β3-GABA₁R; human homomeric α3-GlyR²⁹⁷ and homomeric α1-GlyR from zebrafish (*Danio rerio*). These structures represent various conformations of the receptors that contribute to the overall knowledge of pLGIC structure and dynamics.

A review article by Nemecz et al. characterized several crystal structures into their likely conformational states. The conformational stages a pLGIC goes through, are resting (apo-) when no ligand is bound and the central pore is closed, and thus non-conducting; a transient open or active state, as ligand binds and channel is gated open, and a long lived desensitized state, where ligand is more tightly bound, but the central pore is closed, and the receptor is once again non-conductive. It is important to note, as mentioned in Chapter 1, that crystal structures do not truly represent a native environment of membrane proteins, as they are crystallized in a de-lipidated environment (important especially as pLGICs are transmembrane proteins) with stabilizers such as antibodies or other molecules bound. They also do not represent the dynamics of a protein, as they are static representations, but they do provide structural information, and an understanding of a conformation that represents a low-energy well in a protein’s allosteric transitioning. These structures can provide a starting point to guide towards complete structural knowledge and thus future modulation of function via targeted therapeutics.

As noted, to date, the resting-like structures are: GLIC apo, GLIC LC, GluCla apo, α1-GlyR strychnine, α3-GlyR strychnine. The resting state structures are either apo- or bound with antagonist.
In all of these structures, the pore region shows a constriction point at a hydrophobic residue, which points towards a closed non-conducting conformation. This is due to the unfavorable interactions between the hydrophobic residue and water, which creates an impassable energy barrier for solvated ions.\textsuperscript{243}

The open or active-like structures are: \textit{α1}-GlyR glycine,\textsuperscript{243,296} \textit{α1}-GlyR glycine/ivermectin,\textsuperscript{243,296} GluCl\textsubscript{α} ivermectin,\textsuperscript{243,312} and GLIC pH 4.\textsuperscript{243,308,309} The open state structures are bound with agonist or a potentiator.\textsuperscript{243} The hydrophobic based closure seen in the resting state structures is no longer visible in open state structures, leading to a wider diameter pore. There is some debate to whether this is due to the existence of a hydrophobic “activation gate”, especially in terms of the \textit{α1}-GlyR glycine/ivermectin structure.\textsuperscript{243,296} According to Nemecz \textit{et al.},\textsuperscript{243} though, more work is necessary to further understand the dynamics of the cytoplasmic border in the TMD.

Only one of the current structures is thought to be desensitized-like, which is \textit{β3}-GABA\textsubscript{A}R.\textsuperscript{243,313} And, additionally, not all of the structures were able to be characterized with respect to conformation, notably, ELIC,\textsuperscript{243,306,307} 5-HT\textsubscript{3}A\textsuperscript{243,311} and POPC-GluCl\textsubscript{α}.\textsuperscript{243}

The transitions between conformational states, as well as a discussion of how lipids affect conformational states will be revisited further as it pertains to GlyR in Section 2.3 and in Chapter 3.

2.2.1. The Extracellular Domain (ECD)

The ECD of pLGICs is approximately 190-200 amino acids per subunit and has a \textit{β}-sandwich fold, composed of paired antiparallel inner and outer \textit{β}-sheets.\textsuperscript{301,317} The N-terminal domain is at the top of the protein and consists of a variable-length short \textit{α}-helix, that is absent in prokaryotic relatives.\textsuperscript{243} Ten \textit{β}-strands comprise the \textit{β}-sandwich
The ECD also contains the ligand-binding domain (LBD), in which the binding pocket is between neighboring subunits. The ligand or neurotransmitter binds in between subunits, and interacts with loops and β-strands from subunits on either side, and the C-loop plays a role in binding. The C-loop is thought to move in a hinge-like fashion and cap over the bound ligand, resulting in a long-lived higher affinity desensitized state. Though the C-loop is perhaps the most mobile region within the ECD, proposed to move as much as 7 Å in nAChRs, the other β-strands, particularly β1, β2, β3, part of β5, β6 and β8 are proposed to move as a result of ligand binding (See Figure 5a). The five individual subunits together of the ECD form a water-filled vestibule that provides the ions access to the channel.

2.2.1.1 The Ligand-Binding Domain (LBD) and the C-Loop
The ligand-binding domain (LBD) or the orthosteric site for the endogenous neurotransmitter is located about halfway from the top of the ECD to the membrane, in between two subunits.\(^{243}\) Once ligand binds, the receptor allosterically shifts to a conformation that stabilizes the channel towards the open state. This is the first step towards receptor activation, which is followed by coupling and finally, gating.\(^{238}\) Agonist binding occurs between subunits, at loops A-C on the principal subunit face and loops D-F on the complementary subunit face\(^{238,320–323}\) (see Figure 5b). As discussed previously, the number of binding sites depends on the particular subunit distribution and is diverse among the pLGIC family. Heteromeric receptors require usually at least two ligand-binding sites for full activation,\(^{238,321,324,325}\) but homomers seen in GlyR and GABA\(_A\)Rs, do not require total ligand binding for activation.\(^{238,322,323}\)

The agonist situates itself between the loops; these loops contain critical aromatic residues that allow for cation-\(\pi\) interactions, and hydrogen bonding, which stabilizes the interaction.\(^{317,326}\) This interaction primarily involves loop-B in nAChRs, 5-HT\(_3\)Rs, and GlyRs, and loop-A in GABA\(_A\)Rs.\(^{238,318,327}\) However, the other loops involved in the LBD, also have conserved aromatic residues that are crucial. Loop-A and the C-loop also have conserved aromatic residues, mostly Tyr.\(^{238,328–330}\) On the complementary subunit face, Loop-B and Loop-D contains a conserved tryptophan (Trp) residue in nAChRs and 5-HT\(_3\)Rs; this is replaced for a phenylalanine (Phe) or tyrosine (Tyr) in GlyRs and GABA\(_A\)Rs, respectively.\(^{238,317,318}\) It was later found that loops D and E are actually \(\beta\)-strands.\(^{243}\)

The conformational structural changes due to ligand binding are required to transmit the signal to the other parts of the receptor; the ECD-TMD interface and the
relatively distant gating region (>40 Å away from the LBD\textsuperscript{296,304}) in the TMD, in order to trigger the opening of the channel. The C-loop, as mentioned, is thought to play a role in this signal transmittance by its capping motion, which hinges over the bound ligand. This has been shown in experiments with AChBP, which highlighted that residues within the C-loop (consisting of β-strands 9 and 10, as well as a short loop between them, Figure 5b) may move as much as 7 Å.\textsuperscript{317} Due to its location, being in close proximity to the other loops participating in cation-π interactions with the neurotransmitter, it is a likely candidate for means of transmitting signal to the other loops and the inner β-sheets and activating the channel for gating.\textsuperscript{238,331} Loop-B is shown to rotate to close around the agonist and it also makes contacts with loops A and C, and loop-A also changes conformation due to ligand-binding.\textsuperscript{331–333} Notably, a conserved Asp residue in loop A also signifies its potential importance in ligand binding, and it may also make contacts with loop B.\textsuperscript{333–337} A study conducted by Miller et al.\textsuperscript{338} also shows loop-A in GlyRs can trigger activation through a connection in the hydrophobic core. Furthermore, many studies have mapped mutations in GABA\textsubscript{A}Rs and GlyRs to the loops in the LBD that lead to spontaneously opening channels.\textsuperscript{339–342}

Other studies confirm the importance of the C-loop in triggering activation; a study analyzing electron micrographs of nAChRs showed that in the absence of agonist, the C-loop is in the uncapped formation,\textsuperscript{343} and hydrogen-deuterium exchange mass spectrometry confirmed that in agonist-bound receptors, the C-loop is less accessible.\textsuperscript{344} The proposed movement of the C-loop is a twisting motion of the β9-β10 hairpin that shifts the base of β9 inwards; this likely diminishes the space between the inner β-sheet which forces the sheet to move as well.\textsuperscript{238,310} This movement of the inner β-sheet is thought to signal
downstream activation of the receptor. Similarly, it has also been proposed that movements due to ligand-binding occur early in the activation process and happen close to the binding site, yet still contribute to a downstream effect.\textsuperscript{345,346} These studies highlight the feedback communication between C-loop motion and channel gating.\textsuperscript{238,331}

Functional studies have shown that closure of the C-loop does not have to come from ligand binding, but there is seemingly a connection between how many ligands bind and the duration of the active or open state. Because of this, it is thought that C-loop capping might be a pre-activation step, or a “priming” step before the channel opens.\textsuperscript{238,347,348} This pre-activation step has also been referred to as a ‘flip’ state, in regards to studies performed on nAChRs and GlyRs using agonists of different binding affinities. Agonists that stabilize the ‘flip’ state with high affinity would cause a longer active state, meaning that if the pre-activation and/or ‘flip’ state is equivalent to the capped state, then an agonist can be characterized as full or partial depending on whether it is able to promote C-loop capping.\textsuperscript{238,349} Regardless of the studies conducted, this is still a mechanism that is not fully understood and may not be universally applied to all pLGICs. Though evidence does suggest that C-loop capping is the probable means of transmitting signal, it is not known whether or not this is the initiation step of receptor activation.\textsuperscript{238} It may, instead be a means of providing stabilization to allow other nearby loops to trigger downstream channel activation, as studies on GABA\textsubscript{A}Rs have suggested.\textsuperscript{350,351}

2.2.2 The ECD-TMD Interface

The interface between the ECD and TMD is responsible for communicating the signal of ligand-binding in the ECD to the TMD region where gating will occur. It also houses the Cys-loop, an important region that gives these receptors their name; often a
synonym of pLGIC. As stated previously, the prokaryotic relatives of pLGICs do not contain the disulfide bridge of the Cys-loop which is also known as the β6-β7 loop. Furthermore, this region is wholly important in terms of receptor malfunctions, as mutations to this area are the cause of diseases associated with pLGICs.

The interface consists primarily of three different loops; the β1-β2 loop, the M2-M3 loop, and the Cys-loop (Figure 6), but is also the region where the C-terminal tail, the pre-M1 linker and the loop linking β8 and β9 are located. The β1-β2 loop and the Cys-loop are situated at the base of the ECD; the β1-β2 loop is part of the inner β-sheet, and the Cys-loop connects the inner and outer β-sheets. In studies analyzing nAChRs, these loops form arcs which make contacts using their sidechains with the M2-M3 loop, which is facing upwards from the TMD, connecting M2 and M3. The β1-β2 loop is closest to the M2 channel. What is interesting is that in non-α subunits of nAChRs, these loops are situated approximately 2 to 3 Å away from the pore, which suggests the importance of these loops transmitting signal in active subunits.

The β1-β2 loop in GlyRs has also been shown to be important in signaling. In a single channel study analyzing the mutation A52S in the α1 subunit of glycine receptors in spasmodic rats, results showed a reduction in the affinity of a ‘flipped’ conformation and in apparent cooperativity. This suggests that this loop, like the C-loop discussed in the previous section, may affect ligand binding. Further studies of GlyRs, this time analyzing the M2-M3 loop have shown that different agonists (taurine and glycine) evoke almost identical motions of this loop, perhaps suggesting that this loop moves in a pre-determined fashion. Yet, other studies by Lynch et al. and others have shown that mutations
to the M2-M3 loop in GlyRs, particularly those that cause hyperekplexia (see Section 2.3), affect channel gating, and becomes more accessible during the open state.\textsuperscript{359,360}

In the prokaryotic relatives, ELIC and GLIC, a comparison of the open and closed states revealed movement of the $\beta1$-$\beta2$ loop downwards towards M2, and displacement outwards of both the Cys-loop ($\beta6$-$\beta7$ loop) and the M2-M3 loop.\textsuperscript{238,246,247,306,361} Studies targeting these ancient relatives suggest that they are not only employed in signaling mechanisms long before the origin of the animal nervous system, but that they are functioning in a similar mechanistic manner.\textsuperscript{244,362}

It is not clear whether or not the other loops that are present in the interface, the $\beta8$-$\beta9$ loop and the pre-M1 linker, contribute to the activation of the receptor or if they are communicating with the TMD. Though there are functional studies that suggest this.\textsuperscript{347,363,364}

Overall, the residues that are important for receptor activation interact through hydrophobic and electrostatic

Figure 6. ECD-TMD Interface. The ECD-TMD interface, inset in black, of a pLGIC (GlyR, PDB: 3jad). The primary loops are color coordinated; the $\beta1$-$\beta2$ loop is shown in blue, the Cys-loop is shown in green and the M2-M3 loop is colored orange. Also visible is the C-terminal tail in red. The TMD is also labeled, M1 – M4.
interactions, and the positions of these residues is quite reserved through pLGICs.\textsuperscript{238,303,365–368}

Pharmacologically significant, several diseases and disorders of pLGICs are due to mutations located in this interface. Congenital myasthenic syndromes are a group of disorders that cause decreased surface expression of muscle nAChRs, or slow or fast channel syndrome. Slow channel syndrome is marked by prolonged channel openings and fast channel syndrome by abbreviated openings.\textsuperscript{369,370} Many different mutations cause these syndromes, and more information can be found in Engel et al.\textsuperscript{371} Epilepsy, a common neurological disorder marked by seizures, is linked to malfunctions in GABA\textsubscript{A}Rs,\textsuperscript{372,373} and is caused by several different mutations, often dominant, affecting many of the subunits. Some notable ones that are located in the interface are $\alpha_1 A322D$, $\beta_3 Y302$ and $\gamma_2 R323$.\textsuperscript{374,375} These mutations, and the others that cause epilepsy disorders, often result in reduced whole cell current and decreased surface expression.\textsuperscript{373} Also, hyperekplexia is a disorder that results due to mutated $\alpha 1$ GlyR, most commonly, and will be discussed in Section 2.3.3.

2.2.3. The Transmembrane Domain (TMD)

The last step in the activation process is, of course, channel opening, which occurs in the TMD. There are four transmembrane $\alpha$-helices per subunit, contributing to a TMD region of twenty $\alpha$-helices per receptor. The $\alpha$-helices, M1 – M4, are positioned so that M2 from each subunit is lining the pore. M4 is the outermost helix. The M2 helix, regardless of subunit type, in a single receptor, is very homologous and is composed of identical or very similar residues.\textsuperscript{238} The helices are not aligned perfectly parallel, and are instead wider
at the extracellular end. To stabilize the channel, there are many contacts between the M2 helices of each subunit, as well as minor contacts between M1 and M3.\textsuperscript{238,310}

In order to provide clarity when comparing the residues within the pore to other members of the pLGIC family, a prime number system is used, defined by Miller in 1989, starting at the cytoplasmic end at $0'$ and moving up the helix to reach $20'$ at the extracellular end.\textsuperscript{376} In most of the channels, the middle of the helix, facing inwards contains hydrophobic residues; $9'$ Leu is consistent throughout the pLGIC members, and in the $13'$ position, Ala, Thr, or Val are present.\textsuperscript{238,376} The ends of the M2 helices are most often charged residues, often Glu and/or Asp.\textsuperscript{238,376}

The channel gate has been studied extensively in nAChRs, and was found to be in between the $9' – 14'$ residues, particularly at two places, at the $9'$ and $13'$ Leu.\textsuperscript{310,331} The hydrophobic residues form a hydrophobic bubble, approximately 6 Å in diameter which is sufficient to block hydrated Na$^+$ and K$^+$ ions from displacing further down the channel.\textsuperscript{304}

Analyses of the members of the pLGIC family, specifically nAChR, GABA$\textsubscript{A}$R and GlyR, using hydrophilic residue substitution have shown an increase in susceptibility for the channel to open in response to agonist.\textsuperscript{377–381} This is consistent with hydrophobicity providing stabilization for a closed channel.\textsuperscript{304,310} Other studies, however, suggest that the channel gate is on the cytoplasmic side of the channel, and not the midpoint. These studies used cysteine-modifying reagents in the absence of agonist, as well as picrotoxin (PTX).\textsuperscript{382–385} There are also evidence that there is two distinct gates; this dual-gate model proposes that there is a gate for the closed to open transition, and also a gate for desensitization.\textsuperscript{386} Several studies analyzing GlyR desensitization have proposed the gate is near the cytoplasmic end,\textsuperscript{354,357,386,387} and other studies have shown that picrotoxin (PTX) blocks
desensitization\textsuperscript{386} and that the M3-M4 loop plays a role in desensitization.\textsuperscript{388} This controversy is still not completely resolved and further work needs to be conducted to resolve this issue.

Much of the research focusing on channel opening, however; agrees that pLGICs open by a tilting of M2 helices.\textsuperscript{304,308,309,389–393} The channel has to widen for opening to occur, and this is most likely done through M2 moving back towards M1 and M3, which is thought possible because M2 only makes limited contacts to these other two helices, and is also partially separated from them due to a space between them that is thought to be filled with water and cholesterol.\textsuperscript{360,394–396} Furthermore, since most of the stabilization comes from M2-M2 interactions, C-loop capping due to ligand binding is thought to provide enough destabilization to the closed gate to allow widening of the hydrophobic bubble and opening the channel.\textsuperscript{238,397–399}

In addition to the TMD region being important in regards to channel gating, the region also plays a huge part in the binding of allosteric modulators. It is modulated by several compounds, such as alcohols and anesthetics, endogenous neurosteroids and zinc ions, which will be further discussed in Section 2.3.1.

2.2.4. The Intracellular Domain (ICD)

The intracellular domain contains both the M1-M2 loop and the M3-M4 loop which is in some ways the hallmark of each individual channel, as this region varies widely between the family members in length and sequence variation.\textsuperscript{388} The M1-M2 loop, in GlyRs, is also important in terms of hyperekplexia, a disorder linked to the receptor.\textsuperscript{400} A commonly studied mutation P250T, located in the M1-M2 loop, is a dominant mutation that causes the disorder,\textsuperscript{357,358,401} and will be discussed in the next section.
The M3-M4 loop, however, is much more diverse in terms of its importance. This intracellular loop has many different functions and modulations, including mediating specific interactions with intracellular binding partners,\(^{388}\) post-translational modifications,\(^ {388,402}\) channel structural integrity,\(^ {403}\) gating,\(^ {404}\) desensitization,\(^ {405-407}\) modulation by positive allosteric ligands\(^ {408}\) and receptor surface expression.\(^ {402}\) Studies of pLGICs have shown that if the loop is severely truncated, or if the receptor is lacking the loop entirely, this leads to non-functional receptors.\(^ {388,409-411}\) The M3-M4 loop is also a defining characteristic of the eukaryotic pLGIC members, as it is very short in prokaryotic ELIC and GLIC.\(^ {308,412,413}\) Because of the variation of size and amino acid sequence in each eukaryotic loop, and based on chimeric studies that showed full pentamers could form with a loop that was not subunit-specific, it is known that the loop determines the specific ion channel properties.\(^ {388,414-416}\) The information available on this defining loop comes from experimental studies and not from structural studies, as members of the pLGIC family can only crystallized successfully by truncating this loop and replacing it with a short peptide.\(^ {296,297,331,417,418}\) This lack of structural information provides a strong rationale into the usefulness and importance of the MS studies undertaken in this work. Furthermore, because this loop is so variable, this section will primarily focus on discussing the details of the M3-M4 loop in the context of only glycine receptors.

The M3-M4 loop of \(\alpha1\)-GlyR is nearly 70 residues, which is about 20% of the mass of a single subunit. It is not as large as the M3-M4 loop in nACHRs, but it is comparable in size to the \(\text{GABA}_\text{A}\)Rs. There are several motifs within the loop of considerable importance, such as the stretch of basic residues near M3. Mutations of the basic residues in this region, conducted by Carland \textit{et al.},\(^ {419}\) lead to a non-functional receptor, which
suggested a role for this region to draw anions through the pore. This region is thought to be important for insertion of M3 into the membrane, and also contains the nuclear localization signal (NLS), which interact with intracellular proteins karyopherins α3/α4 and results in the transport of GlyRs into the nucleus. This stretch, particularly RFRKK along with a di-Lys region later in the loop, binds to the G-protein subunit Gβγ. This interaction has been shown to enhance glycine induced chloride current. The basic stretch is also important in terms of binding allosteric modulators, such as ethanol and endocannabinoids. Another region, located next to the basic stretch, residues 358–372, in α3-GlyRs has been shown to play an important role in desensitization, and also folding stability of the region. There are two other motifs seen in GlyRs, a polyproline region, PPPAPS in α1-GlyRs, and a poly-asparagine region, NNSNTTN. The poly-Pro region is thought to have helical structure, specifically as a poly-proline helix type II (PPII) and has been analyzed via CD-spectroscopy. These PPII helices form the SH3 (Src homology 3 domain) consensus sequence, which is a common binding motif found in many proteins. Syndapin 1 is an intracellular protein that binds GlyRs via their SH3 domains and plays a role in trafficking and possibly even anchoring. Currently, the understanding of the basis of the poly-asparagine domain in GlyRs is severely lacking.

As stated previously, structural information of this region does not come from high resolution crystal structures, but there is some information regarding the secondary structure of certain regions of this loop. As described by Unwin et al., most of the secondary structural features are located at the ends of the loops, leaving the middle of the loop unstructured or disordered. There is evidence of α-helices at the beginning and end of
the M3-M4 loop in pLGICs, including nAChRs and 5-HT3Rs, and this has been proposed to be true in GlyRs as well. CD-spectroscopy studies showed evidence of α-helices near M3 and M4. And there is evidence of a PPII helix, mentioned above, also identified via CD-spectroscopy located between residues 365 and 371 of α1-GlyRs.

The M3-M4 loop is also targeted by many intracellular binding partners. An important one to note is gephyrin. Gephyrin is a peripheral membrane protein that consists of three domains; an N-terminal G domain, a C-terminal E domain and an unstructured domain that links the two together. The E domain interacts with the β-subunit of GlyRs and its function is primarily synaptic anchoring. This was proven due to its copurification with both GlyR and tubulin, studies using gephyrin antisense oligonucleotides that stopped formation of GlyR in the membrane, and via treating rat spinal neurons with strychnine which inhibited GlyR and gephyrin clusters at postsynaptic membranes which suggests that activation of the receptor is essential for the proteins to cluster. Other roles for gephyrin is mediating synaptic inhibition and long-term potentiation. Additional intracellular binding partners were already previously mentioned including G-protein subunits Gβγ, karyopherins α3/α4 and syndapin 1. Vesicular trafficking proteins, Sec8 and vesicular presynaptic marker VGLUT1, also bind α3-GlyR intracellularly, and control its trafficking towards presynaptic terminals. A study by Del Pino et al. used mass spectrometry and identified two more binding proteins that target the β-subunit of GlyRs, Vps35 and neurobeachin. Also, a GTP/GDP exchange factor, CB and neuroligin 2 are known to play a role in formation and function of inhibitory postsynapses via binding to GABAγRs, but this interaction has not yet been
shown in GlyRs. It is clear that while it is known that these intracellular binding proteins are essential to GlyR function, there is still a lack of clarity on the mechanisms behind these interactions. There is also a possibility that more interactions exist and will be identified in the future.

Post-translational modifications (PTMs) also affect the M3-M4 loop, particularly ubiquitination and phosphorylation. Ubiquitin is attached to Lys residues and signals that the protein is ready for degradation; in GlyR, there is approximately ten Lys residues in the M3-M4 loop and three of them are ubiquitinated to allow for protein internalization and degradation. There are recessive mutations that result in hyperekplexia that lead to accumulation of the receptor that then influences the degradation of the receptor through the ubiquitination process. Phosphorylation of GlyR M3-M4 loop results from protein kinase A or C (PKA, PKC), depending on the subtype. Residue S391 in α1-GlyRs is phosphorylated by PKC, which is in a similar location to phosphorylation sites in other pLGICs. β-GlyRs are also phosphorylated by PKC at S403. α3-GlyRs are phosphorylated by PKA at position S346; this serine is not conserved in α1-GlyRs and thus, PKA cannot modulate α1 subunits. The role phosphorylation plays on GlyRs has been studied extensively. Research has shown phosphorylation of α1 receptors mediates interaction with intracellular proteins, such as gephyrin. A study by Huang et al. analyzed PKC after being activated by a myristate and found that this led to increased internalization of GlyR, except when a di-Leu region (L314/L315) of the M3-M4 loop was mutated. Another study examined the effects of phosphorylation of GlyRs on ethanol potentiation. Also, as described in Section 2.3.3.2., chronic pain is linked to GlyRs, and the phosphorylation of α3-GlyRs by PKA plays an important role in pain sensitization.
Other studies show that kinases effect the desensitization of the receptors and also the maximal chloride influx.\textsuperscript{452,453}

In summary, the large intracellular loop in pLGICs serves to distinguish the members from one other, contributing to their specific functions. It plays several roles that affect gating, desensitization, trafficking, degradation, among others. Because there is a lack of structural information on this region, CX-MS studies can lead towards a better understanding of dynamics of this region and pinpoint what specific areas of the loop are interacting with the membrane or are in close proximity to the channel.

2.3 The Structure-Function Relationship of GlyR

The glycine receptor is found in localized regions, such as the retina, the spinal cord and brain stem, and is responsible for facilitating inhibitory neurotransmission. Similar to other members of its family, GlyR is a hot-spot for small molecule interactions, many of them have a potentiating effect on the receptor. GlyR is also modified by post-translation modification (PTM) that leads to further diversity and modulations of the receptor. Mutations or malfunctions of the receptor contributes to diseases and disorders, such as hyperekplexia, chronic pain as well as others.

The link between protein structure and protein function is important towards developing a full understanding of the structure and dynamics of a particular protein. Understanding the interactions between small molecules, as well as the modifications that happen naturally to the receptor \textit{in vivo}, also lead towards this greater volume of knowledge that helps define the importance of a protein of interest.

More importantly, however; obtaining more complete structural knowledge can lead to the development of novel therapeutics that can specifically target the receptor and
treat these channelopathies. Various conditions linked to a protein can be treated more effectively once there is a more complete understanding of protein dynamics and structure. Designing therapeutics that are specific to a protein of interest can also help mitigate off-target binding which causes many of the side effects seen in pharmaceuticals.

2.3.1 Small Molecule Targets of GlyR

The glycine receptor has several allosteric and orthosteric modulators that bind the receptor and affect its function. Strychnine, a plant alkaloid found most commonly in the seeds of the *Strychnos nux-vomica* tree, is a competitive antagonist to GlyR, and due to its toxicity is used as a rodenticide. It can be fatal to many mammals, even humans, if a high enough dosage is ingested. Clinical signs and symptoms of strychnine poisoning appear first as nervousness, muscle twitching, and stiff neck, before progressing to convulsions in skeletal muscles. Death results due to asphyxia caused by paralysis of respiratory muscles. The effect on GlyR is the blockage of chloride current, as strychnine binds in the LBD of the receptor.\(^{454}\) Also, strychnine binds much tighter to GlyR than does its ligand, glycine, as represented by the dissociation constants (\(K_d\)), 2 – 4 nM and 25 \(\mu\)M, respectively.\(^{455}\)

Because of this extremely tight binding, strychnine can be used experimentally to distinguish between it and other members of the pLGIC family, particularly between GlyR and GABA\(_A\)R;\(^{236}\) though studies have shown that strychnine can also bind, with much lower affinity, to nAChRs.\(^{456}\)

Another toxin derived from plants is the convulsant alkaloid picrotoxin (PTX) which is used to discriminate between glycinergic and GABAergic currents.\(^{236}\) GABA\(_A\)Rs are highly sensitive to the toxin (at a range of about 1 – 10 \(\mu\)M), whereas; GlyRs are much less sensitive.\(^{236}\) Furthermore, heteromeric \(\alpha\beta\)-GlyRs are less sensitive than homomeric \(\alpha\)-
GlyRs. PTX, therefore; is an important pharmacological tool to distinguish if there is a heteromeric or homomeric GlyR present. Studies have also shown PTX can be used to inhibit GlyR, by blocking chloride current.458

Ivermectin is a member of the avermectin family of compounds and is a macrocyclic lactone.281,459,460 It is widely used to treat parasitic worm infections in ruminant animals, and pinworms, threadworms and whipworms in mammals, including humans.459,460 It is also an effective treatment against river blindness.461,462 There is ongoing research that shows ivermectin has antiviral properties against positive-sense single-stranded RNA viruses, such as SARS-CoV-2.463–465 Relevant to this work, ivermectin, at low concentrations, is a potentiator of GlyR, that binds in the TMD, and does not allow desensitization of the receptor, and at higher concentrations (≥0.03 μM),236,466 ivermectin is an activator of GlyR.236 Using ivermectin at lower concentrations can effectively trap the receptor in the open state, and thus, allows for study of the open conformation.

Other common small molecules that interact with GlyR are ethanol and caffeine. Ethanol was first shown to cause glycine sensitivity in spinal neurons by Celentano et al. in 1988.236,467 Studies by Lobo et al.468 showed that channel gating allows accessibility to M1, M2 and M3, which are targeted by alcohols. Other studies showed that ethanol, as well as other long chain alcohols, potentiate the receptor.469–472 Caffeine, on the other hand, is a structural analog of strychnine, and thus an orthosteric modulator, targeting the LBD of GlyR.473 It has been shown that caffeine inhibits current in both GlyRs and GABAARs.473,474
Medically relevant compounds that affect GlyR function are anesthetics, benzodiazepines and members of the cannabinoid family. Anesthetics have traditionally been thought to be non-specific drugs that acted by disordering lipid bilayers, but since the 1990s, research has shown that ion channels are a specific target of these drugs. Like alcohols, anesthetics target GlyR in the TMD; specifically I299 in M1, S267 in M2 and A288 in M3. Volatile anesthetics, such as isoflurane, nitrous oxide, propofol, thiopentone, pentobarbitone, ketamine, among others, have potentiating effects on the receptor, as well as on other members of the pLGIC family.

Benzodiazepines are used to treat channelopathies present in GlyR and GABA<sub>A</sub>R such as hyperekplexia and epileptic disorders. Studies analyzing the effects of several benzodiazepines on GABA<sub>A</sub>R have shown that these drugs target the N-terminal domain of the receptors, and the effects of diazepam in particular, are diminished if a H101R mutation occurs. GlyR studies, analyzing the effects of diazepam, flunitrazepam, flurazepam, among others, have also shown that these drugs target the LBD, specifically involving the interactions made by strychnine. By binding in the LBD, and thus mimicking glycine, they are able to produce anti-anxiety, anti-convulsant and muscle-relaxant effects for which they are prescribed.

Cannabis, which contains a psychoactive component Δ<sup>9</sup>-tetrahydrocannabinol (THC), is often prescribed and used to treat chronic pain. Though most of the effects of THC and other members of the cannabinoid family are mediated through the cannabinoid type 1 receptor (CB1), some of its effects are due to binding to GlyR, specifically the analgesic effects. THC, as with many other important molecules,
potentiates the receptor and provides an anesthetic effect.\textsuperscript{497,499,500} Both $\alpha_1$-GlyR and $\alpha_3$-GlyR are linked to cannabis-induced analgesia.\textsuperscript{280,447,501–504} These cannabinoids interact with GlyR in the TMD, specifically S296 and S307 in $\alpha_1$ and $\alpha_3$-GlyR.\textsuperscript{500}

In summary, GlyR is a target for a wide variety of small molecular compounds, many of them with important medicinal and pharmacological relevance. Most of them are commonly used to treat disorders such as hyperekplexia, seizure disorders and pain. Some, caffeine, ethanol, and THC are used for daily consumption.

2.3.2. Post-Translational Modifications (PTMs) and Regulatory Aspects of GlyR

Post-translational modifications (PTM) that target the M3-M4 loop (phosphorylation and ubiquitination) have already been discussed in Section 2.2.4., but there are others that occur in other regions. N-glycosylation, which is the attachment of glycans to an Asn residue, occurs at N-X-S/T motifs. $\alpha_1$-GlyRs have one consensus sequence, \textsuperscript{38}NVS\textsuperscript{40} in the ECD, and $\beta$-GlyRs have two N-glycosylation sites, at position 32 and 220. $\alpha_2$-GlyRs and $\alpha_3$-GlyRs also have similar consensus sequences.\textsuperscript{505} These N-glycosylation sites play a role in GlyR trafficking, particularly the exit from the endoplasmic reticulum (ER), as studies have shown that when this consensus sequence is mutated, GlyR cannot exit.\textsuperscript{506,507}

There are also several sequences located in the ECD known as assembly boxes that were determined through site-directed mutagenesis to promote pentameric assembly of the receptors. In $\alpha_1$-GlyRs, these are \textsuperscript{35}PPVNVSC\textsuperscript{41}, \textsuperscript{74}AYNEYPDD\textsuperscript{81}, \textsuperscript{90}LDSI\textsuperscript{93}, and \textsuperscript{125}NVLY\textsuperscript{128}.\textsuperscript{506,508} There are also sequences within M1, M3 and M4 that also promote pentameric assembly.\textsuperscript{409}
Similar to the discussion of PTMs occurring in the M3-M4 loop, there was also mention of intracellular binding partners in said loop. However, extracellular proteins do interact with the ECD of GlyRs as well, in particular, calnexin. Calnexin is a chaperone protein and a member of the lectin family that binds to mono-glycosylated and N-linked core glycans.\textsuperscript{506,509} Coimmunoprecipitation studies showed the interaction between these two proteins, and mutations to the ECD loop β2-3 and to M4 showed increased interaction compared to WT GlyR.\textsuperscript{411}

Another important regulatory molecule that interacts with GlyR outside of its M3-M4 loop is zinc. Zinc is thought to reach a concentration of $>100 \mu M$ during synaptic stimulation,\textsuperscript{236} and is able modulate many post-synaptic and pre-synaptic ion channels, including GlyR.\textsuperscript{236,510} High concentrations of zinc ($> 10 \mu M$) inhibit glycinergic current, whereas; lower concentrations of zinc (0.1 – 10 μM) potentiate the receptor.\textsuperscript{236,511–514} $\text{Zn}^{2+}$ ions can be coordinated by nitrogen, sulfur or oxygen atoms that are in the side chains of histidine, cysteine, aspartate and glutamate residues.\textsuperscript{515} Studies have shown that in GlyRs, $\text{Zn}^{2+}$ ions coordinate to D80,\textsuperscript{516,517} studied extensively as D80A is a prominent hyperekplexia mutation.\textsuperscript{401,516} Zinc coordinating to this site also increases ethanol potentiation.\textsuperscript{517} $\text{Zn}^{2+}$ ions also coordinate with several His residues, including H107, H109, H210, H215, and H419,\textsuperscript{518,519} though most prominently at H107 and H109, as studies have shown that the latter His residues fail to prevent inhibition of the receptor when H107 and H109 were mutated.\textsuperscript{518}

There are other molecules and other regulators that modulate the receptor. Calcium ions cause elevation in the magnitude of glycinergic current that may be mediated by calcium activation of calmodulin-dependent protein kinase II and calcineurin.\textsuperscript{236,520–522}
Also, pH appears to play a role in increasing the glycine EC$_{50}$, shown by studies that mutated various ECD residues, H109A, T112A and T112F. 236,523

2.3.3. The Channelopathies of GlyR

2.3.3.1. Hyperekplexia

Hereditary hyperekplexia, also known as familial/human startle disease, is characterized by extreme startle responses due to unexpected auditory or tactile stimulations. 400,401 It is further marked by prolonged muscle hypertonia, mostly in infancy, which can lead to infant death, due to apnea or asphyxia. Furthermore, gross motor development is delayed in the first two to three years of life. 400 As patients age, hypertonia is still present, though not as severe as during infancy, and injuries due to falling are most prominent. There can also be delayed speech acquisition and intellectual disability. 401

Onset of the startle response results in forceful closure of eyes, rising of bent arms over the head and a stiffness or flexion of the neck, trunk, hips as well as elbows and knees. 400,401 Consciousness is maintained during these episodes, which distinguishes this disorder from epileptic seizures seen due to mutation and malfunction of GABA$_{A}$Rs. 373,401 The disorder is rare, affecting about 1 in 40,000 people in the United States. 524

Benzodiazepines are most often used to treat this disorder, including clonazepam, and diazepam, 525 divalproex sodium, valproic acid, and the anticonvulsant levetiracetam. 525,526 These drugs work effectively to treat hyperekplexia, 527,528 however; side effects of these drugs vary widely, and include, increased concentration and energy, sleeplessness, loss of control, aggression, and suicidal thoughts. 529–531

Hyperekplexia is primarily caused due to mutations in $\alpha$ subunits in GlyRs and in the presynaptic glycine transporter, GlyT2. 400,401,532 As of 2014, a study by Bode and Lynch
reported 55 mutations in α1-GlyR that cause the disorder. There are also mutations of the β-subunit, but these are not as common or as numerous. The mutations are located throughout the receptor. There are many mutations found in the ECD and TMD regions (see Bode and Lynch), and there are some notable ones in the interface, which along with those located in M2 are most often dominant mutations. There are also hyperekplexia causing mutations in the ICD; one in the short M1-M2 linker, P250T, which has been studied extensively, and several in the M3-M4 loop; R316X, G342S, E375X, D388A and R392H, which are important to this work as structure of this region is not well-resolved.

The dominant mutations (located primarily in M2, the ECD-TMD interface as well as P250T) cause functional issues, such as spontaneous activation, reduced single channel conductance, impaired channel gating, an overall lower maximal current, a decrease in ligand binding efficacy and enhanced desensitization (caused primarily by P250T). Some examples of spontaneous activation are: Q226E, located at the top of M1 which also causes reductions in surface expression; V280M in the M2-M3 loop which causes drastic enhancement of glycine sensitivity; this may be due to destabilization of the closed channel, and R414H, which is located in the M4 domain and the mechanism to which it may cause activation is currently unknown.

Impaired channel gating is most often caused by two common dominant mutations, R271Q and R271L. Both cause a significant decrease in glycine sensitivity, but do not cause reduced surface expression. The most likely cause of the decrease in channel gating and conductance is due to the loss of the positive Arg residue.
The recessive-type mutations, which are located throughout the receptor are responsible for affecting receptor biogenesis, surface expression, trafficking and receptor stability. Reduced surface expression is caused by recessive mutations that are located primarily in the TMD; S231R and I244N in M1, R252H in M2 and R392H in M4. However, there are also mutants located in the M3-M4 loop (see above) that also contribute to reduced surface expression as well as decreased receptor stability. Two of them, R316X and E375X lead to truncated receptors, and results in significantly decreased surface expression due to protein misfolding and abnormal receptor trafficking. A similar truncation in the M3-M4 loop is also seen in γ2 subunits in GABA_A Rs, which leads to epileptic disorders.

According to Bode and Lynch, these mutations, either dominant or recessive can be grouped into three categories: dominant mutations located in or around M2 that do not impair expression but do cause misfunction of the receptor, recessive missense mutations located throughout the receptor that cause a decrease in surface expression, and recessive nonsense and deletion or frameshift mutations that result in disruption of full length receptor formation.

Despite understanding much about the many mutants and categories of mutant that cause this disease, and regardless of the advancements modern medicine has made in treating this disease, there is still work to be done. A complete structure of this receptor, particularly focusing on unresolved areas, such as the M3-M4 loop, will enhance the ability of targeted drug design, leading to fewer negative side effects and a more normal life for those that have hyperekplexia.

2.3.3.2. Chronic Pain
Chronic pain is a complex disorder that affects approximately 50 million Americans, according to a 2016 study.\textsuperscript{542} It is found widespread over many conditions, prevalently featured in diagnoses such as fibromyalgia,\textsuperscript{543,544} vulvodynia,\textsuperscript{545–547} trigeminal neuralgia,\textsuperscript{548,549} irritable bowel disorder,\textsuperscript{550,551} arthritis,\textsuperscript{552,553} nerve damage,\textsuperscript{554,555} and many others. It is also often associated with cancer patients.\textsuperscript{556,557} Symptoms, too, vary widely, depending on the area most affected; but often are joint pain, burning pain, and fatigue; depression and anxiety are comorbidly linked as well.\textsuperscript{546,558–560}

Pain, because it is often a part of the innate immune response to a plethora of different underlying conditions – tissue damage, infections, inflammatory disease, cancers and others\textsuperscript{561–564} – chronic pain syndromes are often difficult to diagnose, and there is stigma associated with the disease as the chronic pain patients suffer may not always be believed.\textsuperscript{565–567} Treatment, traditionally, is limited to over-the-counter treatment such as ibuprofen, naproxen and acetaminophen\textsuperscript{560} or prescription drugs such as opioid painkillers,\textsuperscript{566} which are stronger, and thus more effective. However, these latter drugs are habit-forming, and can lead to dangerous overdose.\textsuperscript{568–570} Other less traditional treatment options recently seeing more widespread use, include acupuncture,\textsuperscript{571–573} low-level laser therapy\textsuperscript{574} and THC and CBD as inhalants, oils or pills.\textsuperscript{282,492,493,575–577}

The link between chronic pain and GlyR exists through its connection to the cannabinoid family potentiating the receptor.\textsuperscript{280,447,501–504} As noted in Section 2.3.1., most of the effects of the cannabinoid family are mediated through the cannabinoid type 1 receptor (CB1),\textsuperscript{282,494–496} but evidence shows that some of the effects are also mediated through GlyR, specifically the analgesic effects.\textsuperscript{283,497,498,578,579} Furthermore, studies by Goss \textit{et al.} in 2011, and Harvey \textit{et al.} in 2004, also show a unique link between GlyR and
pain. In the former study, a gene therapy was established that used a herpes simplex virus vector to express GlyR to manage pain in nociceptive neurons.284 In the study by Harvey et al., α3-GlyRs were inhibited by prostaglandin-E2 (PGE) induced phosphorylation and results showed that this mechanism is important in inflammatory-based pain sensitization.388,447 This study also resulted in a pharmacological study that further delved into α3-GlyRs as a potential therapeutic target.580

The implications of this link between GlyR and pain are exciting, and given the shifting of opinions on alternative therapy for pain management, this could lead to better therapies and a more normal way of life for so many that struggle with chronic pain.

2.3.3.3. Other Implicated Diseases and Disorders

Though hyperekplexia and chronic pain are the most well-studied in the context of GlyRs, other diseases and disorders have been linked to the receptor, including, stiff person syndrome (SPS) and progressive encephalomyelitis with rigidity and myoclonus (PERM), autism spectrum disorders, panic and anxiety disorders and epileptic disorders.

SPS and PERM are glutamic acid decarboxylase (GAD)-positive disorders581,582 that share similarities with hyperekplexia in that these disorders also have muscle rigidity and spasms as symptoms.402,582–584 SPS also is characterized with heightened sensitivity to touch and sound.581,582 Both conditions are rare, but PERM is a much more severe diagnosis as it is also marked by brain stem and autonomic dysfunction and breathing problems.583,585 The most effective treatment for both SPS and PERM is immunotherapy; Rituximab, a drug used in immunotherapy has been used to treat both disorders.586–588 Though effective, patient relapses are common, and thus, these disorders are still incurable.402,589
Autism spectrum disorders (ASD) are a group of neurological disorders that affect both behavior and communication and usually the first signs of development are within the first two years of life. Though the underlying mechanism of the disorder is unknown, it is thought that there is a multifaceted response with various genes influenced by the environment to cause the disorder. One of the genes affected is GLRA2, a gene that codes for the α2 subunit of GlyRs. A missense mutation, R153Q, reported in patients with ASD caused a decrease in surface expression of GlyRs. This research suggests that there is a link between social and cognitive impairment and GlyRs, and also that α2-GlyR may play a role in learning and memory. Another study has connected a mutation in a protein neuroligin-3, R451C, to increased inhibitory neurotransmission. This study proposes that this gain-of-function mutation caused impaired social interactions, a hallmark of ASD, and also unexpectedly showed increased inhibitory neurotransmission without affecting excitatory transmission. The link to GlyR is unknown at this time.

Panic disorders, including anxiety disorders and agoraphobia have also been linked to GlyRs. One study found allelic variations in GLRB, the gene encoding for the β subunit of the glycine receptor, in patients with panic disorders. Another study looked at GlyRs in the hippocampus and found that decreased excitatory inhibition due to gain-of-function of GlyR lead to an anxiety phenotype in mice.

Epileptic disorders, commonly linked to GABAₐRs, can also be found in GlyRs. Temporal lobe epilepsy, in particular, is linked to specific glycine receptors with high affinity located in the hippocampus. These gain-of-function glycine receptors contribute to increased glycinergic transmission and cause epileptic seizures. Glycine receptors
located in the supramedullary region of the brain are also thought to contribute to hyperexcitability disorders, such as epilepsy.\textsuperscript{595}

A study by van den Eynden \textit{et al.}\textsuperscript{279} in 2009 researched the non-neuronal locations of GlyRs, and showed the presence of the receptor in immune cells, endothelial cells, renal cells and hepatocytes. Though it is not yet known why GlyRs are found here, it does suggest that this receptor may prove to be a crucial piece of other diseases and disorders in the future.

2.4 Summary

The glycine receptor is a pLGIC which plays an important biological role in inhibitory neurotransmission, and is strongly linked to pain and analgesia. The structure of the receptor is lacking details of certain areas important in surface expression, trafficking, and gating of the receptor. There is also a lack of understanding of protein dynamics associated with channel gating and protein mobility. By using CX-MS (detailed in Chapter 1) and applying it to structure-function studies using a developed network of single, site-specific cysteine mutations (see Chapter 3), the body of knowledge surrounding GlyR can grow and be applied to the future development of more specific therapeutics.

Overexpressing GlyR with designated single Cys mutants, purifying and reconstituting the receptor into a native-like lipid environment, enriching the receptor into its allosteric states and then crosslinking it with a heterobifunctional crosslinker will provide information on structural details in the vicinity of mutants chosen. This information, when placed in the context of a network of various protein-protein interactions can refine the structural details that are missing in crystal structures as well as give information about protein dynamics as the receptor allosterically shifts between its states.
CHAPTER 3. MATERIALS AND METHODS

3.1 Introduction

This chapter will discuss the materials and methods outlined for both Chapter 4 (resting state) and Chapter 5 (open and desensitized states). Though much of what is done for resting state mutants is also done for open and desensitized state mutants is identical, any changes will be noted.

3.1.1 Preparation and Overexpression of Mutant GlyR using the Baculovirus Overexpression System

The following methodologies as described below contribute to the overexpression of a single site-specific cysteine mutated GlyR from our network into *Spodoptera frugiperda* (Sf9) cells.

3.1.1.1. Introduction of Single Site-Specific Cysteine Mutations via Mutagenic Primer Design

Primers were designed according to site-directed mutagenesis protocols (purchased from Operon). Both mutagenic primers (forward and reverse) were designed to contain the desired mutation and to allow for annealing to the same sequence on opposite sides on the plasmid (see Table 1). Each primer was designed to be 25-45 base pairs in length, have a minimum guanosine/cytosine (GC) content of 40%, terminate with a G or a C and have a melting temperature of approximately 78°C. The following equation was used to determine each primer’s melting temperature:

\[ Tm = 81.5 +0.41(\%GC) – (675/N) - \%mismatch, \]
where N is the length of the primer in base pairs (Quikchange Lightning Site-Directed Mutagenesis Kit, Instruction Manual).

Table 1. Forward Oligo Sequences. The forward oligo sequences for each mutation in the network, M287C, K116C and K206C are detailed in the table below.

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Sequence 5’ to 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>M287C</td>
<td>GCCATTGACATTTGGTGCGGAGTTGCCCTGCTCTTTGTG</td>
</tr>
<tr>
<td>K116C</td>
<td>ATCACCACAGACAACTGCTTGCTAAGGATC</td>
</tr>
<tr>
<td>K206C</td>
<td>TACAACACAGGTTGCGCCACCTGCATTGAGGCC</td>
</tr>
</tbody>
</table>

3.1.1.2. Mutant Strand Synthesis Reaction (PCR Cycling) and Digestion of Amplified Products

In thin-walled PCR tubes, the following reaction mixtures were prepared in two tubes: 10x reaction buffer (5 μL), pWhitescript 4.5 kb control primer (5 μL), oligo control primers 1 and 2 (1.25 μL) each, dNTP mix (1 μL), Quik solution reagent (1.5 μL), and deionized (dI) H₂O to bring total volume to 50 μL. In the positive control tube, Quikchange Lightning Enzyme was added, and forward and reverse primers were substituted for the control primers. The temperatures to allow denaturation, annealing and elongation were optimized by the Cascio Lab (Veeramachaneni) and are as follows: 95°C (2 cycles, 30 sec), 55°C (1 cycle, 1 min), 68°C (1 cycle, 7 min). The cycle was repeated 16 times. The desired mutation was then filtered out using Dpn 1 restriction enzyme (2 μL) to digest the non-mutated parental DNA.

3.1.1.3. Transformation of XL10-Gold Ultracompetent Cells

XL10-Gold cell aliquots (45 μL, Agilent) were pipetted into Falcon tubes. 2 μL β-mercaptoethanol as well as Dpn-treated DNA from control and samples were also added. pUC 18 control was used to test transformation efficiency. Tubes were incubated on ice and then heat-pulsed for 30 sec at 42°C. After subsequent ice incubation, liquid SOC media
was added and the tubes were incubated at 37°C with gentle shaking for one hour. 10 μL of each mixture was plated using the spread-plate method onto LB-ampicillin agar plates and incubated overnight at 37°C. Plasmid DNA of colonies were isolated by miniprep using UltraClean Standard Mini Plasmid Prep Kit (MO Bio, Short Protocol), and after determining the concentration of the DNA in the samples via Nanodrop, a 0.8% agarose DNA gel was run, stained and imaged. The presence of mutation was verified by sequencing.

3.1.1.4. Transformation of DH10α Cells and Extraction of Recombinant Bacmid DNA

DH10α cells (100 μL aliquot per mutation, Manufacturer) was pipetted into a 14 mL round bottom Falcon tube. 5 μL mutated plasmid and/or 2 μL control plasmid (pUC19) were added to the cells. After 30 min of ice incubation, the cells were heat-pulsed for 45 sec at 42°C. Samples were then chilled on ice for 2 min before addition of 900 μL of SOC media was added. Serial dilutions were made into two other tubes using 100 μL of the original, more concentrated sample. 100 μL of each tube (original and diluted tubes) were plated via spread-plate method onto LB-kanamycin-gentamycin-tetracycline-X-gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) plates. The plates were then incubated for 40 hr at 37°C to allow formation of blue and white color. An isolated, large white colony was removed from the plates with a pipette tip and placed into a round bottom Falcon tube containing 2 mL LB media, 10 μL kanamycin (10 mg/mL), 1.4 μL gentamycin (10 mg/mL) and 1.66 μL tetracycline (12 mg/mL). The samples were incubated for 16 hr at 37°C. 1 mL of sample was transferred to 2 mL collection tube and centrifuged for 2 min at 13,000 x g. The remaining contents were added to the collection tube and centrifuged. The supernatant was discarded and the pellet was treated with 300 μL Solution 1: 15 mM
Tris (pH 8), 10 mM EDTA, 100 μg/mL RNAse A; and 300 μL of Solution 2: 0.2 M NaOH, 1% SDS. The samples were incubated at room temperature until the solution turned translucent at which time 300 μL of potassium acetate was added slowly and the solution was mixed gently to produce a white precipitate. Samples were incubated for 5 min at -20°C and centrifuged for 10 min (13,000 x g). The supernatant was transferred to a collection tube containing chilled isopropanol and incubated overnight at -20°C. Samples were centrifuged for 15 min at 13,000 x g. The supernatant was discarded and the pellet was washed with ethanol (repeated twice). Samples were again centrifuged (5 min, 13,000 x g), supernatant was discarded and once again centrifuged for an additional minute. Pellets were air dried with lids open, and the resulting recombinant DNA was dissolved in Tris-EDTA buffer before determining concentrations via Nanodrop.

3.1.1.5. Transfection of Sf9 Insect Cells with Recombinant Bacmid DNA

2 mL of Sf9 cells (approximately 3 x 10^5 cells/mL) were added to a 6 well plate and incubated at 27°C to allow for cell adhesion. The following was added to two 1.5 mL centrifuge tubes; Tube 1: 100 μL Grace’s insect media and 1 μg of bacmid DNA; Tube 2: 100 μL Grace’s insect media and 6 μL of cellfectin. The two solutions were then mixed together slowly before 15 min incubation at room temperature. The media was removed from the 6 well plate carefully so as not to disturb the cells and 2 mL of supplemented Grace’s insect media (+ 10% FBS, + 2% penicillin/streptomycin) was added. 800 μL of Grace’s insect media was added to the Grace’s/cellfectin/bacmid mixture to dilute and 160 μL of that was then transferred to the well plate with the cells. The cells were incubated overnight to allow the bacmid DNA to enter the cells. The media was removed from the wells containing cells carefully and 2 mL of supplemented Grace’s insect media (+ 10%
FBS, + 2% penicillin/streptomycin) was added. This was incubated for approximately 72 hours.

3.1.1.6. Determination of a Viral Titer

In a 96-well plate, 50 μL of poly-D-lysine (5 mg in 50 mL) was added to necessary wells. For each mutation, 2 rows and all wells within those rows was utilized. The well plate was incubated for 30 minutes at 28°C before removal of excess poly-D-lysine. 10 μL of sterile PBS was added to each well and the plate was tapped gently to evenly distribute the PBS. The excess was removed and the plate was allowed to dry in a sterile hood for an hour. Sf9 cells were added to the wells at a concentration of 3 x 10^5, approximately 300 μL of cells was added per well. The plate was covered with parafilm and incubated for 45 minutes to an hour at 28°C. The excess cells were carefully removed, so as not to scratch the bottom of the well plate and disturb adhered cells.

In separate 15 mL conical tubes, the following mixtures were prepared, corresponding to the wells used:

A1: 990 μL Grace’s Insect media supplemented with FBS and penicillin and streptomycin (referred to as complete media).

A2: 990 μL complete media + 10 μL control virus (WT GlyR can be used here).

A3/A4: 990 μL complete media + 10 μL of virus to be tested

A5: 990 μL complete media + 10 μL virus from tube 3

A6: 990 μL complete media + 10 μL virus from tube 4

A7: 900 μL complete media + 100 μL virus from tube 5

A8: 900 μL complete media + 100 μL virus from tube 6

A9: 900 μL complete media + 100 μL virus from tube 7
A10: 900 μL complete media + 100 μL virus from tube 8
A11: 900 μL complete media + 100 μL virus from tube 9
A12: 900 μL complete media + 100 μL virus from tube 10

100 μL of the mixture was added into the corresponding wells and incubated at room temperature for an hour before removal. 200 μL of fresh complete media was added to each well, the well plate was sealed with parafilm and placed into a zip-closure bag with a moist paper towel to prevent evaporation. This was stored at 28ºC in an incubator, overnight.

The following day, 4% paraformaldehyde (PFA) in PBS solution was prepared or thawed from the freezer in a water bath. The well plate was checked for the presence of adhered cells. The liquid media was removed gently from the well plate and 50 μL of PFA was added to each well, without touching the bottom of the plate and was incubated for 20 minutes at room temperature. The PFA was removed from the wells and 50 μL of blocking buffer containing 0.05% Tween (10% stock), 5% normal goat serum (NGS), and PBS was added to each well, and incubated at room temperature for 1 hour. In a 1.5 mL centrifuge tube, 1.5 mL of blocking buffer and 1.5 mL of primary antibody (mouse anti-gp64, ab91214, Abcam) was added. After incubation, blocking buffer was removed from all wells and replaced with 50 μL of primary antibody solution (1:1000 dilution) and incubated at room temperature for an hour. A wash buffer containing 1% NGS and 0.05% Tween in PBS was used to wash the wells three times. 100 μL of secondary antibody solution from a solution containing 2.5 mL blocking buffer, 2.5 mL goat-anti-mouse chain B galactosidase was added to each well and incubated at room temperature for an hour. After washing wells three times as before, 100 μL of prepared color solution containing 6 μL X-gal (0.05 g in 1 mL N,N dimethylformamide, DMF), 6 μL PNBT (0.083 g in 1 mL DMF),
500 mM magnesium chloride and PBS was added to each well, incubated at 37°C and checked every 10 minutes for color production. The cells containing color were counted and the viral titer was calculated.

3.1.1.7. Infection of Sf9 Insect Cells

After determining a viral titer of a particular mutation, the following equation was used to determine how much of the amplified virus (resulting from transfection and subsequent amplification of virus) should be added to a volume of approximately 750 – 1000 mL of Sf9 cells at a concentration of 1 x 10⁶.

\[
\text{Inoculum required (mL)} = \frac{\text{(desired MOI)} \times \text{(cells)/(pfu/mL virus)}}{1000}
\]

where the desired MOI (multiplicity of infection) is greater than or equal to 5, the cells are greater than 1 x 10⁸, and the viral titer results (in plaque forming units (pfu)/mL should be approximately, or greater than, 1 x 10⁷.

3.1.2. Purification of GlyR from Sf9 Insect Cells

The glycine receptor was prepared through purification following published protocols. Sf9 cells were harvested three days post-infection, cultured, washed with ice-cold PBS (pH 7.4), suspended for 1 hr in hypotonic solution containing 5 mM Tris (pH 8), 5 mM ethylenediamine tetraacetic acid (EDTA), 5 mM ethylene glycol-bis(β-aminoethylether)-N,N,N’,N’-tetraacetic acid (EGTA), 10 mM dithiothreitol (DTT) to cause cells to swell. Before lysis an antiproteolytic cocktail was added such that final concentrations were 1.6 microunits/mL aprotinin, 100 mM phenylmethylsulfonylfluoride (PMSF), 1 mM benzamidine, and 100 mM benzethonium chloride. Probe-tip sonication (15 sec pulse intervals at 50% duty cycle with 30 sec rest in between) was used to lyse the cells, and then the sample was centrifuged using a Beckman Coulter Ultracentrifuge for 30
min at 290,000 x g\textsuperscript{606} (at 4°C). The resulting supernatant, the cytosolic fraction, was stored and the protein pellet was placed in a resuspension buffer (hypotonic solution and 640 μL 300 μM NaCl) and sonicated and centrifuged as before. The pellet was added to a digitonin/deoxycholate buffer containing: 1% digitonin; mixed lipids (9:1 plant extract (95% phosphatidycholine purity): egg extract (60% phosphatidycholine purity) at 1.5 mg/mL; 0.10% deoxycholate; 25 mM potassium phosphate monobasic; 25 mM potassium phosphate dibasic; 1M potassium chloride; 5 mM EDTA; 5 mM EGTA; 10 mM DTT and the antiproteolytic cocktail previously used; to solubilize the membrane proteins, sonicated as before and incubated overnight (4°C). After centrifugation as before, the resulting supernatant was added to a 2-aminostrychnine resin (approximately 3 – 5 mL of resin was used) to bind the GlyR and incubated overnight. To remove the GlyR from the resin, an elution buffer containing: mixed lipids as before, 1% cholate, 25 mM potassium phosphate monobasic; 25 mM potassium phosphate dibasic; 1 M potassium chloride; 5 mM EDTA; 5 mM EGTA; and 2 M glycine was added and equilibrated for 48 hr. The resulting slurry was pelleted and the supernatant containing the eluted, purified GlyR was collected.

For open state mutants, which contain two additional mutations (Cys null + A288G/F207G) to allow binding to IVM, the elution step differs. The double mutation causes IVM to act as an agonist, but decreases GlyR binding affinity for glycine\textsuperscript{66,605} and thus mobile strychnine (1.5 mM) rather than glycine, is used to elute purified GlyR from the 2-aminostrychnine resin. This was incubated during affinity chromatography for at least 48 hr and the resulting slurry was pelleted and the supernatant containing the eluted, purified GlyR was collected. The purification process for desensitized states is identical to resting state mutants.
3.1.3. Reconstitution of GlyR into Lipid Vesicles

Reconstitution of eluted, purified GlyR was completed using protocols described in previous study\textsuperscript{428}. Briefly, mixed lipids (9:1 plant extract (95\% phosphatidycholine purity): egg extract (60\% phosphatidycholine purity) at 15 mg/mL, stored as suspended vesicles) were added to purified GlyR/lipid/detergent micelles to yield a final concentration of 1.5 mg/mL. Cholesterol (15.07 mM in methanol) was included to yield approximately 33 mol percent. GlyR/lipid detergent mixture was placed into a dialysis cassette (3500 MW cutoff, Thermo) and the resulting protein/vesicle pellet was dialyzed against 3 x 1 liter volumes of 25 mM potassium phosphate (KPi, pH 5.8). GlyR purification was verified by Coomassie staining and Western blotting analysis, and GlyR concentration was quantitated using a Modified Lowry assay.

3.1.4. GlyR-MTS-Benzophenone Crosslinking

100 μL of purified protein was used for crosslinking studies. 5mM of MTS-benzophenone crosslinker was added to the protein sample and allowed to bind for one hour at 4°C with rotation. The sample was placed into a quartz cuvette and exposed to high-intensity UV light (200-300 nm) for two ten-minute increments; the cuvette was flipped in between to expose both sides, and the sample was allowed to rest in between exposure for ten minutes. Non-reducing SDS-PAGE (15\% resolving, 5\% stacking) was performed with the sample and gel plugs were excised from the gel above the 66 kDa BSA migration line and between the 37 and 66 kDa migration line to encompass the oligomeric and monomeric fractions of GlyR, respectively.
Specific to the M287C mutation, 1 μL of glycine was added to allow the crosslinker to bind. The sample was dialyzed against 3 x 1 liter volumes of 25 mM potassium phosphate (KPi, pH 5.8) in the dark to remove the glycine before flash photoactivation.

To study GlyR in the open state (Cys null + IVM protein), 30 nM IVM was added to the sample. To study the protein in the desensitized state, excess glycine (10 mM) was added to the protein sample. 5 mM of MTS-bzp crosslinker (Figure 7) was added to the protein sample and allowed to form disulfide crosslinks for one hour at 4⁰C with nutated mixing. The sample was placed into a quartz cuvette and exposed to high-intensity UV light (200-300 nm) for two ten-minute increments; the cuvette was flipped in between to expose both sides, and the sample was allowed to rest on a bed of ice in between exposure for ten minutes. Non-reducing SDS-PAGE (15% resolving, 5% stacking) was performed with the sample and gel plugs were excised from the gel above the 66 kDa BSA migration line and between the 37 and 66 kDa migration line to excise the oligomeric and monomeric bands of GlyR, respectively.

3.1.5. In-Gel Trypsin Digestion

Both the monomeric and oligomeric gel plugs were placed into labeled, non-stick microcentrifuge tubes and 100 mM ammonium bicarbonate (ambic, pH 7.8) and 50% acetonitrile (ACN) was added to each tube (enough to cover the plugs) and incubated on a nutator (VWR Thermal Shake Touch) at 37⁰C for fifteen minutes. This was repeated three times. The supernatant was removed and 10mM of DTT was added to each tube and incubated at 56⁰C for one hour. After removing the supernatant, 100 mM ambic and 200 mM iodoacetamide (IAM) was added to each tube and incubated at 37⁰C for one hour in the dark. After washing the plugs with 100 mM ambic, 100% ACN was added to each tube
and dried in an Eppendorf vacufuge for 15 to 20 minutes, depending on plug size. Enough ampic was added to cover the plugs and 1 mg/mL liquid trypsin was added to each tube and incubated on the nutator overnight. The supernatant, which contains the eluted, crosslinked peptide pieces, was removed and placed into new, labeled non-stick tubes. To the gel plugs, elution buffer (0.1% formic acid, and ultrapure water) was added to each tube and incubated for 30 minutes before removal and being placed into newly labeled tubes containing the eluted, crosslinked peptide pieces. The tubes containing the total volume of eluted, crosslinked peptide pieces were placed on the Eppendorf vacufuge and dried.

3.1.6. Mass Fingerprinting of Crosslinked MTS-Benzophenone to GlyR Peptides

50 μL of 0.1% formic acid in ultrapure water was added to tubes containing dried tryptic fragments and incubated on the nutator for at least 2 hours and vortexed. Samples were filtered using Phenex™-NY Syringe Filters (4mm diameter, Phenomenex) before being placed into MS vials (Agilent). Electrospray ionization quadrupole time-of-flight mass spectrometry (ESI-Q-TOF-MS) measurements were taken using an Agilent 6530 Q-TOF-MS equipped with an Agilent HPLC-Chip II G4240-62006 ProtID-Chip-150, with a 40 nL enrichment column and a 75 μm x 150 mm separation column packed with Zorbax 300SB-C18 5 μm material. The mass spectrometer was run on positive mode using internal standards (1221.9906 and 299.2944, Agilent) for calibration. Mobile phase compositions were Solvent A (95% H₂O, 5% acetonitrile, 0.1% formic acid) and Solvent B (95% acetonitrile, 5% H₂O, 0.1% formic acid). The nanoflow elution gradient was developed as follows at 0.50 μL of Solvent A (minute: percent A): 0.00:95%, 4.00:10%, 6.00:70%, 9.00:50%, 11.50:95%, and 13.00:95%. Data were processed using Agilent Qualitative
Analysis Software v6.0. MTS-Benzophenone crosslinked peptides within a 10 ppm accuracy window were identified, accounting for possible peptide modifications (oxidation, acrylamidation, alkylation).

For MS/MS studies, crosslinked samples were run again on the Agilent 6530 Q-TOF-MS, targeting the specific $m/z$ ratio, charge, and retention time (RT) of crosslinked peptides identified in previous runs. Collision-induced dissociation (CID) was used for MS/MS fragmentation following a linear increase in collision energy by $m/z$ using the formula: $y = 3.7x + 2.5$. CID was performed at ±0.2 min from the initial MS scan RT of each crosslinked precursor ion identified. Data were processed using Agilent Qualitative Analysis Software v6.0 in conjunction with ProteinProspector v5.14.3 (University of California, San Francisco) and a python script utilized to match exported spectra data (kindly provided by Amanda Dumi). The list of benzophenone masses, depending on whether protonated, potassiated, sodiated, or alkylated are as follows: benzophenone $m/z$ shift, +312.391823, 350.482184, 334.373652, 370.452009, 408.542369, 392.433838 amu.
CHAPTER 4. PROTEIN-PROTEIN INTERACTIONS WITHIN THE RESTING (APO-) STATE NETWORK

4.1 Introduction

The coupling of CX-MS to a network of single, site-specific Cys mutations can provide a network of protein-protein interactions within GlyR which will help elucidate and refine structure. In this chapter, this network will be applied to analyze the structure of this receptor in its resting state. GlyR, like other pLGICs, transitions from a closed state to a short-lived open state after ligand binding, in which the gate is open and conducting ion flow, and then transitions to a longer-lived desensitized state, where the pore is closed, and ligand is more tightly bound. The dynamics of ligand binding, and the structural changes to accommodate this will be discussed in the next chapter.

4.1.1. The Resting State

Structures of α1-GlyR and α3-GlyR homopentamers have been crystallized in the presence of their competitive antagonist, strychnine. These structures represent a closed channel, as indicated by constriction points in the pore and unfavorable interactions between hydrophobic regions and water, thus leading to a large energy barrier that ions cannot overcome. In the α1-GlyR structure, the ion channel pore shows two sites of constriction, at position 2’ (Pro 266) and 9’ (Leu 277). The 9’ position is narrowest in the strychnine bound state, at a radius of 1.4 Å, which is too small to allow for a hydrated chloride ion to pass through. However, though both are closed structures, they may not represent a true apo-state, as ligand is bound to the orthosteric site. Other non-mammalian pLGICs have been successfully crystallized in their apo-states, including GLIC (at pH 7) and GluCl. However, it is unclear if any known structure represents a good
homolog of apo GlyR, and thus, structural information of a true apo-state is necessary (details of the transition from the resting state to the active state and comparison to other pLGICs can be found in Chapter 5). Furthermore, as with many crystal structures of pLGICs, the large intracellular loop is truncated in known structures, resulting in a gap in structural knowledge. Even if a true apo-state of GlyR were to be determined, this gap would still exist. Since the loop is important in affecting function, is a structure of a pLGIC lacking the intracellular loop reflect the physiological state of the full-length receptor?

Another concern is the role membrane lipids play in structure, stability and function. Current crystal structure data is developed from analyses in a delipidated environment, or in lipid nanodiscs, and thus, may not represent a native receptor environment, and possibly affect the structure of the receptor. Lipid interactions are common to the TMD region of pLGICs. According to Unwin et al.303,304,332,597 and others,598 a three-ring organization of the TM α-helices exists. The M2 region makes up the innermost ring and has no contact with membrane lipids. The second ring is composed of M1 and M3, and in the outermost ring lies M4. These three helices contact the bilayer and may bind selective lipids.598

Analyses of binding general anesthetics to various pLGICs showed the importance of lipids. Many general anesthetics bind in the TMD region, and displace bound lipids. For example, cannabinoids are thought to form direct contacts with an exposed Ser residue in M3 of GlyRs as well as with lipids.283,598 Propofol and desfluorane, which were crystallized with GLIC, appear to compete for binding with lipids that are bound to apo-state GLIC structures.413,598 This is also seen in the binding pocket for IVM.598
Lipids also appear to play a role in gating, conformational dynamics and stability.\textsuperscript{67,598–600} Cholesterol has been shown to provide supports between the agonist-binding domain and the pore of nAChRs.\textsuperscript{394} A molecular modeling study has shown that the displacement of lipids in the cavities from motions by M2 and M3 shows that the shape and volume of the cavities are coupled to gating.\textsuperscript{413} Cholesterol, PA and PC also have an influence on function and transitions between conformational states in both nAChRs and GlyRs.\textsuperscript{14,320} When nAChRs are in the presence of PC alone, they tend to favor a desensitized state, but in the presence of PA and cholesterol, a resting state is favored.\textsuperscript{601} For \(\alpha_1\)-GlyRs, the concentration of cholesterol distinguishes two distinct apo-states.\textsuperscript{14} The biggest implication from these studies, as pertaining to this work, is that the apo-state of GlyRs is flexible, dependent on lipids for stabilization, function and conformation, and cannot be studied by crystallized methods in a conformational manner. The studies outlined in this chapter focuses on methodology that allows the analysis of the native receptor, in a lipid-environment, without any truncations, functioning as a true resting state receptor.

4.1.2. Network Rationale

The Resting State Network built in this work (and used in subsequent conformational states as well) is composed of three site-specific single cysteine mutations: M287C, K116C and K206C. Each mutation was inserted into a plasmid that contained GlyR cDNA that had been modified; three cysteines were mutated (C41S/C290A/C345S), that do not participate in disulfide bridges when the protein is folded; this is referred to as \textit{Cys null} GlyR. \textit{Cys null} GlyR has similar activity to WT GlyR.\textsuperscript{602} After overexpression, the mutant GlyR was purified and crosslinking studies with MTS-bzp (Figure 7) were conducted.
These three sites were chosen for their potential to study areas of the receptor not well-resolved, and due to ease of analysis. The extracellular mutants, positions 116 and 206, were analyzed previously in the Cascio lab, using lysine-lysine crosslinking.\textsuperscript{32} Position 116 is located on the interior of the receptor, on a loop behind beta sheet 8 (Figure 5, 8) and position 206 is located on beta sheet 10, which is on the exterior of the receptor; and so the inclusion of both of these locations in the network was valuable. Furthermore, position 206 is located within the C-loop region, which is theorized to cap over bound agonist during ligand binding (Section 2.2.1.1).

The TMD mutant, position 287 is located in M3, but not far enough into the membrane that it cannot be reached by the crosslinker in the presence of glycine.\textsuperscript{603,604} A study by Williams et al.\textsuperscript{604} showed accessibility of M3 in a state dependent manner. The failure of previous studies conducted at 290°C were attributed to inaccessibility. Furthermore, it was hypothesized that studying a TM mutant may lead to information about the elusive M3-M4 loop region, and since this region is truncated in current crystal structures and models, insights from crosslinking studies would prove to be useful when
refining the structure. It was also hypothesized that
the crosslinker used, once bound to the thiol, would
be hydrophobic enough to penetrate the membrane
and thus, access the intracellular M3-M4 loop.

These three site-specific points will later be
combined with several other positions that are either
completed (A41C and H419C) or are currently in
progress (K6C, G23C I412C; see future work). It is
hypothesized that these mutants will provide vital
information to discuss receptor dynamics as well as
complement high-resolution techniques which will
lead to more complete structures, with a particular
focus on the regions in the receptor that are currently
not well resolved.

4.3. Results and Discussion

Crosslinking-mass spectrometry (CX-MS) is a sensitive and useful tool to
characterize spatial relationships of proteins, and by mapping these distance restraints,
more highly resolved and accurate molecular models can be obtained. This, in turn, can
be used to complement other high-resolution structures and will help bridge the gap of
understanding between protein structure and modulation of function. The designed
network contains one mutant located in the transmembrane region (M287C) and two
mutants located in the extracellular domain (K116C, K206C). In each of these cases, the
Cys mutation was introduced in a Cys null background – a GlyR mutated to remove all free

Figure 8. Network of Mutations. The three mutations chosen as primary
attachment sites for MTS-bzp are M287C, K116C and K206C, represented as
colored spheres on a single subunit of GlyR (PDB: 3jad). M287C is in blue, and
resides in M3. K116C is in pink and is located on a loop behind beta sheet 8.
K206C is colored green and is on beta sheet, which is a constituent of the C-loop.
thiols such that each subunit in the pentamer only has a single reactive thiol. Each will be discussed separately, relating to the results found in the resting state and then integrated together in terms of refining structure and analyzing dynamics of the resting state. In Chapter 4, the open and desensitized states will be discussed separately, integrated together, and then will include the resting state to discuss the dynamics of the receptor relative to each position.

After purification of each single Cys receptor, respective GlyRs were reconstituted into lipid vesicles with the addition of cholesterol, which is an essential membrane component and important to the function of pLGICs. The purified, reconstituted receptor was then crosslinked in the dark with only the addition of MTS-bzp (Figure 7), ensuring the receptor was in its apo-state, allowing the formation of a covalent disulfide to the heterobifunctional crosslinker at the site of insertion of the single reactive Cys residue. After photoactivation, the reactive benzophenone non-specifically forms another covalent crosslink with the area proscribed by the length of the crosslinker (~25 Å). Non-reducing SDS-PAGE was run and bands containing crosslinked GlyR, both monomeric and oligomeric, were separately excised. Bands were then reduced before trypsinolysis to release the crosslinker from its known original site of disulfide linkage. Bands were alkylated and extracted for MS and MS/MS analysis. This was repeated for an \( n \geq 3 \) independent infections/preparations and mass shifted peptides are reported for an \( n \geq 2 \) for all states and subsequent mutations. Each of these biological replicates were subjected to at least two technical replicates.

Unique peptides found in samples where no ligand was added, IVM was added or excess glycine was added correspond to unique areas that are approximately 25 Å from the
single cysteine mutation in the resting, open or desensitized state, respectively. Peptides that were found in oligomeric band (intramolecular + intermolecular crosslinks) that were identical to peptides found in monomeric bands (intramolecular crosslinks) were identified as intramolecular crosslinks. This is because non-reducing SDS-PAGE was utilized, to only separate GlyR oligomers from monomers; membrane proteins commonly run as both single subunits and oligomeric subunits even without crosslinking. Also, due to the inability to distinguish between five potential crosslinking events per pentamer this results in crosslinks identified from oligomeric bands can be either intra- or intersubunit, and thus, only those found in oligomeric bands that were not in monomeric bands were considered to be intermolecular and reported as such in the tables.

For MS analysis, two missed cleavages were allotted as crosslinking at Lys or Arg sites may result in reduced trypsin proteolysis. Mass ions consistent with MTS-bzp crosslinking from the initial MS run were identified using a 10ppm mass cutoff. The sequence coverage for all mutations had a large range of about 40 – 80%, averaging about 60% sequence coverage; therefore, though the tables are expansive, they may not represent all of the crosslinked peptides in the receptor. Peptides that were shifted by MTS-bzp were targeted for a second run if they met the following conditions: only one MTS-bzp was identified in a particular retention time, the peptide was greater than three amino acids in length and it did not match with a m/z found in the blank, gel blank or control runs. In order to refine the crosslinking results to a specific area within the peptide, ideally a single amino acid residue, a second round of MS was run, which targeted the identified precursor ion by its mass, retention time and charge.
Using collision induced dissociation (CID), each product ion was fragmented, resulting in b, y or a product ions (Figure 1). These resulting fragment masses were extracted from each spectra obtained and placed into a matching program, which matched theoretical results from ProteinProspector to results found. Match parameters were within 0.1 Da. Any identified b, y, or a ion was plotted onto the peptide to refine the crosslinking position (Figure 1). Depending on how many times the particular peptide ionized it could be found in multiple retention times, meaning either isobars are linked at different sites or
different m/z but same chemical structure, and thus, could have various crosslinking events within it. After \( n \geq 3 \) trials were completed, the refined crosslinking results that were found in at least two trials were tabulated and mapped onto the crystal structure determined by Du et al.\textsuperscript{296} The strychnine-bound version (PDB: 3JAD) was used, as it represents a non-conducting state, and it avoided using the crystal structure bound with glycine and ivermectin because it is not certain whether that structure represents an allosterically modulated desensitized state or a partially open state.\textsuperscript{296} It was subsequently used for all mutations in all states for uniformity purposes, as this would allow for easier detection of mobile regions and changes between the states.

Distance measurements were taken between the C\( \alpha \) of the mutation site and the C\( \alpha \) where the crosslinker bound. If only a single site was resolved in MSMS studies as the site of attachment, the measurement was recorded with no standard error measurement. If crosslinking could not be resolved to a single site, but rather to a stretch of continuous amino acids, an average of the respective C\( \alpha \) distances were reported with its standard error. These distance measurements can be used to examine the accuracy of the model used as well as the model consistency or lack thereof. It also can be used as a means of interrogating dynamics and receptor mobility.

4.3.1. Position 287

A total of 39 sites of crosslinking relative from 287C have been identified; 24 of these are intramolecular crosslinks (Table 2). These sites are distributed throughout most of the receptor but are primarily located in the intracellular M3-M4 loop region and the TM domain (with the exception of M1). There are also crosslinking sites in the preM1 region of the ECD and the M2-M3 loop, found in the ECD-TMD interface (Figure 10,
Table 2. The C-terminal tail, which is also not resolved in crystal structures has also been crosslinked intramolecularly (Table 2).

Table 2. M287C Resting State Crosslinks.
The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The blue underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>M287C Resting State</th>
<th>Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intramolecular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17LMGR<strong>T</strong>SGYDARIRPNFK<strong>3</strong></td>
<td>Pre M1</td>
<td>73±1.72, 61</td>
</tr>
<tr>
<td>60VNIFLRQWNDPR<strong>72</strong></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>66QQW<strong>NDPR</strong>LAYNEYPDDSLDLDPSMLD</td>
<td></td>
<td>63±3, 35</td>
</tr>
<tr>
<td>SIWKPDLFFANKE<strong>104</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105GA<strong>H</strong>FEITDNKLRISR<strong>122</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>191EE<strong>K</strong>DLRYC<strong>T</strong>K</td>
<td></td>
<td>43±2, 49, 50, 54±9</td>
</tr>
<tr>
<td>201HY<strong>T</strong>GKFCIEAFFL<strong>ER</strong>218</td>
<td></td>
<td>37±2, 41</td>
</tr>
<tr>
<td>272ASLPKV<strong>S</strong>YVK<strong>281</strong></td>
<td>M2-M3 loop</td>
<td>50±1, 48±1, 30±2</td>
</tr>
<tr>
<td>282<strong>A</strong>IDIW<strong>C</strong>AVALLEFVSALL<strong>E</strong>YAAVNF</td>
<td>M3</td>
<td>16, 34</td>
</tr>
<tr>
<td>VRQ<strong>H</strong>K</td>
<td></td>
<td></td>
</tr>
<tr>
<td>322<strong>R</strong>RIHK<strong>325</strong></td>
<td></td>
<td>25±1, 22±0.3</td>
</tr>
<tr>
<td>356GANN<strong>S</strong>IFFTNPAP<strong>S</strong>KSP<strong>E</strong>EM<strong>R</strong>K</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td>378<strong>K</strong>LF<strong>I</strong>QR<strong>A</strong>K</td>
<td></td>
<td>34, 30±0.2</td>
</tr>
<tr>
<td>381<strong>I</strong>DK<strong>I</strong>SR</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>411<strong>R</strong>ED<strong>V</strong>HNO<strong>421</strong></td>
<td>C-term tail</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Intermolecular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34<strong>G</strong>PPV<strong>N</strong>VSANIFINSFGSIAETTMD<strong>Y</strong>R</td>
<td>Pre M1</td>
<td>63±2, 32±0.4</td>
</tr>
<tr>
<td>120<strong>I</strong>SRN<strong>G</strong>NVLYSIRITTLACPM<strong>D</strong>L</td>
<td>M2, M2-M3 loop</td>
<td>64±2, 26</td>
</tr>
<tr>
<td>25<strong>V</strong>GLGTT<strong>V</strong>LTMT<strong>Q</strong>SSGSR</td>
<td></td>
<td>19±0.3</td>
</tr>
<tr>
<td>ASLP<strong>K</strong>Y<strong>S</strong>Y<strong>V</strong>K</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>310<strong>Q</strong>HKE<strong>L</strong>LRFR<strong>R</strong>K</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td>320<strong>E</strong>DELETEG<strong>R</strong>FNFSAYGM<strong>P</strong>SLQAK</td>
<td>M4</td>
<td>32±1, 26</td>
</tr>
<tr>
<td>DG<strong>S</strong>Y<strong>K</strong></td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>356<strong>I</strong>SV<strong>K</strong>GAN<strong>S</strong>NTTN<strong>P</strong>P<strong>A</strong>P<strong>S</strong>K</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>378<strong>S****P</strong>E<strong>E</strong>M<strong>R</strong>KLF<strong>I</strong>QR</td>
<td></td>
<td>N/A, 39±1</td>
</tr>
<tr>
<td>411<strong>T</strong>SRIFGPMAFLIFNM<strong>F</strong>YWIIY</td>
<td>M4</td>
<td>32±1, 26</td>
</tr>
</tbody>
</table>
Given the position of the mutation, in M3, it was hypothesized that crosslinking to this site might allow crosslinking to regions deeper, or even across the membrane. It was considered possible that once the disulfide bond is made to the ionized thiolate at position 287, the crosslinker is no longer membrane impermeable and its benzophenone group and spacer are hydrophobic enough to penetrate the cell membrane. Other studies have also shown the hydrophobicity of the MTS-bzp crosslinker and its ability to penetrate the membrane.\textsuperscript{609,610} This proves to be the case in this study as well, as a large proportion of the crosslinking sites, 38\% of them, are in the M3-M4 loop. This is particularly noteworthy as previous studies have not been able to visualize crosslinking data to this region,\textsuperscript{602} and this loop is truncated in current crystal structures, and thus this work provides new information relating to structural details and dynamics of this elusive region. The approximately 70 residue long loop, which is characterized by its sequence and length diversity among pLGICs, and is not resolved in current structures of any pLGIC, shows crosslinking throughout, but is mostly seen in the beginning and the end of the loop, as represented by the peptides, \textsuperscript{310}QHKELLRFRRK\textsuperscript{320} and \textsuperscript{387}IDKISR\textsuperscript{392}, respectively (Table 2, Figure 10).

Another area that was anticipated to be crosslinked was the TMD region and the loops in the interface as they are in close proximity to position 287. M3 was crosslinked intramolecularly, and M4 and M2 were crosslinked intermolecularly (Table 2, Figure 10). The fragmentation data based on the different retention times led to the inability to refine this region to any particular area. The M2-M3 loop was also crosslinked heavily. The Cys-loop, the loop located in the ECD-TMD interface, the hallmark loop of the pLGICs, was crosslinked at position K143. The Cys-loop, and the other loops that make up the ECD-
TMD interface are predicted to be mobile, and so, it is hypothesized that seeing these regions in other states, refined at different positions indicate mobility of these regions (see Chapter 4).

Of particular note, it was unexpected that crosslinking would occur at the top of the ECD, seen in the peptide $^{17}$LMGRTSGYDARIR$^{33}$. This could be indicative of the “bloomed” appearance used to describe the resting state, because the ECD is wider at the top, and more flexible as an apo-state protein, there could be the possibility that it would come into close proximity of the membrane, and thus position 287. This may also be due to local unfolding events of the protein in this region, a concept that will be discussed more thoroughly in later sections. Previous studies showed that affinity purification of homomeric GlyRs was functional after reconstitution.$^{428}$

Given the dimensions of MTS-bzp and its
flexibility, the distance between the Cα at position 287 to the crosslinked sites should be approximately 25 Å. The measured distances using the available crystal structure bound to strychnine (PDB: 3jad) shows a large discrepancy from this number, with an overall range of 14.5 to 72.5 Å. It is important to note that there is no available measurement data for the crosslinks to the M3-M4 loop region or C-terminal tail, as those features are not present in available crystal structures. It is also relevant that these distance measurements are taken on a static image based on a dynamic experimental study, and so, they are most useful to show dynamics and protein flexibility, and to visualize what is most congruent with available crystal-based structural data. These seemingly excessive crosslinking lengths have been seen in other crosslinking studies and is still a phenomenon that requires more analysis. In a CX-MS study analyzing the entire human mitochondria, approximately 10% of the crosslinks found exceeded their 30 Å Eulerian crosslinker distance. This study used lysates and thus were not purified, so it became difficult to distinguish intra- from intermolecular crosslinks. They also allowed up to 20 ppm error cutoff. Our studies, in contrast, uses purified protein and minimizes the potential for protein aggregates, and thus, distinguishing both intra-subunit and inter-subunit crosslinks is made more manageble. Also, the error cutoff in this work is 10 ppm, allowing more stringency. The major point here is that currently, without a true set of standards for conducting CX-MS studies, each study sets its own standards for what is a confident match, what is an artifact, and what error cutoff should be used. This will change in time as CX-MS becomes a more widespread technique.

Overall, with respect to position 287 in the resting state, there are a few important discoveries. This is the first look at the M3-M4 loop in the context of a full-length protein
in a native environment. These studies suggest that the loop is intimately associated with the membrane when the protein is in an apo-state. Also, the protein appears to be quite flexible, which is expected, as there is crosslinking seen throughout the protein, and in the upper ECD. The lack of contact to M1 may mean that M3, or the top of it, is more closely associated with M4 and M2 in this state. Finally, due to the diversity of structures observed in cryo-EM studies, this too, suggests flexibility.\textsuperscript{296,297}

4.3.2. Position 116.

Position 116 is located on a loop behind $\beta$-sheet 8, on the $\beta$5-$5'$ loop, in the ECD (Figure 5). Previous studies using Lys-Lys crosslinking suggests its surface accessibility.\textsuperscript{32} The study by Liu \textit{et al.}\textsuperscript{32} examined intra- and intermolecular crosslinking in GlyBP (a truncated form of GlyR consisting of its ECD), with dimethylsuberimidate (DMS), an amine-reactive reagent. GlyBP has a total of 11 Lys that could be potentially crosslinked with DMS, 12 intramolecular and 2 intermolecular Lys-Lys crosslinks were identified. Similar to our studies, intermolecular crosslinks were only considered to be between two subunits if they were not identified in monomeric bands. Importantly, the top portions of the subunits (residues 1-33), including the $\alpha$-helix were found to be flexible, as shown by the crosslinking between K6 and K166.\textsuperscript{32}

In this work, crosslinking results after $n \geq 3$ trials, resulted in 29 crosslinking events. Similar to what was seen in crosslinking studies to position 287, some sites could be refined to single sites of attachment. Quantitation studies that examine the propensity for various sites attached and how often will help to further understand the significance of these sites to position 116 and others. Of the 29 individual crosslinking sites, the majority of them were found intramolecularly. Most of the crosslinks were found in the preM1 region of the
receptor which is expected as position 116 is located in the ECD. In comparison to position 287, crosslinking data to position 116 crosslinks to more of the top of the ECD, including the peptide seen in 287, $^{17}$LMGRTGYDARIR$^{33}$ as well as the peptide $^{1}$ARSATKPMSPSDFLDK$^{16}$ (Table 3, Figure 11). Position 116 also crosslinks to the C-loop, which when analyzed in the context of all three states will give better perspective on how the C-loop may be moving relative to this position (Chapter 4). Crosslinking identified in regions containing K6 and the C-loop have also been found in the study by Liu et al.,$^{32}$

Table 3. K116C Resting State Crosslinks.

The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The pink underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>K116C Resting State</th>
<th>Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intramolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{1}$ARSATKPMSPSDFLDK$^{16}$</td>
<td>Pre M1</td>
<td>20±1 (9-14 only)</td>
</tr>
<tr>
<td>$^{17}$LMGRTGYDARIRNFK$^{33}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{60}$VNLFLRQQWNDPR$^{72}$</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>$^{132}$TLTLACPMDLK$^{143}$</td>
<td></td>
<td>39±2</td>
</tr>
<tr>
<td>$^{194}$DLRYCTKHYNTGK$^{206}$</td>
<td></td>
<td>30±1</td>
</tr>
<tr>
<td>$^{214}$HLER$^{218}$</td>
<td></td>
<td>31±2</td>
</tr>
<tr>
<td>$^{253}$VGLGITTVLTMTTQSSGSR$^{271}$</td>
<td>M2</td>
<td>51±1</td>
</tr>
<tr>
<td>$^{272}$ASLPKYSYVK$^{281}$</td>
<td>M2-M3 loop</td>
<td>40±1, 40, 44</td>
</tr>
<tr>
<td>$^{310}$QHIKELRRK$^{320}$</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td>$^{322}$RHIK$^{325}$</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>$^{372}$PEEMRKLFIQRR$^{383}$</td>
<td></td>
<td>91 (378 only)</td>
</tr>
<tr>
<td>$^{378}$KLFIQRK$^{385}$</td>
<td></td>
<td>85±1</td>
</tr>
<tr>
<td>$^{384}$AKIDK$^{389}$</td>
<td></td>
<td>78±1</td>
</tr>
<tr>
<td>$^{386}$KIDKISR$^{392}$</td>
<td></td>
<td>76±1</td>
</tr>
<tr>
<td>Intermolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^{1}$ARSATKPMSPSDFLDK$^{16}$</td>
<td>Pre M1</td>
<td>N/A</td>
</tr>
<tr>
<td>$^{120}$ISRGNVLSIR$^{131}$</td>
<td></td>
<td>30±2, 24±1</td>
</tr>
<tr>
<td>$^{123}$NGVLYSIRILTTLACPMDLK$^{143}$</td>
<td></td>
<td>31±2, 20±1, 27±2</td>
</tr>
<tr>
<td>$^{193}$EKDLR$^{196}$</td>
<td></td>
<td>31±2</td>
</tr>
<tr>
<td>$^{253}$VGLGITTVLTMTTQSSGSR$^{271}$</td>
<td>M2</td>
<td>58±2</td>
</tr>
<tr>
<td>$^{272}$ASLPKYSYVK$^{281}$</td>
<td>M2-M3 loop</td>
<td>43, 38±1</td>
</tr>
</tbody>
</table>
but K16 to K116 was not found. The ECD-TMD interface is also crosslinked, though in a different pattern than seen in position 287. The Cys-loop has more crosslinking to it, but the M2-M3 loop has fewer sites of attachment. In both position 287 and 116, the M2-M3 loop is crosslinked both intramolecularly and intermolecularly, meaning that loops from adjacent subunits can interact with both the middle of the ECD where 116 is and the top of M3 where 287 is. Again, in the context of multiple conformational states, it will be clearer as to how these loops may be moving.

The crosslinking results in the TM domain is very different compared to what was seen in 287. Only M2 is crosslinked to position 116, and as seen previously, the fragmentation data for this region does not allow for refinement to a single amino acid. This may be due to how well this peptide ionizes, or because multiple isobars are not chemically distinct enough to be
separated by LC and precluding product ion assignment. This will be better resolved once quantitation studies are conducted.

The most unexpected result for this position is the crosslinking that is seen in the M3-M4 loop. The intracellular loop is expected to be too far away from the ECD to crosslink to it, and yet in every trial conducted these crosslinking events were reproducibly observed. Crosslinking to this area is only seen intramolecularly, and mostly at the beginning and the end of the loop, so very close to the inner membrane surface. Many of the peptides crosslinked have a number of Lys and Arg residues, including; 310QHKELLRFRRK320, 378KLFIQRAK385, and 386KIDKISR392 (Table 3, Figure 11). The significance of this is that poly-Lys and poly-Arg peptides have been shown to penetrate the cell membrane,611–621 which might be the reason that these peptides are able to be crosslinked from position 116 in the ECD. This will be discussed in more detail in the next chapter.

The range of distance measurement are even larger here than what was seen for position 287, considering crosslinks were identified in the M3-M4 loop (Some peptides from this region are part of the crystal structure though shifted to be included in M4) and M2 (Table 2 and 3). However, it is likely that the apo-state is highly flexible, perhaps more so than other states,238,243,598 and thus, seeing a large amount of out-of-bounds crosslinking is not that concerning.

Overall, with respect to position 116, the ECD appears to be flexible as much of it is crosslinked. Areas reported to be mobile, such as the C-loop, the Cys-loop and the M2-M3 loop are crosslinked. The TMD region, however, is not heavily crosslinked, which indicates that position 116 does not come into contact with the TM, with the exception of
M2. It could also mean that crosslinking to this region was detected by the MS. This too can be better explained in the context of the open and desensitized states.

4.3.3. Position 206.

Table 4. K206C Resting State Crosslinks.
The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The green underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>Location</th>
<th>K206C Resting State</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intramolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1ARSATKPMSPSDFLDRk</td>
<td>Pre M1</td>
<td>39±1 (9-16 only)</td>
</tr>
<tr>
<td>132TTLACPMDLK</td>
<td></td>
<td>27±1, 34</td>
</tr>
<tr>
<td>227ASLPKVSYVK</td>
<td>M2-M3 loop</td>
<td>45±1, 38±1</td>
</tr>
<tr>
<td>372SPEEMRKLFIQRK</td>
<td>M3-M4 loop</td>
<td>84 (378 only)</td>
</tr>
<tr>
<td>Intermolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17LMGRTSGYDARIPNFK</td>
<td>Pre M1</td>
<td>31±2, 35±0.1</td>
</tr>
<tr>
<td>165GAHFHEITDNK</td>
<td></td>
<td>19±1</td>
</tr>
<tr>
<td>177LRLRISNGNVLYSIR</td>
<td>M2</td>
<td>69±1, 61±1</td>
</tr>
<tr>
<td>253VGLGTVVLMTTQSSGSR</td>
<td>M2-M3 loop</td>
<td>44, 48</td>
</tr>
<tr>
<td>272ASLPKVSYVK</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td>310QHKEFRFRK</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>321RRHHK</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>334FNFSAYGMGPASLQAK</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>350DGISYKGANNSTTNPPAPSK</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>372SPEEMRKLFIQR</td>
<td></td>
<td>88±1 (378-388 only), 84±2</td>
</tr>
<tr>
<td>384AKKIDKISR</td>
<td></td>
<td>77±1</td>
</tr>
<tr>
<td>412IVRREDVHNNQ</td>
<td>C-term tail</td>
<td>45±2 (412-413 only)</td>
</tr>
</tbody>
</table>

Position 206 is located on the C-loop and so gaining perspective on what is crosslinking to this site is advantageous to analyzing its movement during gating and channel activation. Studies have shown this area to be quite mobile, moving about 7 Å, to cap over the bound ligand.238,317,318,343,344 There were 26 individual crosslinking sites identified, and unlike those for positions 287 and 116, most of these were intermolecular.
crosslinks. Similar to 116, position 206 comes into contact with the upper preM1, in residues 1-16 and residues 17-33. It also crosslinked to position 116 intermolecularly (refined from \textsuperscript{105}GAHFHEITTDNK\textsuperscript{116} to \textsuperscript{113}TDNK\textsuperscript{116}). This is not reciprocated by results seen in 116, either intra- or intermolecularly. The loops in the ECD-TMD interface, the M2-M3 loop and the Cys-loop are also crosslinked. The proximity of position 206 to these loops could be significant in triggering downstream channel activation.

Similar to position 116, very little crosslinking is seen in the TM domain, and is only seen in M2 (Table 4; Figure 12 shows crosslinking to M4 but this area is actually M3-M4 loop peptides). The significance of the ECD mutants being in closer proximity to the pore and not the other \(\alpha\)-helices, may have something to do with how the \(\alpha\)-helices align themselves into a three-ring system.\textsuperscript{304,332,389,598} And also similar is crosslinking to the M3-M4 loop. What is interesting though, is the peptides crosslinked do

![Figure 12. K206C Resting State Crosslinks. Two subunits of GlyR (PDB: 3jad) are shown to visualize both intramolecular crosslinks (left subunit) and intermolecular crosslinks (right subunit) to position 206 (represented as a yellow sphere). All crosslinked areas are highlighted in green. The bead model below the structure shows the M3-M4 loop, both intra- and intermolecularly. Again, the crosslinking data is highlighted in green. The C-terminal tail is not shown. Refer to Table 3.](image)
not have an abundance of poly-R/K regions, and thus, membrane penetration, seems less likely, unless the poly-R/K regions pull other parts of the loop into the membrane as well. Two of the peptides, $^{334}$FNFSAYGMGPSL$^{349}$QAK, refined to the last three amino acids, and $^{350}$DGISVKGANNSNTTNPPPAPSK$^{371}$ are in the middle of the loop (See Section 4.5 for discussion of the M3-M4 loop).

4.3.4. Integration of Network in the Resting (apo-) State

Analyzing the resting state as a whole, with respect to the network of mutations can provide some unique perspectives into its structure and dynamics. All three mutations crosslinked to the top of the ECD, to the peptide $^{17}$LMGRTSGYDARIR$^{33}$, position 287 to residues 20-21, 23, and 27; position 116 to residues 19-24, 27-28 and 31-32; and position 206 to 18-19 and 27-28 (Tables 1-3). K116C and K206C crosslinked to the peptide $^{1}$ARSATKPMSPSDFLDK$^{16}$, but M287C did not. It is expected that the ECD mutants may crosslink to the upper ECD, given the approximate distance of the crosslinker of 25 Å, but it was unexpected that position 287 would. This is indicative of protein flexibility in the resting state. In order for the receptor to contact these three mutants in its apo-state it must be sampling other energy wells as it searches for the lowest resting state conformation. To confirm these movements, future work involves analyzing two upper ECD mutants, K6C and G23C. If these mutants crosslink to all three mutants in this network, it is more definitive of protein flexibility.

Another part of the ECD, $^{105}$GAHFHEITTDNKLLRISR$^{122}$ which contains position 116, was crosslinked to both 287 and 206, but was not seen in 116 results. This could be due to crosslinker length. However, what is interesting is that this network of mutations is
able to “see” one another, and thus provide a sense of reciprocity that increases the confidence of the results.

The ECD-TMD interface is known to play an important role in transducing the effects of ligand binding to the TMD region, particularly the M2-M3 loop and the Cys-loop, the hallmark loop of these channels. An analysis of the resting state alone cannot provide enough details into the dynamics of this region, but what is apparent is that all three mutants are crosslinking to this region, and thus, this network provides important vantage points. The Cys-loop in α1-GlyRs is $\text{138CPMDLKNFPMDVTC152}$. Position 287 only crosslinks to K143 on the complementary subunit; whereas, position 206 crosslinks to K143 intramolecularly. Position 116 sees the most crosslinking to this loop, at residues $\text{141DLK143}$ intramolecularly, and $\text{138CPM140}$, intermolecularly. All of the residues that are crosslinked are on the exterior part of the loop, and primarily near the bottom, closest to the TMD region, except for position 116, which also crosslinks to the top of the exterior loop. It is important to note that 287 is crosslinking to the complementary Cys-loop at the bottom, whereas; both ECD mutant are crosslinked to the bottom of the principle Cys-loop. Position 116 crosslinks to the top of the complementary loop. It is thought that the Cys-loop displaces outward during activation, so this might show a different pattern of interaction in the active state, perhaps to the other side of the loop. However; given the length of the crosslinker used, it may be too large to detect subtle changes that are taking place, as there is mostly likely only a 2-3 Å difference in their displacement. The M2-M3 loop interacts with these three mutations slightly differently, as there is no clear indication of where the loop is relative to these positions. However,
analyzing this loop in the context of other conformational states may be more helpful in explaining the dynamics of this loop.

The C-loop, as stated previously, has a larger displacement during activation, as much as 7 Å, though still much smaller than the length of MTS-bzp. With respect to position 206, there is no crosslinking seen in either the principle or complementary subunit. This could be due to crosslinker length. However; the other two mutants, which are not located on the C-loop as 206 is, do crosslink to the principle subunit. Position 116 crosslinks to residues LRYCTKHYN, 198-203 are part of the loop; and position 287 crosslinks more sporadically, to C198, NT204, and KF207.

Crosslinking results in the TM domain were also somewhat surprising, as both ECD mutants only crosslinked to M2, the α-helix that lines the pore; whereas, 287 located in M3, had more uniform distribution of crosslinking results in this region. The results for 287 are expected as 287 is located in this region, but it is interesting to see that position 116 and 206 are similar in what they contact.

The C-terminal tail, which is not resolved in crystal structures is crosslinked to positions 206 on the complementary subunit and 287 on the principle subunit, but not to 116. This is somewhat expected as 116 is quite far away, and hypothesized to be in a less mobile region than 206.

The M3-M4 loop is also unresolved in crystal structures, and what was most surprising to see crosslinked to all three mutants. Both positions 287 and 206 crosslinked more so than position 116. Most of the crosslinking sites were located at the beginning and ends of the loop, but the middle of the loop was crosslinked to 287 and 206. In the context
of resting, open and desensitized states, a better perspective can be gained about what role this loop may play in gating and desensitization.

The distance data for all mutations shows that PDB:3jad, a strychnine-bound non-conducting state,\textsuperscript{296} has inconsistencies in terms of receptor flexibility. With respect to all of the mutants, a maximum of 32% of the crosslinks were found within an expected range of $\leq 30$ Å (Table 1). These numbers can only take into account the distances able to be measured on the crystal structure, and thus, M3-M4 loop and C-terminal tail crosslinking were omitted. Based on this data, it can be interpreted that this non-conducting crystal structure cannot truly represent an apo-state GlyR structure.
CHAPTER 5. PROTEIN-PROTEIN INTERACTIONS WITHIN THE OPEN AND DESENSITIZED STATE NETWORKS

5.1. Introduction

Representing protein dynamics is difficult to do with high-resolution techniques such as cryo-EM and x-ray crystallization. The majority of current structural knowledge comes from NMR and x-ray structures; the popularity of cryo-EM is increasing, and more structures determined by this technique are being added to the Protein DataBank (PDB). However, cryo-EM and x-ray structures cannot represent protein dynamics. NMR does have this capability, but it lacks the ability to analyze large membrane proteins. Thus, using CX-MS to help refine the particularly dynamic parts of a membrane protein, as well as provide information on proteins in near-native environments in various conformations is essential. Available crystal structures have helped to speculate how the receptor may transition from one state to the next, and CX-MS experiments can provide additional details as to whether these speculations have merit.

5.1.2. Receptor Activation and Desensitization

The Monod-Wyman-Changeux (MWC) theory was originally used to model the behavior of allosteric enzymes.\textsuperscript{623} It is also an accepted theory to explain the agonist-induced activation seen in pLGICs.\textsuperscript{623} It is based on the idea that the receptor spontaneously shifts between pre-existing states, and that ligands shift the equilibrium towards high affinity states.\textsuperscript{623} It also provides an explanation for reorganization of multiple subunits, meaning that two states; a low-affinity resting state and a high-affinity ligand bound state are shifted towards the higher bound state as more ligand binds. This theory works well for active and resting states, but less so for desensitization.\textsuperscript{324}
The transition between the resting state and the active state, according to available structures shows that an apo-state is less rigid, especially in the ECD region. The interface between subunits at the ECD is larger, and more flexible, resulting in fewer interactions and creating more solvent-exposed areas. As ligand binds to the receptor, the β-sandwich folds in each subunit are postulated to move as a rigid body in response, which results in the ECD to become closer together. There is an overall anti-clockwise twist of the ECD, and it is referred to as ‘un-blooming’. The C-loop caps over the orthosteric site, and the tip of the loop appears to move with the complementary subunit. This un-blooming and anti-clockwise twist is observed in transitions of GLIC, GluCl and GlyRs though there are some differences in mechanisms. GluCl does not appear to have much movement of the C-loop in the same manner as GLIC, but does show tightening of loops A, B and C, pushing them towards the complementary subunit which allows for ligand binding. A residue important for binding agonist in the F loop, K171, also shifts from a downward to an upward position for better hydrogen binding. In the crystal structure of α1-GlyRs, since they are already bound to strychnine, they appear already un-bloomed, instead of going through an un-blooming conformation as it shifts to the active state. It is important to note, that GlyRs are not crystallized in a true apo-state, and so the competitive antagonist may be locking the receptor in this un-bloomed configuration. The transition of an apo-state GlyR may in fact, resemble the transitions seen in GLIC or GluCl. The C-loop of GlyRs, however; moves similarly to what is predicted in GLIC structures.

The transition at the level of the ECD-TMD interface in GLIC involves mostly the M2-M3 loop. There is a revolving motion towards M3 resulting in a 4Å shift. This shift appears to suggest that there is allosteric coupling of the movements in the ligand-bound
ECD region to the TMD gate region. The interface in GluCl structures also involves the M2-M3 loop, but it moves differently, in a lateral motion, instead of a revolving motion.\textsuperscript{312} The result here causes a tilt in the M1 and M3 helices to allow IVM to bind.\textsuperscript{312} Because the GlyR structure is already bound at the orthosteric site with strychnine the resulting transitions to the active state are blunted. What is interesting is that the transitions in the TMD differ between $\alpha_1$ and $\alpha_3$-GlyRs. The former behaves similarly to what is seen in GluCl, but $\alpha_3$-GlyRs have a more radial twist. A voltage-clamp fluorometry study in 2009 showed that the addition of strychnine caused conformational changes to the ECD, but did not open the pore which indicates that the dynamics caused by strychnine binding are different than the dynamics caused by glycine binding.\textsuperscript{624} The movement of M2 is similar among pLGICs, as it responds to the displacement of the Cys-loop, and the M1-M2 loop. The top of the helix rotates outward, towards the M1 and M3 helices.\textsuperscript{596}

According to the structure determinations from Du \textit{et al.},\textsuperscript{296} there is a 4-5 Å difference between the strychnine-bound structure and the glycine and glycine/IVM-bound structures in the pore, due to the rotation of the 9′ Leu position. Du and colleagues also conducted two-electrode voltage clamp electrophysiology experiments to analyze these conditions (glycine v. glycine/IVM) in a physiological context. They concluded that the glycine-bound structure was an agonist-bound open state, and they suspected that the glycine/IVM-bound structure was an allosterically modulated desensitized state or a partially open state.

As previously mentioned, current crystal structures lack dynamics information and are missing pieces of the receptor, particularly the intracellular M3-M4 loop. But another concern based on these available structures is the source of the protein and the lack of a
lipid bilayer. Du et al. crystallized GlyR from zebrafish, and not human, and the main difference between the two is how they respond to the binding of IVM. Human α1-GlyR is potentiated by IVM and it prevents the channel from desensitizing, whereas; in zebrafish GlyR, IVM may shift the receptor towards desensitization. This difference in interaction could implicate minor structural or mechanistic details that need to be better understood.

In this work, IVM will be used to analyze the open state as the binding keeps the receptor from desensitizing. This is possible due to a double mutation in GlyR, A288G/F207G, which allows IVM to bind GlyR with higher affinity. A similar mechanism for preventing desensitization has also been shown in α1-GlyRs with the use of picrotoxin. Similarly, as previously discussed, the role of lipids is extremely important in membrane protein structure and function. All of this leads to uncertainty in proposed mechanistic insights provided by high-resolution structures and thus, the need for complementary information from CX-MS studies.

It is now understood that in addition to an active, open state, there are also pre-active states, a ‘flipped’ state, which exhibits only partial agonist stabilization and response, and a ‘primed’ state that results in a shorter-lived open state. Neither of these pre-active states have been investigated structurally, and so it is unknown how they contribute to the dynamics of an open state configuration.

Further speculation and controversy arise when understanding the transition from the open, active, state to a desensitized one. Desensitization is defined by a closed, ligand-bound state that is refractory to any activation and decreases ion flow. It is thought to prevent over-activation. Much of the information regarding desensitization came from early experiments studying nAChRs. Kinetic studies showed that there are both ‘fast’ and
‘slow’ desensitization states\textsuperscript{324,625–627} and that desensitization happens when fully ligand-bound.\textsuperscript{324,628,629}

Original theories suggested a single gate opened and closed, but in the 1950s, Del Castillo and Katz\textsuperscript{630} proposed a dual-gate model, one that opened after ligand binding to produce an open channel, and another that closed as the channel desensitized.\textsuperscript{324,631–633} The activation gate, sometimes called the hydrophobic gate, is proposed to be located near the center, at position Leu 9’, and the desensitization gate is proposed to be near the intracellular end of the TMD.\textsuperscript{243,324,631,632} Many crystal structures of resting states or agonist-bound closed states have shown constriction at the central position in the pore,\textsuperscript{243,296,297,308,309,311,312} giving evidence towards this activation gate. The dual-gate model also proposes that a desensitized state would have a widening of the upper part of the TM \(\alpha\)-helices and a constriction at the intracellular end.\textsuperscript{324,632} NMR studies of ELIC have shown a contraction at the intracellular end of the TMD in an apparent desensitized state.\textsuperscript{632} Some suggest that several crystal structures (GLIC, GluCl, and \(\alpha\)-GlyR)\textsuperscript{296,324,634} may also represent desensitized states; whereas, others suggest that only the crystal structure of \(\beta\)-GABA\(_{A}\)R is a true representation of a desensitized state.\textsuperscript{243}

Based on the experiments analyzing desensitization rates thus far, some mechanistic details have emerged. Site-directed mutagenesis experiments showed several residues in pLGICS have an effect on desensitization. The 9’ residue that is the hydrophobic gate for activation, is conserved among pLGICS (often a Leu), when mutated almost completely suppresses desensitization.\textsuperscript{324,635,636} Several mutations in the ECD and the ECD-TMD interface also affect desensitization.\textsuperscript{637,638} Relevant to \(\alpha\)-GlyRs, mutations
in the intracellular M1-M2 loop, near the selectivity filter, I244A and P250T, strongly increase desensitization and lead to loss-of-function phenotype.\textsuperscript{354,357,387}

Analysis of phosphorylation sites in the intracellular loop have also provided insight into desensitization. In nAChRs and GABA\textsubscript{A}Rs, the desensitization rate was decreased in the presence of phosphorylation of the M3-M4 loop.\textsuperscript{639,640} The loop also seems to play a role in desensitization in α1-GlyRs, as splice variations and mutations have shown,\textsuperscript{388,414} though this has not been shown due to effects of phosphorylation.

Despite the wealth of kinetic data that exists for desensitization, there is still a gap in knowledge regarding the dynamics and structural configuration of α1-GlyRs and other pLGICs. There are countering ideas on which available crystal structures represent a truly desensitized state; this ambiguity can likely be resolved with more structural experiments.

In these studies, we propose to examine allostery by CX-MS. Open and Desensitized State Networks are composed of three site-specific single cysteine mutations: M287C, K116C and K206C. Each mutation was inserted into a plasmid that contained GlyR cDNA that had been modified, the resulting protein is referred to as $Cys\textsuperscript{null}$ GlyR (C41S/C290A/C345S, Chapter 4). In addition to using $Cys\textsuperscript{null}$ GlyR described previously, to study the desensitized state, two additional mutations are needed to study the open state (A288G/F207G), by allowing IVM to bind, and keep GlyR from desensitizing. After overexpression, the mutant GlyR was purified and crosslinking studies with MTS-bzp (Figure 7) were conducted. After MS and MS/MS analysis, sites of crosslinking were refined for each mutation in open and desensitized states and were then compared to resting state crosslinking data to gain insight into receptor dynamics. These three site-specific sites were chosen for their potential to study areas of the receptor not well-resolved, and due to
ease of analysis. It is hypothesized that these mutants will provide vital information to
discuss receptor dynamics as well as complement high-resolution techniques which will
lead to more complete structures, with a particular focus on the regions in the receptor that
are currently not well resolved.

5.2. Results and Discussion

The use of CX-MS and site-specific single cysteine mutations can provide useful
information on protein structure as it can complement high-resolution techniques and study
proteins in native environments.\textsuperscript{5,7} It also has the ability to provide information of protein
dynamics, as the protein is studied as a fully-functional protein and multiple
conformational states can be compared.\textsuperscript{10} The site-specific network of Cys mutations is
identical to what was proposed in Chapter 3: one mutant located in the transmembrane
region (M287C) and two mutants located in the extracellular domain (K116C, K206C).
Each of these mutant-containing GlyR is purified, reconstituted into lipid vesicles (+
cholesterol) which is essential for the function of pLGICs,\textsuperscript{14,394,601,607,608} and then enriched
in either resting, open or desensitized states before being crosslinked with MTS-bzp. By
using MS and MS/MS analysis, both intra- and intermolecular sites of crosslinker
attachment are identified. By comparing crosslinking results from open and desensitized
states to crosslinking results identified in resting states, dynamic information regarding
conformational changes in GlyR can be determined. Furthermore, by comparing the
resting, open and desensitized networks as a whole, with respect to each of the mutants in
the network gives structural information about the receptor in each state.

5.2.1. Position 287

5.2.1.1. Open and Desensitized State Results
Analysis of the open state revealed 25 individual sites of crosslinking. There is
distribution throughout the receptor, as was seen in the resting state, though the top of the
ECD is not observed to be crosslinked (Table 5, Figure 12). There were much fewer
crosslinks identified in the desensitized state (11 overall), but the distribution of
crosslinking in the ECD is similar to the open state, in that the top of the ECD (residues 1-33)
are not detected as being crosslinked (Table 6, Figure 12). The ECD interface is
crosslinked in both open and desensitized states, mostly the M2-M3 loop, though the Cys-
loop is crosslinked in the open state. In the open state, the M2-M3 loop is crosslinked

<table>
<thead>
<tr>
<th>M287C Open State Crosslinks.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The blue underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M287C Open State Intramolecular Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNIFLROQWNDPR</td>
<td>Pre M1</td>
</tr>
<tr>
<td>GAHFHEITDDNKLPR</td>
<td>51±1, 58±1</td>
</tr>
<tr>
<td>LLRISRGNVLYSIR</td>
<td>45, 36±0.3, 24±2</td>
</tr>
<tr>
<td>NGNVLYSIRITLIALPMDLK</td>
<td>45, 36±0.3, 24±2</td>
</tr>
<tr>
<td>ASLPKVSYSVK</td>
<td>M2-M3 loop</td>
</tr>
<tr>
<td>GANNSNHTNPPAPSKSPEEMRK</td>
<td>M3-M4 loop</td>
</tr>
<tr>
<td>LFQRAK</td>
<td>32±1</td>
</tr>
<tr>
<td>AKKIDK</td>
<td>26±0.3</td>
</tr>
<tr>
<td>IDKISR</td>
<td>23±2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M287C Open State Intermolecular Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPPVNVSANIFNSFGSIAETTMDYRVNIFLR</td>
<td>Pre M1</td>
</tr>
<tr>
<td>QOWNPRLAYNEYPPDSLPLDPSMLDSWKPDLFFANEK</td>
<td>64±2</td>
</tr>
<tr>
<td>EEKDLRYCTKHYNUTGK</td>
<td>M3</td>
</tr>
<tr>
<td>AIDIWCAVALLVFSALLEYAAVNFVSRQHK</td>
<td>M3-M4 loop</td>
</tr>
<tr>
<td>OHIKELLRR</td>
<td>M4</td>
</tr>
<tr>
<td>IGFPMAFLIFENMEYWIIYK</td>
<td>411</td>
</tr>
</tbody>
</table>
intramolecularly to the beginning of the loop $^{272}$ASLP$^{275}$, and the middle, $^{278}$SY$^{279}$, and not at all to the complementary subunit.

Table 6. M287C Desensitized State Crosslinks.
The table shows intramolecular and intermolecular mass-shifted peptides from $\geq 3$ separate trials, compiling only what was found in $\geq 2$ trials. The blue underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>M287C Desensitized State</th>
<th>Intramolecular Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>253 VGLGTTVLTMTTQSSGSR$^{271}$</td>
<td>M2</td>
<td>13±</td>
</tr>
<tr>
<td>384 AKKD$^{389}$</td>
<td>M3-M4 loop</td>
<td>29, 29, 25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intermolecular Location</th>
<th>Pre M1</th>
<th>37, 56±2</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 VNIFLRQOWNDPR$^{72}$</td>
<td>48±1</td>
<td></td>
</tr>
<tr>
<td>105 GAHFHEITDNKLLR$^{119}$</td>
<td>39±1</td>
<td></td>
</tr>
<tr>
<td>191 EKDLR$^{196}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>272 ASLP$^{275}$KSYV$^{281}$</td>
<td>M2-M3 loop</td>
<td>22, 18±1</td>
</tr>
<tr>
<td>310 QHK$^{316}$ELL</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The desensitized state shows intermolecular crosslinking to this region primarily near the end, $^{277}$VS$^{280}$V, and also to P275. Though it is difficult to determine whether the differences seen in these two states is indicative of this region’s involvement in gating, it does show mobility, which is congruent with what is seen in the literature.$^{238,304,310}$

The TM domain is heavily crosslinked in both M3 and M4 in the open state and in M2 in the desensitized state, as seen previously, M2 appears to be a region of GlyR that does not ionize well, and thus, the lack of fragmentation data leads to the inability to refine the position to a smaller area of amino acids. This is evident due to the number of counts seen in peptides from M2 in comparison to other peptides. Furthermore, the amount of fragmentation spectra is also fewer; we allow refinement data for a peptide that has four or more fragmentation ions per spectra. In both M3 and M4, there was evidence of multiple
binding sites, each species had different matched product-precursor sites at a particular retention time, and thus led to the entire region not resolving well.

The M3-M4 loop region, with the exception of the peptide \( ^{359} \text{NSNTTNPPPA}^{368} \) in the open state, is mostly crosslinked at the beginning and end of the loop, regions which were discussed previously, contain poly-R/K regions. The C-terminal tail is not crosslinked in either of these states, which differs from what was seen in the resting state, relative to position 287.

5.2.1.2. Dynamics Relative to Position 287.

Dynamic information can be revealed by analyzing the resting, open and desensitized states relative to position 287. Starting at the region furthest from the extracellular membrane surface (and using Figure 10 and 13), residues 1-33, are crosslinked much more so in the resting state than in ligand-bound states. It is likely that this is because in a true apo-state, GlyR is “bloomed”, and thus the ECD is wide at the top, perhaps bringing the top of the ECD closer towards the membrane, and 287C. It may also be due to relatively increased flexibility of the resting state; as this unliganded state may have a shallow energy profile and sample more conformational space. The peptides \(^{60} \text{VNIFLRQQWNNDPR}^{72}\), and the beginning of \(^{66} \text{QQWNDPRLAYNEYPDD…ANEK}^{104}\), which is part of \( \beta \)-sheet 2, is found crosslinked in all three states. In the resting state, it is found intramolecularly at L64 and \(^{69} \text{NDPR}^{72}\); whereas, in the open state, is found intramolecularly at \(^{62} \text{IFLRQQ}^{67}\) and \(^{69} \text{NDP}^{71}\) and intermolecularly at \(^{69} \text{NDPR}^{72}\). In the desensitized state it is only seen intermolecularly, at N61, and \(^{67} \text{QW}^{68}\). Non-conducting states interact with this area on different subunits, and the open state has the most contacts to this area. One explanation could be the that the receptor, in its active state, has ‘un-
bloomed’ and thus brought the middle β-sheets, where this peptide is, closer to position

Figure 13. M287C Resting, Open and Desensitized State Crosslinks. All states are represented by two subunits of GlyR (PDB: 3jad). The left subunit of each visualizes intramolecular crosslinks, and the right subunit, intermolecular crosslinks to position 287 (represented as a yellow sphere). All crosslinked areas are highlighted in blue. The bead models below each structure represent the M3-M4 loop, both intra- and intermolecularly. The C-terminal tail is not shown. Refer to Tables 2, 4 and 5.
287, then it would be in the resting state or the desensitized state, especially since both the principle and complementary subunits are crosslinked in the open state but not in non-conducting states.

The C-loop, which is known to be an area that is particularly mobile, also shows dynamics relative to position 287. In the resting state, the C-loop on the principle subunit is crosslinked. In the ligand-bound states, the complementary subunit’s C-loop is crosslinked. This could be indicative of the C-loop closing over the ligand in these states. Also, in the ligand-bound states, the position of crosslinking changes; in the active state, the very tip of the C-loop is crosslinked, and in the desensitized state, the middle and bottom of β9 is crosslinked. Because the length of the crosslinker is approximately 25 Å, small details are difficult to elucidate, but this does show the dynamics of this particular region, and though speculative, the tip of the C-loop moves closer to M3 in the open state. Another area known to be dynamic is the interface between the ECD and the TMD. Differential crosslinking patterns in the M2-M3 loop and the Cys-loop between the conformational states indicate mobility. The Cys-loop appears to be moving away from M3 as ligand binds, and the M2-M3 loop of principle and complementary subunits is closer to the pore in the resting state, as there is much more crosslinking to this loop in the resting state.

The TM domain also indicates movement when comparing the conformational states. In the resting state, M2 is crosslinked intermolecularly, and there is minor crosslinking seen in M3 in the principle subunit. In the open state, there is an increase in crosslinking seen in M3, and also the addition of crosslinking seen in M4. This suggests that relative motions in M3 and M4, as M4 is either moving closer to M3, or regions
become accessible to crosslinking upon channel gating, which correlates to other studies by Han et al.\textsuperscript{641} and Ferraro et al.\textsuperscript{14} Since this study is measuring Euclidian distances, connecting residues by a straight line, it does not take into account regions that are blocked sterically. Because of this assumption the actual distances between residues may differ. In the desensitized state, M2 is crosslinked to the principle subunit. This may suggest that the pore in the desensitized state could be wider than it is in the resting state, since in the resting state, the complementary subunit’s M2 α-helix is closer to position 287. The difference, according to structures determined by Du et al.\textsuperscript{296} suggests a change of 4-5 Å between a closed state and a possible desensitized state; however; the desensitized state is not well characterized. Also, the desensitized state, according to the dual-gate model, predicts that the TMD is wider at the top in the desensitized state. Based on these crosslinking results, there is no evidence to support or refute this.

The importance of using position 287 is highlighted in its ability to examine the intracellular M3-M4 loop. Crosslinking patterns during conformational changes can provide details into the role this loop may play in gating and desensitization. In the resting state, the loop is intimately associated with the membrane, more so than in any other state. The beginning of the loop, \textsuperscript{310}QH\textsuperscript{311} and \textsuperscript{315}LRFRK\textsuperscript{320} are crosslinked in the complementary subunit, and \textsuperscript{321}RR\textsuperscript{322} and \textsuperscript{324}HK\textsuperscript{325} in the principle subunit. The end of the loop, including the part of the loop that is part of M4 in available crystal structures is also crosslinked in the resting state, at \textsuperscript{369}PSK\textsuperscript{374} SPE, \textsuperscript{382}Q, \textsuperscript{384}AK\textsuperscript{385} and D\textsuperscript{388} to the principle subunit and \textsuperscript{371}K, \textsuperscript{377}EEM\textsuperscript{377}, and \textsuperscript{383}IQR\textsuperscript{383} to the complementary subunit. These crosslinks to the beginning and the end of this intracellular loop are within reach of the crosslinker, however; what is also seen in the resting state are crosslinks to the middle of
the loop; at $^{361}$NT$^{362}$ (intra-), R333, $^{352}$ISVK$^{355}$, and $^{356}$GAN$^{358}$ (inter-). This suggests that the middle of the loop is associating closer to the membrane as well in order to come close enough to interact with 287C. In the open state, there is intersubunit crosslinking to $^{359}$NSNTNPPPA$^{368}$, which is close to residues seen in the resting state. All of the other crosslinks seen in the open state are at the beginning or the end of the loop. The desensitized state only has minimal crosslinking to the loop, at A384, K386, and K388 on the principle subunit and $^{311}$HK$^{312}$ on a neighboring subunit. These data suggest that as GlyR transitions from its resting state to the open state, the M3-M4 loop moves away from the membrane. It also suggests that with respect to position 287, there is not much change in the desensitized loop structure from what is seen in the open state. It could also be because one of the desensitized state samples had a lower sequence coverage and thus, might have not included parts of the M3-M4 loop.

In summary, position 287 provided a look into the M3-M4 loop, an unresolved area in current crystal structures. The apo-state receptor is flexible, ‘bloomed’ and its intracellular loop is associating with the membrane. The conformational shift as the receptor transitions to a high-affinity ligand bound state involves ‘un-blooming’ of the ECD and capping of the C-loop. The ECD-TMD interface is dynamic and may move away from the pore. In the TMD region, M4 moves closer to M3. The open state to desensitized state transition is represented by mobility of the C-loop and mobility of the M2-M3 loop. With respect to position 287, there may not be much change detected during this transition.

5.2.2.2. Position 116.

5.2.2.1. Open and Desensitized State Results
Position 116 is located on the loop between β5 and β5’ in the ECD, and was chosen to be part of this network because it has been previously studied in Lys-Lys crosslinking studies and so has proven to not hinder the functionality of the receptor when crosslinked.\textsuperscript{32} There is a decrease in the amount of crosslinking data in the open state in comparison to the resting state of position 116. This could either be due to data matching protocols or perhaps the ionizability of the non-identified peptides, but it could represent that the apo-

Table 7. K116C Open State Crosslinks.
The table shows intramolecular and intermolecular mass-shifted peptides from $\geq 3$ separate trials, compiling only what was found in $\geq 2$ trials. The pink underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Ca of the residue crosslinked to the Ca of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre M1</td>
<td>N/A, 20±1, 20±2</td>
</tr>
<tr>
<td></td>
<td>11±1, 20±1</td>
</tr>
<tr>
<td>Pre M1</td>
<td>6, 16±2, 32±2</td>
</tr>
<tr>
<td></td>
<td>27±2, 24±2</td>
</tr>
<tr>
<td></td>
<td>28±1</td>
</tr>
<tr>
<td>M2-M3 loop</td>
<td>41±1, 45±1</td>
</tr>
<tr>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>84±1</td>
</tr>
<tr>
<td></td>
<td>80, 76</td>
</tr>
<tr>
<td></td>
<td>75±1</td>
</tr>
<tr>
<td>Pre M1</td>
<td>22±1</td>
</tr>
<tr>
<td></td>
<td>31±1</td>
</tr>
</tbody>
</table>

state is more flexible than the open state and/or the attached crosslinker has greater conformational flexibility.

A total of 19 intramolecular sites and 2 intermolecular sites were identified. The majority of them are found in the ECD, but as seen in the resting state, there is also
crosslinking identified in the M3-M4 loop. The Cys-loop, the M2-M3 loop and the C-loop are crosslinked; since all of these areas are known to be mobile, these crosslinks provide insight into dynamics. In the desensitized state, there are 32 identified crosslinks, and are more evenly distributed throughout the receptor, similar to the resting state though there is more crosslinking seen in the M3-M4 loop, primarily at the beginning and end of the loop.

As seen in the open state, there are also identified crosslinks to the Cys-loop, M2-M3 loop and the C-loop. M2 and M4 are crosslinked which was not present in the open state.
5.2.2.2. Dynamics Relative to Position 116.

Figure 14. K116C Resting, Open and Desensitized State Crosslinks. All states are represented by two subunits of GlyR (PDB: 3jad). The left subunit of each visualizes intramolecular crosslinks, and the right subunit, intermolecular crosslinks to position 116 (represented as a yellow sphere). All crosslinked areas are highlighted in pink. The bead models below each structure represents the M3-M4 loop, both intra- and intermolecularly. The C-terminal tail is not shown. Refer to Tables 3, 7 and 8.
Comparing the resting state to the open and desensitized states highlights some important information regarding receptor transitioning. The non-conducting states are more alike than the open state in the upper part of the ECD, residues 1-33. Resting state crosslinks include intramolecular $^6$KPMSPSDFL$^{14}$, $^{19}$GRTSGY$^{24}$, $^{27}$RI$^{28}$, and $^{31}$NF$^{32}$; and intermolecular $^1$ARS$^3$. In the desensitized state, there is crosslinking to A4, M18, intramolecularly and $^{21}$TSGYDARIRPNFK$^{23}$ intermolecularly. In the open state, there is only intramolecular crosslinking, $^3$SATK$^6$, $^9$SPS$^{11}$, $^{15}$DK$^{16}$. The lack of intermolecular crosslinking in the open state suggests that position 116 is farther away from the complementary subunit in the open state and then moves closer to that subunit in the desensitized state. $^{60}$VNIFLRQQWNDPR$^{72}$, of β2, is crosslinked within and between subunits, and is only observed intramolecularly in non-conducting states (Figures 11 and 14, Tables 3, 7 and 8).

Crosslinking to the C-loop is consistent with observations that this region is mobile.$^{238,304,318,343,344,348,642}$ In the resting state, β9 and half of the C-loop connected to it...
are crosslinked to position 116. In the open state, crosslinking to β10 and the connecting part of the C-loop are identified. In the desensitized state, crosslinking is only found at the end of β10, closest to the C-loop and throughout the C-loop (Figure 15). This suggests that the C-loop does move during ligand binding and may cap over the ligand in the open state. The shift in crosslinking in the desensitized state may be due to better refinement of the region in this state or that the bottom of β10 is no longer close enough to crosslink to position 116.

The ECD-TMD interface also shows dynamic changes between the states. In the resting state, both the Cys-loop and the M2-M3 loop are crosslinked intramolecularly and intermolecularly (Table 3). In the ligand-bound states, however; there is solely intrasubunit crosslinks to both loops. With respect to the Cys-loop, in the open state, $^{138}$CPMD$^{141}$ is crosslinked, and in the desensitized state, P139 is crosslinked. The resting state is shifted slightly to $^{141}$DLK$^{143}$, but also includes $^{138}$CPM$^{140}$ on the complementary subunit. In the M2-M3 loop, the lack of crosslinking to the complementary subunit in both ligand-bound states, suggests that either position 116 has moved farther away from the loops or the M2-M3 loop has moved farther away from position 116. Han et al. analyzed taurine and glycine binding to α1-GlyRs and showed near-identical motion of the M2-M3 loop in response to binding of either agonist, and they concluded that the loop moves in a predetermined fashion, regardless of the ligand bound.

M2 is heavily crosslinked in the resting state, on both principle and complementary subunits. M2 is crosslinked in the desensitized state, as is M4. However, in the open state, there is no crosslinking to the TMD (Figure 14 shows M4 being colored in the open state, but those residues are technically M3-M4 loop residues). This crosslinking pattern could
suggest that position 116 is not able to contact the TMD in the open state, due to the ‘un-
blooming’ conformation that results. Since position 116 is an interior mutant, in may be in
a position that is somehow blocked from reaching the TMD.

Another difference seen between the states is the involvement of the C-terminal
tail, 412IVRREDVHNQ421. This is an important region, since it is not resolved in current
crystal structures and is involved in the modulation of activity by Zn2+.644 It is only
intermolecularly crosslinked in the desensitized state. In position 287, this region was only
crosslinked in the resting state and position 206 sees this in the resting and desensitized
state. Future work involves analyzing I412C, so this mutant might help to reciprocally
discern allosteric motions in this C-terminal tail of the receptor. However, the fact that it is
never seen in the open state may have something to do with the ‘un-blooming’
conformation, or the lack of involvement of the tail in gating.

The intracellular loop also involves the desensitized state more than the other states,
as it has crosslinking to this region both intra- and intermolecularly. All states show
crosslinking to the beginning of the loop. The non-conducting states also show crosslinking
to the end, but on different subunits. A detailed discussion of the intracellular loop can be
found in Section 5.4.

Position 116, since it is located in the interior of the ECD, provides a unique
perspective into what may happen as ligand binds. This data is consistent with C-loop
mobility, and suggests that as glycine binds, the C-loop may cap over it. The open to
desensitized state transition also shows mobility of this region. The receptor has higher
affinity to glycine in the desensitized state,236,631,637 and interactions made with the C-loop,
as well as the other loops in the LBD may contribute to this. A shorter crosslinker, as well
as the examination of position 206 may help to provide more details into the differences between the open and desensitized state in terms of movement of the C-loop region.

5.2.3. Position 206

5.2.3.1. Open and Desensitized State Results

Table 9. K206C Open State Crosslinks.
The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The green underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>K206C Open State</th>
<th>Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intramolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3^SATKPMSPDFLDKLMGR20</td>
<td>Pre M1</td>
<td>41±1 (9-12 only)</td>
</tr>
<tr>
<td>10^GAHFHEITTDNKLIRISR122</td>
<td></td>
<td>33±1</td>
</tr>
<tr>
<td>12^NGNVLYSIRITLTLACMDLK143</td>
<td></td>
<td>27±1</td>
</tr>
<tr>
<td>32^RRHHK125</td>
<td>M3-M4 loop</td>
<td>N/A</td>
</tr>
<tr>
<td>378^KLQRAKK386</td>
<td></td>
<td>80±1</td>
</tr>
<tr>
<td>Intermolecular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17^LMGRTSGYDARIIRPNFK33</td>
<td>Pre M1</td>
<td>34±1</td>
</tr>
<tr>
<td>132^ITLTLACMDLK143</td>
<td></td>
<td>28±2</td>
</tr>
<tr>
<td>201^HYNTGCATCIEARFHILER218</td>
<td></td>
<td>41±2, 30</td>
</tr>
</tbody>
</table>

Similar to position 116, there are less crosslinks identified in the open state than in the other two states. Only 9 crosslinks were identified in the open state, compared to 31 in the desensitized state and 25 in the resting state. In the open state, the crosslinking is primarily located in residues before M1, in the ECD and ECD-TMD interface. The upper part of the ECD, residues 1-33 are crosslinked. The region where 206 is located is also crosslinked, 201HYNTGCATC209, though only to the complementary subunit (Table 9). The only other crosslinks identified are found in the M3-M4 loop. With respect to the distance measurements, there are none that were found less than the crosslinker length,
which suggests that the open state is not well-represented by the crystal structure. Results from the desensitized state are more consistent to what was seen in the resting state, as

Table 10. K206C Desensitized State Crosslinks. The table shows intramolecular and intermolecular mass-shifted peptides from ≥3 separate trials, compiling only what was found in ≥2 trials. The green underlined areas represent the sites of refinement. The average distance ± standard error measurement (SEM) column shows the distance from the site of mutation to the refined areas. Areas that are refined to a single amino acid are represented by a single distance; whereas, refinement to multiple amino acids is represented as a mean and the standard error. Residues from 310-377 are not resolved in crystal structures, and thus no measurement data is available. Note that the distance between Cα of the residue crosslinked to the Cα of the mutation is approximately 25 Å.

<table>
<thead>
<tr>
<th>K206C Desensitized State</th>
<th>Intramolecular</th>
<th>Location</th>
<th>Avg. Dist. ±SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARSATKPMSPSDFLDKLMGR20</td>
<td>Pre M1</td>
<td>N/A, 40, 32±1</td>
<td></td>
</tr>
<tr>
<td>TSGYDARIRPNFK32</td>
<td></td>
<td>22±1</td>
<td></td>
</tr>
<tr>
<td>GAHFHEITTDNKLRRISR122</td>
<td></td>
<td>32±2</td>
<td></td>
</tr>
<tr>
<td>ISRNGNVLYSIRITLTLACPMDLK143</td>
<td></td>
<td>26±1, 25±1</td>
<td></td>
</tr>
<tr>
<td>EEKDLRyCTK200</td>
<td></td>
<td>20±2</td>
<td></td>
</tr>
<tr>
<td>ASLPKVSYYVK281</td>
<td>M2-M3 loop</td>
<td>38±1</td>
<td></td>
</tr>
<tr>
<td>QHKELLREFR318</td>
<td>M3-M4 loop</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>RHKKEDEAQER333</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>FNSAYGMPASLQAK349</td>
<td></td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>DGISVKGANNSSNTPPAPSK371</td>
<td></td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>SPEEMR377</td>
<td></td>
<td>81, 79±1, 73</td>
<td></td>
</tr>
<tr>
<td>KLFIORAKK386</td>
<td>M3-M4 loop</td>
<td>73±1, 67±1</td>
<td></td>
</tr>
<tr>
<td>IvrRedvHnoQ421</td>
<td>C-term tail</td>
<td>37 (412-413 only)</td>
<td></td>
</tr>
<tr>
<td>ARSATKPMSPSDFLDKLMGR20</td>
<td>Pre M1</td>
<td>N/A, 32±2</td>
<td></td>
</tr>
<tr>
<td>EEKDLRyCTK200</td>
<td></td>
<td>36±1</td>
<td></td>
</tr>
<tr>
<td>VGLGITTIVLMTTQSSGSRASLPKVSYYVK281</td>
<td>M2, M2-M3 loop</td>
<td>64±2, 48±1, 49</td>
<td></td>
</tr>
<tr>
<td>RRHK325</td>
<td>M3-M4 loop</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

more regions of the receptor are represented. Similar to the resting and open states, the upper part of the ECD (residues 1-33) is crosslinked. Again, as seen in the open state, the C-loop is crosslinked to the complementary subunit. This is interesting, as the resting state does not crosslink to this area, so there is similarity between the ligand-bound states. In
Figure 16. K206C Resting, Open and Desensitized State Crosslinks. All states are represented by two subunits of GlyR (PDB: 3jad). The left subunit of each visualizes intramolecular crosslinks, and the right subunit, intermolecular crosslinks to position 206 (represented as a yellow sphere). All crosslinked areas are colored green. The bead models below each structure represents the M3-M4 loop, both intra- and intermolecularly. The C-terminal tail is not shown. Refer to Tables 4, 9 and 10.

terms of the ECD-TMD interface, the desensitized state is more similar to the resting state
than to the open state. Both non-conducting state show intramolecular crosslinking to the Cys-loop, and crosslinking to the M2-M3 loop, though the desensitized state interacts with both subunits in this area. As mentioned previously, the C-terminal tail is crosslinked to the principle subunit. With respect to position 116, the C-terminal tail of the complementary subunit was crosslinked in the desensitized state. And unlike position 116, the tail is also crosslinked in the resting state, similar to position 287. The most striking result from this state, though, with respect to this position are the unique crosslinking sites identified in the M3-M4 loop region. There are regions crosslinked here that are not found in any other state or mutant in the analyzed network (See Section 4.5).

5.2.3.2. Dynamics Relative to Position 206

Position 206 is located on β10, part of the β-sheets that connect the C-loop, and cap over the ligand once bound. Thus, it is hypothesized to be a highly mobile region, and so there was an expectation for visualizing protein dynamics. The first 16 residues, 1ARSATKPMSPSDFLDK16 are crosslinked to the principle subunit in all three states; unrefined to any particular residues in the resting state, 3SATKPMSPSD12 in the open state and 7PM8, and P10 in the desensitized state. The complementary subunit only has crosslinking to this region in the desensitized state, at 1ARSA4. The peptide 17LMGRTSGYDARIR33, is seen crosslinked to the complementary subunit in both the resting and open states, 18MG19, 31RI32 and 17-32, respectively; but in both principle (17LMG19, 21TSGYD25) and complementary subunits (18MGR20) in the desensitized states. Though this region has not been reported to be mobile, it is consistently seen interacting in various patterns with all of the analyzed mutants. Position 206 is in a mobile location, but
positions 287 and 116 are not and so, this region is hypothesized to be an area that may have mobility (Section 5.5).

Another important part of the ECD is position 116. Position 206 crosslinks to 116 in the complementary subunit of the resting state and the principle subunit of the open state, but is not seen in the desensitized state at all. An explanation for this might be the length of the crosslinker; perhaps in the desensitized state, and the principle subunit of the resting state, the two residues are too close to one another. This may also be why the C-loop, the region where position 206 is located is not seen in the resting state, but is crosslinked in the complementary state in both ligand-bound states. According to studies involving the transition between the resting state to the pre-active state, the ‘flip’ state, the motions of the C-loop extend to the β1-β2 loop (50SIAETT55) in the ECD-TMD interface in GlyRs.238,352 What is interesting is that none of the mutants analyzed crosslink to this area; though position 287 comes closest, crosslinking to 57DYR59 and Y58 in the complementary subunit of the resting and open states, respectively. And so, there is no evidence from this data to show involvement of this loop in the transition between resting and open states.

The absence of crosslinking in the open state to the TMD is similar to what was seen with respect to position 116. Because both ECD mutant do not contact the TMD in the open state, this suggests that the orientation of these mutants is somehow blocked from reaching the TMD or is out of the reach of the crosslinker. The interaction to the TMD in non-conducting states, for both positions 116 and 206 suggests that there is a conformational change from resting to open and from open to desensitized that involves the accessibility and inaccessibility of the TMD with respect to these mutants.
The M3-M4 loop, though unexpected to be crosslinked to, may suggest an interesting role for the loop. Like position 116, there are few crosslinks in the open state. However, dissimilar to position 116, there is more crosslinking found in the non-conducting states. More so, the crosslinking found is not just at the beginning and end of the loop, which may make sense to be crosslinked given a larger tolerance of the crosslinker length, but is found in the middle, in unique areas not seen when analyzing other mutants. The resting state crosslinking sees the most unique areas, and is primarily crosslinked intermolecularly. In the desensitized state, the crosslinking is primarily found to the principle subunit. In position 287, the majority of crosslinking was found in the resting state, leading to the suggestion that the loop’s intimate association with the membrane in this state may play a structural role in a non-conducting receptor. Position 206 seems to suggest similarly, that the non-conducting states involve the loop, more so than the active state does.

5.2.4. Distance Data

After the crosslinking data was mapped onto the crystal structure (PDB:3jad), the distance between Ca of either position 287, 116 or 206 to the Ca at the site of crosslinking was measured. This information was useful in determining model accuracy with respect to the mutations analyzed. Using the same crystal structure for all states and all of the mutations provided uniformity, however; the structure used was the strychnine-bound closed state, and so it is unlikely to match entirely to the data, as the structure is not a true apo-state. The measurements were grouped into three categories: <20 Å, 20 – 30 Å, and >30 Å, to show crosslinks that were less than the length of the MTS-bzp crosslinker, within a suitable range based on the approximate 25 Å length, and greater than the length of MTS-
bzp, respectively. Generally, a tolerance of approximately 3-5 Å is applied to the length of the crosslinker used to account for molecule flexibility.\textsuperscript{5,128,645,646} A measurement that was out of the range of the crosslinker, the third category, is perceived to show model inaccuracy, and also protein dynamics.

What was immediately obvious was that many of the crosslinks were out of the range of the crosslinker’s length (Table 11), greater than 40% of them for all positions and states. While it is expected that some of the distances would be out-of-bounds due to the capability of CX-Ms to study protein dynamics, it was unexpected that the majority of the measurements (in most cases) would be. However, studies by Merkley \textit{et al.},\textsuperscript{646} Kahraman \textit{et al.},\textsuperscript{123} and Ryl \textit{et al.}\textsuperscript{20} have noted a considerable range of out-of-bounds crosslinking distances in their studies. According to these studies, it is likely the out-of-bounds crosslinking is due only to protein dynamics that are unable to be detected in experimental structures and not the formation of protein aggregates.\textsuperscript{20,123,646,647} It is also suggested that a larger length tolerance should be considered, thus decreasing the out-of-bounds crosslinking percentage.\textsuperscript{646}

Since the structure used does not adequately define any of the analyzed states, but is closest, perhaps, to the resting state, the data was expected to closely fit this structure.
However, that was not the case. The resting states of all analyzed mutations had the highest percentage of out-of-bounds crosslinks. This is most-likely due to the resting state being a more flexible state than the ligand-bound states.\textsuperscript{243} Because the structure is bound to strychnine, it is far less flexible than it would have been.

By totaling up the crosslinks that are <20 Å and those that are between 20 – 30 Å, the desensitized states, of all mutants has the highest percentage. Again, this is not to say that the structure used to map these results is a desensitized structure, but that the results from the desensitized state crosslinks have the best fit. What this could mean is that a desensitized state, from the perspective of position 287, 116 and 206 best fits an antagonist-bound closed state, which is what PDB:3jad is representative of. Though studies have shown that dynamics due to glycine versus strychnine binding differ,\textsuperscript{642} there may be a similar end result that was captured conformationally in the crystal structures, and so, an antagonist-bound state may be similar to what can be visualized by the positions analyzed.

The most important point is that the high percentage of out-of-bounds crosslinks is indicative of protein dynamics and that dynamic information cannot be properly mapped onto a static image. Crystal structures cannot represent dynamics, and thus, to understand how a protein is transitioning between its states, structural studies that include dynamics should be performed.

5.3. Network Integration

In Chapter 3, the resting state was integrated with respect to position 287, 116 and 206. Of note, the resting state showed contacts to the upper part of the ECD, residues 1-33 with respect to all mutants, lending credibility that this region may be especially mobile, or that the resting state is more flexible. Analysis of regions where the ECD mutants were
located showed that they were crosslinked by each other and by 287, but not themselves. This is most likely due to crosslinker length, but it also is indicative of a network that provides an important perspective into studying structure, and a good starting point.

Investigating the resting state with CX-MS showed compatibility with what is reported in the literature of dynamic areas. The C-loop, Cys-loop and M2-M3 loop show a difference in refinement patterns that is consistent with the mobility of these loops. The resting state also showed similarity in the TMD region from the ECD mutants, and the identified crosslinking found in the unresolved M3-M4 loop region suggest an intimate association with the membrane. Integrating the open and desensitized states will provide more context into overall receptor dynamics.

5.3.1. Open State

Analysis of the open state reveals several important differences from what was seen in the resting state. The upper ECD, residues 1-33, are no longer crosslinked with respect

Figure 17. Comparison of Open State Crosslinking. To show the differences between crosslinking data the open state crosslinking results for network mutants is shown. On the left, in blue, are crosslinking data for M287C. In the middle, in pink, are the results from K116C, and on the right in results from K206C. Larger images with detailed explanation can be found in Figures 13, 14 and 16, respectively.
to position 287, but are still crosslinked to the ECD mutants. It is speculated that the ‘bloomed’ configuration of the resting state changes to an ‘un-bloomed’ configuration as glycine binds, and this may mean that the top of the ECD is no longer close enough to crosslink to M3, where 287 is located.

There are also differences seen in the TMD and the M3-M4 loop. M2, M3 and M4 were crosslinked to position 287 in the resting state, but positions 116 and 206 only saw crosslinking to M2. In the open state, there is still heavy crosslinking to position 287 to M3 and M4, but not M2. From the perspective of the ECD mutants, however, there is no crosslinking to the TMD. It is unclear why these mutants are able to crosslink to the helix that lines the pore in an apo-state, but not in an active state. Could it be that somehow the crosslinker is blocked from reaching the pore in an active state? Further work must be conducted to gain more insight. It is also interesting that position 287 no longer crosslinks to M2, but is heavily crosslinked to the complementary M3 and M4. According to studies that proposed mechanisms for channel opening, there is a tilting of the M2 helices to widen at the top to allow ion flow. The ability for the pore to widen from approximately 2 Å to 6–7 Å is due to limited contacts M2 makes with M1 and M3, and a cavity in between the helices that holds water or cholesterol. One perspective from this data, with respect to position 287, could be that if M2 is widening at the top and also removing the water and/or cholesterol from the cavity to make room for ion flow, the distance between the top of M2 and the top of M3 becomes so small that the crosslinker is unable to crosslink; where in the resting state, the crosslinker was able to crosslink to M2. Specific studies should be conducted on MTS-bzp to see both the minimum and maximum distance it can reach.
The M3-M4 loop shows major differences also. In the resting state, there were many crosslinks to this region, from all mutants in the network, and there is a dramatic decrease in the amount of crosslinking to the loop in the open state. With respect to the ECD mutants, there is crosslinking to the loop only at the beginning of the principle loop; 311HKELL 315 and 321RRHH 324 in position 116, and 321RRH 323 in position 206. Position 287 sees more crosslinking than the ECD mutants but still much less than it did in the resting state; to the end of the loop in the principle subunit, and the beginning of the loop in the complementary subunit. It appears that the loop is moving away from the membrane as the receptor is activated. Whether this means that the loop does play a role in the structural stability of an apo-state receptor is unclear. Further work on this loop is necessary to discern its importance in stability, gating and activation.

The ECD-TMD interface, because it is known to play a role in transducing signal from the LBD to the TMD, is an important area for visualizing dynamics in an active state. The Cys-loop, with respect to the ECD mutants, shows a decrease in crosslinking from the resting state to this region. In fact, position 206 does not crosslink to this region at all in the open state. Position 116 interacts with both the principle and complementary loops in the resting state, but only the principle loop in the open state. Position 287 only sees the principle subunit in both states. The M2-M3 loop shows a similar decrease in crosslinking from the resting to open states for all positions. It is unclear whether this is indicative of transducing signal, but it does show an overall mobility of this region.

Lastly, studies have shown that the β-sheets of the ECD move as rigid bodies pulling the ECD closer together as the receptor is activated. Examination of β-sheets 1 through 6 of the ECD in both resting and open states does show movement of these regions,
but it is inconclusive if the ECD is becoming closer together. With respect to position 287, crosslinking is identified in β2, β4 and β5 in the principle subunit, and β1, β2 and β5 in the complementary subunit. In the open state, however; there is crosslinking only to β2, β5 and β6 in the principle subunit and β2 in the complementary subunit. This could suggest that the ECD is coming closer and thus, the crosslinker length, being approximately 25 Å is too long and thus insufficient to crosslink to these sheets, but it could mean these areas were not identified in these studies. Similar results can be seen for both position 116 and 206, and thus, this shows mobility of the region, but is not conclusive towards a constriction of the ECD.

Overall, the major changes between the resting and open states are found in the upper ECD, the TMD and the M3-M4 loop. Fewer contacts are made to these regions, with respect to all mutants. This may suggest an ‘un-bloomed’ conformation due to the changes in the upper ECD with respect to position 287. This may also suggest that the M3-M4 loop moves away from the membrane during the active state.

5.3.2. Desensitized State

Analysis of the desensitized state reveals similarities to the open state, which is interesting as both are ligand-bound states, for instance, the upper ECD, residues 1-33, are crosslinked only to positions 116 and 206, but not position 287. This may suggest that with respect to this region, there is not much change as the receptor transitions between the open and desensitized state.

An area that was expected to change overall would be the C-loop, and the β-sheets that surround it, β9 and β10. This region consists of $^{185}$PGFILKEEKDLRY$^{197}$ for β9, $^{198}$CTKHYNTGKFTC$^{209}$ for the loop, and $^{210}$IEARFHLER$^{218}$ for β10. In the resting state,
there were multiple contacts in the principle subunit for both positions 287 and 116; 193KDLR196, C198, 203NT204, 206KF207, 213RFH215 and 195LRYCTKHYN203, 215HLER218, respectively. However, there was no contact in position 206, though this could be because K206 resides in this region. Only position 116 crosslinked to the complementary subunit, at 192EKDLR196 in the resting state. In the open state, where it is proposed that the C-loop caps over the ligand, there was no change seen by position 206. For 287, however; all the contacts seen in the resting state were no longer present, but the complementary subunit saw crosslinking to 200KH201 and 204TG205. In position 116, the complementary subunit crosslinking is identical, but in the principle subunit, there are subtle differences, 204TGKATC209 (note that position 207 is one of two mutations that are added to allow IVM to bind and thus trap GlyR in its open state), 212AR213 and 214FHLER218. In the desensitized state, GlyR has a higher affinity to its ligand,386,649 and so it is expected that there may be some change in proximity of certain residues in the loop with respect to the ligand resulting in stronger hydrogen bonding. According to Brejc et al.,417 the ligand makes contacts at Y202, and close to K200. With respect to position 116, crosslinking is seen to both principle and complementary subunits, at N203, 205GK206, and 191EEKDLR196, 210IE211, respectively. In 287, similar to the open state, no contacts are made to the principle subunit, but there is intermolecular crosslinking found at 191EEKDLR196. The biggest change is position 206 which crosslinks both intra- and intermolecularly, which wasn’t seen in the other two states, at 193KDLRY197 and 198CT199. Though these data cannot provide evidence of minute changes that might result in higher affinity binding, the fact that there is a difference in crosslinking patterns does show this loop is moving in all states.
The ECD-TMD interface also shows dynamics in this state as well. In the open state of position 206, there was no crosslinking seen in the Cys-loop or the M2-M3 loop, but in the desensitized state, there is contact to both. In contrast, for position 287, the Cys-loop crosslinking is no longer present, and in the M2-M3 loop, there is a shift in crosslinking seen from the principle subunit to the complementary one. Position 116 shows only subtle changes in both loops. The ability to crosslink to these regions, particularly from 206 and 287 is closer to what was seen in the resting state. This may suggest that the interface, with respect to this network, looks similar in non-conducting receptors, despite the fact that the ligand is still bound. Perhaps, the energy added due to ligand binding initially causes the loops in the interface to move, but then they revert back to a more original location once the receptor desensitizes.

Figure 18. Comparison of Desensitized State Crosslinking. To show the differences between crosslinking data the desensitized state crosslinking results for network mutants is shown. On the left, in blue, are crosslinking data for M287C. In the middle, in pink, are the results from K116C, and on the right in results from K206C. Larger images with detailed explanation can be found in Figures 13, 14 and 16, respectively.

This similarity to the resting state is also seen in the TMD. Once again, there is crosslinking to M2, which was seen earlier. In the open state, the ECD mutants did not
crosslink to the TMD at all. In context of all the conformational states, it is interesting that M1 was never able to be crosslinked, by any of the mutants in the network. Is it because the region doesn’t ionize well? If a mutation could be placed here, and result in a functional receptor, it would be valuable to expand the network to include this region.

What is most striking about the desensitized state are the unique crosslinking patterns seen in the M3-M4 loop region. Position 206 sees unique sites in the loop that are not identified in any other state or by any other mutant. The reasons for this are currently unknown, but will be further discussed in the next section.

Integrating the network in all of the states has identified dynamic regions within the receptor, The M3-M4 loop, the C-loop, the loops in the ECD-TMD interface, and perhaps even the upper residues of the ECD. This integration has also provided proof that this methodology is able to show dynamics. As the receptor transitions between resting, open and desensitized states, unique patterns emerge for each conformation, even in the resting state. This shows that the receptor can be effectively trapped in each state.

5.4. The Intracellular Loop

The M3-M4 loop is multifaceted in terms of its functions and modulations; it mediates specific interactions with intracellular binding partners, is targeted by PTMs, modulated by positive allosteric ligands, and it plays a role in channel structural integrity, gating, desensitization and receptor surface expression. It distinguishes the pLGICs from one another, and determines the specific ion channel properties. Because this loop is truncated in crystal structures, its structure and the interaction it may have with the rest of the receptor is lacking. This work has been able to visualize the loop in resting, open and desensitized states, and this information can be used
to help better understand how the loop interacts with the receptor, as well as with the membrane in its conformational states.

The M3-M4 loop of α1-GlyR is approximately 70 residues, which is about 20% of the total mass of the receptor (Figure 19). Each of the mutations in the network was able to crosslink to the loop in all states. From the perspective of position 287, this was expected. However, the crosslinking of the ECD mutants to this loop was surprising. One explanation of the results could be due to poly-R/K regions. Studies have shown that peptides containing Arg and Lys residues have the ability to penetrate the membrane.614,616,619–621,650 These peptides have been classified as cell-penetrating peptides (CPPs),613,616,617,619–621 and are commonly grouped into three classifications; protein derived CPPs, which are sometimes also called protein transduction domains; model CPPs and designed CPPs.616

With respect to this work, only protein-derived CPPs will be discussed here, for information on other types see Zorko et al.,616 Lindgren et al.651 and Gräslund et al.652. Protein-derived CPPs come from short sequences within proteins that have translocation

Figure 19. Intracellular M3-M4 Loop of α1-GlyR. The bead model depicts the amino acid residues that makes up the intracellular loop of α1-GlyR. The loop is not resolved in available crystal structures. The blue lines represent commonly found peptide pieces, with the exception of 317FRRKRHRHK625 which has multiple R and K residues and thus, multiple tryptic pieces could result. It is important to note that there are other peptide pieces that could produce different tryptic pieces, as well. The last two peptides, KLFIQRAK and KIDKISR are represented on structure determinations PDB: 3jad, 3jae and 3jaf (strychnine-bound, glycine-bound and glycine/IVM-bound) but are part of the M4 helix and not as part of the loop due to a 16 amino acid shift of the sequence. Because of this, the bead model for the crosslinking results seen in M287C, K116C and K206C resting, open and desensitized results only include up to residue R377.
properties. The significance of CPPs, other than providing a path for information to be brought into the cell, has evolved into a method for drug delivery.

A wealth of knowledge regarding CPPs came from studying the HIV-Tat protein. Tat (trans-activator of transcription) is a protein that is known to be involved in the replication of human immunodeficiency virus type-1 (HIV-1). In order for the virus to replicate itself it needs a way to enter the plasma membrane and the nuclear membrane of the cell, and studies have shown that the Tat protein is responsible for facilitating cell uptake. The specific sequence is, CGRKRRQRRPPQC. Another early identified CPP was derived from Antennapedia homeodomain of *Drosophila*, called penetratin; its sequence is RQIKIWFQNRRMKWKK. Note that both peptide sequences contain Arg and Lys residues, but do not have any other specific motif. This holds true for all protein-derived CPPs; they are all short (< 30 residue) peptides that are both basic and amphipathic, but other than these two characteristics can be quite variable.

The mechanism of how these basic, amphipathic sequences penetrate the membrane is unclear, but analysis of both Tat and penetratin suggests that the positively charged peptide interacts with negatively charged phospholipids and that creates an inverted micellar structure that then opens the membrane inward or outward. Other researchers have proposed that poly-R groups interact more strongly than poly-K groups with phospholipid bilayers, particularly those containing PC and PG. This may be due to the guanidinium group and its ability to form up to five strong hydrogen bonds to carboxyl, phosphoryl and sulfuryl groups that are in carbohydrates and phospholipids.
Keeping this in mind, the crosslinking results for the M3-M4 loop suggest the possibility that sequences within the loop may act as CPPs. If the loop is divided purposely into commonly digested trypsin fragments, there are a total of 9 peptides (Figure 18), but two have been put together and referred to as the poly-R/K region; they are as follows: (1) \(310^\text{QHKELLR}^{316}\), (2) \(317^\text{FRRKRRHHK}^{325}\) (poly-R/K region), (3) \(326^\text{EDEAGEGR}^{333}\) (4) \(334^\text{FNFSAYGMGPACLQAK}^{349}\), (5) \(350^\text{DGISVK}^{355}\), (6) \(356^\text{GANNSNTTNPPPAPS}^{371}\), (7) \(372^\text{SPEEMR}^{377}\), (8) \(378^\text{KLFIQRAK}^{385}\) and (9) \(386^\text{KIDKISR}^{392}\). Note that peptides 8 and 9 are part of M4 in available crystal structures. Also, though peptide 2 shows the highest concentrated area of Arg and Lys residues, there are other peptides that also have these basic residues and thus, may display cell-penetrating behavior.

Figure 19 plots the distribution of crosslinking seen in the M3-M4 loop in resting, open and desensitized states for M287C, K116C and K206C, with respect to each commonly trypsin-digested peptide fragment. Relative to all mutants, there was the least amount of crosslinking seen in the open state, which could be due to the loop moving away from the membrane as the receptor activates. There was the most crosslinking seen in the resting state; this could be due to the loop providing stability to an apo-state receptor.

Analyzing the results with respect to each mutation identifies differences between patterns seen. Crosslinking results for K206C is broadly distributed throughout the loop, and is seen at least once in each peptide region. M287C crosslinking results are skewed more towards the resting state, but it does crosslink uniquely to two regions in the open state, \((356-371 \text{ and } 372-377)\) which are not seen with respect to the other mutants. K116C sees the fewest crosslinking results overall, but most of them are found in the desensitized
state. It is likely that position 116 is in a less mobile region that position 206, and this could

Figure 20. Total Intracellular M3-M4 Loop Crosslinks. This chart represents the total number of crosslinks found in the M3-M4 loop and the distribution pattern of the crosslinking results. Each commonly trypsin-digested peptide region is represented by residue number on the x-axis, as are resting, open and desensitized states. The colored bars indicate where crosslinking sites were identified, blue for M287C, pink for K116C and green for K206C. The bar size indicates how much crosslinking was identified per region and per mutation. If the crosslinking was identified in either the principle or complementary subunit, the bar is shorter; if crosslinking was identified in both principle and complementary subunits, the bar doubled in size. The maximum count a region could receive is 6, for both subunits having crosslinking for all mutations. If crosslinking was not seen in a particular region at all, that state received a ‘zero’. Figure 18 identifies the residues for all peptide regions.

<table>
<thead>
<tr>
<th>Residue Region</th>
<th>Resting</th>
<th>Open</th>
<th>Desensitized</th>
</tr>
</thead>
<tbody>
<tr>
<td>303-356</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>317-325</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>316-333</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>348-409</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>350-405</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>356-535</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>357-531</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>372-377</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>373-385</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>385-509</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>396-509</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M3-M4 Loop Crosslinking
be why it sees less crosslinking to the intracellular loop.

With respect to the peptide regions, there is a greater amount of crosslinking identified at the beginning (\textsuperscript{310}QHKELL\textsuperscript{316} and \textsuperscript{317}FRRKRRHHK\textsuperscript{325}) and the end of the loop (\textsuperscript{378}KLFIQRAK\textsuperscript{385} and \textsuperscript{386}KIDKISR\textsuperscript{392}), for all mutations and in all states. The regions that have the highest amount of crosslinking overall is \textsuperscript{372}SPEEMR\textsuperscript{377} and \textsuperscript{378}KLFIQRAK\textsuperscript{385}. The reasoning behind this, based on the sequence, could be due to CPP-like behavior. Peptide 2, in particular, consists of most of the basic stretch in \textalpha\textsubscript{1}-GlyRs, \textsuperscript{316}RFRRKRR\textsuperscript{322} which according to Sadtler et al.\textsuperscript{420} contains the NLS and binds to intracellular proteins to allow its entry into the nucleus. It is also binds to the G-protein subunit G\textbeta\gamma, as does a di-Lys region, \textsuperscript{385}KK\textsuperscript{386}.\textsuperscript{388,422} Interestingly enough, this interaction has been shown to enhance glycine induced current,\textsuperscript{423} and mutations to these residues results in a non-functional receptor.\textsuperscript{419} The basic stretch also binds allosteric modulators, like ethanol and cannabinoids,\textsuperscript{424-427} These modulators have also been proposed to bind to the TMD region at M1, M2 and M3,\textsuperscript{390,462,475-479} and though this does not mean the M3-M4 loop is binding to these modulators in a hydrophobic environment like the membrane, but it is interesting and requires further study.\textsuperscript{479} and though this does not mean the M3-M4 loop is binding to these modulators in a hydrophobic environment like the membrane, but it is interesting and requires further study.

Another interesting peptide, parts of peptides 5 and 6, \textsuperscript{358}NNSNTTNPPAPSK\textsuperscript{372} has been proposed to play a role in desensitization, and folding stability in \textalpha\textsubscript{3}-GlyRs.\textsuperscript{405} These regions are also identified as common motifs, the poly-asparagine (\textsuperscript{358}NNSNTTN\textsuperscript{364}),\textsuperscript{388} and the polyproline motifs (\textsuperscript{366}PPAP\textsuperscript{371}), which may have helical
structure\textsuperscript{387,428} and also forms the SH3 consensus sequence.\textsuperscript{429,430} Crosslinking to these regions was identified by positions 287 and 206 in the resting state, only 287 in the open state, and 116 and 206 in the desensitized state. It is unclear why these regions are visualized in varying patterns by the network, and could simply mean due to poly-R/K regions in the loop, they are brought into close enough contact.

K206C also saw crosslinking to two unique peptides, \textsuperscript{326}EDEAGEGR\textsuperscript{333} and \textsuperscript{334}FNFSAYGMGPACLQAK\textsuperscript{349}, not identified by any other mutant, and only seen in desensitized state results. Those these two peptides do not have poly-R/K regions, they are relatively close, sequentially to peptide 2, and that could be why they were able to be brought into close enough proximity to crosslink to position 206. But why only in the desensitized state? This is currently unclear and needs to be studied further.

Phosphorylation of the loop may also play a role in how the loop interacts with the membrane and the rest of the receptor. \textalpha 1-GlyRs have a phosphorylation site at S391.\textsuperscript{445} Studies that analyze phosphorylation of this site may shed light on how the loop changes with the addition of a large negative charge. The addition of single site-specific cysteine mutations to the M3-M4 loop may also increase the knowledge of how the loop is moving as the receptor shifts allosterically and how it interacts with the membrane.

5.5. Mapping Mobility

CX-MS, because full-length, functional proteins are used in its analysis, can study protein dynamics. This cannot be visualized with other high-resolution techniques and so provides additional information that should be taken into account as current structural models are refined, or as new ones are built. Multiple studies have shown dynamic areas to be the C-loop, the ECD-TMD interface, and likely, the M3-M4 loop.\textsuperscript{238,304,388,389} This
work is in agreement with these studies (Figure 21). The C-loop, and the β-sheets that are

Figure 21. Proposed mobile regions of α1-GlyR. Analysis of all regions with respect to all mutations and conformational states reveals various regions that show mobility. In order to be scored as mobile, the region had to be found in each mutation, and at least one state, and have a different pattern of crosslinking refinement between the conformational states. These areas were then mapped onto PDB:3jad, in orange. Represented below are clockwise from the right, the cartoon schematic of α1-GlyR, with the bead model of the M3-M4 loop, the bottom-up view and the top-down view of the receptor.

Top-down

Bottom-up

connected to it show mobility with respect to all mutants in the network. The ECD-TMD interface, but particularly, the M2-M3 loop, and the Cys-loop also show dynamic changes as the receptor transitions between resting, open and desensitized states. The intracellular loop, which is not resolved in crystal structures is likely mostly unstructured, but appears to be dynamic at the beginning and the end of the loop. Lastly, the upper ECD may also be a dynamic region, as it has varying crosslinking with respect to the mutants in this network. Due to the considerable length of MTS-bzp, and that the length tolerance should most likely
be increased more than the additional 3-5 Å,\textsuperscript{20,123,224} it is difficult to fully describe the dynamics on a static crystal structure. It also becomes difficult to truly understand how this relates to model inconsistency. Does one structure fit the data more so than another? Is the apo-state receptor more flexible? These questions can be better answered by using a larger network, and the addition of varying-length crosslinkers.
CHAPTER 6. SCIENTIFIC IMPACT AND FUTURE DIRECTIONS

6.1 Overall Conclusions

The major takeaways from this work can be split into five categories; (1) the utility of CX-MS for protein structure studies, (2) the capability of the methodology to discern various conformational states of resting, open and desensitized, (3) the compatibility and complementarity of CX-MS to other high-resolution techniques, (4) the ability to resolve missing pieces in current crystal structures, and (5) that protein dynamics and mobility can be examined using CX-MS.

By using only small gel plugs excised from SDS-PAGE, and analyzing the sample via MS, dozens of unique crosslinking sites were identified. CX-MS does not require large concentrations of sample and can provide valuable information with only femtomolar quantities. The analysis of three single, site-specific cysteine mutations led to the identification of 145 intramolecular crosslinks and 80 intermolecular crosslinks.

Enriching the receptor in its conformational states: resting (MTS-bzp), open (MTS-bzp + IVM) and desensitized (MTS-bzp + excess glycine) resulted in unique crosslinking constraints for all states. It was originally thought that there would be carryover from resting state data into the other conformational states and that the resting state would not have unique sites, but that was not seen. Instead, each conformational state could be effectively enriched and unique information for the three states could be identified. This is of course, most important in terms of understanding the apo-state, since no crystal structures accurately define this state for GlyR, but is also useful for studying the desensitized state, as there is disagreement in what structurally defines a desensitized state.
CX-MS, though low-resolution, has the ability to complement high resolution techniques. The information gained from CX-MS studies is valuable as it has the ability to examine a full-length, functional protein of any molecular weight in native conditions. It cannot provide minute details, as cryo-EM and x-ray diffraction can, but it does provide compatible data that helps understand a protein or protein complex in a dynamic way. Furthermore, CX-MS is incredibly valuable for the study of membrane proteins.

Truncations, and mutations for stability often render a structure that has missing pieces or unresolved regions. CX-MS can provide information on these regions, since full-length, functional proteins are used in these studies. With respect to this work, the M3-M4 loop was able to be visualized, which gives valuable insight into its structure and position in various conformational states of GlyR. It is known that the M3-M4 loop in pLGICs is variable and distinguishes them from one another. It affects gating, desensitization, trafficking, and degradation. Discovering what parts of the loop are interacting with the membrane or with other parts of the channel can help better understand how this loop modulates function of the receptor. It appears that non-conducting states, the resting and desensitized states, involve a M3-M4 loop that is associated with the membrane, but the open state seems to lack this involvement. Studies analyzing M3-M4 loop mutants may help to clarify this.

Experiments on pLGICs have reported areas known to be mobile, including the C-loop, the M2-M3 loop, the Cys-loop and likely, also the M3-M4 loop. This work has come to an agreement that these areas are indeed dynamic and contribute to the transitions between resting, open and desensitized states. This work has also shown that the upper ECD region may be mobile as well, though the reason behind this is unclear.
6.2 Scientific Impact

Because the structure of a protein is fundamentally linked to its function, having a complete understanding of the structure and all of its parts, as well as a deeper understanding of how the protein is shifting allosterically as it transitions through resting, open and desensitized states, will lead to advancement in therapeutics. Therapeutics are derived from protein structural knowledge. For instance, chronic pain, a debilitating, and quite common disorder, is linked to the glycine receptor. These studies are a step forward towards understanding this receptor, developing drugs based on this knowledge and alleviating chronic pain symptoms.

This methodology can be applied to all of the members of the pLGIC family, and proteins in general, which will lead to similar discussions of other proteins in this family, as well as other transmembrane proteins that are particularly difficult to purify and crystallize. Transmembrane proteins make up the large majority of therapeutic targets but are not represented well in crystal structures, as the majority of structures are of soluble proteins. With the addition of CX-MS to the structural biochemist’s toolbox, a multitude of diseases, disorders and conditions can see treatment improvements as structure determinations improve.

6.3 Future Work

This work is an ongoing project. The network of mutants to which the structural roadmap is built is increasingly getting larger. The three mutants analyzed (M287C, K116C and K206C) will be added to two prior mutations, A41C and H419C\textsuperscript{602} in future publications. And, currently, other mutants are being analyzed to grow the network further: two upper ECD mutants K6C and G23C, and a C-terminal tail mutant, I412C.
As addressed in Chapter 1, the utilization of different length crosslinkers can be crucial to a more complete structure as they provide different radii in which various points of interest are connected. They can help provide reciprocity; if A links to B, and B to C, does A link to C? And if the crosslinking distance between A, B or C is shortened, are they still interacting? A diazirine crosslinker is being proposed to crosslink to the network already described (Figure 22). Using other crosslinker functionalities may also help alleviate any binding bias the crosslinker may have. One study has proposed that MTS-benzophenone crosslinkers have a tendency to bind to Met at a higher rate, but there is no evidence of this reported in this work.

Further work will also be conducted on the M3-M4 loop. Though this work has made important discoveries into its mobility and dynamics, there is much more to address. Analyzing mutants within the loop will help determine reciprocal linking between mutants already in the network as well as provide new information into the dynamic positions of the loop during conformational states. It may also help answer the question of membrane permeability by the loop, given its poly-R/K regions.

Another important piece of studying membrane proteins like GlyR is its interaction with various lipids within the membrane. CX-MS is a valuable tool to study protein-lipid interactions. Studies have already been conducted analyzing cholesterol concentrations, and are currently ongoing with the intention of studying how varying concentrations of PC and PA affect the structure of the glycine receptor.
Lastly, quantitation of crosslinking is particularly useful. The peaks in MS spectra do not give information on how abundant a peptide is, just its ability to be ionized. Quantitation would aid in determining which sites of attachment are more favorable or common than others versus which sites might be less abundant. This is also a currently ongoing investigation in the Cascio lab.
REFERENCES


(67) Yeagle, P. L. Non-Covalent Binding of Membrane Lipids to Membrane Proteins.


(77) Mehmood, S.; Marcoux, J.; Hopper, J. T. S.; Allison, T. M.; Liko, I.; Borysik, A.


(92) Eschweiler, J. D.; Frank, A. T.; Ruotolo, B. T. Coming to Grips with Ambiguity:


(127) Hofmann, T.; Fischer, A. W.; Meiler, J.; Kalkhof, S. Protein Structure Prediction Guided by Crosslinking Restraints - A Systematic Evaluation of the Impact of the


(149) Williams, K. R.; Konigsberg, W. H. Identification of Amino Acid Residues at


(172) Aramayo, R.; Sherman, M. B.; Brownless, K.; Lurz, R.; Okorokov, A. L.; Orlova, E. V. Quaternary Structure of the Specific P53-DNA Complex Reveals the


(208) Gomes, A. F.; Gozzo, F. C. Chemical Cross-Linking with a Diazirine Photoactivatable Cross-Linker Investigated by MALDI- and ESI-MS/MS. J. Mass


168


Hénault, C. M.; Baenziger, J. E. Functional Characterization of Two Prokaryotic


(249) The pharmacological profile of ELIC, a prokaryotic GABA-gated receptor


(273) Gilmore, J.; D’amato, S.; Griffith, N.; Schwartzberg, L. Recent Advances in


On the Nature of Partial Agonism in the


(361) Velisetty, P.; Chalamalasetti, S. V; Chakrapani, S. Structural Basis for Allosteric Coupling at the Membrane-Protein Interface in Gloeobacter Violaceus Ligand-


(400) Shiang, R.; Ryan, S. G.; Zhu, Y. Z.; Hahn, A. F.; O’Connell, P.; Wasmuth, J. J.


(422) Yevenes, G. E.; Moraga-Cid, G.; Guzmán, L.; Haeger, S.; Oliveira, L.; Olate, J.; Schmalzing, G.; Aguayo, L. G. Molecular Determinants for G Protein By


(433) Burgos, C. F.; Castro, P. A.; Mariqueo, T.; Bunster, M.; Guzmán, L.; Aguayo, L. G. Evidence for α-Helices in the Large Intracellular Domain Mediating...


(458) Lynch, J. W.; Rajendra, S.; Barry, P. H.; Schofield, P. R. Mutations Affecting the...


(471) Mascia, M. P.; Machu, T. K.; Harris, R. A. Enhancement of Homomeric Glycine


(483) Mihic, S. J.; Ye, Q.; Wick, M. J.; Kolchtein, V. V.; Krasowski, M. D.; Finn, S. E.; Mascia, M. P.; Valenzuela, C. F.; Hanson, K. K.; Greenblatt, E. P.; Harris, R. A.; Harrison, N. L. Sites of Alcohol and Volatile Anaesthetic Action on GABAA and


(495) Costa, B.; Trovato, A. E.; Comelli, F.; Giagnoni, G.; Colleoni, M. The Non-


(519) Nevin, S. T.; Cromer, B. A.; Haddrell, J. L.; Morton, C. J.; Parker, M. W.; Lynch,


(524) Acosta, M.; McClintock, W.; Pearl, P. *Hyperekplexia - NORD (National Organization for Rare Disorders)*; 2003.


(555) Maguire, M. F.; Latter, J. A.; Mahajan, R.; Beggs, F. D.; Duffy, J. P. A Study


(581) Sriwastava, S.; Srinivas, M.; Kanna, A.; Yarraguntla, K.; Jowkar, A.; George, E. Anti-Glutamic Acid Decarboxylase Antibody (GAD) Syndromes May Have More Aggressive Disease Course in African Americans and Early Onset of Presentation


Scharfenort, R.; Schumacher, J.; Forstner, A. J.; Baumann, C.; Schiele, M. A.;
Gajewska, A.; Glotzbach-Schoon, E.; Mühlberger, A.; Alpers, G.; Fydrich, T.;
Fehm, L.; Gerlach, A. L.; Kircher, T.; Lang, T.; Ströhle, A.; Arolt, V.; Wittchen,
H. U.; Kalisch, R.; Büchel, C.; Hamm, A.; Nötten, M. M.; Romanos, M.;
Domschke, K.; Pauli, P.; Reif, A. GLRB Allelic Variation Associated with
Agoraphobic Cognitions, Increased Startle Response and Fear Network
Activation: A Potential Neurogenetic Pathway to Panic Disorder. *Mol. Psychiatry*

Lehmann, T. N.; Gloveli, T.; Grantyn, R.; Meier, J. C. Glycinergic Tonic
Inhibition of Hippocampal Neurons with Depolarizing GABAergic Transmission

Eichler, S. A.; Förstera, B.; Smolinsky, B.; Jüttner, R.; Lehmann, T. N.; Fähling,
M.; Schwarz, G.; Legendre, P.; Meier, J. C. Splice-Specific Roles of Glycine

Meier, J. C.; Henneberger, C.; Melnick, I.; Racca, C.; Harvey, R. J.; Heinemann,
U.; Schmieden, V.; Grantyn, R. RNA Editing Produces Glycine Receptor
744. https://doi.org/10.1038/nn1467.

Burgos, C. F.; Yévenes, G. E.; Aguayo, L. G. Structure and Pharmacologic

*Neuron* 1993, 10 (SUPPL. REV.), 31–41. https://doi.org/10.1016/0896-
6273(93)90239-N.

Barrantes, F. J.; Fantini, J. From Hopanoids to Cholesterol: Molecular Clocks of
https://doi.org/10.1016/j.plipres.2016.03.003.

Salari, R.; Murlidaran, S.; Brannigan, G. Pentameric Ligand-Gated Ion Channels:
https://doi.org/10.1080/08927022.2014.896462.

Dacosta, C. J. B.; Baenziger, J. E. Gating of Pentameric Ligand-Gated Ion
Channels: Structural Insights and Ambiguities. *Structure.* August 6, 2013, pp

Dacosta, C. J. B.; Ogrel, A. A.; McCardy, E. A.; Blanton, M. P.; Baenziger, J. E.
Lipid-Protein Interactions at the Nicotinic Acetylcholine Receptor. A Functional
Coupling between Nicotinic Receptors and Phosphatidic Acid-Containing Lipid
https://doi.org/10.1074/jbc.M108341200.

Veeramachaneni, R. J.; Donelan, C. A.; Tomcho, K. A.; Aggerwal, S.; Lapinsky,
D. J.; Cascio, M. Site-Specific Crosslinking Coupled with Mass Spectrometry as a


Mishra, A.; Lai, G. H.; Schmidt, N. W.; Sun, V. Z.; Rodriguez, A. R.; Tong, R.; Tang, L.; Cheng, J.; Deming, T. J.; Kamei, D. T.; Wong, G. C. L. Translocation of...


