AN ABSTRACT OF THE THESIS OF

Callia K. Palioca for the degree of Honors Baccalaureate of Science in Biology presented on May 31st, 2013.
Title: The Crystal Structure of the Potential Drug Target *Mycoplasma pneumoniae* Glycerol 3-Phosphate Oxidase.

Abstract approved:

_____

P. Andrew Karplus

Mycoplasma pneumoniae is the primary cause of community-acquired pneumonias, including what is commonly known as walking pneumonia. This disease affects people from all demographics, but especially children and older adults. Outbreaks are a significant public health concern and work to develop new pharmacological agents is currently being researched. How *M. pneumoniae* causes disease is not fully understood, but studies have pointed to hydrogen peroxide as a pathogenicity factor. It is produced as a byproduct of glycerol metabolism by the enzyme glycerol 3-phosphate oxidase (GlpO). Using X-ray crystallography, we determined the three-dimensional structure of this enzyme in order to elucidate its binding properties and guide structure-based drug design efforts. Here, we report the crystallization of *M. pneumoniae* GlpO along with the native structures of oxidized and reduced GlpO at resolutions of to 2.4 Å and 2.5 Å, respectively. We compared GlpO from *M. pneumoniae* to another GlpO, a glycerol 3-phosphate dehydrogenase, a glycine oxidase, and the most structurally similar protein which is a protein of an unknown function from *Bordetella pertussis*.

Key Words: *Mycoplasma pneumoniae*, X-ray crystallography, walking pneumonia, GlpO, structure-based drug design

Corresponding email address: callia.palioca@gmail.com

The Crystal Structure of the Potential Drug Target

Mycoplasma pneumoniae Glycerol 3-Phosphate Oxidase

By

Callia K. Palioca

A PROJECT

Submitted to

Oregon State University

University Honors College

in partial fulfillment of

the requirements for the

degree of

Honors Baccalaureate of Science in Biology (Honors Scholar)

Presented May 31st, 2013

Commencement June 15th, 2013

Honors Baccalaureate of Science in Biology project of Callia K. Palioca presented on May 31st, 2013

APPROVED:

Mentor, representing Biochemistry and Biophysics

Committee Member, representing Biochemistry and Biophysics

Committee Member, representing Zoology

Chair, Department of Biology

Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College. My signature below authorizes release of my project to any reader upon request.

Callia K. Palioca, Author

ACKNOWLEDGMENTS

This thesis has been the culmination of 4 years of hard work but I would not be here today without the help of so many people. To the following people, I thank you for all that you have done for me and will never forget how much you have supported me.

A large thank you to Andy Karplus for teaching me about science and about the world. You are highly accomplished yet dedicate your life to mentoring students. You believed in me from the beginning and gave me my first real job. You taught me how to think like a scientist and have been a huge support in editing this thesis. You never gave up on me and have mentored with patience and compassion. I have enjoyed being a part of the lab and I thank you for all the fun lab dinners we have had. I could not have asked for a better thesis mentor!

To the Karplus lab, I never would have learned so much or gotten so far without your guidance. Thank you for all the helpful discussions, the fun, and for being so willing to help me. I would have been clueless without your help.

To Kevin Ahern and Indira Rajagopal for supporting me and advising me throughout my years at OSU. You have been an inspiration from the start and I have enjoyed getting to know you and learn many of life's lessons from you. Thanks for the laughs and the yummy food.

To Devon Quick for reading this very long thesis. Thank you for being such a wonderful anatomy and physiology professor. I have learned so much from you that will help me in my future career. Thank you for also being role model to me and fellow students.

To Ms. Balogh, for first showing me proteins. You did not just teach me about structure and function, but you let me figure it out on my own. I will never forget the epiphany of learning how hemoglobin binds oxygen, watching the structure change before my eyes. Your effort and dedication to teaching are an inspiration and I am lucky to have had such an effective and fun teacher like you.

To my parents for making me weed the yard and clean the dishes. You have taught me how to be a hard worker and have been an absolute supporter of everything I do. Mom you have been to every soccer game and Dad, I thank you for instilling a love of nature in me. Thank you for your hard work, love and support.

To my fiancé, Issac for believing in me throughout everything. You have been there for every hardship and struggle and helped me see the good even in the toughest times. You have pushed me to do what I needed to do and never put up with my excuses. I look forward to all the fun times we will have together. I love you very much!

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

The Crystal Structure of the Potential Drug Target *Mycoplasma pneumoniae* Glycerol 3-Phosphate Oxidase

CHAPTER 1
INTRODUCTION

Overview of the Project

Mycoplasma pneumoniae is the bacterium that causes primary atypical pneumonia, otherwise known as walking pneumonia. A type of community-acquired pneumonia (CAP), this condition is a worldwide health threat for vulnerable populations, which include children and the elderly. Understanding the spread of disease and the mechanisms of pathogenicity are essential to providing better treatment and prevention. This introductory chapter largely provides background information about the genus *Mycoplasma*, the diseases caused by *M. pneumoniae*, their impact, and theories into the pathogenicity. The remainder of this thesis then describes the process of determining the structure of a protein thought to be a main pathogenicity factor causing the symptoms of walking pneumonia.

Mycoplasmas as a Significant Cause of Human Suffering

Mycoplasmas are a unique genus of bacteria which were first isolated in culture in 1898 as a bovine pleuropneumonia agent. Prior to that, they were described as a type of fungus. Indeed, the name mycoplasma derives from mykes (fungus) and plasma (formed).[1] Interestingly, these organisms have the smallest cell and genome sizes of any self-replicating, autonomous organisms.[2] For comparison, the *Escherichia coli* genome is five times larger than the 816 kb genome of one species from this genus, *M. pneumoniae*.[2,3] Many members of this genus have been the subject of research into identifying the minimal number of genes required for cell existence. Current evidence

exists that these bacteria are continually undergoing reductive genome evolution.[3] Mycoplasmas have also been extensively because of their unique lack of a cell wall. Structurally, they are able to compensate for this lack of cell wall by having sterols in the cell membrane that provide physical support. Functionally, they also compensate by having a close adaptation with the eukaryotic host. The lack of a cell wall has also meant that the cell is extremely susceptible to desiccation and therefore, close contact is necessary for transmission by airborne droplets.[2] Because all members of this genus lack a cell wall and do not have the ability to synthesize peptidoglycan, they are unaffected by typical antibiotics such as penicillin which work by targeting cell wall synthesis.

These bacteria play a role in diseases affecting humans, other mammals (including cows and pigs), reptiles, fish, arthropods, and plants.[4] One specific example is arthritis in mice, rats, cattle, swine, and poultry. Many mycoplasma species cause respiratory diseases. Pneumonias caused by mycoplasmas are found in humans, (*M. pneumoniae*), calves (*M. bovis*), sheep and goats (*M. capricolum*), swine (*M. hyorhinis*), and turkeys (*M. gallisepticum*). A variety of mycoplasmas, both harmless and pathogenic can be isolated from the human body. In one study group, *M. orale* was isolated in 84% of gingival crevices. *M. hominis* can be found in the human urogenital tract but when it turns pathogenic, it causes 5% of cases of pyelonephritis (a urinary tract infection that has reached the kidneys) and even causes some cases of pelvic inflammatory disease.[5]

One particularly well-studied organism, *Mycoplasma mycoides,* subsp. *mycoides SC,* where SC stands for small colony and the organism is commonly called *M. mycoides*, is responsible for causing contagious bovine pleuropneumonia (CBPP). As the most common form of cattle disease in Africa, eradication of this virulent pathogen is of

ecological and economic importance.[6] *M. pneumoniae,* the causative agent of walking pneumonia, has similar proteins as those found in *M. mycoides*, including a sequence-similar glycerol 3-phosphate oxidase thought to be pivotal in causing both CBPP and atypical pneumonia. Eradicating the diseases caused by these organisms would enhance the quality of life of many individuals, as well as save money by avoiding healthcare costs of hospitalizations, the loss of cattle, and the loss of productive hours.

M. pneumoniae and its Associated Diseases

In 1944, Eaton et al.[7] isolated and described what they called an 'Eaton agent' from the sputum of a patient with primary atypical pneumonia. This is a type of pneumonia which does not respond to therapy with sulfonamides or penicillin.[5] While it was originally thought to be a virus, it was reclassified as a pleuropneumonia-like organism in 1961 and given the taxonomic designation of *Mycoplasma pneumoniae* in 1963.[2] Since then, it has been the subject of numerous studies and research articles. For those interested in more in-depth information, this topic has been extensively reviewed by Waites and Talkington.[2]

Symptoms and overall effects of M. pneumoniae *infection on the host*

M. pneumoniae infections can occur in the upper and lower respiratory tracts. The severity and longevity of symptoms varies, but the symptoms typically consist of a sore throat, hoarseness, sinus congestion, headache, middle ear infection, and a persistent, hacking cough commonly associated with atypical pneumonia.[2] Many patients also experience flu-like symptoms, which differentiates these infections from typical pneumonia.[5]

Local damage of the respiratory tract causes many of these symptoms. As an example, the *M. pneumoniae* bacteria reach the bronchi and bronchioles causing vacuolization and, in more severe cases, total destruction of bronchi cilia.[8] It is thought that these manifestations contribute the most to the hacking cough which subsequently spreads the bacteria.[2] Further, many cells involved in the immune response, such as macrophages and neutrophils, can accumulate in the tissues. These can lead to lesions that continue to harm the host, showing how the host immune system can exacerbate the disease. However, many infected individuals never progress to the more severe stages of this disease involving a lower respiratory infection, and up to 20% are asymptomatic.[9]

Demographics and epidemiology of community-acquired pneumonia

While *M. pneumonia* infection is sometimes regarded as moderately harmless and more of an inconvenience, it has been known to cause hospitalizations, sometimes affecting large parts of a community, loss of productivity, and even death. These atypical pneumonias fall in the category of community-acquired pneumonias (CAPs), and in one study, *M. pneumoniae* accounted for 20.7% of adult CAP cases,[2] making it the single most frequently identified pathogen for these conditions.

In general, atypical pneumonia affects the elderly and children. *M. pneumoniae* was the second most common pathogen found in hospitalized adults with CAP-like symptoms in a survey of two Ohio counties in 1991, where a significantly higher percentage of people over the age of 65 were hospitalized in comparison to younger adults.[10] These bacteria were also recently shown to cause more than 100,000 hospitalizations in the U.S. each year.[2] Children are also susceptible to infection. In Finland, this bacterium was found in over 50% of children aged 5 years and older with

CAP.[11] The rest of the population is not immune, though, as many parents acquire the disease from their children.[2] Fortunately, however, it can take many weeks for family members to become infected, giving families time to begin treatment.[5]

In the United States, most outbreaks occur in late summer and early fall[8], but they can happen anywhere and at any time. Outbreaks often start in closed or semi-closed areas such as schools, military bases, or hospitals.[2] Based on serological studies done in Denmark from 1946 to 1995[12], endemic disease transmission, or transmissions localized in a small population, were interspersed with cyclic epidemics every 3-5 years. The long incubation period, slowly progressing symptoms, and the ability of most patients to continue daily activities (which gave rise to the colloquial term "walking pneumonia") may help explain the cyclical pattern of outbreaks.[2]

Connections to other conditions

Besides atypical pneumonias, *M. pneumoniae* has been implicated in chronic asthma, encephalitis and rheumatoid arthritis. Since the 1980s, this organism has been seen as a trigger for acute asthmatic attacks. Patients for whom asthma symptoms have persisted for years, consistent with a stubborn infection, tend to have had *M. pneumoniae* infections.[8] In one study, patients saw significant improvement in symptoms after a 6-week treatment trial of clarithromycin[13], an antibiotic commonly used against this bacterium.

In addition, *M. pneumoniae* was the most common infectious agent identified in one study of 2000 people with encephalitis, an inflammation of the brain with associated headaches and seizures in more extreme cases.[8] Clusters of encephalitis conditions tend to occur during outbreaks of *M. pneumoniae*. In one example, the death of an elementary

school student with encephalitis occurred during a mycoplasma outbreak in the local area.[14,15] Rheumatoid arthritis has similarly been linked *M. pneumoniae* in some studies. Consistent with this association, rheumatoid arthritis can often be successfully treated by tetracyclines, a group of drugs to which *M. pneumoniae* is also sensitive.[16] Both this condition and encephalitis have complex and still not fully understood etiologies. It has been suggested that because the adhesins on the bacterial cells have similarities to host structures, molecular mimicry (an autoimmune response based on similar antigens present on both the bacterium and host cells,[16,17]) may be important in developing inflammation that would be associated with both conditions.[8,16]

Current strategies of detection

Since the initial discovery of *M pneumoniae*, detection in a clinical setting has proven to be inconsistent and difficult. Each of the available approaches such as culture, serology, and Polymerase Chain Reaction (PCR) has inherent limitations. As such, current diagnosis relies on a combination of PCR and serology. IgM serology has been shown to be most useful in children while IgA serology (not yet universally available) works best for older adults.[2]

Culturing has the unique advantage in that, if positive and done properly, the identification is 100% specific to *M. pneumoniae*. However, *M. pneumoniae* is a fastidious and slow-growing organism that can take up to 6 weeks to grow on microbiological medium, an impractical amount of time for clinical diagnosis. Supplies are also very expensive and technical expertise is required. As such, culturing has not been utilized as an effective clinical diagnostic tool for *M. pneumoniae*.

Tests for *M. pneumoniae* using serology have been inconsistent in detecting *M. pneumoniae* in patients. Serologic tests utilize serum, the liquid portion of blood after the cells have been removed from whole blood. Serum contains antibodies, which are essential for immune system function and also are markers that can indicate what type of infection is present in the body. One type of antibody, IgM, indicates when there is an acute infection, while IgG indicates a past infection or immunity. Tests are available that detect each of these antibodies or both. IgA is another antibody class that is typically produced early in an infection. Serology is inadequate in immunosuppressed patients because it relies on the immune system to produce antibodies. Antibody production can also be delayed in some people, and thus the date of serum collection can alter the result. As such, serology is considered a very useful tool for epidemiological studies, but less so for the diagnosis of individual patients.[8]

Since their development in the late 1980s, PCR assays have been essential tools of the biological sciences. The sensitivity is high and only a single organism is needed to make a diagnosis. It seems to be highly effective with identifying *M. pneumoniae* in patients with extrapulmonary syndromes such as encephalitis. Nevertheless, a common difficulty with identifying *M. pneumoniae* as the pathogen responsible for someone's symptoms is the similarity it has with other pathogens such as *C. pneumoniae* and *Legionella pneumophila*. Development and commercialization of PCR assays that can differentiate between these microorganisms would be very beneficial in both research and healthcare settings.

As stated previously, studies now point to the combined application of PCR assays and serological tests for detecting *M. pneumoniae* in suspected cases of atypical

pneumonia. However, standardization and distribution of a cost-efficient, multiplex detection technology is still needed.[2]

Current strategies of treatment

Due to the lack of a consistent and standardized method for detecting of *M. pneumoniae,* evaluation of the efficacy of available drugs against this pathogen is difficult. Generally, antibiotics are the primary treatment option. Macrolides are considered the treatment of choice for atypical pneumonia.[8] Erythromycin and azithromycin are examples of this group of drugs. While use has been effective in reducing symptoms in patients, there is concern over macrolide-resistant *M. pneumoniae.* In 2001, researchers from Japan described and isolated a resistant strain from children with pneumonia and bronchitis. Resistance to both azithromycin and erythromycin was observed. Overall, it is estimated that 10-33% of isolates obtained between 2001 and 2006 were macrolide-resistant. The CDC reported that 27% of *M. pneumoniae* in an outbreak in the northeastern region of the United States from 2006 to 2007 were macrolide-resistant. In fact, the predominant resistant strain in the U.S. is the same as one of those found in Japan.[18] Patients with macrolide-resistant *M. pneumoniae* typically have to be changed to a different type of drug group because of persistent fever, cough, and no resolution of symptoms.[8]

The other two main types of antibiotics useful for atypical pneumonia are tetracyclines and fluoroquinolones. However, the ability to effectively apply these drugs is limited. In particular, tetracycline is not approved either for use in children under the age of 8 or in pregnant women. Similarly, fluoroquinolones are not recommended for children because their bones have not fully ossified. This antibiotic seems to cause

toxicity in cartilage.[2] While natural isolates of *M. pneumoniae* bacteria have not been found to be resistant to these drugs, *in-vitro* experiments have developed resistant mutants.[8]

Due to antibiotic resistance, it would be beneficial to have a type of primary prevention technique. The development of a vaccine against *M. pneumoniae* would be appropriate for prevention because of its propensity for outbreaks in schools, hospitals, and military bases. However, there has so far been little success in vaccine development. In some cases, symptoms were exacerbated, with a trial vaccine sensitizing the host in some way and resulting in a more severe illness after being exposed to *M. pneumoniae*.[2]

One method of combating *M. pneumoniae* would be to synthesize a drug that specifically targets this pathogen. This can be done through structure-based drug design.[19] With the aid of X-ray crystallography, a three-dimensional structure of a protein drug target can be determined. By selecting and optimizing molecules based on drug-binding properties and algorithms, a drug can be 'tailor-made' for a specific protein. The very specific binding of the drug molecule can reduce side interactions with other molecules, reducing patient side effects. Drugs may also be prescribed at lower dosages if efficacy and specificity are increased.[19] However, to effectively choose a lead molecule and drug target, more information must be acquired about the mechanisms of pathogenicity of *M. pneumoniae*.

Proposed Mechanisms of Pathogenicity of *M. pneumoniae*

Many processes come together for *M. pneumoniae* to cause virulence. The bacteria must get to the proper spot next to the lung epithelium (using gliding motility),

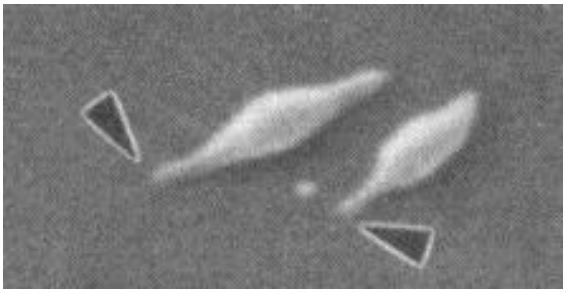
adhere to the cells (via the attachment organelle), and produce toxins such as cytotoxins, peroxides, or both. The precise mechanism and the relative influences of peroxide versus cytotoxins are still not known. Current knowledge on these topics is extensively reviewed by Waites and Talkington.[2]

Current models for how M. pneumoniae causes disease – Gliding motility & Attachment

For this pathogen to cause disease, it must first reach the target cell through a process known as gliding motility. Most organisms move using flagella. However, *M. pneumoniae* does not have flagella and instead glides across surfaces, never changing the leading end.[20] While the exact mechanisms are not understood, it is essential to cause virulence. A mutant organism that lacks a protein involved in gliding motility, was shown to reduce gliding velocity to 5% of that of the wild type and, in a separate study, the patient infected by this mutant organism had reduced lung lesions when compared to a patient infected with a wild type *M. pneumonaie*.[21] Gliding motility relates to attachment in that it uses the attachment organelle[2], a structure seen in the image of *M. pneumoniae* in Fig. 1.

Once the parasite is in the proper location of the respiratory tract of an individual, *M. pneumoniae* must adhere to the epithelial lining. This adhesion step is also essential to pathogenicity. Specific proteins that are concentrated on the tip of an attachment organelle reach out from this extension to bind to host receptors for fibronectin, allowing the bacterium to gain access to cell proteins and nutrients.[2] Inhibition of the main protein P1 by specific antibodies results in a 75% reduction of respiratory epithelium attachment by virulent *M. pneumoniae*.[22]

Fig 1: **Morphology of *M. pneumoniae***. This scanning electron micrograph shows that *M. pneumoniae* has a thick body with two thinner ends. Arrows point to the attachment organelle which is essential in adherence to host epithelium. This figure was adapted from Fig. 4 of Waites and Talkington.[2] Original credit assigned to Krause and Taylor-Robinson.[5]



Current models for how M. pneumoniae causes disease – Production of a cytotoxin

Another virulence factor recently discovered is the CARDS TX, or community-acquired respiratory distress syndrome toxin.[18] This protein, encoded within the *M. pneumoniae* genome, has a 27% amino acid sequence identity to pertussis toxin in over 239 residues.[23] Key similarities include NAD binding and ADP-ribosylating activity residues seen in pertussis toxin. It is not yet clear whether these toxins act via the same mechanism.[23,24] A recombinant CARDS TX caused vacuolization of mice cells in a dose-dependent fashion and caused reduced ciliary movement in baboon trachea[23]. Both mice and baboon models treated with this toxin had inflammatory responses and reduced airway function similar to those observed in *M. pneumoniae* infection.[25] These studies indicate that the CARDS TX can play a significant role in the pathogenesis of *M. pneumoniae*. Studies are currently underway to determine the structure of this toxin.[24]

Current models for how M. pneumoniae causes disease – Production of peroxide

Generation and secretion of hydrogen peroxide have also been identified as key factors of *M. pneumoniae* pathogenicity.[3] The peroxide is primarily produced as byproduct of glycerol metabolism from the enzyme α-glycerol 3-phosphate oxidase (GlpO). As a reactive oxygen species (ROS), the peroxide can cause damage to the host's cells by inducing oxidative stress, a state of a greater number of oxidants than antioxidants.[26] If the peroxide is partially reduced, the radical hydroxyl ion produced can react with lipids on the host plasma membranes, changing the shape of polyunsaturated lipids.[26] This can puncture the membrane and affect membrane fluidity, thus impacting homeostasis of the cell. Peroxide can also enter the host cell and induce inflammatory processes.[6] The inflammation can, on the one hand, minimize disease by triggering host

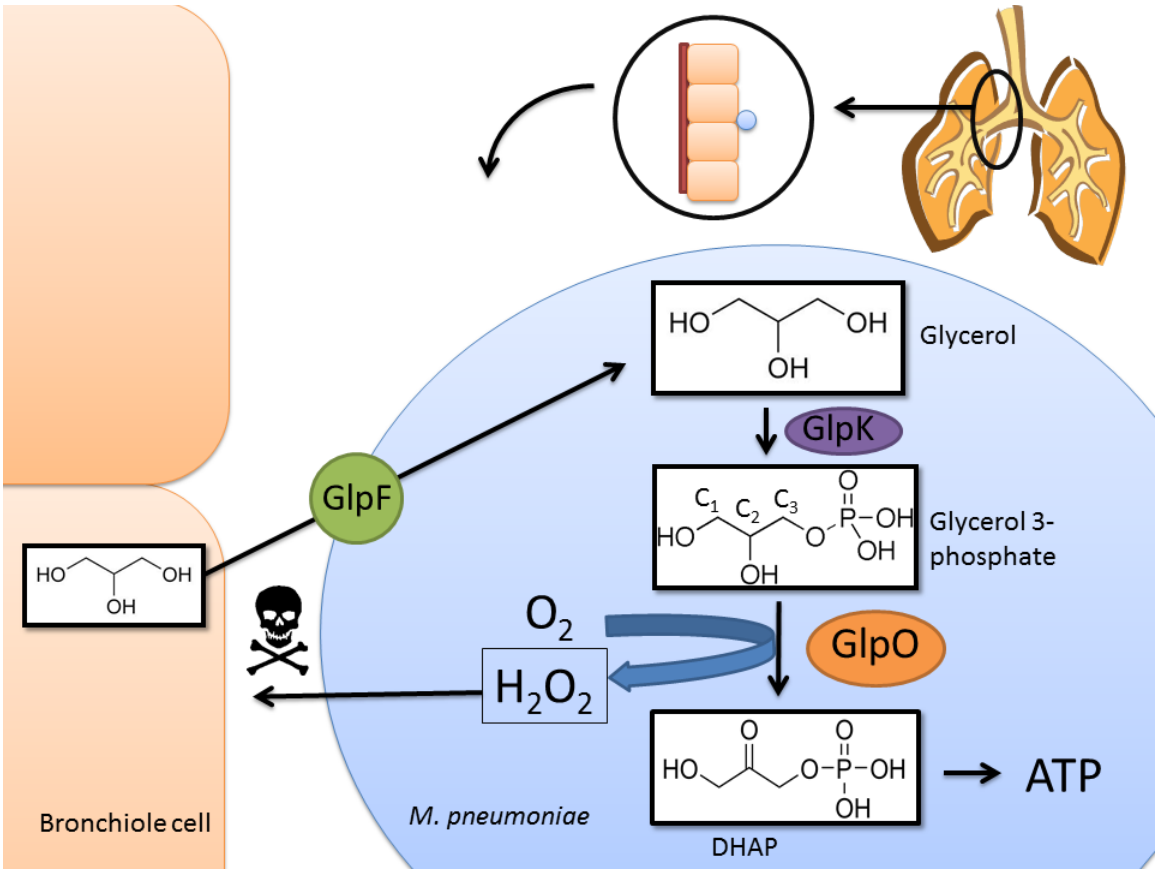
defense mechanisms that eliminate the organism. However, on the other hand, the peroxide induced expression of host pro-inflammatory genes can exacerbate disease though damage of the respiratory epithelium and surrounding tissues.[6,18] Although humans have rather robust antioxidant enzymes, it appears that *M. pneumoniae* produces superoxide anion that inhibits the host enzyme catalase, that normally protects it from hydrogen peroxide.[8, 27] This makes the host more susceptible to oxidative damage.

Glycerol Phosphate Oxidase (GlpO) as a Drug Target

For this project, we focused on blocking the production of peroxide as a way to decrease *M. pneumoniae* pathogenicity. As noted above, the peroxide that causes pathogenicity is a product of glycerol metabolism. The glycerol is derived from host phospholipids and lung epithelia surfactant. As can be seen in Figure 2, glycerol is brought into the bacterial cell with the help of a glycerol facilitator protein (GlpF). It is then phosphorylated to glycerol 3-phosphate by glycerol kinase (GlpK), and finally oxidized by GlpO (sometimes referred to as a GlpD based on naming conventions formulated prior to its characterization as an oxidase). This final oxidation step forms H_2O_2 as a byproduct. The mechanism of this process in the better studied virulent related organism *M. mycoides* can give insight into the biochemical pathway for cytotoxicity of *M. pneumoniae.* [3]

In *M. mycoides,* H_2O_2 is formed via a GlpO. This enzyme oxidizes α-glycerolphosphate (Glp) (also known as glycerol 3-phosphate) to dihydroxyacetone phosphate (DHAP) by using a flavin adenine dinucleotide (FAD) molecule and oxygen to produce hydrogen peroxide. This is in contrast to most glycerol metabolism mechanisms,

Figure 2: **Overview of *M. pneumoniae* Pathogenicity**. This extracellular organism adheres to tracheal and lung epithelia. There it uptakes glycerol from host phospholipids, bringing this molecule into its cell using the transporter GlpF. A phosphate is enzymatically added by GlpK to give glycerol 3-phosphate (Glp). Carbon numbers are indicated on this molecule. Glp acts as a substrate, along with O_2, for GlpO which produces both DHAP and toxic H_2O_2. DHAP is utilized in glycolysis to yield ATP for the bacteria. Due to the bacterium's adherence to the host, the peroxide is shuttled out of the bacterial cell and contacts the epithelium, causing vacuolization and ciliary destruction. These changes contribute to the symptoms of walking pneumonia.



which utilize a dehydrogenase enzyme that reduces NAD^+ to NADH.[3] As expected, DHAP can then enter into glycolysis to produce ATP, which can be utilized for energy. Based on sequence and protein structure similarities, it is hypothesized that *M. pneumoniae* works by an identical mechanism. Due to the cytotoxic effects of peroxide, *M. pneumoniae* GlpO (*Mp*GlpO) is a potential target for structure-based drug design.

What is Currently Known about the Structure and Functional Mechanism of GlpO

While the three-dimensional structure was not known prior to this report, some insights were drawn from previous studies on both the structure and the mechanism of function. Based on searches that identify similar proteins, *Mp*GlpO is part of the D-amino acid oxidase (DAAO) family of flavoenzymes. These proteins use the flavin of an FAD cofactor to carry out a two-step reaction. First, they oxidize carbon-nitrogen bonds of amino acids, primary amines, or secondary amines while the enzyme bound FAD is reduced to $FADH_2$. Then, the reduced flavin is re-oxidized by O_2 and this forms H_2O_2 as a byproduct.[28] The mechanism by which the flavin is reduced has been under debate since the discovery of DAAO. Structural studies of DAAO at high resolution indicate that a hydride transfer occurs from the substrate α-carbon to the reactive N5 of the flavin.[29] The structure showed that amino acid ligands were bound in the correct position for a hydride transfer.[29] Specifically, the α-hydrogen is pointed towards the flavin N5 atom, which is the site of reactivity. Because the flavin in DAAO enzymes reacts via a hydride transfer, it is expected that the flavin in *Mp*GlpO will react in a similar fashion, and that it is the C2-atom of Glp (Fig. 2) that corresponds with the α-carbon of the substrates of the related enzymes.[30]

Solving the Crystal Structure of *M. pneumoniae* GlpO

Having new, more effective drugs for treating *M. pneumoniae* infections and disease would be beneficial for the public health, by providing more efficient forms of treatment for walking pneumonia, an example of a community-acquired pneumonia that afflicts hundreds of thousands of people every year. As described above, the production of hydrogen peroxide by GlpO as a part of glycerol metabolism is a crucial contributor to *M. pneumoniae* pathogenicity. Elucidation of the three-dimensional structure of GlpO is a foundational step required to guide the structure-based development of new drugs, and the focus of this thesis work is the successful use of X-ray crystallography to solve the *M. pneumoniae* GlpO structure. In the next two chapters, I present the crystallization techniques, the native, oxidized structure of GlpO to 2.4 Å resolution, and the reduced form to 2.5 Å resolution. I also discuss the process by which we arrived at this structure, as well as comparisons to similar structures.

CHAPTER 2

MATERIALS AND METHODS

A scientific article presents a condensed story of key experiments and observations leading to logical conclusions that represent advances in knowledge. However, the deeper story, that of the detours, the dead ends, and the strategies implemented to try to navigate around problems, is less frequently discussed. In contrast with typical Materials and Methods sections, which include brief technical descriptions providing only necessary information needed for others to understand and repeat the work done, in this section I will also describe some of the above-mentioned broader meandering aspects of this project. These were a very impactful part of my experience in scientific research. I make this attempt to reveal more of the entire story, because it was through my experience of these meanderings that I now better understand how scientific research really works. I hope that readers will find this presentation enjoyable, and that they also will find it gives them a more accurate picture of the kinds of challenges faced by scientists as they pursue a research project.

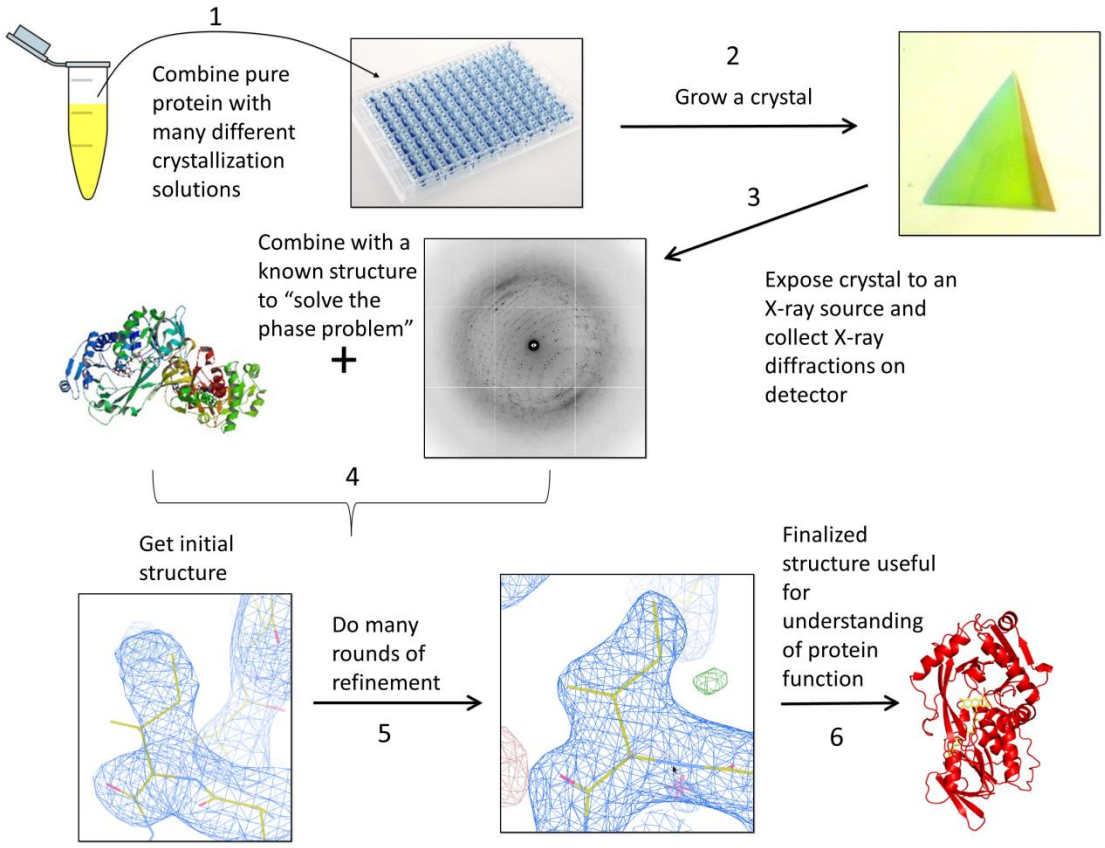
In this chapter, I will first provide a brief overview of the steps typically taken in determining the structure of protein using X-ray crystallography. I will then describe my process of taking a protein solution and growing a crystal, collecting data from that crystal, and solving the structure and refining it to get a three-dimensional model of *Mp*GlpO.

Brief Overview of X-ray crystallography Experiments in this Work

Before delving into the details of my research, I provide some background on the X-ray crystallography methods used to solve the structure of *Mp*GlpO. X-ray crystallography is, along with nuclear magnetic resonance (NMR), one of the two ways to solve a high resolution protein structure. As outlined in a flow chart (Figure 3), in this method, a protein crystal must be grown and then exposed to X-rays while the crystal is rotated through many different orientations. The X-rays used in these experiments are simply light (i.e. photons) with a wavelength of near 1 Å (10^{-10} m). During X-ray diffraction, the electrons in the protein interact with the X-rays to scatter them and create a diffraction pattern, which is collected by an X-ray sensitive detector (for instance, a CCD detector). When the sample is a crystal, the scattered X-rays typically create a pattern of lunes (see Figure 3 step 3)**.** The data needed to solve the structure are derived from this pattern by assigning a unique identification index (in the form of the integers h, k, and l) and measuring the intensities to each of the diffraction spots (also known as 'reflections').

The assignments and intensity calculations are done using computer programs. The CCP4 (Collaborative Computational Project Number 4; [31]) package is a collection of computer programs that we used for many of the X-ray crystallographic calculations, from analyzing the data to applying final refinements to a completed protein model. The program we used to process the diffraction data is called iMosflm.[32,33] It takes a subset of the reflections and determines the space group, an inherent characteristic of the crystal that describes the symmetry of protein molecules in the crystal. This space group

Fig. 3: **Overview of X-ray Crystallography.** The process of determining the 3-dimensional structure of a protein involves obtaining a purified protein solution. Due to the FAD cofactor, the solution of *Mp*GlpO was yellow. This protein is mixed with chemical solutions and placed in a sitting-drop well plate. If crystals form, these can be exposed to X-rays where a detector collects the diffraction pattern. The information from the diffraction pattern is combined with phases from a structurally similar model to solve the phase problem. Many rounds of refinement are then performed until the structure very closely matches the density.



identification is used to predict the location of the reflections that make up the diffraction pattern. For each reflection, the program integrates the spot to determine the intensity. Because the complete sets of diffraction data that are collected contain many independent measurements of each unique reflection, once the spots are all integrated, a program called Scala within the CCP4 suite of programs can be used to scale the reflections to one another and average together the multiple observations of each kind of reflection. In my crystal's space group, P23, there was a high level of symmetry. This meant that, as the crystal rotates, many symmetry-related reflections are duplicated and must be merged. The merging function of Scala creates both average reflection intensities and variance statistics. It is with the statistics generated from this program that it is possible to decide which data are accurate enough to use and which should be discarded. The major decision in this regard is choosing a resolution cut-off, which defines the highest resolution (e.g. 2 or 2.5 or 3 Å) at which the data are still of acceptable quality.

After getting the intensities, the CCP4 Truncate program converts them into structure factors, also called F_{obs} (for 'F-observed'), by, in most cases, simply taking the square root of the intensities. The F_{obs} is the amplitude of the wave form of the light ray that generated the given reflection. Light acts as a wave and from the experimental intensities, as noted, we can obtain the height (or amplitude) of the wave. We cannot, however, obtain the phase of the wave. Because this phase information is lost, it leads to the so-called 'phase problem.' Solving this problem, that is figuring out the phases, is important because both phase and amplitude information are necessary to apply a Fourier transform that converts the data from the diffraction pattern into a form that recreates the

real space electron density distribution. This is the density in which we model a protein structure.

In my case, to solve the phase problem and be able to piece the puzzle together, I used molecular replacement. In general, a protein structural homolog that has already had its structure determined (that is, a known protein structure that looks like the protein structure to be solved) is repositioned at multiple possible places in the unit cell until it can roughly account for the observed diffraction data. Using this best position of the search model, the placed protein model can be used to calculate initial phases.[34] In determining a protein structure via molecular replacement, there is the inherent issue of model bias. The goal is to get a structural representation of the desired protein (*Mp*GlpO) based on the data, and to not have that model be skewed based on the structure of the search model. To be able to detect bias that may occur, 5% of the reflections are set aside prior to molecular replacement and the model building and refinement steps, and these set aside reflections are used for cross-validation – the process of validating a model based on data that were not used to determine it. Refinement is the iterative process of improving the phases and model by making slight adjustments that improve the model's fit to the observed data. The statistic R_{free} is the measure of agreement of the model with the 5% of the reflections that are not utilized in the refinement. Thus, this statistic is independent of manual remodeling and is referenced to ascertain if refinement has brought the model closer to the real solution. R_{free} is reported along with the R-factor in refinement statistics. Lower values are better and R-values typically range from 0.6 (indicating total disagreement of the model and data) to as low as 0.10 to 0.20. These

lower values indicate that the model is sufficiently close to the observed data to be acceptable as a final model.

Typically, the real differences between the unknown structure and the search model can be modeled into the electron density computed from molecular replacement by running an 'AutoBuild' program and refinement through which the computer is able to optimize the changes made to the model. However, if the protein search model is too dissimilar, manual model building is required in place of AutoBuild. This is a process in which each individual position in the protein is looked at by the researcher to determine if the electron density supports a change in the structure. This can involve a change in the amino acid or side chain itself or in the pathway of the main chain. This manual rebuilding process is continued until no additional changes supported by the density can be made.

Cloning and Protein Purification

Concentrated and purified His_6-tagged *Mp*GlpO was acquired from collaborators in the laboratory of Dr. Al Claiborne at Wake Forest University. The protein was sent at a concentration of 10 mg/mL in a buffer of 50 mM potassium phosphate, pH 7.0, and 0.1 mM EDTA. Having the purified and concentrated protein available allowed me to proceed with the next step of growing crystals.

Crystallization, Data Collection, and Data Processing

Crystallization was carried out using the sitting drop and hanging drop methods, which allow the buffer conditions in which a protein is dissolved to slowly change as a

drop of solution containing the protein slowly evaporates to achieve equilibrium with a reservoir solution. Upon being trained in how to use the Phoenix crystallization robot developed by Art Robbins [35], I set up initial crystallization trials using 96 different buffer conditions that are commercially available in the 'Hampton Index' crystallization screen.[36] These experiments are set up in special 96-well plates, which for each buffer condition, allows for three different mixing ratios of protein stock and the reservoir. This first screen used protein:reservoir mixtures as follows: drop 1 - 0.25:0.50 μL ; drop 2 - 0.25:0.25 μL ; drop 3 - 0.50:0.25 μL. Within a week at 4℃, yellow crystals formed in some conditions, but since a large portion of the drops had precipitated protein, we concluded the protein concentration was higher than would be ideal. We thus diluted the protein by 50% and prepared further 96-well plates using the above ratios of drops, again using the Hampton Index for a control, as well as using other commercial reservoir varieties (Hampton Crystal Screen I and II, Precipitant Synergy, Wizard I and II, and Wizard III and IV). A representative crystal (originally designated cpaj), grown in a reservoir of 2.68 M NaCl, 3.35% v/v isopropanol, and 0.1 M HEPES pH 7.5 (from Precipitant Synergy Screen #7) at 4℃, was harvested.

The crystal was then exposed to X-rays at a laboratory X-ray source at Oregon State University using Cu-K_α X-rays from a Rigaku rotating anode generator set at 50 kV and 100 mA and a Raxis IV X-ray detector. The initial diffraction pattern indicated that the crystal was in fact protein (as inferred from the spots being relatively close together) but that it was useable only out to ~6 Å resolution, insufficient to determine a protein structure. An assessment of crystal formation gave us leads to pursue. There was a distinction made in the shape of the crystals because this affected diffraction quality.

Those that were sharp and pyramidal gave better resolutions and statistics than those that were rounded.

Lead reservoir conditions were optimized by using 24-well hanging drop plates with a 400 μL reservoir volume. Optimization consisted of making new reservoir solutions that were altered slightly from the original conditions and comparing the quality of the new crystals formed versus the initial lead from the 96-well plate. This produced numerous useable crystals. This process was repeated for the top 3 leads. Yellow, tetrahedral-pyramidal crystals (like that shown in Fig. 3) were grown from one lead optimization tray within a week. The largest and sharpest crystals (what we hoped would be of the best diffraction quality) were grown in 2.5 M NaCl, 0.1 M Imidazole (pH 7.15) at 4℃ with a ratio of 1 μL of protein: 2 μL of reservoir.

One crystal (named cpaz in files, here called GlpOox1), of size 0.35 mm x 0.35 mm x 0.35 mm, was retrieved out of this solution. Collected in a small loop, the crystal was cryo-protected with oil, submerged in liquid nitrogen, packaged, and sent to a synchrotron source: the Advanced Light Source at Lawrence Berkley National Laboratory in Berkley, CA. There, a complete dataset was collected (360 images, 300 mm detector distance, $\lambda = 1.0$ Å). This was a case where I was fortunate to have well-behaved crystals. They grew quickly (within a week) and were not destroyed upon my manipulation to retrieve them with the loop or with exposure to X-rays. No observable flavin reduction was observed as crystals remained yellow after data collection. A reduced crystal would appear pale yellow or even opaque. Further, they lasted long in solution if kept at 4℃, allowing us to extend our resources and keep the crystals over many months for use in later experiments.

The initial diffraction patterns had diffraction spots that were observed to be out to ~2.7 Å resolution. The images were input into iMosflm and processed. The crystal GlpOox1 was indexed as space group P23 and integrated with a unit cell of a=b=c = 112.18 Å and α, β, γ = 90°. The resulting integrated data were then input into the CCP4 program Scala to scale and merge the symmetry related reflections to give the final intensities. Based on data statistics given by the program and general recommendations for resolution cut-offs, we determined that GlpOox1 data set to have a resolution limit of 2.5 Å. For data collection statistics, see Table 1.

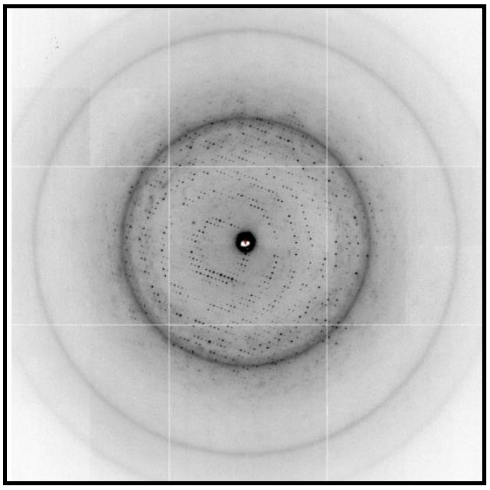
An initial structure was determined using data from GlpOox1. However, because of the limit of data quality, we tried collecting data from additional native crystals in order to get higher resolution data. This would allow us to get a more accurate representation of the structure of *Mp*GlpO. Yellow, trigonal-pyramidal crystals of ca. 0.3 mm x 0.3 mm x 0.3 mm were obtained within a week in reservoir conditions of 2.68 M NaCl, 0.1 M HEPES (pH 7.5), and 2% v/v isopropanol (optimized from Precipitant Synergy Screen #7) at 4℃. Ice content can also affect data quality so we optimized the cryo-protectant which serves to protect the crystal while in liquid nitrogen. Solutions of AML mixed with varying concentrations of glycerol were prepared and similar crystals were allowed to soak in each condition for three minutes. Results of this experiment are shown in Figure 4. 15% glycerol in an AML solution was chosen as the optimal cryo-protectant because it was the smallest concentration in which few ice rings were detected. A representative crystal (called cpck in files, here called GlpOox2) was scooped out of the drop and placed in artificial mother liquor (AML) of 3.0 M NaCl, 0.1 M BisTris (pH 7.0) for one hour. The crystal was then stored for 3 minutes in the AML plus 15%

Table 1: **Data Collection and Refinement Statistics for GlpOox and GlpOred**

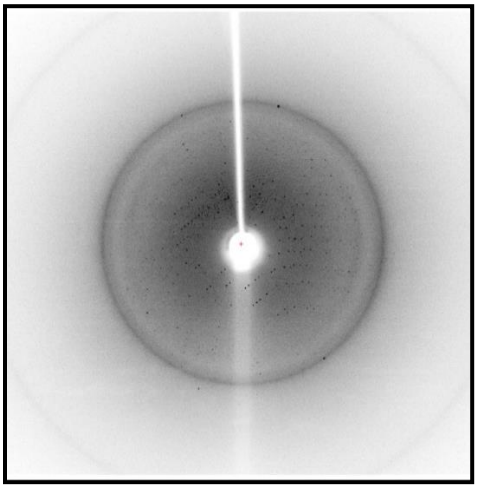
	GlpOox	GlpOred
Data quality statistics		
X-ray wavelength (Å)	0.9765	0.9765
Space group	P23	P23
Unit cell axis length (Å)	111.59	111.61
Resolution range(Å)	49.90-2.40	55.80-2.50
High resolution bin range (Å)	2.53-2.40	2.64-2.50
No. of reflections	430576	836296
No. of unique reflections	18401	16368
Completeness (%)	99.79	100
Multiplicity	23.4 (23.6)	51.1 (28.6)
R_{pim}	0.022 (0.185)	0.072 (0.281)
R_{meas}	0.109 (0.900)	0.533 (1.514)
I/σ	5.7 (0.90)	0.7 (0.50)
Refinement statistics		
R_{factor} (%)	15.8	16.4
R_{free} (%)	21.4	22.6
Number amino acid residues	384	384
Number solvent atoms	201	202
Total number atoms	3305	3285
Average B ($Å^2$) protein atoms	30.8	46.1

* Values in parentheses are for the high resolution bin

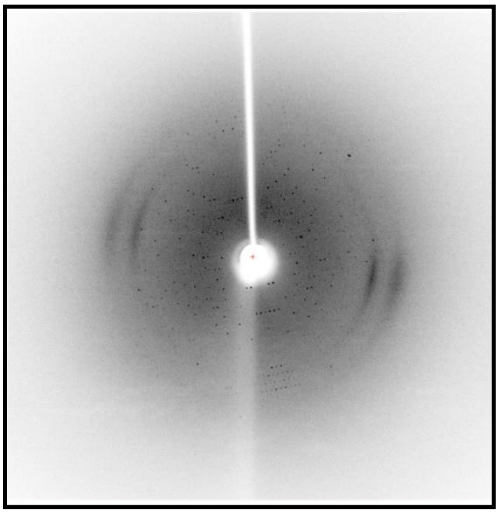
Figure 4: ***Mp*GlpO Cryo-protectant Optimization Images**. Crystals were soaked in a variety of glycerol concentrations to determine the optimum conditions for data collection. A) Initial harvesting was performed with oil and no glycerol. Ice rings circle the data and can reduce data quality. B) Ice rings still persist when the crystal was soaked in 10% glycerol. C) No large ice rings are visible when soaked in 15% glycerol. D) The diffraction pattern when the crystal was soaked in 20% glycerol looks similar to that with 15% glycerol. We chose to use 15% glycerol for harvesting because it produced the cleanest diffraction pattern with the least amount of glycerol.



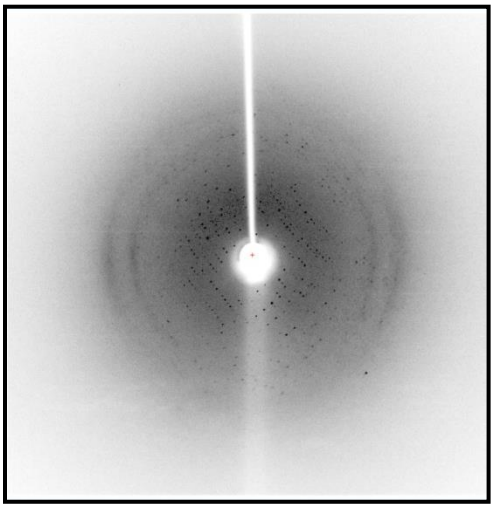
A: Crystal cpaz dragged through oil- 0% glycerol



B: Crystal cpbk soaked in 10% glycerol for 3 min



C: Crystal cpbm soaked in 15% glycerol for 3 min



D: Crystal cpbq soaked in 20% glycerol for 3 min.

glycerol. This crystal diffracted to 2.4 Å resolution, meaning it gave us more useful information. The unit cell dimensions and data quality statistics for GlpOox2 can also be found in Table 1. Reduced crystals and substrate soak crystals were grown in similar conditions. This experiment was performed in order to elucidate how the structure may change with FAD reduction or ligand binding. Ligand binding in particular is important for structure-based drug design because it may give insight into how a potential drug binds to inactivate the enzyme. We soaked crystals in 10 mM dithionite in a degassed AML or in 10 mM of the respective substrate in a 15% glycerol AML for one hour. A total of 24 crystals were either soaked in one of four substrates or reduced via two different methods. We collected 18 datasets. For substrate soak conditions and their resolution limits see Table 2. Two data sets of dithionite-reduced crystals (designated cpcy and cpcz in files, here called GlpOred1 and GlpOred2 respectively) were collected out to 2.5 Å resolution (240 images, 360 mm detector distance, λ = 1.0 Å) and the reduced data from the two crystals were merged together to produce a single data set with better data quality. The substrate-soaked crystals were similarly exposed to X-rays at the synchrotron source and data were collected, although none of them showed any ligand binding.

Structure Determination and Refinement of Oxidized *Mp*GlpO

Molecular replacement

The intensities of the diffraction data were converted into structure factor amplitudes (F_{obs}) using Truncate from CCP4. However, subsequent attempts using these structure factors to run molecular replacement using the GlpO model from *Steptococcus*

Table 2: **Overview of *Mp*GlpO with Soaked Substrates and Results**. The results of numerous ligand soaks are presented. Closely identical native crystals were grown and soaked in a variety of conditions in an attempt to get ligands to bind. All harvesting conditions mirrored that of the native crystal, having both AML and 15% glycerol with no ligands present. Box shading indicates the color of the crystals after soaking. No bound substrates were seen in the protein structure.

Soak	Crystal ID	Harvesting Conditions	Substrate Bound?	Resolution
Native	cpck	1 hour in AML with 15% glycerol	No	2.4
2phophoglycerate	cpdg	10 mM 2-phosphoglycerate for 1 hour, 3 minutes in AML + 15% glycerol	No	2.5
	cpdh			2.6
	cpdj			2.6
G3P	cpcl	10 mM G3P for 1 hour, 3 minutes in AML + 15% glycerol	No	2.7
	cpcn			3.0
PEP	cpch	10 mM PEP for 1 hour, 3 minutes in AML + 15% glycerol	No	2.7
	cpcj			2.9
Dithionite+mv	cpdc	10 mM dithionite, 0.5 mM methyl viologen for 1 hour in degassed AML, 15% glycerol	No	3.2
	cpdd			2.7
	cpde			2.9
	cpdf			2.7
Dithionite	cpcy	10 mM dithionite for 1 hour in degassed AML, 15% glycerol	N/A	2.7
	cpcz			2.5
	cpda			2.6
	cpdb			2.6
Tartaric Acid	cpcr	10 mM tartaric acid for 1 hour, 15% glycerol	No	2.8
	cpct			2.8
	cpcv			2.7

sp. (PDB entry 2RGH) were unsuccessful. We initially thought that the problem was the parameters set in the program, so we adjusted those to no avail. We then tried modifying our search model of *Ss*GlpO. Sometimes, replacing each amino acid residue with an alanine can provide enough structural similarity to allow the program to find an acceptable initial model. However, this was a very inefficient procedure and we decided to tackle it if other strategies did not help. Another idea was to increase the number of molecules to search for in the asymmetric unit (ASU). The ASU is the basic unit of the crystal that, when translational and some rotational operations are applied, forms the complete crystal. To solve a protein structure one must only model the unique information that is contained in the ASU. The programs that initially failed were directed to try and find one *Mp*GlpO molecule in the ASU based on a prediction made by the Matthew's Probability Calculator.[37,38] Since this was just a prediction, we tested whether increasing the number of search models would be sufficient to calculate the phases. Again, this was of no success. A final issue was addressed when we explored the output log files of the two failed programs. As it turned out, the molecular replacement program was not detecting the presence of reflections. This means that the program was working without any diffraction data. Working backwards, we determined that the error was in the Truncate program from CCP4, which due to a bug was outputting an empty file. We thus used an older version of the program and then were successful in getting the molecular replacement programs to recognize the reflections. This really taught me to look at the log output files in troubleshooting.

One important lesson from this detour was the importance of having a structurally similar search model. We were skeptical that the level of sequence similarity with

*Ss*GlpO was enough, so I ran a database BLAST[39] search to look for any more similar protein homologs. The solved structure that was the most similar to *Mp*GlpO in sequence was one from *Bordetella pertussis* (PDB entry 3DME) with a similarity score of 89.4 bits and a level of sequence identity of 27%. Throughout this thesis, this structure is referred to as *Bp*3DME. *Ss*GlpO wasn't even on the top 100 list and was thus presumed to not be the best choice to use as a search model for molecular replacement. An alignment of the sequences of *Ss*GlpO and *Mp*GlpO reveal a similarity score of only 21.2 bits. A high sequence identity was given but is only valid for 34% of the protein. However, we did suspect that *Ss*GlpO could be a useful search model if we could extract just the core GlpO protein. We compared *Ss*GlpO and 3DME using Pymol[40] and were able to create a modified *Ss*GlpO model that was missing the C-terminal domain (amino acids 453-606), the first α-helix (amino acids 1-17), an extraneous loop (amino acids 241-251) and waters to create '*Ss*GlpO-truncated.' Sequence alignment of *Ss*GlpO-truncated to *Mp*GlpO still, however, did not produce a reasonable alignment without numerous gaps in the sequences.

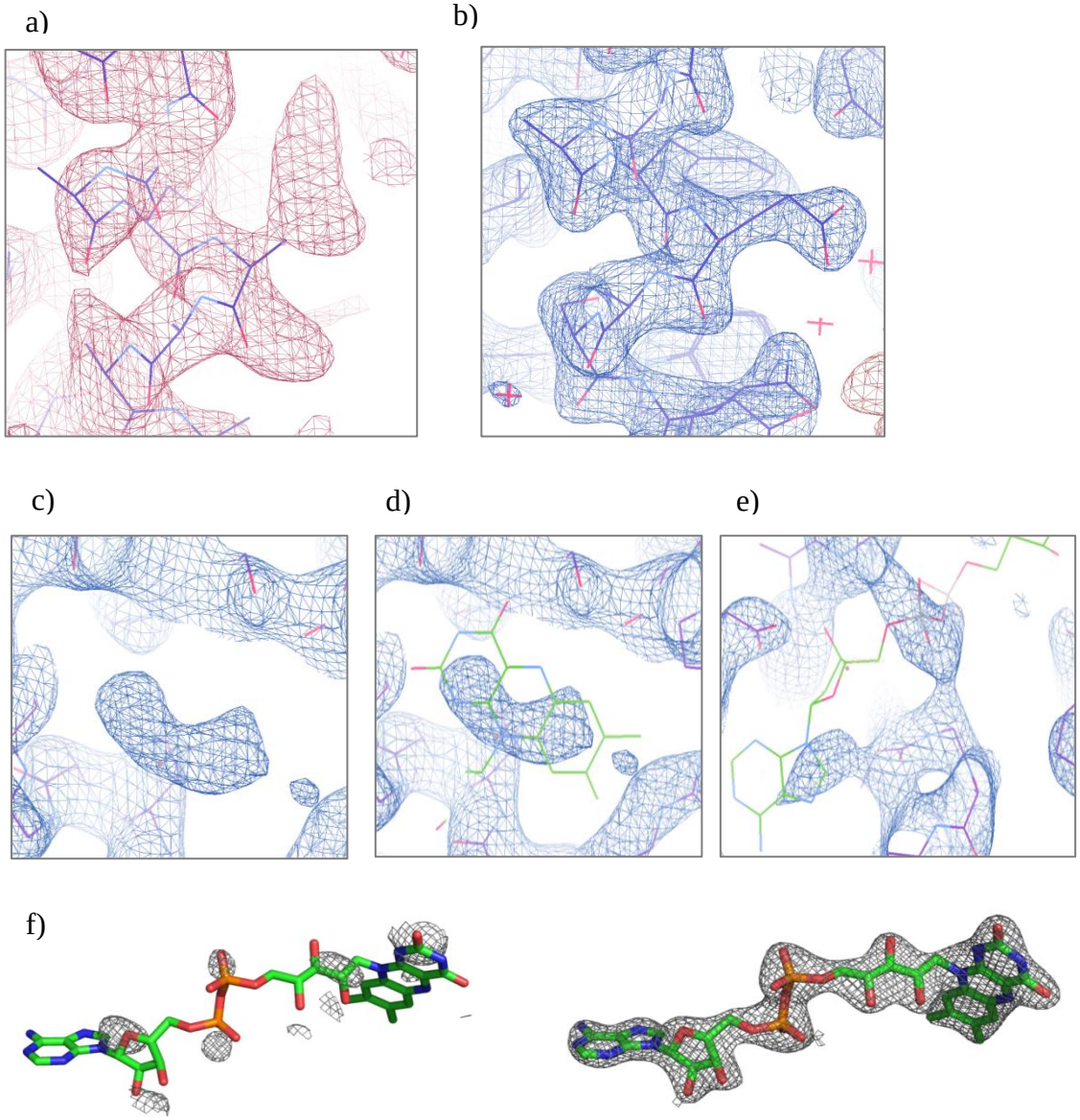
Molecular replacement was tried using either *Ss*GlpO-truncated or *Bp*3DME with one, two, or three molecules in the ASU as parameters. However, it was not until we switched to the Auto-MR program of Phenix (another computer suite of programs in X-ray crystallography; [41]), that we were successful in getting an initial model that was a true solution. Chain A of *Bp*3DME was used as a search model in which to calculate initial phases based off the structure factors of the *Mp*GlpO data. The log-likelihood gain (LLG) was 57.7. This is a measure of how much better the model is in comparison to a random distribution of the same atoms.[34]

It was essential to verify our solution and we did this by having an internal control. The GlpO activity of *Ss*GlpO is due to a FAD molecule.[30] *Bp*3DME also contains an FAD molecule, even though the function of the protein is currently unknown. By deleting the FAD from the model in refinement, we could be sure that any density that was around the FAD region was not due to model bias. Further, we identified a key location where a larger amino acid (such as a glutamate at residue 21 in *Mp*GlpO) was supposed to be in the location of a smaller amino acid (residue alanine 20 in *Bp*3DME), based on the sequence alignment (see Figure 13 in results/discussion). Rather than just accepting the model from the first refinement, we ran the same refinement program multiple times but with slightly different parameters and were able to pick the model with the best density to build into and refine. This process continued until specific parameters were narrowed down as most beneficial for refinement. Pictures of both the FAD and glutamate 21 in their electron density from the initial refinements and the final structure can be seen in Figure 5. The large peaks of electron density in the location where the *Mp*GlpO should have had more atoms gave us confidence that the molecular replacement solution was real. We could then proceed to refinement of the model to complete it and improve its fit to the density.

Manual modeling and refinement – stage 1: refinement of GlpOox1 model

Once a molecular replacement solution has been verified, in many cases, the model refinement can be quickly completed if there is high similarity between the search model and the structure being solved. 3DME, however, had only a 27% sequence identity with *Mp*GlpO. As such, attempts to automatically convert the amino acids of 3DME into the amino acids of *Mp*GlpO were unsuccessful. Manual modeling of *Mp*GlpO was

Fig. 5: **Ala 22 Comparison and FAD Internal Control**
a) Initial model of MpGlpO for comparison of Glu 21 (*Mp*GlpO numbering) before manual remodeling. Side chain of Glu 21 was modeled into the large empty peak of the $2F_o$-F_c.electron density map. b) Same region of the completed MpGlpO model. c) A region of electron density when FAD was not in the input model. d) A flavin was placed over this electron density to show that this piece of the FAD molecule roughly fit within the shape of electron density. e) Region of strong electron density. When FAD was placed in the model to align with the electron density in the first panel, the pyrophosphates of the FAD aligned exactly within this strong electron density. This indicated that FAD belonged in that place in the model and that our molecular replacement solution was valid. f) Poor electron density for the FAD in an initial structure is contrasted with the strong electron density for the FAD in the final structure.



therefore the only viable option to pursue. This involved investigating each individual amino acid position, assessing which amino acid should be substituted in, and then doing so only when the electron density map gave evidence for this change. Both visualizing and modifying of the model were done in a program called COOT (Crystallographic Object-Oriented Toolkit; [42]). Because the experimental electron density map was the best indicator of what model features would satisfy the observed data, it was essential to follow where the density guided. While this was more laborious and time-consuming, walking through the protein allowed me to appreciate the intricacies of the structure and understand how each individual amino acid residue plays a role in determining the three-dimensional structure. This process also taught me about protein chemistry and geometry. One important lesson was not to put full trust in a sequence alignment. There were many instances where a gap or insertion in the alignment I had originally created was not verified by the density of the structure. A table summarizing the many rounds of model building and subsequent computational refinement statistics is given in Table 3. After determining that *Bp*3DME was a suitable search model, six more rounds of simulated annealing and minimization refinements were performed with GlpOox1 (with both Cartesian and Torsion Simulated Annealing [41]) coupled with manual model building using COOT, and leading to an improved model with R and R_{free} values of 0.32 and 0.45 respectively.

Manual modeling and refinement – stage 2: addition of GlpOox2 data

As mentioned previously, we continued to collect data on additional crystals in order to get better diffracting crystals. One such crystal, GlpOox2, yielded data useful to a resolution of 2.4 Å. The intensities were converted to structure factors, and R_{free} flags

Table 3: **Rounds of Refinement**. A table of the pathway of model building and subsequent computational refinement statistics is presented. Starting with round 9, the GlpOox2 data were used. As refinement progressed, the number of atoms placed increased while R and R_{free} decreased. Starting at round 13, a variety of interpretations were tried for the strong electron density at the crystallographic 3-fold (listed in the last column labeled "Metal").

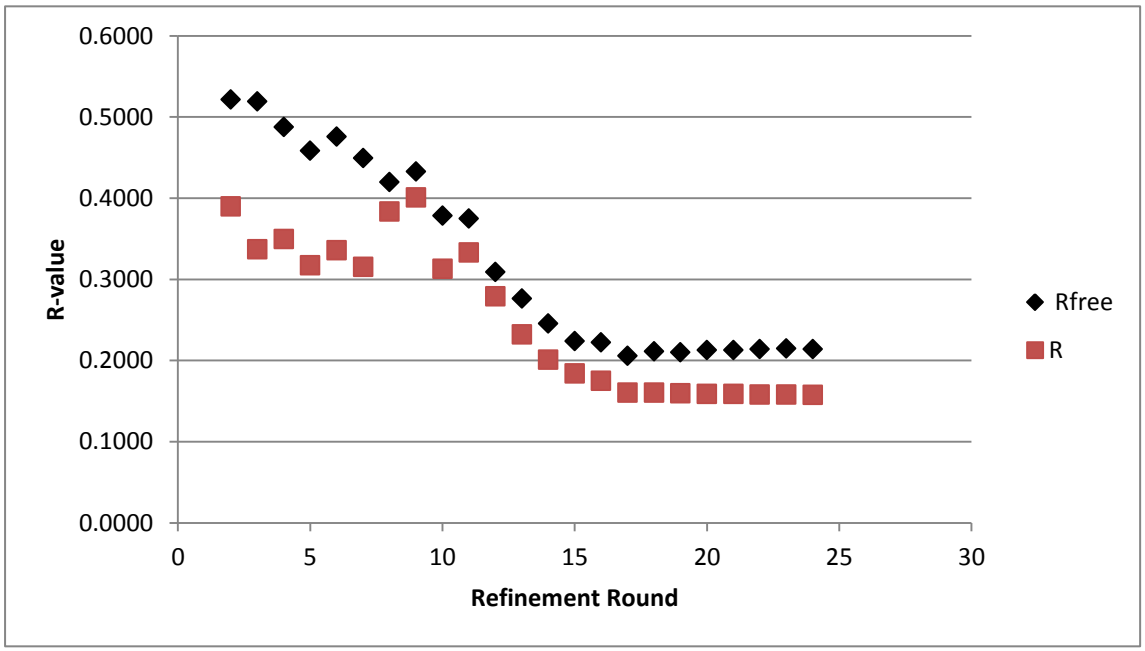
Rounds of Refinement Table												
Refinement Round Number	Description	File Name	Program	Resolution limit (Å)	Number of residues	Number of atoms	Solvent Molecules	R initial	R final	Rfree initial	Rfree final	Metal
1	Auto-Molecular Replacement	AutoMR_run_2	CCP4	N/A	363	2688	0					-
2	Cartesian Simulated Annealing	refine_10	Phenix	3.0	366	2688	0	0.5574	0.3897	0.5538	0.5215	-
3	Cartesian Simulated Annealing	refine_21	Phenix	3.0	366	2688	0	0.4291	0.3370	0.5356	0.5191	-
4	Cartesian Simulated Annealing	refine_37	Phenix	3.0	366	2741	0	0.6869	0.3498	0.5235	0.4877	-
5	Torsion Simulated Annealing	refine_45	Phenix	3.0	366	2693	0	0.3998	0.3172	0.4800	0.4583	-
6	Minimization	refine_63	Phenix	2.8* resolution cut off within program	347	2570	0	0.3925	0.3359	0.4731	0.4758	-
7	Torsion Simulated Annealing	refine_72	Phenix	3.0	347	2570	0	0.3754	0.3153	0.4442	0.4494	-
8	Buster refinement with new 2.4Å data	Buster-10	Buster	3.0	326	2294	0	0.3928	0.3835	0.4364	0.4198	-
9	Buster refinement - no manual rebuild in Buster-10	Buster-11	Buster	2.4	326	2294	0	0.4116	0.4008	0.4370	0.4328	-
10	Cartesian Simulated Annealing	refine_92	Phenix	2.4	363	2702	0	0.3947	0.3131	0.3830	0.3786	-
11	Buster refinement	Buster-15	Buster	2.4	363	2702	0	0.3092	0.3333	0.3801	0.3751	-
12	Buster refinement	Buster-16	Buster	2.4	372	2899	0	0.3157	0.2790	0.3299	0.3091	-
13	Buster refinement	Buster-18	Buster	2.4	384	3045	0	0.2685	0.2324	0.2817	0.2763	Na
14	Buster refinement	Buster-19-3	Buster	2.4	384	3028	65	0.2271	0.2010	0.2421	0.2456	Na
15	Buster refinement	Buster-20	Buster	2.4	384	3068	150	0.2138	0.1840	0.2192	0.2238	Na
16	Buster refinement	Buster-21-1	Buster	2.4	384	3059	185	0.2050	0.1752	0.2133	0.2222	Na
17	TLS and Restrained Refinement	Refmac_83	Refmac	2.4	384	3104	191	0.2225	0.1603	0.2411	0.2057	K
18	TLS and Restrained Refinement	Refmac_87	Refmac	2.4	384	3112	195	0.2165	0.1605	0.2536	0.2113	SO_4^-
19	TLS and Restrained Refinement	Refmac_89	Refmac	2.4	384	3107	198	0.2145	0.1595	0.2567	0.2103	H_2O
20	TLS and Restrained Refinement	Refmac_91	Refmac	2.4	384	3108	198	0.2152	0.1588	0.2510	0.2130	Zn-S
21	TLS and Restrained Refinement	Refmac_104	Refmac	2.4	384	3109	200	0.2169	0.1590	0.2554	0.2128	Zn-H_2O
22	TLS and Restrained Refinement	Refmac_105	Refmac	2.4	384	3109	203	0.2110	0.1581	0.2559	0.2141	Ni-Cl
23	TLS and Restrained Refinement	Refmac_110	Refmac	2.4	384	3109	199	0.2128	0.1581	0.2595	0.2150	Ni-Cl
24	TLS and Restrained Refinement	Refmac_122	Refmac	2.4	384	3109	200	0.2097	0.1578	0.2589	0.2139	Ni-H_2O

were imported from the previous work done using GlpOox1. We did not want to start back at the beginning of molecular replacement since we had already made real changes in the search model structure that were validated by the density of *Mp*GlpO. We were able to extend the work we had already accomplished by merging the data and importing the R_{free} flags. Refinement using these new data, however, continued using the same strategy involving rounds ofmanual model rebuilds. With the improved data quality, the model quality indicators R and R_{free} decreased quickly, as shown in Figure 6. Sometimes in manual refinement, large changes to the protein structure do not change the refinement statistics substantially, leaving values around the 0.4 to 0.5 region. Incorporation of this improved data from GlpOox2 helped me to get over the hump of manual refinement and see drastic improvements in the phases and electron density that enabled me to model *Mp*GlpO more effectively.

Manual modeling and refinement – stage 3: final adjustments and metal chemistry

Using the techniques described above, the main chain and most side chains of *Mp*GlpO were modeled. Water molecules were added after round 13 with the following criteria: (1) a peak of ≥ 1 ρ_{rms} in the electron density map, and (2) a distance of ≥ 2.2 Å and ≤ 3.6 Å between the water and nearby hydrogen-bond donor or acceptor. On the last round of refinement, water molecules were renumbered in accordance with their density with water-1 having the highest electron density. Residue Lys79 was left in a stubbed state (i.e. modeled as an Ala). This is a specific example of where we didn't have evidence for where to place the side chain because little to no density was present for atoms beyond the Cβ-atom. MolProbity [43] and other verification tools in COOT were utilized in later stage refinements. One large puzzle was figuring out the meaning of an

Figure 6: **Progress of Model Building and Refinement as Monitored by R and R_{free}.** The R and R_{free} values are plotted against the round of refinement. With the iterations of manual model building, both R values decreased until leveling out around 0.15 and 0.21, respectively.



unexpected electron density peak at a three-fold symmetry axis, near the side chain of His59. An anomalous difference map calculated using the CCP4 suite of programs gave a peak suggesting it was a metal. Then, remotely we conducted a set of fluorescence scans using the synchrotron at Lawrence Berkeley National Laboratory that were near wavelengths appropriate for various metals, helping us identify the metal as a nickel ion (see Figure 10). Manual modeling and Refmac refinements were continued to give a final R and R_{free} values of 0.158 and 0.214, respectively, for the oxidized, native *Mp*GlpO after 24 rounds of refinement. Alternate side-chain or Cα conformations were modeled for residues Gln40, His244, Trp375, Asn376, and Gly377.

Diffraction data from a crystal chemically reduced by 10 mM sodium dithionite was combined and processed, with the R_{free} test set imported from the oxidized GlpO data set. Rigid body refinement was performed to give R and R_{free} of 0.2059 and 0.2426, respectively. A similar method was performed with dithionite reduced data with added methyl-viologen. An F_o-F_o difference map of the dithionite reduced data minus the oxidized GlpO data was created and analyzed to find structural changes caused by the reduced state. Minor changes were made to the model. Manual modeling and Refmac refinements were carried out to give a final R and R_{free} values of 0.164 and 0.226, respectively.

For each substrate soak data set collected, an F_o-F_o difference map was created to visualize the density differences between GlpOox2 and the soaked crystal. A search for bound ligands was primarily centered on the active site, above the flavin.

Structural Comparisons

Once the completed oxidized model was built, we compared *Mp*GlpO to other proteins with similar structures. Using the Dali server[44] we identified four structures of interest. We compared *Bp*3DME, *Ss*GlpO, *Ec*GlpD (an aerobic glycerol-3-phosphate dehydrogenase with PDB code of 2QCU), and *Bs*GlyOx (a representative glycine oxidase with PDB code of 1RYI). The Dali results were also used in creating a structure-based sequence alignment. I tried multiple programs, but with Dali, only some modification was needed to create an alignment that worked well for all five structures. DSSP[45] and Pymol[40] were used for defining the secondary structure assignments that were used to make a topology diagram.

CHAPTER 3

RESULTS AND DISCUSSION

In this section, I describe the overall structure and make comparisons between *Mp*GlpO and proteins with similar structures, aiding us in understanding how this enzyme works and leading to evolutionary and drug design implications.

Enzymatic Activity and Oligomeric Structure in Solution

Recombinant *Mp*GlpO was successfully expressed and purified as an N-terminally His-tagged protein, and certain solution properties of recombinant *Mp*GlpO have be assessed by collaborators in the laboratories of Dr. Al Claiborne (Wake Forest University) and Dr. Pimchai Chaiyen (Mahidol University, Thailand). Although these experiments are not documented here, I summarize them briefly as they provide an important part of the context of the structural studies. One FAD resides in each enzymatic chain. The native molecular mass of *Mp*GlpO is 42 kD based on gel filtration analysis, and compared with the calculated subunit molecular mass of 45 kD, this indicates that *Mp*GlpO is a monomeric enzyme. This is intriguing because members of the structurally and sequentially similar DAAO family are dimeric enzymes.

The first published data about the activity level of *Mp*GlpO indicated that it is in fact an oxidase instead of a dehydrogenase.[3] Unpublished, preliminary data from the laboratory of Dr. Al Claiborne indicates that *Mp*GlpO has an activity level of 12.8 units/mg of protein. This is approximately 25% of the activity of *Ss*GlpO, for which the specific activity is 60.3 units/mg of protein. Since the structure of *Bp*3DME was solved by the NorthEast Structural Genomics group, there was no publication reporting on the

enzymatic activity in this enzyme. The Claiborne group has shown it has no GlpO activity (also unpublished data).

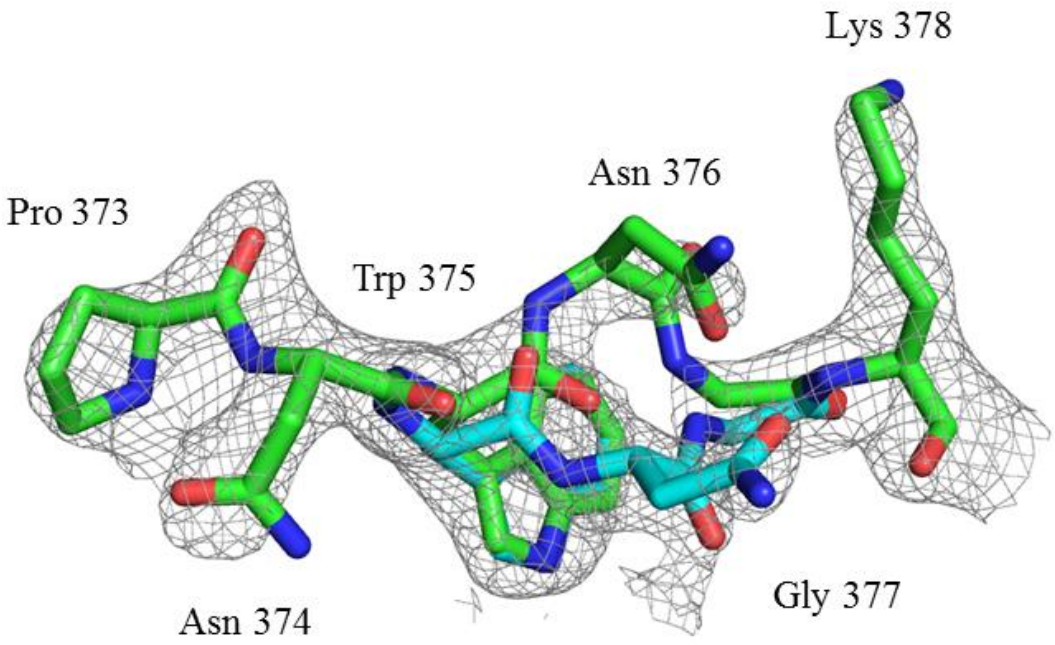
Solution of the *Mp*GlpO Structure

The structures of native *Mp*GlpO in both the oxidized and the reduced flavin states have been determined at 2.4 Å and 2.5 Å resolution, respectively. The oxidized structure was determined by molecular replacement (see methods) and refined to a final R of 0.158 and R_{free} of 0.214, and the reduced structure was solved by a simple difference Fourier analysis and refined to a final R of 0.164 and R_{free} of 0.226 (see Table 1).

The modeled structure of oxidized *Mp*GlpO consists of 384 amino acids, including residues 1-384 modeled for the single chain in the asymmetric unit, one FAD, 201 water molecules, and one nickel ion. The His-tag was not visible in the structure. There are a few places in the structure in which an alternate backbone pathway or side chain conformation was supported by the electron density. One such alternate pathway was modeled from Trp-375 to Gly-377 (Fig. 7). As model building progressed, a strong F_o-F_c density peak near Asn-376 matched with a shifted position of Asn-376. In order to properly fit this density, Trp-375 and Gly-377 were also duplicated and allowed to adjust. This alternate pathway was important to model because the protein, as it moves and functions, has an equal chance of being in either of these paths.

In a protein structure, it is also common to see alternate conformations for some amino acid side chains. This can best be seen in *Mp*GlpO with His-244. Due to bond geometry and atomic interactions, there are only certain local side chain conformations that are favorable. The electron density gave evidence that the imidazole ring of the

Fig. 7: **An Alternate Backbone Conformation Example**. Because proteins are mobile, they can take a variety of equally likely paths. In *Mp*GlpO, there was electron density to support 376-Asn in both the green and cyan positions. In order to place the cyan amino acid, 375-Trp and 377-Gly were both slightly adjusted. As a result the protein path appears to diverge and then converge.



histidine residue, the bulk of this amino acid, could take on either of two conformations (Fig. 8). Since the $2F_o$-F_c density for each of the two positions had roughly the same peak height, we concluded that each of the alternate side chains occur 50% of the time in this crystal form of the protein. In addition to the alternate conformations, the Lys-79 side chain was truncated to an alanine because, while the main chain could confidently be modeled, the side chain had very weak electron density and couldn't be modeled. Since Lys-79 is a surface residue, we presume that it is too mobile to model.

Interestingly, at the crystallographic 3-fold axis, three symmetry-related imidazole side chains of His-59 coordinate a peak of strong density (Fig. 9a and 9b). Analysis of an anomalous difference map provided significant evidence that this is a metal ion (Fig. 9b). In order to identify this metal, we performed a fluorescence scan on the crystals and saw a signal at the K edge for nickel (Fig. 10). As shown in Figure 9a, the atomic distances of Ni at this site and with a water molecule as a fourth ligand fit within typical Ni-coordination guidelines. As this was a His-tagged protein purified over a Ni-affinity column, we suspect some Ni could have come from the column and remained in the protein solution after purification. This contamination was fortunate because, based on the location of the metal at the crystal contact, crystal formation may have only happened because of its presence.

Overall Structure

In terms of the tertiary structure, *Mp*GlpO is, as expected, very similar to the *Bp*3DME protein used as a search model and to the Streptococcal GlpO (Figure 11).

Fig. 8: **An Alternate Side Chain Conformation Example**. Proteins can exhibit alternate side chains where pieces of the side chain reside in equally likely rotamer positions. 244-His of *Mp*GlpO has one conformation of the imidazole head in green and another one in cyan.

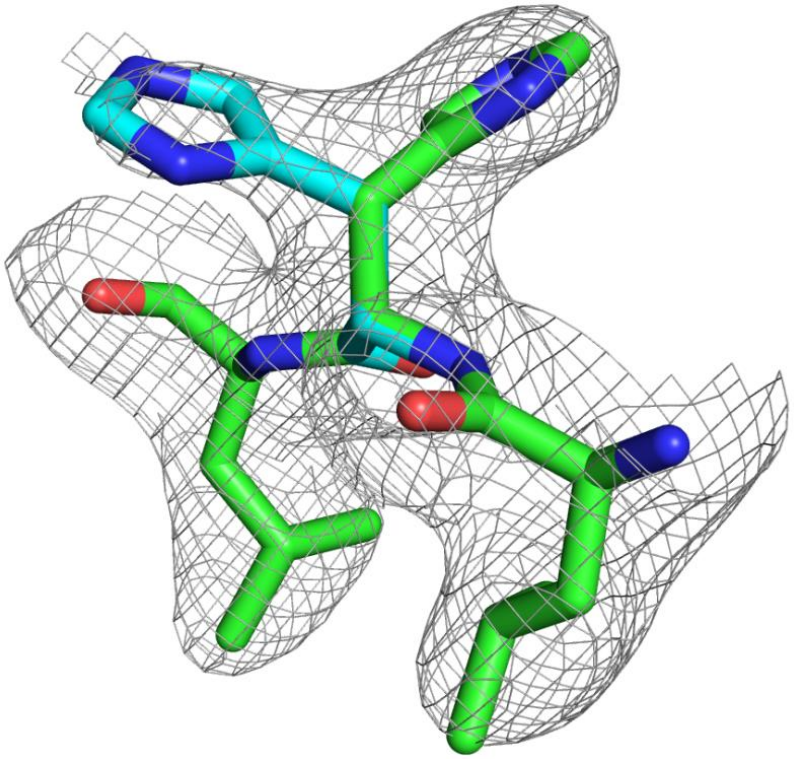


Fig. 9: **Nickel Ion Coordination and Potential Dimer Interface**. A) Ni is bound between three imidazole rings of 59-His in three different molecules. The distance between the Ni and all three imidazole rings is 1.98 Å. A water ion also coordinates to the Ni. This distance is 2.17 Å. B) Packing of the three molecules of *Mp*GlpO around the Ni ion at a three-fold axis in the crystal is shown along with the anomalous difference map peak for the metal (contoured at 14.50 ρ_{rms}). Helix α3 is near the interface. C) A stereo image of the most extensive crystal packing interface that represents a potential dimer interaction. The interface combines two beta sheets to make a new, larger beta sheet interface. The Ni ion (shown with His side chain) is on the opposite side of the potential dimer interface.

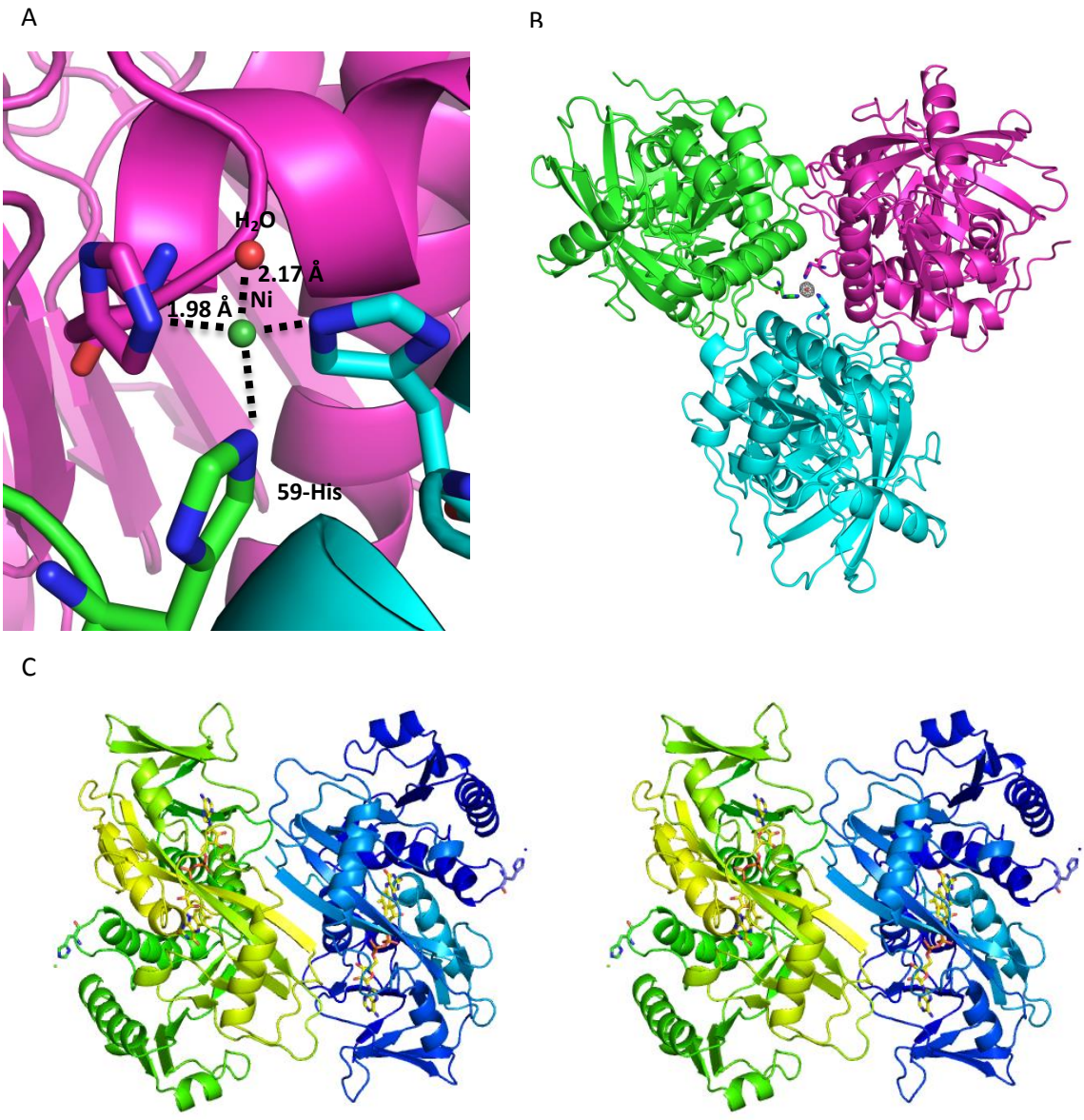


Fig. 10: **Synchrotron Source Scan of *Mp*GlpO aids Identification of the Metal Ion**. Shown is the fluorescence scan with the X-ray wavelengths tuned to the absorption edge of nickel. After trying scans at the absorption edges of various metals, Ni was the only metal which had a clear signal. The known value for Ni is 8339 eV (indicated by the arrow).

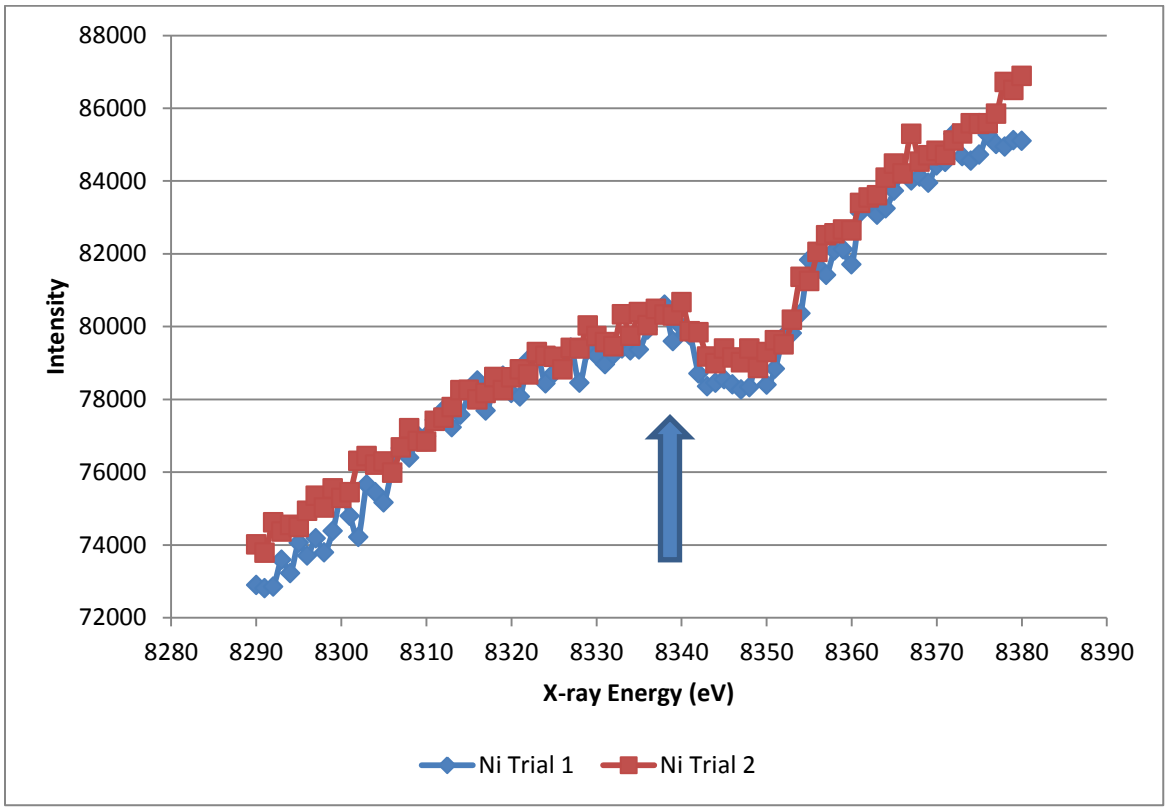
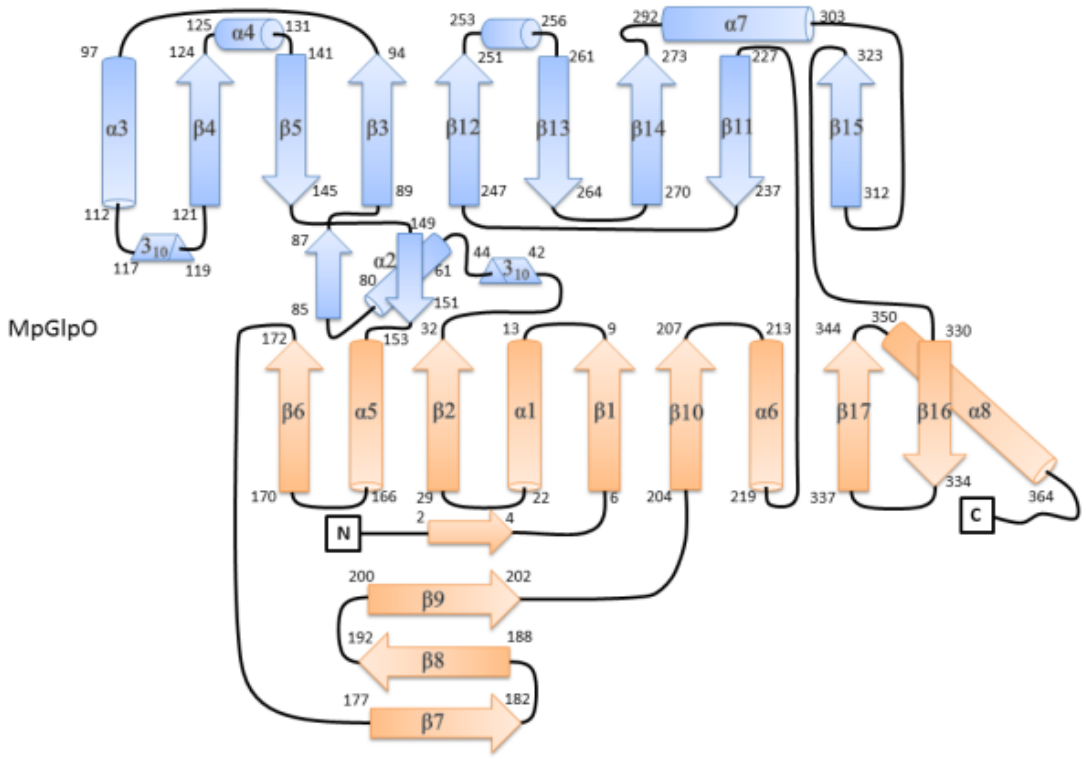
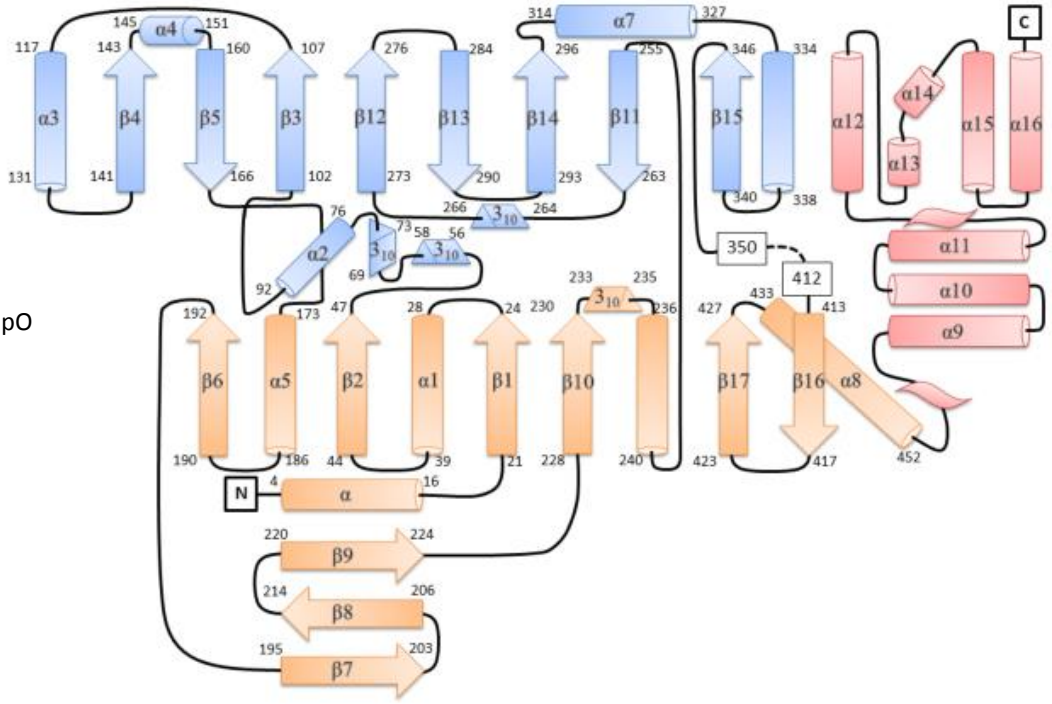


Fig. 11: **The Topology of *Mp*GlpO, *Ss*GlpO, and *Bp*3DME.** A) *Mp*GlpO folds into two discontinuous domains, with a 9-stranded β-sheet in one domain (blue in the topology) and a 6-stranded β-sheet in the other (orange). Both of these β-sheets are mostly anti-parallel. The latter domain also has an additional 3-stranded β-meander. B) *Ss*GlpO has a similar topology to *Mp*GlpO but it has an additional large C-terminal domain also found in similar glycerol 3-phosphate dehydrogenase enzymes. *Ss*GlpO lacks two beta strands that bridge the blue and orange domains. C) *Bp*3DME also lacks the large C-terminal domain and closely resembles the structure of *Mp*GlpO.

A

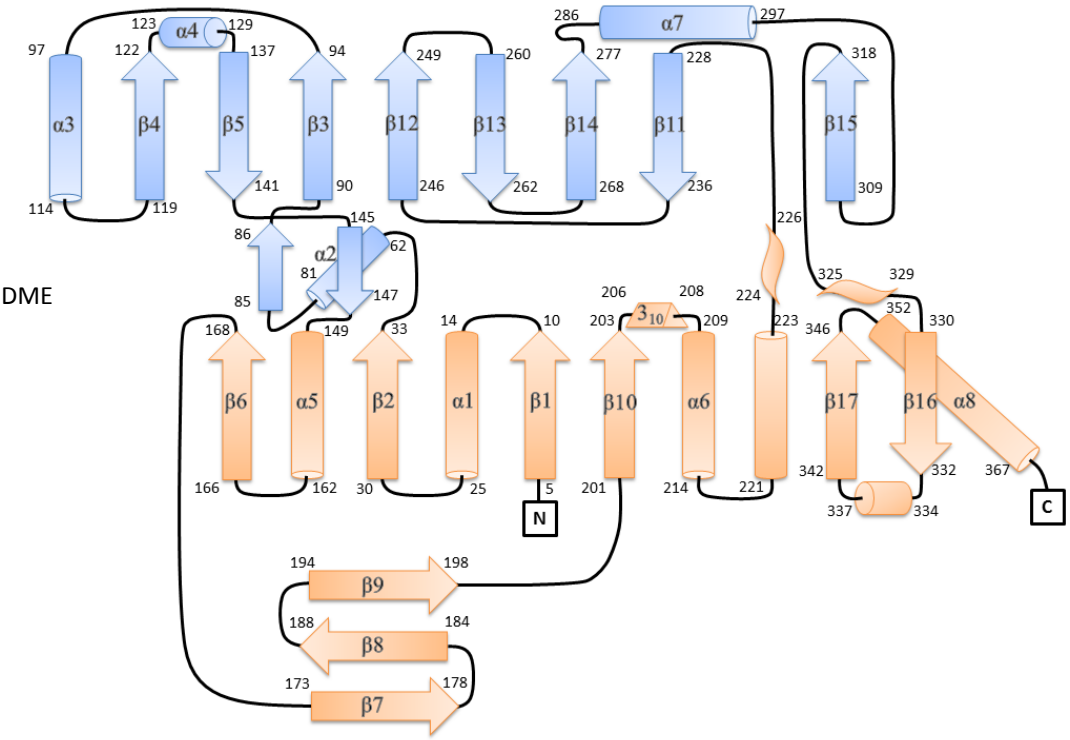


B



*Ss*GlpO

C



*Bp*3DME

*Mp*GlpO folds into two discontinuous domains, with a 9-stranded β-sheet in one domain (blue in Fig. 11A) and a 6-stranded β-sheet in the other (orange in Fig. 11A). Both of these β-sheets are mostly anti-parallel. The latter domain also has an additional 3-stranded β-meander. The domains are connected in two main regions, one more ordered and one more mobile. The more ordered region consists of helix α2 and surrounding secondary structures. Both *Mp*GlpO and *Bp*3DME have two beta strands around the α2 that are not present in *Ss*GlpO. *Ss*GlpO has an overall similar topology but has an additional helical domain at its C-terminal end (red in Fig. 11C). As observed by the three topology diagrams, most of the 2-domain structure of *Mp*GlpO is conserved in *Bp*3DME and *Ss*GlpO. In the protein structure, the domains are wrapped around each other in a way that leaves a big pocket for binding of FAD (Fig. 12).

Using the PISA server[46], we analyzed the packing interactions between *Mp*GlpO molecules in the crystal to see if any of the interfaces could be physiologically relevant. The expected FAD: *Mp*GlpO interface was observed. If *Mp*GlpO were a monomer, this would be expected to be the only significant interaction. However, one very large protein:protein interaction in the crystal represents a potential *Mp*GlpO:*Mp*GlpO dimer interface with approximately 1600 $Å^2$ interaction surface (Figure 9c). While inconsistent with the gel filtration analyses described above, it is consistent with both *Bp*3DME and *Ss*GlpO, both of which are dimers. Members of the DAAO family are also known to be dimers. Comparing which secondary structures interacted to make up the dimer gave intriguing results. Strand β15, part of the 8 stranded β sheet on one *Mp*GlpO molecule, interacts with the same β15 strand on a rotated molecule of *Mp*GlpO (the -x, y, -z+1 symmetry mate) to make a 16 stranded β sheet across the dimer (Fig. 9c). In contrast,

Fig. 12: **Stereo Ribbon Diagram of *Mp*GlpO**. Shown is a ribbon diagram of *Mp*GlpO where the α-helices are represented by spirals and the β-sheets are represented by arrows pointing in the C-terminal direction. The FAD cofactor is shown in yellow. In this view, an open pocket for flavin interaction is easily seen.

*Bp*3DME forms a dimer (with a second molecule in the asymmetric unit of the crystal) via interactions involving helix α6, and *Ss*GlpO forms a dimer via interactions involving helices α2 and α3 (with the -x+1, -y, z symmetry mate). Since the potential dimer interface of *Mp*GlpO is not conserved in either of these most closely related proteins, and since solution studies indicate it is a monomer, we conclude that the dimer packing interface seen in the *Mp*GlpO crystals is not physiologically relevant. However, if gel filtration data showing *Mp*GlpO to be a monomer were inaccurate, this would be the most favorable interface for a dimer.

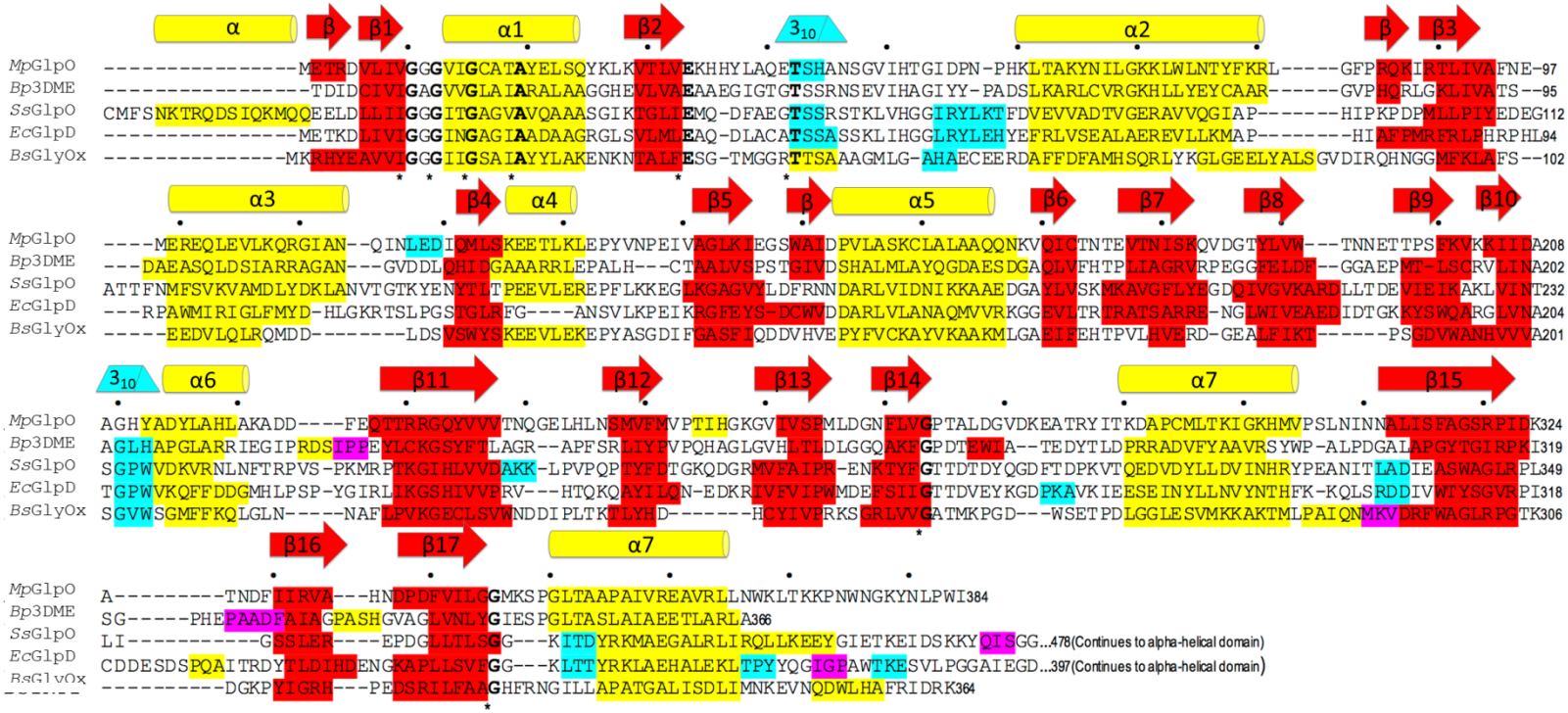
*Mp*GlpO as a Member of the DAAO Superfamily

Looking more closely into the structures of *Mp*GlpO, *Bp*3DME, and *Ss*GlpO, conserved regions in the amino acid sequence were determined. In order to get a bigger picture of the relationships between *Mp*GlpO and other proteins, the program DALI[44] was used to identify the structures available in the Protein Data Bank (PDB) that were similar to *Mp*GlpO. The DALI results are ranked by 'Z-score' which reports how many standard deviations a given comparison is above the mean of all comparisons; Z-scores below 6 are considered uninteresting. Consistent with the original sequence-based identification of *Bp*3DME as the known structure most similar to *Mp*GlpO, the DALI server reported *Bp*3DME as the most structurally similar protein (Z-score = 44.0). *Ss*GlpO was also identified by DALI as being similar, but was actually 171st on the list (Z-score = 27.8). A more closely similar glycerol-phosphate dehydrogenase (pdb code 2QCU) is from the organism *Escherichia coli* (*Ec*GlpD) and was 75th on the DALI list (Z-score = 35.6). Surprisingly, much more similar than these GlpO structures were the

glycine oxidases, for which Z-scores ranged from 41.1 to 41.9. These enzymes made up the top 5 most similar structures along with *Bp*3DME. We selected the highest resolution structure, 1RYI from *Bacillus subtilis* (*Bs*GlyOx), from the top class of glycine oxidases to compare with *Mp*GlpO. Sarcosine oxidases were also very structurally similar and only after a large number of these enzymes did any of the glycerol 3-phosphate dehydrogenases like *Ec*GlpD appear on the list. Glycine oxidases, sarcosine oxidases, and D-amino acid oxidases (DAAO), all make up the DAAO family of flavoproteins. It is interesting that *Mp*GlpO is structurally more similar to more functionally distinct members of the DAAO family than to any of the glycerol 3-phosphate dehydrogenases or oxidases with which it shares a similar function.

An alignment of the amino acid sequences of *Ec*GlpD and *Bs*GlyOx along with *Mp*GlpO, *Bp*3DME, and *Ss*GlpO is presented in Fig. 13. Of particular importance are the secondary structure markers above each row of five aligned sequences. All of the structures align well with regard to these main secondary structural elements and they are labeled with designators (such as α2 and β2) that indicate a consensus secondary structure, the first description of such for glycerol phosphate dehydrogenases and oxidases enzymes. One of the functions of the sequence alignment is that it shows which residues are conserved and where they are found. As can be seen in Fig. 13, the regions that have the most conservation of sequence tend to be in the active site and in certain secondary structures. Fully conserved (i.e. identical) residues among the five aligned structures can be found surrounding and within α-helix 1 (α1), β-strand 2 (β2), the first 3_{10}-helix, β14, and β17. α1 packs against the pyrophosphate of the FAD. β2 is right next to and interacting with the adenine of FAD. The 3_{10} helix is unique in that it contains a

Fig. 13: **Structure-based Sequence Alignment of *Mp*GlpO and Similar Enzymes**. Aligned sequences are shown for *Mp*GlpO with a similar GlpO, GlpD, glycine oxidase, and *Bp*3DME structures with secondary structure types highlighted as red: β-strand, yellow: α-helix, cyan: 3_{10}-helix, and magenta: P_{II}-spiral. Fully conserved (i.e. identical) residues are indicated by asterisks and bold lettering. Every tenth residue of *Mp*GlpO is indicated by a dot above the sequence.


α β β1 α1 β2 3₁₀ α2 β β3
MpGlpO ----------METRDVLIVGGGVIGCATAYELSQYKLKVTLVEKHHYLAQETSHANSGVIHTGIDPN-PHKLTAKYNILGKKLWLNTYFKRL-----GFPRQKIRTLIVAFNE-97
Bp3DME ----------TDIDCIVIGAGVVGLAIARALAAGGHEVLVAEAAEGIGTGTSSRNSEVIHAGIYY-PADSLKARLCVRGKHLLYEYCAAR------GVPHQRLGKLIVATS--95
SsGlpO CMFSNKTRQDSIQKMQQEELDLLIIGGGITGAGVAVQAAASGIKTGLIEMQ-DFAEGTSSRSTKLVHGGIRYLKTFDVEVVADTVGERAVVQGIAP-------HIPKPDPMLLPIYEDEG112
EcGlpD ----------METKDLIVIGGGINGAGIAADAAGRGLSVLMLEAQ-DLACATSSASSKLIHGGLRYLEHYEFRLVSEALAEREVLLKMAP--------HIAFPMRFRLPHRPHL94
BsGlyOx ----------MKRHYEAVVIGGGIIGSAIAYYLAKENKNTALFESG-TMGGRTTSAAAGMLG-AHAECEERDAFFDFAMHSQRLYKGLGEELYALSGVDIRQHNGGMFKLAFS--102
α3 β4 α4 β5 β α5 β6 β7 β8 β9 β10
MpGlpO ----MEREQLEVLKQRGIAN--QINLEDIQMLSKEETLKLEPYVNPEIVAGLKIEGSWAIDPVLASKCLALAAQQNKVQICTNTEVTNISKQVDGTYLVW---TNNETTPSFKVKKIILDA208
Bp3DME ---DAEASQLDSIARRAGAN---GVDDLQHIDGAAARRLEPALH---CTAALVSPSTGIVDSHALMLAYQGDAESDGAQLVFHTPLIAGRVRPEGGFELDF--GGAEPMT-LSGRVLINA202
SsGlpO ATTFNMFSVKVAMDLYDKLANVTGTKYENYTLTPEEVLERFPFLKKEGLKGAGVYLDFRNNDARLVIDNIKKAAEDGAYLVSKMKAVGFLYEGDQIVGVKARDLLTDEVIEIKAKLVINT232
EcGlpD ---RPAWMIRIGLFMYD-HLGKRTSLPGSTGLRFG----ANSVLKPEIKRGFEYS-DCWVDDARLVLANAQMVVRKGGEVLTRTRATSARRE-NGLWIVEAEDIDTGKKYSWQARGLVNA204
BsGlyOx -----EEDVLQLRQMDD--------LDSVSWYSKEEVLEKEPYASGDIFGASFIQDDVHVEPYFVCKAYVKAAKMLGAEIFEHTPVLHVERD-GEALFIKT------PSGDVWANHVVVA201
3₁₀ α6 β11 β12 β13 β14 α7 β15
MpGlpO AGHYADYLAHLAKADD----FEQTTRRGQYVVVTNQGELHLNSMVFMVPTIHGKGVIVSPMLDGNFLVGPTALDGVDKEATRYITKDAPCMLTKIGKHMVPSLNINNALISFAGSRPIDK324
Bp3DME AGLHAPGLARRIEGIPRDSIPPEYLCKGSYFTLAGR--APFSRLIYPVPQHAGLGVHLTLDLGGQAKFGPDTEWIA--TEDYTLDPRRADVFYAAVRSYWP-ALPDGALAPGYTGIRPKI319
SsGlpO SGPWVDKVRNLNFTRPVS-PKMRPTKGIHLVVDAKK-LPVPQPTYFDTGKQDGRMVFAIPR-ENKTYFGTTDTDYQGDFTDPKVTQEDVDYLLDVINHRYPEANITLADIEASWAGLRPL349
EcGlpD TGPWVKQFFDDGMHLPSP-YGIRLIKGSHIVVPRV---HTQKQAYILQN-EDKRIVFVIPWMDEFSIIGTTDVEYKGDPKAVKIEESEINYLLNVYNTHFK-KQLSRDDIVWTYSGVRPI318
BsGlyOx SGVWSGMFFKQLGLN-----NAFLPVKGECLSVWNDDIPLTKTLYHD-------HCYIVPRKSGRLVVGATMKPGD---WSETPDLGGLESVMKKAKTMLPAIQNMKVDRFWAGLRPGTK306
β16 β17 α7
MpGlpO A---------TNDFIIRVA---HNDPDFVILGGMKSPGLTAAPAIVREAVRLLNWKLTKKPNWNGKYNLPWI384
Bp3DME SG-----PHEPAADFAIAGPASHGVAGLVNLYGIESPGLTASLAIAEETLARLA366
SsGlpO LI----------GSSLER----EPDGLLTLSGG---KITDYRKMAEGALRLIRQLLKEEYGIETKEIDSKKYQISGG...478(Continues to alpha-helical domain)
EcGlpD CDDESDSPQAITRDYTLDIHDENGKAPLLSVFGG---KLTTYRKLAEHALEKLTPYYQGIGPAWTKESVLPGGAIEGD...397(Continues to alpha-helical domain)
BsGlyOx ----------DGKPYIGRH---PEDSRILFAAGHFRNGILLAPATGALISDLIMNKEVNQDWLHAFRIDRK364

conserved threonine along with a slightly less conserved region consisting of threonine, serine, and histidine. Three residues C-terminal to His44 is a serine that is particularly important in the active site. This serine (Ser47) interacts directly with the N5 of the flavin in FAD. Interestingly, as can be seen in Fig. 13, this serine is not one of the conserved residues across these five structures. *Ss*GlpO has a threonine in that position and *Bs*GlyOx has an alanine residue. Since serine and threonine have similar properties, this result is consistent with the expectation that the two GlpO should have active site groups that are more similar to each other than to glycine oxidases.

Active Site Characterization

Besides the area around the FAD, another feature we looked at was in the FAD itself. In comparing these structures, we discovered that overlays of the structures based on the whole protein chain did not provide optimal alignments of the active site regions. To solve this problem, we used the program hcore (PAKarplus unpublished) which allowed us to define specific chain regions surrounding the FAD and use those segments to guide the overlays. This allowed us to better visualize the true differences among these structures in the active site region (Figure 14). Once aligned with hcore, it was easy to see that there were no substantial differences in the ways in which the FAD was arranged in the proteins. Small deviations in one direction correlated with a similar deviation in the same direction of the entire main chain.

More substantial was the difference in the flavin, or the three-ringed structure of the FAD. The classic model of the flavin is planar. However, it was readily apparent from the electron density of *Mp*GlpO that there was a kink or a bend across the three rings. As

Fig. 14: **Overlay of the Active Sites of *Mp*GlpO, *Ss*GlpO, and *Bp*3DME**. The general active site of *Mp*GlpO (magenta) is similar to that of *Bp*3DME (cyan). The active site of *Ss*GlpO (green) appears to have a deletion in a loop above the flavin, shown with an arrow.

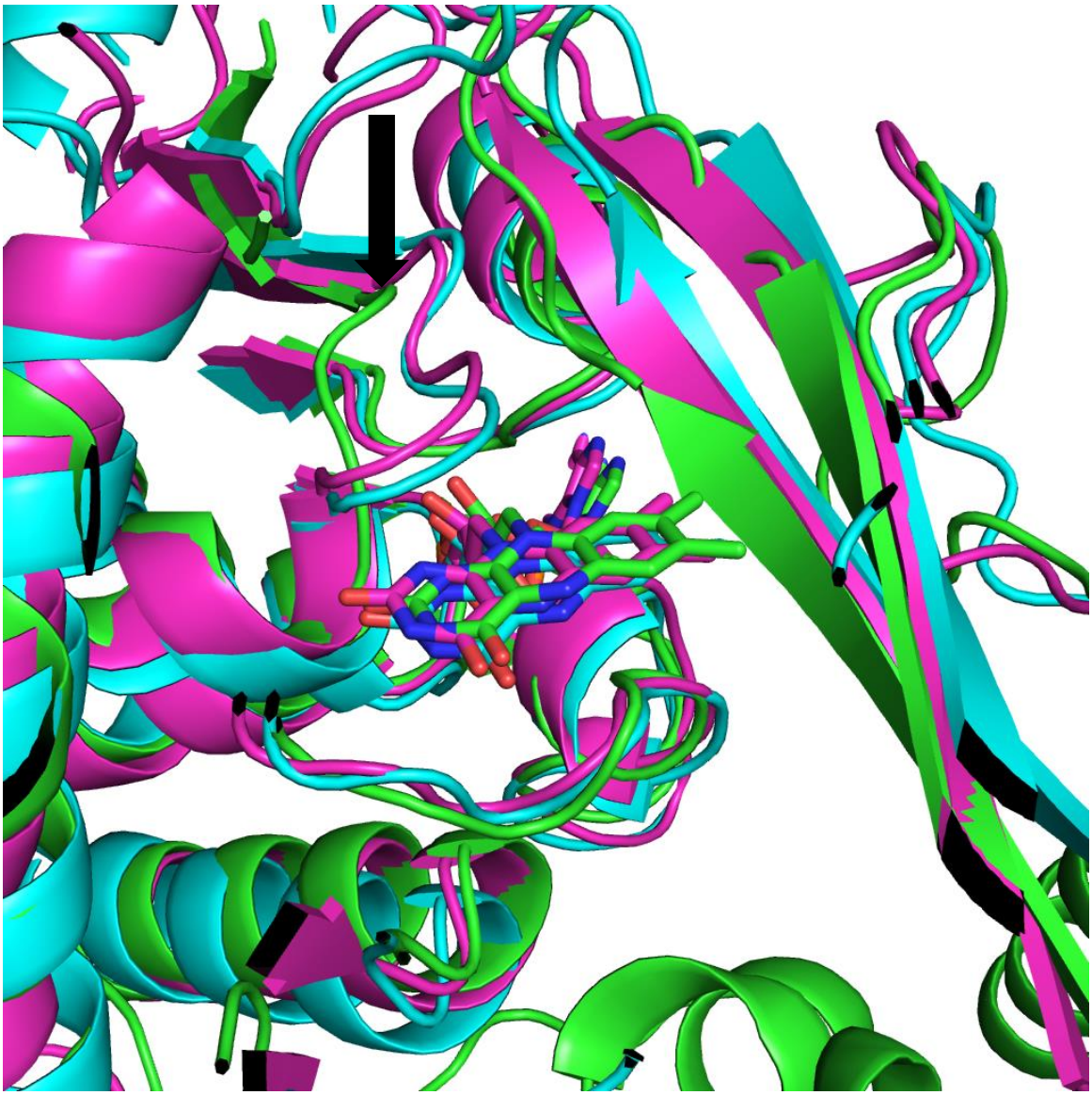
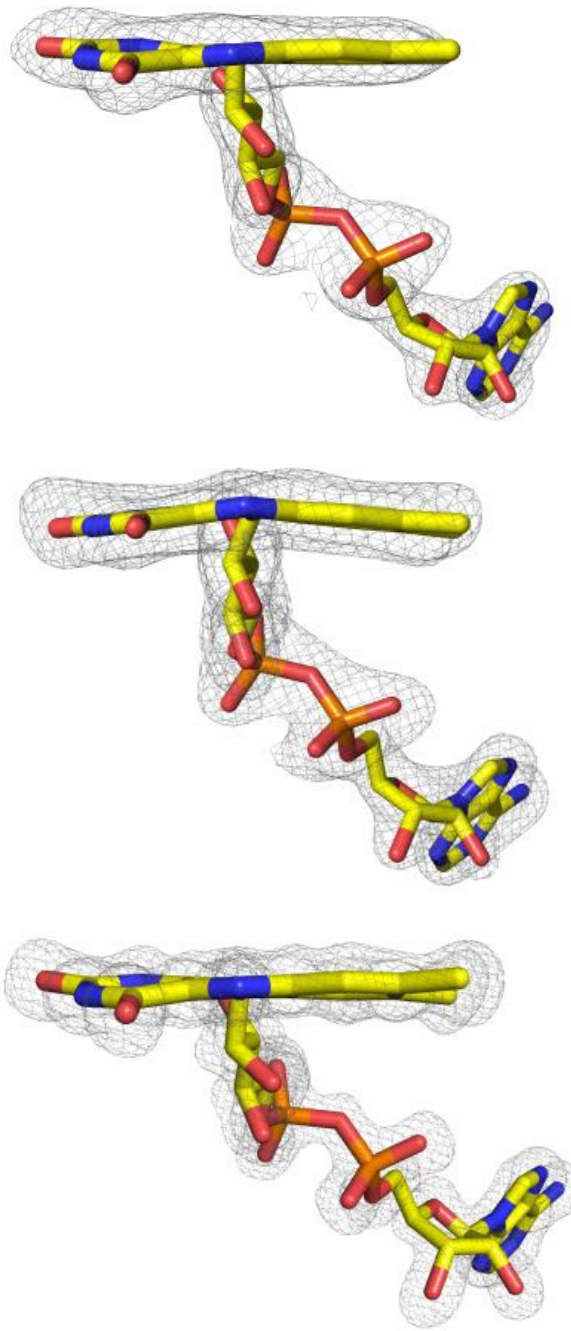


Fig. 15: **Comparison of FAD Bending**. Both models of GlpO have non-planar flavins (the 3-ring structure of FAD). The flavin directly interacts in substrate binding and transfers electrons to produce the product. *Mp*GlpO has a downward twist on one side and an almost planar ring on the other. The flavin of *Bp*3DME has a similar twist to *Mp*GlpO while the flavin of *Ss*GlpO is a downward butterfly bend.



*Mp*GlpO

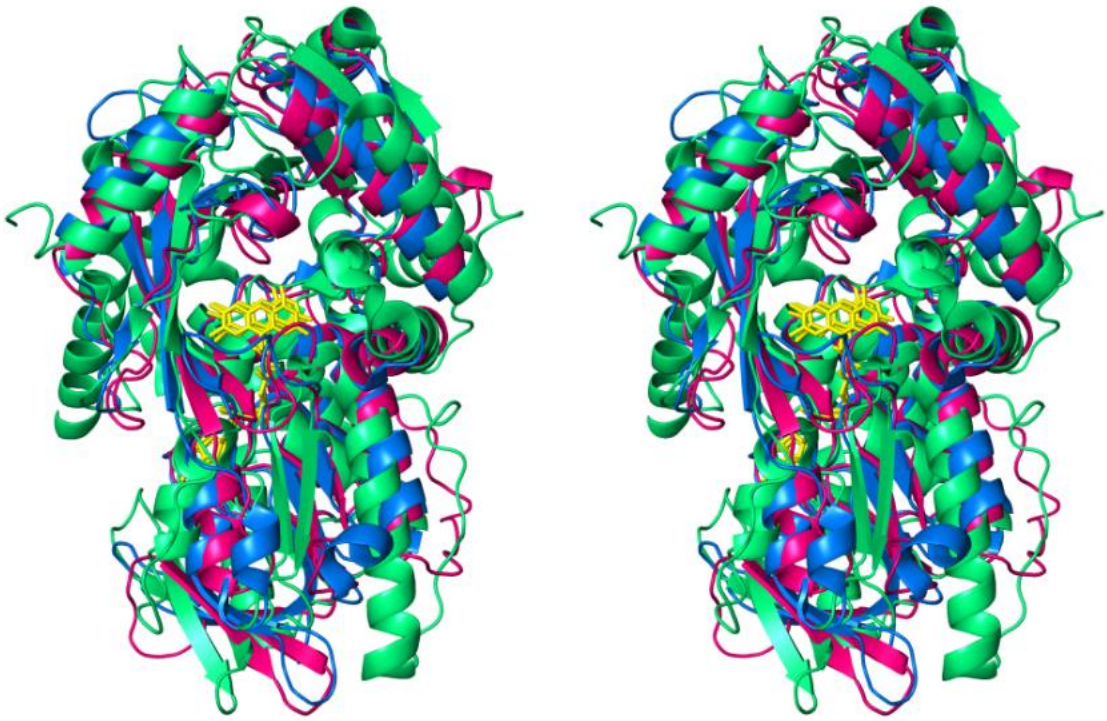
*Ss*GlpO

*Bp*3DME

seen in Figure 15, there is an asymmetric twist in which the two carbonyl groups lean towards the viewer and the methyl groups are relatively planar. An almost identical twist occurs in *Bp*3DME, but in this case, the dimethyl end also has a small twist. The flavin of *Ss*GlpO has a so called butterfly bend in which there is symmetrical deviation from planarity on both sides of the bend at the median of the middle ring.[30] While characteristic of reduced flavins, it is only occasionally observed in oxidized flavins. In *Ss*GlpO, this butterfly bending directs the reactive nitrogen of the flavin towards the Glp substrate.[30] *Ec*GlpD and *Bs*GlyOx (not shown) have similar twists to each other, with an almost planar flavin on the carbonyl side and a slight twist on the dimethyl side.

In all structures except *Bp*3DME, a loop of main chain resides beneath the FAD (on the si face of the flavin) and can be seen in Fig. 14. This contains the active site serine, threonine, or alanine residues mentioned above. After necessary refinements, one of the chains of *Bp*3DME had the loop beneath the flavin in a similar location as the other structures while the other chain's loop was much farther out of the reach of the N5 of the flavin. Above the flavin (on the re side), there is a loop of amino acids believed to be important in stabilizing the substrate.[30] Interestingly, these loops extend much farther in *Mp*GlpO, *Bs*GlyOx, and *Bp*3DME than in *Ss*GlpO and *Ec*GlpD. It looks like an insertion event has occurred in an area very close to the active sites. These active site differences are interesting, particularly for *Mp*GlpO, *Ss*GlpO, and *Bp*3DME because we expected to see the first two showing the most similarities than either to *Bp*3DME. A stereo overlay of these three structures is shown in Fig. 16.

Fig. 16: **Stereo Overlay of *Mp*GlpO, *Ss*GlpO, and *Bp*3DME**. *Mp*GlpO (magenta), *Ss*GlpO (green), and *Bp*3DME (blue) have roughly the same overall structure. FADs and many alpha helices roughly align. Behind the FAD are additional alpha helices that make the C-terminal domain. The FAD binding pocket is also roughly conserved.



The Reduced *Mp*GlpO Structure

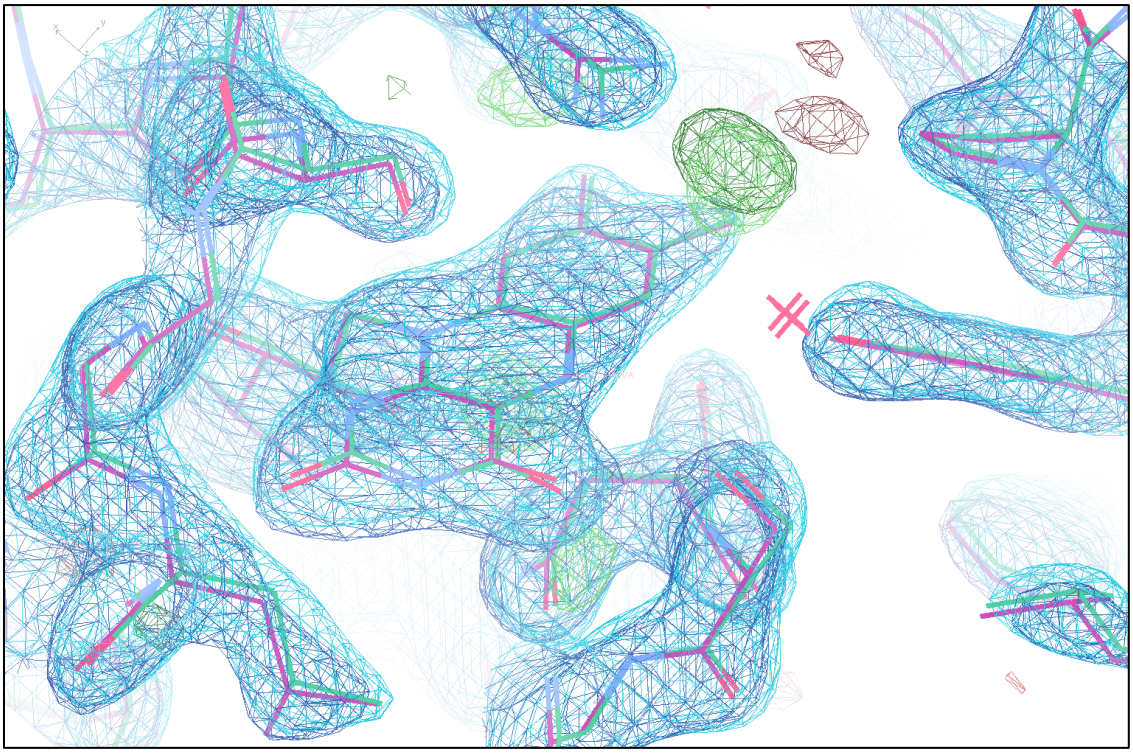
After solving the structure of oxidized *Mp*GlpO, the structure of the chemically reduced enzyme was determined to visualize how the structure changes when the FAD is reduced to $FADH_2$. This was performed by soaking the crystals in the reductant sodium dithionite. Within seconds of placing the crystals in the dithionite solution the crystals changed from a bright yellow to a pale blue color. This color change indicated a reduction of the FAD. Data were collected from several reduced crystals and the data were combined to yield a preliminary reduced *Mp*GlpO (*Mp*GlpOred) structure that could refined at 2.5 Å resolution (Table 1). Surprisingly, the results show that there appear to be no significant differences between the oxidized and reduced structures (Figure 17).

Conclusions and Outlook

Through this thesis project, I have determined the structure of *Mp*GlpO and compared it to other enzymes. This is an important contribution to the scientific knowledge-base and could be utilized in the design of a new drug against walking pneumonia. In addition, I have learned far more about scientific research than I expected to. I have learned about the interesting process of X-ray crystallography, getting to grow a tiny crystal and find out its structural mysteries. I learned that scientific research has many detours and ‘dead-ends’. These were never truly ends, because with persistence and teamwork, one can get back on track and solve a structure.

The work presented here is not complete. There are many questions left unanswered. One of the interesting questions worthy of further exploration is the evolutionary relationship of GlpO enzymes from various organisms and how there came

Fig. 17: **Reduced GlpO Difference Map.** There are no significant differences between the electron densities of reduced *Mp*GlpO (dark blue) and oxidized *Mp*GlpO (light blue). Positive difference electron density is shown in green and negative difference electron density is shown in red. Darker greens and reds are from reduced data and lighter greens and reds are from oxidized data. Both sets roughly align with no large changes to the model when *Mp*GlpO was reduced.



to be such distinct versions. It would also be beneficial to bind ligands into the active site and observe any structural changes. While my work on *Mp*GlpO may be complete, I know that I take the lessons I have learned and hope that others will continue to unravel its mystery.

BIBLIOGRAPHY

1. Krass, C.J. & Gardner, M.W. (1973). Etymology of the Term Mycoplasma. *Int. J. Syst. Bacteriol.* **23**, 62–64

2. Waites, K.B. & Talkington, D.F. (2004). Mycoplasma pneumoniae and Its Role as a Human Pathogen. *Clin. Microbiol. Rev.* **17**, 697–728

3. Hames, C., Halbedel, S., Hoppert, M., Frey, J. & Stülke, J. (2009). Glycerol Metabolism Is Important for Cytotoxicity of Mycoplasma pneumoniae. *J. Bacteriol.* **191**, 747–753

4. Razin, S., Yogev, D. & Naot, Y. (1998). Molecular Biology and Pathogenicity of Mycoplasmas. *Microbiol. Mol. Biol. Rev.* **62**, 1094–1156

5. Alain, B. & Browning, G. (CRC Press: 2005). *Mycoplasmas: Molecular Biology, Pathogenicity and Strategies for Control*.

6. Pilo, P., Vilei, E.M., Peterhans, E., Bonvin-Klotz, L., Stoffel, M.H., Dobbelaere, D. & Frey, J. (2005). A metabolic enzyme as a primary virulence factor of Mycoplasma mycoides subsp. mycoides small colony. *J. Bacteriol.* **187**, 6824–6831

7. Eaton, M.D., Meiklejohn, G. & Van Herick, W. (1944). STUDIES ON THE ETIOLOGY OF PRIMARY ATYPICAL PNEUMONIA A FILTERABLE AGENT TRANSMISSIBLE TO COTTON RATS, HAMSTERS, AND CHICK EMBRYOS. *J. Exp. Med.* **79**, 649–668

8. Atkinson, T.P., Balish, M.F. & Waites, K.B. (2008). Epidemiology, clinical manifestations, pathogenesis and laboratory detection of Mycoplasma pneumoniae infections. *Fems Microbiol. Rev.* **32**, 956–973

9. Clyde, W.A., Jr (1983). Mycoplasma pneumoniae respiratory disease symposium: summation and significance. *Yale J. Biol. Med.* **56**, 523–527

10. Marston, B.J., Plouffe, J.F., File, T.M., Jr, Hackman, B.A., Salstrom, S.J., Lipman, H.B., Kolczak, M.S. & Breiman, R.F. (1997). Incidence of community-acquired pneumonia requiring hospitalization. Results of a population-based active surveillance Study in Ohio. The Community-Based Pneumonia Incidence Study Group. *Arch. Intern. Med.* **157**, 1709–1718

11. Korppi, M., Heiskanen-Kosma, T. & Kleemola, M. (2004). Incidence of community-acquired pneumonia in children caused by Mycoplasma pneumoniae: Serological results of a prospective, population-based study in primary health care. *Respirology* **9**, 109–114

12. Lind, K., Benzon, M.W., Jensen, J.S. & Clyde, W.A., Jr (1997). A seroepidemiological study of Mycoplasma pneumoniae infections in Denmark over the 50-year period 1946-1995. *Eur. J. Epidemiol.* **13**, 581–586

13. Kraft, M., Cassell, G.H., Pak, J. & Martin, R.J. (2002). Mycoplasma pneumoniae and chlamydia pneumoniae in asthma*: Effect of clarithromycin. *Chest J.* **121**, 1782–1788

14. Walter, N.D., Grant, G.B., Bandy, U., Alexander, N.E., Winchell, J.M., Jordan, H.T., Sejvar, J.J., Hicks, L.A., Gifford, D.R., Alexander, N.T., Thurman, K.A., Schwartz, S.B., Dennehy, P.H., Khetsuriani, N., Fields, B.S., Dillon, M.T., Erdman, D.D., Whitney, C.G. & Moore, M.R. (2008). Community outbreak of Mycoplasma pneumoniae infection: school-based cluster of neurologic disease associated with household transmission of respiratory illness. *J. Infect. Dis.* **198**, 1365–1374

15. Zezima, K. (2007). School Is Shut After Outbreak of Encephalitis Kills a Pupil. *New York* at <http://www.nytimes.com/2007/01/04/us/04warwick.html>

16. Ramírez, A.S., Rosas, A., Hernández-Beriain, J.A., Orengo, J.C., Saavedra, P., de la Fe, C., Fernández, A. & Poveda, J.B. (2005). Relationship between rheumatoid arthritis and Mycoplasma pneumoniae: a case-control study. *Rheumatol. Oxf. Engl.* **44**, 912–914

17. Science | Direct-MS. at <http://www.direct-ms.org/molecularmimicry.html>

18. Waites, K.B., Balish, M.F. & Atkinson, T.P. (2008). New insights into the pathogenesis and detection of Mycoplasma pneumoniae infections. *Future Microbiol.* **3**, 635–648

19. Lundstrom, K. (2007). Structural genomics and drug discovery. *J. Cell. Mol. Med.* **11**, 224–238

20. Radestock, U. & Bredt, W. (1977). Motility of Mycoplasma pneumoniae. *J. Bacteriol.* **129**, 1495–1501

21. Szczepanek, S.M., Majumder, S., Sheppard, E.S., Liao, X., Rood, D., Tulman, E.R., Wyand, S., Krause, D.C., Silbart, L.K. & Geary, S.J. (2012). Vaccination of BALB/c Mice with an Avirulent Mycoplasma pneumoniae P30 Mutant Results in Disease Exacerbation upon Challenge with a Virulent Strain. *Infect. Immun.* **80**, 1007–1014

22. Baseman, J.B., Cole, R.M., Krause, D.C. & Leith, D.K. (1982). Molecular basis for cytadsorption of Mycoplasma pneumoniae. *J. Bacteriol.* **151**, 1514–1522

23. Kannan, T.R. & Baseman, J.B. (2006). ADP-ribosylating and vacuolating cytotoxin of Mycoplasma pneumoniae represents unique virulence determinant among bacterial pathogens. *Proc. Natl. Acad. Sci.* **103**, 6724–6729

24. Pakhomova, O.N., Taylor, A.B., Becker, A., Holloway, S.P., Kannan, T.R., Baseman, J.B. & Hart, P.J. (2010). Crystallization of community-acquired respiratory distress syndrome toxin from Mycoplasma pneumoniae. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **66**, 294–296

25. Hardy, R.D., Coalson, J.J., Peters, J., Chaparro, A., Techasaensiri, C., Cantwell, A.M., Kannan, T.R., Baseman, J.B. & Dube, P.H. (2009). Analysis of pulmonary inflammation and function in the mouse and baboon after exposure to Mycoplasma pneumoniae CARDS toxin. *Plos One* **4**, e7562

26. Castro, L. & Freeman, B.A. (2001). Reactive oxygen species in human health and disease. *Nutr. Burbank Los Angeles Cty. Calif* **17**, 161, 163–165

27. Almagor, M., Kahane, I. & Yatziv, S. (1984). Role of superoxide anion in host cell injury induced by mycoplasma pneumoniae infection. A study in normal and trisomy 21 cells. *J. Clin. Invest.* **73**, 842–847

28. Fitzpatrick, P.F. (2010). Oxidation of amines by flavoproteins. *Arch. Biochem. Biophys.* **493**, 13–25

29. Umhau, S., Pollegioni, L., Molla, G., Diederichs, K., Welte, W., Pilone, M.S. & Ghisla, S. (2000). The x-ray structure of D-amino acid oxidase at very high resolution identifies the chemical mechanism of flavin-dependent substrate dehydrogenation. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12463–12468

30. Colussi, T., Parsonage, D., Boles, W., Matsuoka, T., Mallett, T.C., Karplus, P.A. & Claiborne, A. (2008). Structure of alpha-glycerophosphate oxidase from Streptococcus sp.: a template for the mitochondrial alpha-glycerophosphate dehydrogenase. *Biochemistry (Mosc.)* **47**, 965–977

31. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A. & Wilson, K.S. (2011). Overview of the *CCP* 4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242

32. Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. & Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281

33. Leslie, A.G.W. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 48–57

34. Rupp, B. (Garland Science: 2010). *Biomolecular crystallography: principles, practice, and application to structural biology*.

35. Protein Crystallography - ARI. at <http://www.artrobbins.com/protein-crystallography/>

36. Hampton Research. at <http://hamptonresearch.com/product_detail.aspx?cid=1&sid=24&pid=5>

37. Kantardjieff, K.A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871

38. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497

39. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410

40. Citing PyMOL, JyMOL and AxPyMOL | www.pymol.org. at <http://www.pymol.org/citing>

41. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. & Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221

42. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132

43. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S. & Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383

44. Holm, L. & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549

45. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637

46. Krissinel, E. & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797