Structural analysis of UDP-*N*-acetylgalactopyranose mutase from *Campylobacter jejuni* 11168

A Thesis Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Biochemistry

University of Saskatchewan

Saskatoon, SK

By

Carla Protsko

PERMISSION TO USE

ABSTRACT

UDP-galactopyranose mutase (EC 5.4.99.9; UGM), the product of the *glf* gene, is an enzyme that catalyzes the conversion of uridine diphosphate galactopyranose (UDP-Gal*p*) to UDP-galactofuranose (UDP-Gal*f*). UGM activity is found in bacteria, parasites and fungi, however is absent in higher eukaryotes. This enzyme is essential for the viability of many pathogenic organisms, such as *Mycobacterium tuberculosis* and *Escherichia coli*, due to the broad distribution of Gal*f* in crucial structures such as the cell wall or capsular polysaccharide. Not surprisingly, galactofuranose biosynthesis has become an attractive antimicrobial target due to the absence of these sugars in higher eukaryotes. The UGM homologue, UDP-*N*-acetylgalactopyranose mutase (UNGM), was identified in *Campylobacter jejuni* 11168, encoded for by the *cj1439c* gene. UNGM is known to function as a bifunctional mutase, which catalyzes the reversible ring contraction between the pyranose-furanose forms of UDP-galactose (UDP-Gal) and UDP-*N*-acetylgalactosamine (UDP-GalNAc). UNGM is essential for the virulence of *C. jejuni,* due to the incorporation of UDP-*N*-acetylgalactofuranose into the capsular polysaccharide. We report the first structure of UNGM determined by X-ray crystallography, to a resolution of 1.9 Å. Analysis of the dimeric, holoenzyme structure of UNGM has identified that the cofactor flavin adenine dinucleotide is bound within each monomer of the enzyme. Comparative analysis with UGM homologues has confirmed the conserved active site residues involved in the binding of various substrates. Docking studies suggest that UNGM binds its natural substrates in a productive binding mode for catalysis with the flavin cofactor, which is consistent with the proposed mechanism for UNGM. The mobile loops are essential for substrate binding, and we have identified that the conserved arginine residue, Arg169, and the neighboring Arg168, function to stabilize the diphosphate region of UDP, although not concurrently. The non-conserved arginine residue, Arg168, appeared to favor the stabilization of *N*-acetylated sugars, which is in agreement with the enzyme's higher binding affinity for UDP-GalNAc over UDP-Gal by a factor of 0.9. We have also identified that the active site Arg59 exists in two conformations in the structure of UNGM, with one conformation directed toward the active site. Arg59 is 2.5 to 3.0 Å from the acetamido moiety of GalNAc, which is favorable for stabilization and is believed to confer specificity for this substrate.

ACKNOWLEDGEMENTS

DEDICATION

To my parents, John and Brenda, I thank for their encouragement in my every endeavor. I would not have been able to accomplish this work without your constant support. To Louis, you have inspired me in many ways, and you will always be missed. To my friends who make me laugh, you are my family and I love you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

EQUATION

LIST OF ABBREVIATIONS

afUGM	*Aspergillus fumigatus* UDP-galactopyranose mutase
CCP4	Collaborative Computing Project Number 4
C. jejuni	*Campylobacter jejuni* 11168
E. coli	*Escherichia coli*
ecUGM	*Escherichia coli* UDP-galactopyranose mutase
drUGM	*Deinococcus radiodurans* UDP-galactopyranose mutase
FAD	flavin adenine dinucleotide, oxidized form
FADH$^-$	flavin adenine dinucleotide, anionic reduced form
GA	genetic algorithm
Gal*f*	galactofuranose
Gal*f*NAc	*N*-acetylgalactofuranose
GalNAc	*N*-acetylgalactosamine
Gal*p*	galactopyranose
Gal*p*NAc	*N*-acetylgalactopyranose
HPLC	high-performance liquid chromatography
IDH	inositol dehydrogenase
IolG4	putative inositol dehydrogenase 4
kpUGM	*Klebsiella pneumoniae* UDP-galactopyranose mutase
LOS	lipooligosaccharide
LPS	lipopolysaccharide
mtUGM	*Mycobacterium tuberculosis* UDP-galactopyranose mutase
NAD(H)	Nicotinamide adenine dinucleotide
NADP(H)	Nicotinamide adenine dinucleotide phosphate
SDS PAGE	sodium dodecyl sulfate polyacrilamide gel electrophoresis
UNGM	UDP-*N*-acetylgalactopyranose mutase
UDP	uridine diphosphate
UDP-Gal*f*	UDP-α-D-galactofuranose
UDP-Gal*p*	UDP-α-D-galactopyranose

CHAPTER 1: INTRODUCTION

The major objective of this study is to structurally characterize UDP-*N*-acetylgalactopyranose mutase (UNGM). The holoenzyme structure of UNGM will illuminate the active site residues and cofactors that are bound to the enzyme. Crystals of UNGM in complex with one of its substrates will allow for identification of structural changes upon substrate binding. Previous kinetic studies have shown that this enzyme has bifunctional activity for the sugar nucleotides UDP-galactose and UDP-*N*-acetylgalactosamine; however the lack of a crystal structure has prevented the identification of the amino acid residues that confer specificity for UDP-GalNAc. We hypothesize that the structure of UNGM will identify the residues involved in substrate binding as well as residues that differ from other UDP-galactopyranose mutase (UGM) homologues. Comparative studies with UGM homologues will also be used to identify the functions of conserved residues within the substrate-binding cleft. Structures of UGM in complex with substrate have been very difficult to obtain, as only the UGM homologues from *Deinococcus radiodurans* and *Klebsiella pneumoniae* have formed stable complexes in crystalline form. If no UNGM-complex structure is available, docking experiments will determine the dependence of substrate-binding on cofactor reduction, loop conformations, and specific residues that are known to contribute to the dual specificity of the active site. The structure of this enzyme will be an invaluable tool for future drug design and research on this pathogen.

CHAPTER 2: LITERATURE REVIEW
2.1 Virulence in Prokaryotes

Antibiotic resistance is an escalating problem in the treatment of many infections, and therefore it is essential to characterize essential enzymes of the causative pathogens. The cell envelope consists of the plasma membrane, cell wall, and the outer membrane (in gram negative bacteria). These cellular structures have long been targets for antimicrobial intervention, due to their essential function in cell survival. Bacteria acquire antibiotic resistance through acquisition of resistance genes, which may result in the expression of a protein that is able to destroy the antibiotic, or modify the metabolism to bypass the pathway targeted by the drug. This could be through the expression of a multidrug transporter that

expels drugs out of the bacteria. Another mechanism of resistance is through *de novo* mutations, which may destroy the drug's binding site. A common target for antibiotic agents is the cell wall as it is essential for bacterial viability and is not found in mammalian cells. Two major classes of drugs have been successful in preventing cell wall synthesis, β-Lactams (i.e. penicillins and cephalosporins) and glycopeptides (i.e. vancomycin). The increase in bacterial resistance against these drugs, such as the broad expression of β-lactamase in *E. coli*, has rendered many of these drugs ineffective against infection (Pitout *et al*., 2004). Therefore, drugs that differ in their methods of action are essential in the fight against emerging pathogens.

2.1.1 *Campylobacter jejuni*

The gram-negative bacterial pathogen *Campylobacter jejuni* is the leading cause of food- and water-borne inflammatory enterocolitis worldwide, causing approximately 500 million human infections annually (Allos, 2001; Friedman *et al*., 2000). *C. jejuni* are curved, helical bacteria, and are generally non-clonal organisms with a high degree of gene transfer (Parker *et al*., 2007). Prevalent strains of this organism are no longer susceptible to most β-lactams, and an emerging trend of resistance has been observed for the topoisomerase II and IV inhibitor Ciprofloxacin, which belongs to the second generation of fluoroquinolone class of antibiotics (Allos and Blaser, 2009). Symptoms of Campylobacteriosis include diarrhea, bloody stools, fever, and vomiting. Gastrointestinal infections of this pathogen are associated with a diversity of sequelae, which are the pathological conditions resulting from a disease. *Campylobacter* infection is the causative antecedent in the development of the autoimmune neurological disorders Guillian-Barré syndrome and Miller-Fisher syndrome, resulting in abnormal muscle cooridination and paralysis of eye muscles (Kaldor and Speed, 1984; Jacobs *et al*., 1995). This pathogen is associated with the development of Reiter's syndrome, also known as reactive arthritis, as well as gall bladder inflammation and irritable bowel syndrome (Mølbak and Havelaar, 2008; Vaughan-Shaw *et al*., 2010). Recently, the bacterial pathogen was found to cause spondylodiscitis, a post-infectious complication in which infection follows a hematogenous route from a distant site to the intervertebral discs of the spinal cord (Tappe *et al*., 2012). An important virulence attribute of this bacterial pathogen is its ability to enter and survive within host epithelial cells, although little is known about the pathogenic mechanism (Young *et al*., 2007). Understanding the crucial pathways for the survival of *C. jejuni* will be

paramount for developing new therapies and targets. Specifically, targeting the pathways that allow for host cell adhesion and invasion will alleviate the ability of this pathogen to cause infection.

2.2 Cell wall biosynthesis / capsular polysaccharide biosynthesis

Cell walls are important structures that serve a structural and protective role for many microorganisms. Universally, the bacterial cell wall confers rigidity and shape. The cell wall primarily consists of a layer of peptidoglycan in both gram-positive and gram-negative bacteria, and prevents osmotic lysis of the cell. Peptidoglycan is a polymer of long glycan chains that are cross-linked via flexible isopeptide bridges to form a strong, flexible structure, and it protects the underlying protoplast from lysing due to the elevated internal osmotic pressure (Höltje, 1998). The basic unit of peptidoglycan is a disaccharide-pentapeptide composed of *N*-acetylglucosamine and *N*-acetylmuramic acid linked together by β-1,4 glycosidic bonds. The glycan chains are synthesized by transglycosylation reactions, and the glycan chains are cross-linked by peptide bridges via the transpeptidation reaction (Scheffers and Pinho, 2005). Transglycosylation attaches the modified *N*-acetylmuramic acid to the peptidoglycan chain through the sugar backbone, whilst transpeptidation occurs through the formation of bond between diaminopimelic acid of the pentapeptide of one chain and an alanine residue from the pentapeptide of another chain.

In gram-positive bacteria, the cell wall is a thick layer of peptidoglycan outside of the cell membrane. In contrast, gram-negative bacteria have a relatively thin layer of peptidoglycan, although an additional outer membrane protects these organisms. The outer membrane is composed of phospholipids, proteins, and lipopolysaccharides (LPS) on the outer leaflet (Figure 2.1A) (Raetz and Whitfield, 2002). The lipopolysaccharide consists of several regions including: Lipid A (anchors the LPS to the outer membrane), the inner and outer cores (comprised of glycan chains), and the O-polysaccharide (antigen) repeat region (Figure 2.1B). The O-polysaccharide is part of the outer membrane and is important for recognition by the host immune system. Structural diversity of O-antigens is incredible, with over 60 monosaccharides and 30 non-carbohydrate components recognized (Knirel & Kochetkov, 1994

3

Jannson, P.E., 1999). Mucosal pathogens often lack an O-antigen, instead they produce lipooligosaccharides (LOS) that are anchored to Lipid A in the outer membrane.

The capsular polysaccharide (CPS) is a protective, exterior layer consisting of monosaccharides joined by glycosidic linkages, and is produced by many bacteria (Roberts, 1996). Biosynthesis and assembly of the CPS is a complex process that requires activated precursors (nucleotide sugars) assembled into the polysaccharide by enzymes associated with the inner membrane. Structurally, capsular polysaccharides are very diverse, although they do share some similar characteristics. The CPS is the outermost structure of the bacteria, and therefore these glycoconjugates play a critical role in the interactions between these organisms

A

B
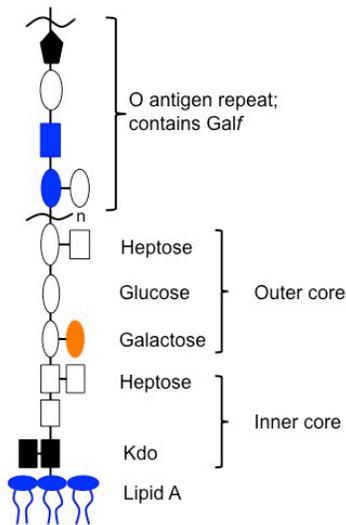
Figure 2.1: Cell envelope of gram-negative bacteria. **A-** External cellular structures in *C. jejuni*. The capsule is present on the outermost surface of these organisms. The lipooligosaccharide (LOS) anchors the capsule to the outer membrane. **B-** Lipopolysaccharide O-antigen repeat contains Gal*f* residues in *E. coli*, although the residues in the O-antigen are highly variable, with over 170 serotypes in *E. coli*. Therefore, the figure above does not show one serotype, it is a general representation of the O-antigen region. Kdo abbreviates 2-deoxy-D-manno-octulosonic acid.

4

and their immediate environment. The CPS aids in prevention of desiccation by forming a highly hydrated gel around the cell wall, as well as protection from harmful chemicals (Whitfield, 2006). The protection against the action of complement system and phagocytosis is likely the most significant function of the CPS (Jann and Jann, 1997). Previous work has shown that the capsular polysaccharide of *C. jejuni* is important for virulence. Specifically, *C. jejuni* 11186 does not incorporate Gal*f* into its glycoconjugates, instead Gal*f*NAc is incorporated into the capsular polysaccharide (CPS).

2.2.1 Galactofuranose (Gal*f*) in the cell wall

Galactose is most commonly found in nature as a six-membered ring in the pyranose form, also known as galactopyranose. Galactofuranose (Gal*f*) is the five-membered ring or furanose form of galactose, and is less thermodynamically stable than the pyranose sugar. Gal*f* is the most commonly found hexofuranose sugar in nature. Other hexofuranose sugars are found in several bacterial saccharide structures, including 6-deoxy-D-galactofuranose and 2-acetamido-2-deoxy-D-galactofuranose (Gal*f*NAc) (Feng *et al*., 2004; St. Michael *et al*., 2002). Gal*f* production is essential for viability of many bacteria, fungi, and parasites. Gal*f* production in these organisms is solely dependent on the ring contraction of galactose catalyzed by UDP-galactopyranose mutase. Galactan chains that contain Gal*f* are important structural components of the cell wall or capsular polysaccharides of many organisms, however they are not observed in higher eukaryotes (Peltier *et al*., 2008). The sugar nucleotide UDP-Gal*f* is the product of the ring contraction catalyzed by UDP-galactopyranose mutase. UDP-Gal*f* is the activated donor species for galactofuranosyltransferases, enzymes that are directly involved in the biosynthesis of galactan chains. These enzymes have been well characterized in *Mycobacterium tuberculosis* (GlfT2) and *Klebsiella pneumoniae* (WbbO). Both of these enzymes have been found to exhibit bifunctional activity, by catalyzing the transfer of Gal*f* residues from β-(1➔5) and β-(1➔6) to the galactan chain (Mikuŝová *et al*., 2000; Guan *et al*., 2001).

Gram-negative bacterial pathogens such as *Escherichia coli* and *Klebsiella pneumoniae* incorporate Gal*f* into the lipopolysaccharide (LPS) O antigens (Knirel and Kochetkov, 1994) (Figure 2.1B). Little is known about the composition and biosynthesis of the O polysaccharide. Galactofuranose also occurs in the T1 antigen of the important pathogen *Salmonella typhimurium,* which is an essential virulence factor of this organism (Sarvas and Nikaido, 1971;

Valtonen *et al.*, 1971). Mycobacterial species, such as *Mycobacterium tuberculosis,* also require Gal*f* residues. Gal*f* is an essential component of the arabinogalactan layer that connects the peptidoglycan and the mycolic acids in their cell walls (Bersa *et al.*, 1995) (Figure 2.2). The mycolic acid layer is highly impermeable and important for the protection of these organisms from environmental stresses. Lower eukaryotes also incorporate Gal*f* into their glycoconjugates. Fungal species including *Aspergillus fumigatus* and *Pencillium* spp. require Gal*f* for cell wall construction (Bernard and Latge, 2001). The protozoan parasites *Trypanosoma cruzi* and *Leishmania* species utilize Gal*f* residues in external cell-surface structures and glycoprotein modification (de Lederkremer and Colli, 1995; McConville *et al.*, 1990).

2.2.2 Glycoconjugates in *C. jejuni*

C. jejuni 11186 incorporates Gal*f*NAc into its capsular polysaccharide (CPS), however the biosynthetic pathway of the CPS is poorly understood. The protective function of the CPS
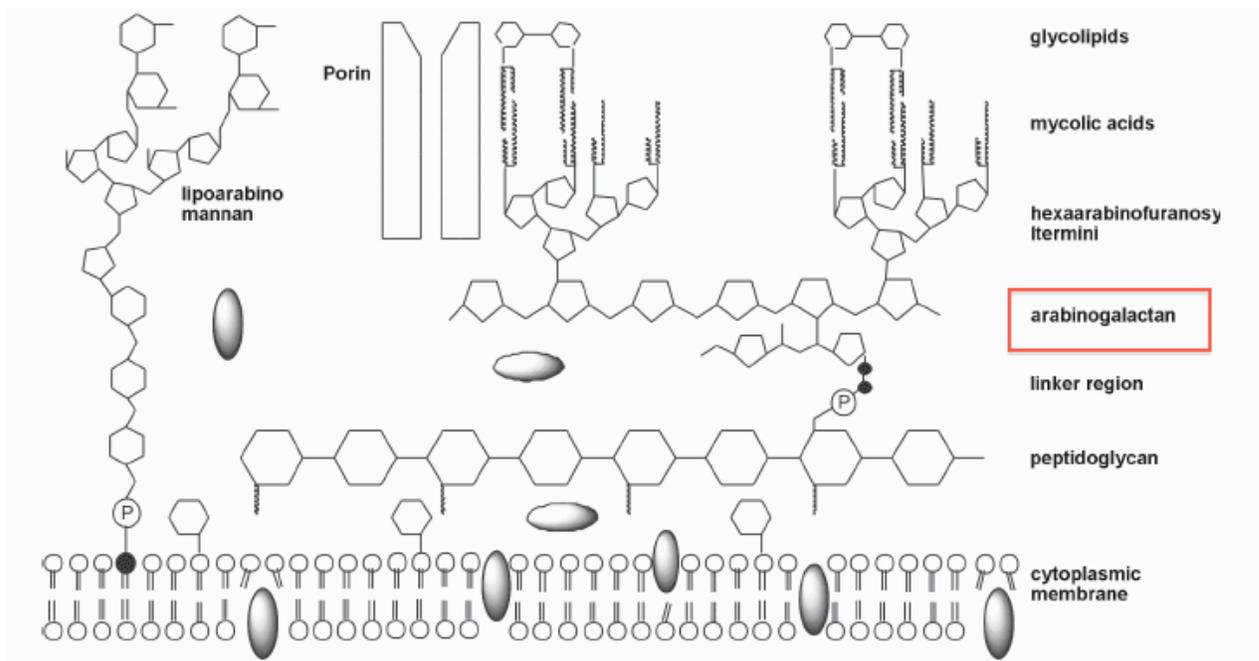
Figure 2.2: The arabinogalactan layer of the cell wall of mycobacterial species. This layer is composed of galactofuranose and arabinofuranose (figure modified with permission from Basso *et al.*, 2005).

6

enables bacteria to inhabit various environments (Corbett and Roberts, 2009). This is a particularly beneficial attribute for pathogenic bacteria such as *C. jejuni*, where environmental changes can occur quickly due to the host immune response to the infection, and survival is dependent on the ability to form a CPS in defense. Knock-out studies for the *glf* gene in *C. jejuni* resulted in no CPS formation, which confirmed that UNGM is an essential enzyme in this biosynthetic pathway (Poulin *et al*., 2009). The CPS structures of *C. jejuni* are highly variable amongst the 70 known serotypes. The CPS of *C. jejuni* 11168 contains 2-acetamido-2-D-galactofuranose, 6-methyl-D-*glycero*-α-L-*gluco*-heptopyranose, α-D-glucouronic acid modified with 2-amino-2-deoxyglycerol, and β-D-ribofuranose (Figure 2.3) (St. Michael *et al*., 2002). The CPS is one of the few virulence determinants of this pathogen, which has been found to modulate the invasion of gastrointestinal epithelial cells (Monteio *et al*., 2009; Bacon *et al*., 2001).

The lipooligosaccharides (LOS) in *C. jejuni* are highly variable glycoconjugates with a role in serum-resistance and adherence of this organism. A genomic study of LOS biosynthesis loci in 11 strains of *C. jejuni* expressing eight different ganglioside mimics identified poly(G), resulting in LOS variation (Gilbert *et al*., 2002). Variations in the number of G's or C's during replication can result in alteration of the reading frame. The LOS is anchored to Lipid A in the outer membrane, and is known to contain Gal*p*NAc residues (Figure 2.1A). Enzymatic activity for a β-1,4-*N*-acetylgalactosaminyltransferase, CgtA, has been detected in this strain and requires the sugar nucleotide UDP-Gal*p*NAc as a donor substrate (Gilbert *et al*., 2000). CgtA transfers *N*-acetylgalactosaminyl (GalNAc) residues via a β-1,4 linkage to the Gal moiety at the non-reducing end of the growing LOS chain. This enzyme may function similarly to other galactofuranosyltransferases, which catalyze the addition of the donor substrate UDP-Gal*f* to the acceptor sugar in the polysaccharide chain (Splain and Kiessling, 2010; Kremer *et al*., 2001). Gal*f*NAc residues are linked via β-1,4 with Glc*p*A6(Ngro) in the CPS, therefore we hypothesize that CgtA (that transfers residues in a β-1,4 linkage) may also function in the construction of the CPS, although this has not yet to be confirmed.

C. jejuni incorporates *N*-acetylgalactosamine or 2-acetamido-2-deoxy-D-galactopyranose (GalNAc) residues into an *N*-linked glycan. *N*-linked glycosylation involves the modification of asparagine residues on proteins found in the membrane or periplasmic space (Young *et al*., 2007). Prior to the discovery of the *N*-linked modification system in *C. jejuni*,

this type of glycosylation had been observed only in eukaryotes and archaea (Szymanski *et al.*, 1999).   An increase in immunoreactivity of some *N*-linked glycosylated proteins has been observed, which indicated a role in evasion of the immune system.   *N*-linked glycans are important for epithelial cell adherence, invasion, and interaction with the host.   Studies have shown that *C. jejuni* isolates from patients with fever and diarrhea correlated with a higher level of binding to the epithelial cells in comparison to isolates from patients without fever and diarrhea (Fauchere *et al.*, 1986).   Therefore, the involvement of *N*-glycosylated proteins in cellular adherence appears to significantly increase the severity of infection.

2.3 UDP-galactopyranose Mutase (UGM)

UDP-galactopyranose mutase (UGM), encoded for by the *glf* gene, is a homodimeric flavoenzyme belonging to the isomerase class of enzymes. UGMs have been identified in bacteria, parasites and fungi.   Galactofuranose (Gal*f*) is an essential sugar for cell wall or capsular polysaccharide biosynthesis in the aforementioned organisms.   UDP-galactopyranose mutase was first identified in *Escherichia coli* K12 as the enzyme responsible for biosynthesis of galactofuranose (Nassau *et al.*, 1996).   Since this discovery, UGM homologues have been identified in many different prokaryotes, including *Klebsiella pneumoniae*, *Deinococcus radiodurans*, and *Mycobacterium tuberculosis*.   These enzymes are solely responsible to the generation of Gal*f* in these organisms, although the incorporation of this sugar into glycoconjugates varies significantly between each species.   Knock-out of the *glf* gene in *Mycobacterium smegmatis* showed that UGM is essential for growth (Weston *et al.*, 1997).

UGM has been found to be essential for the viability of many pathogens and is absent in higher eukaryotes, making this enzyme an attractive antimicrobial target.   The disruption of cell wall biosynthesis through inhibition of UGM has compelled extensive studies to elucidate the enzyme mechanism.   Galactopyranose is converted to galactofuranose at the UDP-sugar level, as this ring contraction is not seen at the sugar-phosphate or free-sugar level (Sarvas and Nikaido, 1971).   UDP-galactopyranose mutase catalyzes the reversible ring contraction of UDP-galactopyranose (UDP-Gal*p*) to UDP-galactofuranose (UDP-Gal*f*) (Figure 2.4).   These sugar nucleotides are the activated precursors used by galactosyltranferases in galactan chain synthesis (Kremer *et al.*, 2001).   UGM from *E. coli* (ecUGM) has been found to convert UDP-Gal*p* to UDP-Gal*f* at a ratio of 93:7 respectively (Zhang and Lui, 2000).
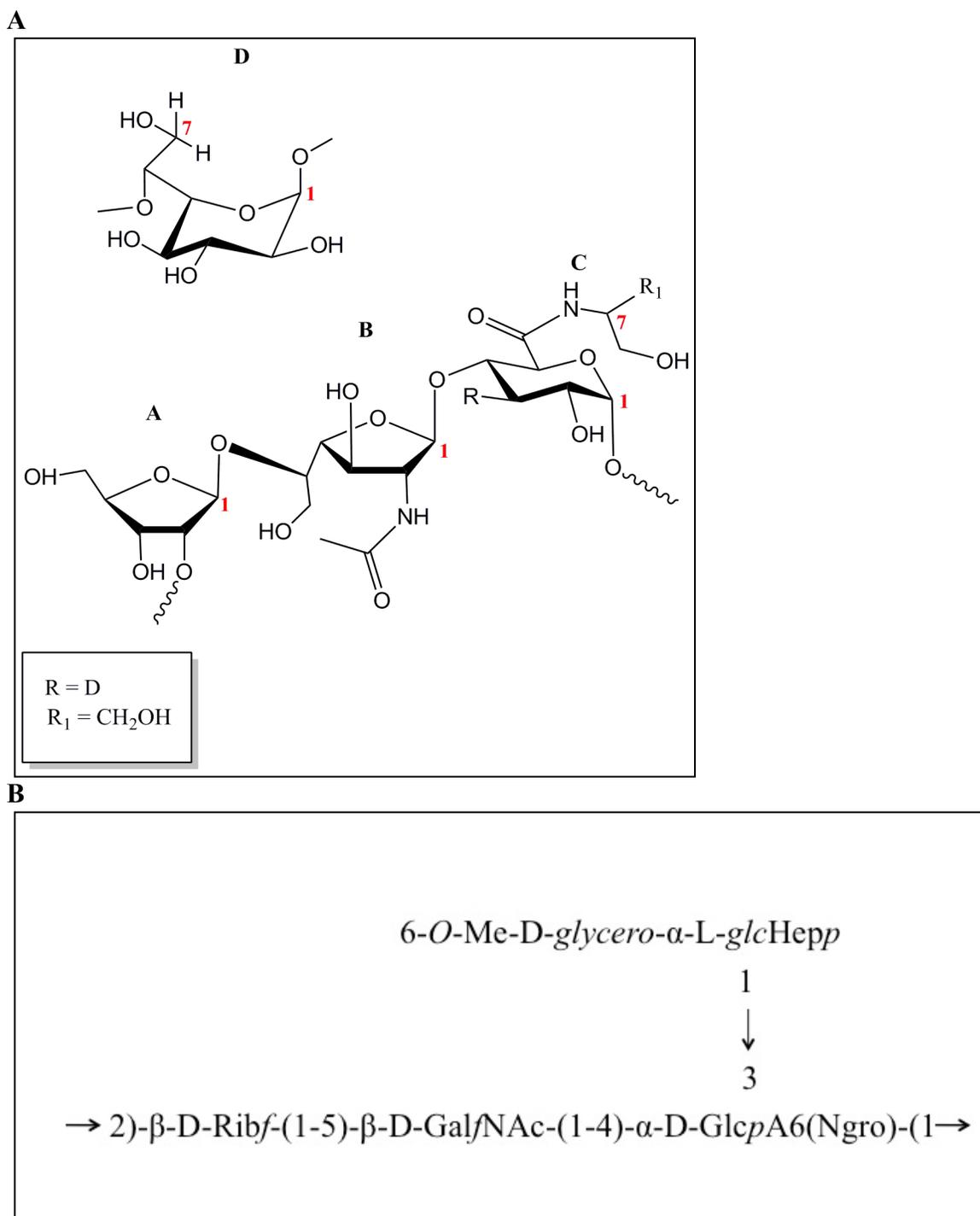
Figure 2.3: The capsular polysaccharide of *C. jejuni*. **A-** The polymer repeat of the CPS in the HS:2 serotype of *C. jejuni* 11168 contains the following sugars: 6-methyl-D-glycero-α-L-glucoheptopyranose (6-O-Me-D-*glycero*-α-L-*glc*Hep*p*)[**D**], β-D-ribofuranose (β-D-Rib*f*) [**A**], 2-acetamido-2-D-galactofuranose (β-D-Gal*f*NAc)[**B**], and β-D-glucouronic acid modified with 2-amino-2-deoxyglycerol at C-6(α-D-GlcpA6(Ngro)[**C**]. Figure adapted from St. Michael *et al*. 2002, generated in ChemDraw (Mills, 2006). **B-** The polymer repeat of the CPS.

UGM is a flavoenzyme, which non-covalently binds a single flavin adenine dinucleotide (FAD) cofactor in each monomer of the enzyme. Flavoenzymes are enzymes that rely on the cofactor (or prosthetic group) FAD for catalysis. These enzymes catalyze a diversity of biological reactions, involving one and two-electron oxidation/reduction reactions, and the cofactor can exist in several different oxidation and ionic states (Mansoorabadi *et al.*, 2007; Massey, 2000). Flavoenzymes catalyze redox chemistry through transformations involving C4a or N5 of the isoalloxazine ring system (Ghisla and Massey, 1989). The inactive form of FAD is oxidized and the isoalloxazine ring is in a planar conformation (Figure 2.5A & 2.6). The reduction of FAD causes a pucker in the isoalloxazine ring at N5 or C4a (Figure 2.5B). Enzymatic activity is observed when the FAD cofactor is in a reduced state, indicating that the cofactor is directly involved in catalysis (Sanders *et al.*, 2001). There is no net gain or loss of electrons during the reaction, as the reduced cofactor is required for the catalysis of both UDP-Gal*p* and UDP-Gal*f*.
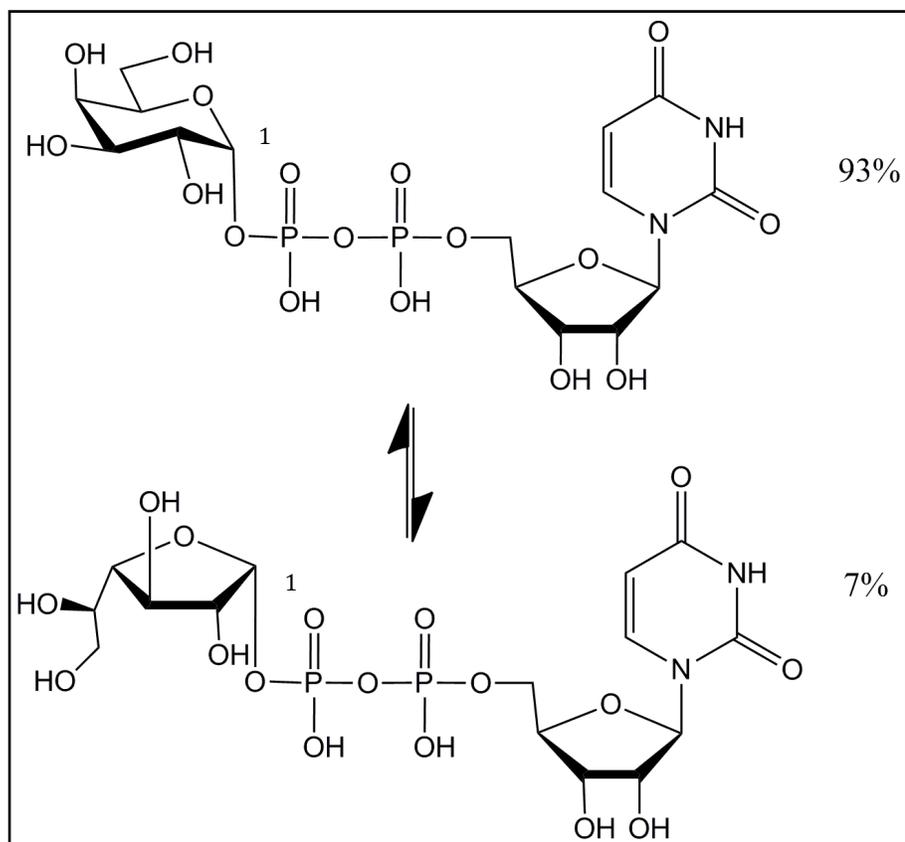
Figure 2.4: Reaction equilibrium between UDP-Gal*p* and UDP-Gal*f*. UGM catalyzes the conversion of UDP-galactopyranose (top) to UDP-galactofuranose (bottom). This conversion ratio is observed for both UNGM and UGM. Figure generated in ChemDraw (Mills, 2006).
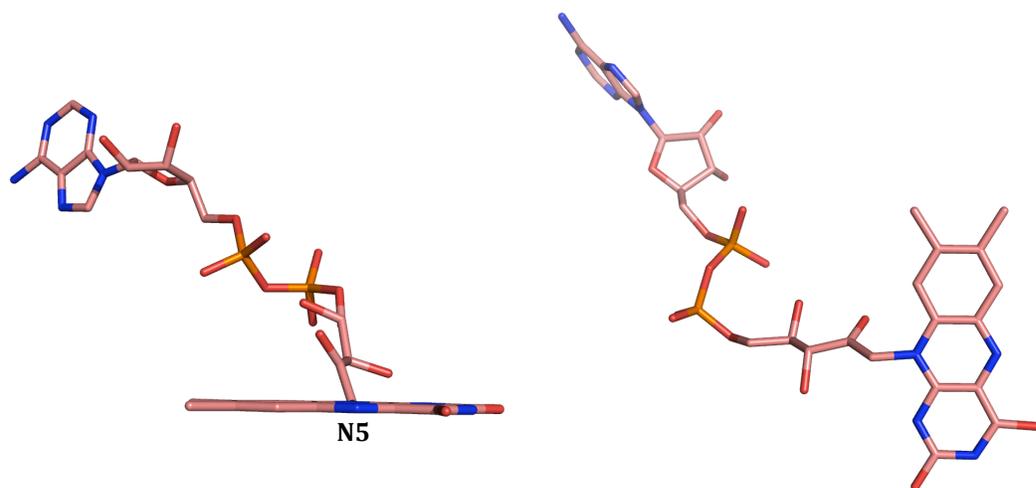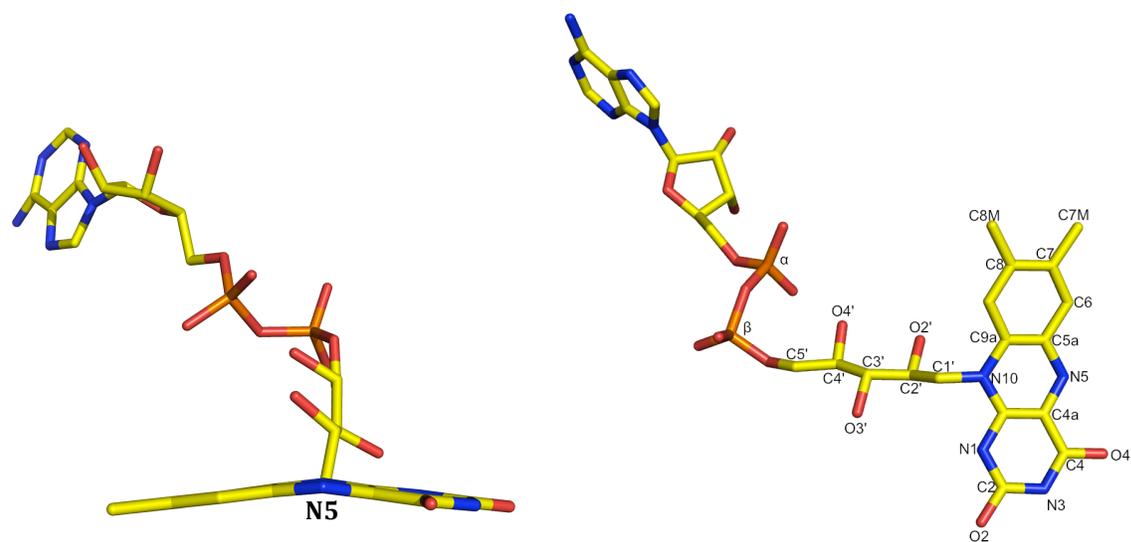
Figure 2.5: Structural conformations of flavin adenine dinucleotide. **A-** The oxidized flavin adenine dinucleotide, the isoalloxazine ring is planar (left) and is shown from the face-on view of the isoalloxazine ring (right) (from PBD 2BI7). **B-** The reduced flavin adenine dinucleotide, the isoalloxazine ring is puckered at N5 (left), and the flavin and phosphates are labeled on the right, from the face-on view (from PDB 3INT). Figures generated in Pymol (Version 1.5.0.4 Schrödinger, LLC).
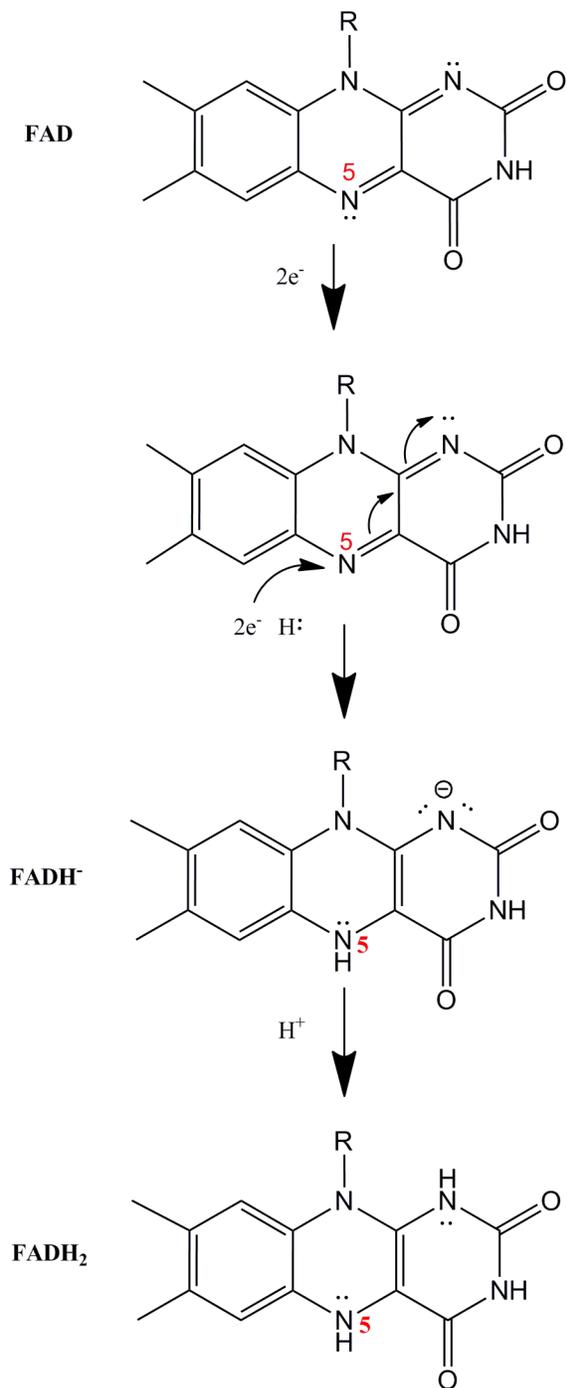
Figure 2.6: Flavin adenine dinucleotide in the oxidized and reduced forms. The transfer of $2e^-$ to N5 of FAD causes the reduction of the isoalloxazine ring. Figure generated in ChemDraw (Mills, 2006).

UGMs are known to catalyze this reaction via a highly conserved mechanism; although the exact mechanism by which the flavin adenine dinucleotide (FAD) cofactor participates in this reaction is still under debate (mechanism of UGM is described in detail in Section 2.5.2) (Soltero-Higgin *et al*., 2004). UGM specificity and recognition is dependent on the axial C4 hydroxyl of UDP-Gal. In comparison, UGM does not recognize UDP-Glucose, although it is identical to UDP-Gal with the exception of the equatorial C4 hydroxyl. Previous studies have indicated that the N5 locus of FAD is important for UGM's catalytic activity, which has also been corroborated by reduction of the isoalloxazine ring at N5 in several structures of UGM (Huang *et al*., 2003; Karunan Partha *et al*., 2009).

2.4 Proposed Reaction Mechanism of UNGM and UGM

Prokaryotic UGMs catalyze the reversible ring contraction of UDP-Gal*p* to UDP-Gal*f* via a highly conserved mechanism. UNGM is expected to adopt the same mechanism as UGM, due to its highly conserved active site and catalytic requirement for FAD reduction. Additionally, UNGM is able to catalyze the interconversion of Gal*p*NAc and Gal*f*NAc, which is expected to occur using the same mechanism that is observed for the reaction between Gal*p* and Gal*f*. The chemistry involved in this ring contraction is unique and has not been observed in other enzymes. The cleavage of the anomeric C-O bond of UDP-Gal*p* was established by oxygen isotope experiments (Barlow *et al*., 1999). The flavin adenine dinucleotide must be fully reduced for catalysis to occur, and is found in the anionic form FADH$^-$, which was

confirmed by thermodynamic analysis (Fullerton *et al*., 2003). Three different mechanisms have been proposed for this enzyme; SN$_2$, SN$_1$, and/or Single Electron Transfer (SET) (Fullerton *et al*., 2003; Soltero-Higgin *et al*., 2004; Itoh *et al*., 2007). All three mechanisms include the formation of the following species: FAD-galactopyranose adduct (I), iminium ion (II), and the FAD-galactofuranose adduct (III) (Figure 2.7). The mechanism by which the FAD-galactose adduct forms is still unknown. The S$_N$1 pathway involves the formation an oxocarbenium ion between C1 and the ring oxygen of galactose with UDP as a leaving group. The electron deficient C1 of galactose then accepts an electron pair from N5 of FADH$^-$ (Figure
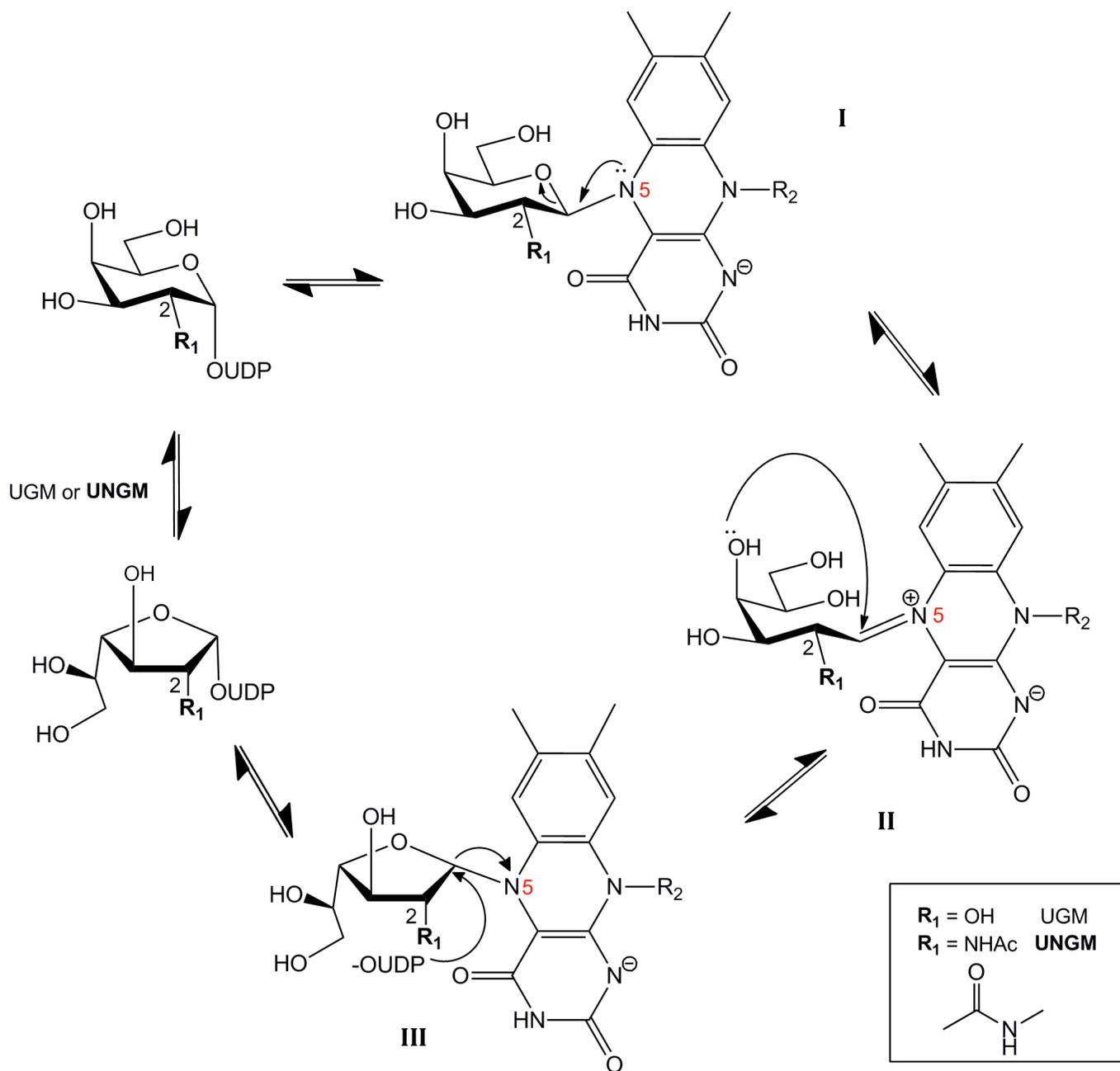
Figure 2.7: The proposed reaction mechanism of UGM and UNGM. UNGM is able to interconvert the pyranose and furanose forms of both UDP-Gal and UDP-GalNAc. All of the proposed mechanisms contain the following species: FADH⁻-galactopyranose adduct (I), iminium ion (II), and the FADH⁻-galactofuranose adduct (III). The figure was generated with ChemDraw (Mills, 2006).

2.7:I). In the S_N2 pathway, N5 of FAD acts as a nucleophile to attack the anomeric carbon, breaking the C-O bond between C1 and UDP and forming an adduct between the cofactor and sugar. Another proposed mechanism involves an electron deficient oxocarbenium ion at C1, resulting in an anomeric radical. The recombination (single electron transfer) between the anomeric radical and the semiquinone FAD results in the FAD-galactose adduct. After formation of the FADH⁻-galactopyranose adduct, N5 of FAD loses its lone pair of electrons to the ring oxygen of galactose, to form an FAD-iminium ion between C1 of Gal*p* and N5 of FAD, which has been identified by mass spectrometry (Soltero-Higgin *et al*., 2004) (Figure 2.7: II). The formation of the iminium ion facilitates the opening of the sugar ring, and the anomeric carbon can then undergo nucleophilic attack by the C4 hydroxyl, which is further stabilized by hydrogen bonding with O4 of FAD. The FADH⁻-galactofuranose adduct is formed upon the ring closure, and release of Gal*f* is thought to occur by the formation of the C-O bond of UDP-Gal*f* (Figure 2.7:III).

2.5 Structural studies of UGM

Several prokaryotic UGMs have been crystallized and structurally characterized, including the mutase enzymes from *E. coli, K. pneumoniae*, *M. tuberculosis*, and *Deinococcus radiodurans* (Sanders *et al*., 2001; Beis *et al*., 2005; Karunan Partha *et al*., 2009). UDP-*N*-acetylgalactopyranose mutase from *C. jejuni* shares the highest sequence identity with *E. coli* UGM (ecUGM), with 59 % shared identity, *M. tuberculosis* UGM (mtUGM) with 44 % identity, *K. pneumoniae* UGM (kpUGM) with 40 % identity, and lastly *D. radiodurans* UGM (drUGM) with 36 % identity. All of the UGM homologues with a crystal structure adopt the same overall fold and secondary/tertiary structure (Figure 2.8A). UGM belongs to the α/β class of proteins, and consists of 3 major domains. Domain 1 is the FAD-binding site, which contains an αβα Rossman fold. Domain 2 is a 5-helix bundle and domain 3 is a six-stranded antiparallel β-sheet.

UGM always forms a homodimeric structure, with each monomer nearly identical to the other with the exception of the mobile loop regions. Dimerization occurs through interactions between domain 2 (the 5 helix-bundle) of each monomer. The mobile loops 1 and 2 undergo
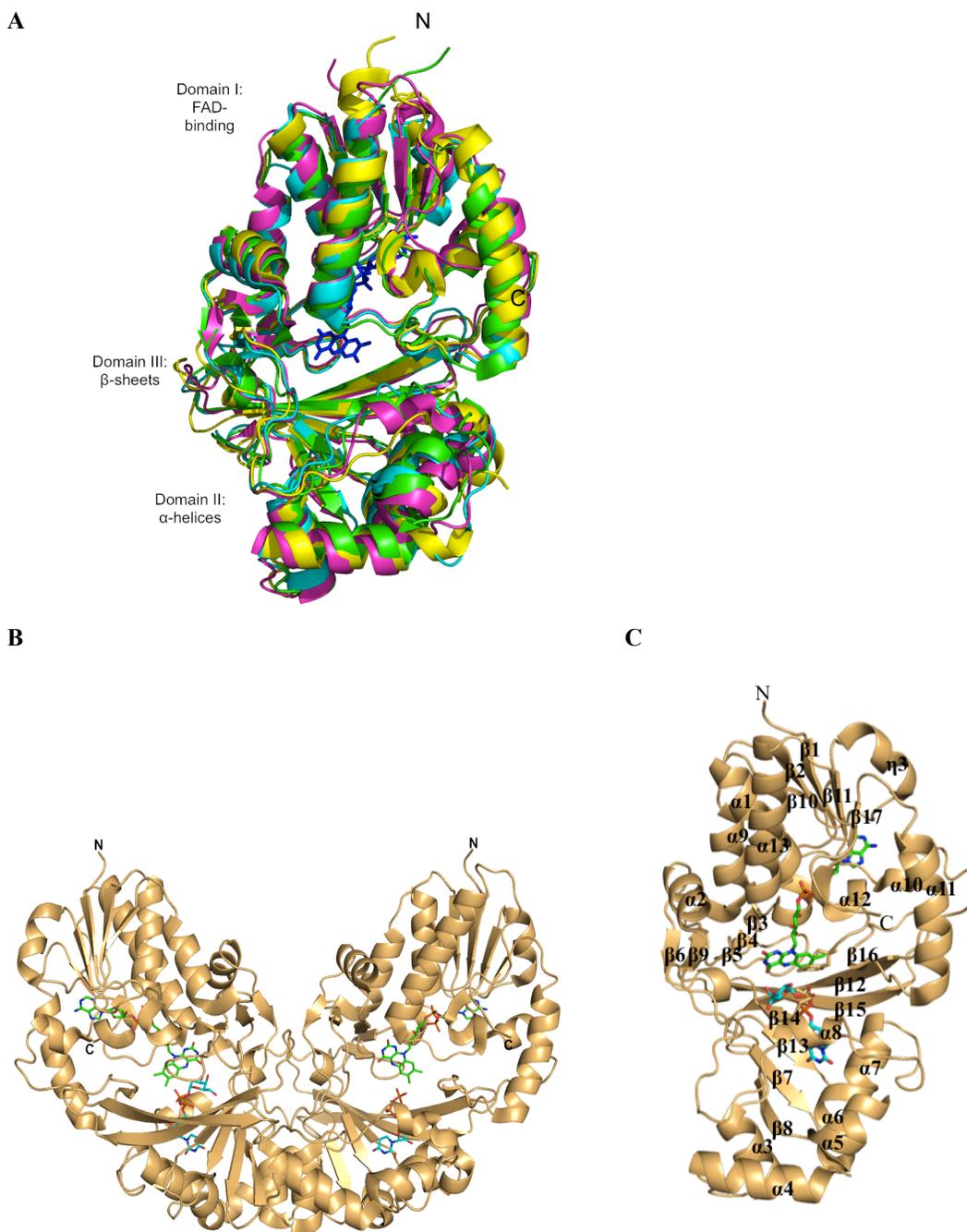
Figure 2.8: Structures of UNGM homologues. A- The tertiary structure of UGM is shown as: drUGM (green; resolution of 2.4 Å, PDB 3HDY), ecUGM (cyan; resolution of 2.4 Å, PDB 1I8T), kpUGM (magenta; resolution of 2.5 Å, PDB 3INT), and mtUGM (yellow; resolution of 2.25 Å, PDB 1V0J); the location of the active site is directly below the FAD cofactor (shown in blue). Superposition of monomers was performed using Secondary Structure Matching (SSM) in COOT (Krissinel and Henrick, 2004; Emsley and Cowtan, 2004). **B-** Homodimer of kpUGM (3INT); FAD shown in green, UDP-Gal*p* (left) and UDP (right) shown in blue. **C-** Secondary structures of kpUGM. Figures generated in Pymol (Version 1.5.0.4 Schrödinger, LLC).

16

conformational changes upon binding of substrate in several UGM structures (Figure 2.9). All UGMs contain mobile loop 2 (α7-β9 loop in kpUGM), which borders the active site cleft and must be in a closed conformation for catalytic activity of the enzyme (Figure 2.8C). This loop has been observed to undergo significant conformational changes, and is known to move from an open loop conformation to a closed loop conformation upon interaction with the substrate (Figure 2.9). In structures of UGM, the open loop conformation is observed primarily when no substrate is bound within the active site, and the closed loop conformation is found when amino acid residues in the loop interact with the substrate in the active site. In several UGM homologues, mobile loop 2 contains a small region of α-helical secondary structure. Mobile loop 1 is also conserved (α4-α5 loop), is often very short, and does not contain any secondary structure. This loop undergoes subtle conformational changes upon substrate binding as well, and allows for the translocation of the α5 helix toward the active site and stabilization of the closed conformation of the enzyme (Figure 2.9).
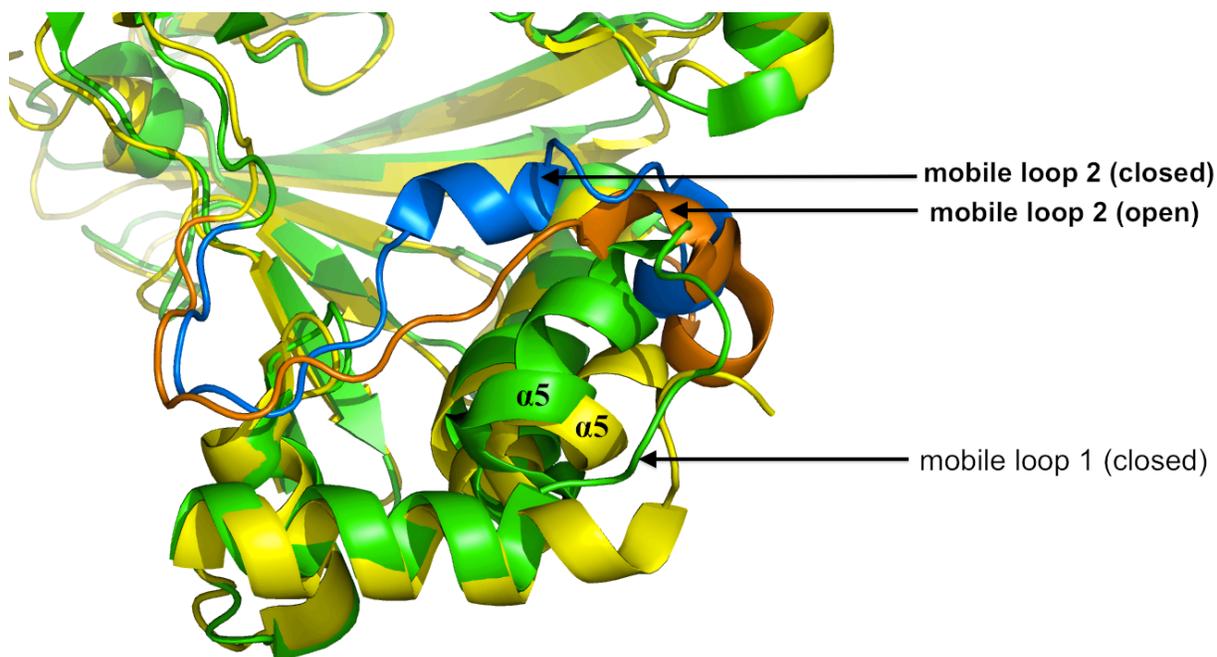
Figure 2.9: Mobile loops of drUGM. The reduced structure of drUGM (green) was superimposed with the oxidized structure of mtUGM (yellow). The blue loop (mobile loop 2) in drUGM indicates the closed conformation of the substrate binding mobile loop, and the orange loop (mobile loop 2) indicates the open conformation of this loop in mtUGM.

Structures of UGM with the oxidized form of FAD most commonly are in the open loop conformation such that the active site is accessible for substrate (PDB codes 2B17, 1VOJ, 1I8T) (Beis *et al*., 2005; Sanders *et al*., 2001). Until recently, structures of UGM in complex with substrate were unattainable despite intensive efforts. Several structures have been solved in which substrate(s) or substrate analogs are bound to the active site but the cofactor is not reduced, including UDP, UDP-Gal*p*, UDP-Glc, uridine diphospho methylene galactopyranose (PDB codes 3HE3, 3INR, 3GF4, 3HDQ, 3MJ4) (Gruber *et al*., 2009a; Gruber *et al*., 2009b; Karunan Partha *et al*., 2009; Karunan Partha *et al*., 2010). Structures of the holoenzyme of UGM with the reduced (puckered) form of FAD are nearly identical to the oxidized structures. The sole reduction of the cofactor does not promote the closed loop conformation (PDB codes 2BI8 and 1WAM) (Beis *et al*., 2005). However, when the reduced flavin is coupled with substrate binding (particularly the UDP region), the mobile loops are more commonly in a closed conformation (PDB codes 3INT and 3HDY). The closed loop conformation has been shown to aid in stabilization of the substrate through interactions with the conserved arginine on mobile loop 2. The FAD cofactor is in a bent (puckered) conformation in the reduced structures, although the pucker does not always occur at N5 of the flavin as predicted from the reaction mechanism. The reduced drUGM:UDP-Gal*p* structure has a variation in the reduction of FAD cofactors. The pucker in the isoalloxazine ring occurs at the flavin C4a or N5, both atoms have been known to participate in redox chemistry in other flavoenzymes. Although the reaction mechanism has suggested that N5 of the isoalloxazine ring forms an adduct with C1 of the sugar moiety, the cofactor is also thought to play a role in substrate stabilization. Co-crystallized structures of UGM with UDP-Gal*p* indicate that O4 of the isoalloxazine ring forms a hydrogen bond with the axial C4 hydroxyl of galactose (Karunan Partha *et al*., 2009).

Prokaryotic UGMs catalyze the reversible ring contraction of UDP-galactose (UDP-Gal) via a highly conserved mechanism, in which substrate binding is stabilized through interactions with several amino acid residues in the active site, as well as hydrogen bonding with the flavin cofactor (Figure 2.10). The roles of these residues have been determined by site-directed mutagenesis in ecUGM and kpUGM (Sanders *et al*., 2001; Chad *et al*., 2007). There are two highly conserved arginine residues (Arg305 and Arg198 in drUGM), which are essential for catalytic activity, and mutagenesis of these residues resulted in an inactive enzyme. Also, the conserved tryptophan residue (Trp184 in drUGM, Trp160 in kpUGM) forms binding

interactions with the uracil moiety of UDP, and had a significant effect on catalysis if replaced with a different residue. There are four tyrosine residues that have been found to participate in substrate stabilization, but do not have a pronounced effect on catalysis (Tyr335 and Tyr370 are shown in Figure 2.10).

2.6 *C. jejuni* UPD-*N*-acetylgalactopyranose Mutase (UNGM)

The UGM homolog, UDP-*N*-acetylgalactopyranose mutase (UNGM), was identified in *Campylobacter jejuni* 11168, encoded for by the *cj1439c* (*glf*) gene. This enzyme has a bifunctional activity, and is able to interconvert the pyranose and furanose forms of galactose and *N*-acetylgalactosamine. *C. jejuni* UNGM has a molecular weight of 43,445 kDa and consists of 368 amino acid residues. UNGM has been found to be essential for the production of UDP-2-acetamido-2-deoxy-D-galactofuranose (UDP-Gal*f*NAc), the precursor for Gal*f*NAc residues. Therefore, UNGM is crucial for the biosynthesis of the capsular polysaccharide in *C. jejuni*, by providing the activated precursors of Gal*f*NAc. This enzyme is an important antibiotic target, as the CPS is an important virulence factor in this pathogen.

2.6.1 Bifunctional Activity of UNGM

It is not uncommon for bacteria with compact genomes to express enzymes with more than one activity. In the pathogen *C. jejuni*, many bifunctional enzymes have been characterized such as IspD/IspF (MEP cytidylyltransferase/synthase), GalE (UDP-Glc/UDP-GlcNAc epimerase, and GlfT2 (β-1➜5- and β-1➜6-galactosyltransferase). Expression of the *glf* gene (which codes for UGM) in this organism is essential for virulence. UDP-*N*-acetylgalactopyranose mutase (UNGM) is the product of this gene, and also has a relaxed substrate-specificity in comparison to other known UGMs. Kinetic analyses have demonstrated a bifunctional enzymatic activity for UNGM, allowing for the recognition and conversion between the pyranose and furanose forms of UDP-galactose and UDP-*N*-acetylgalactosamine (Poulin *et al*., 2009). Complementation studies in *E. coli* confirmed that UNGM has UGM activity, and is therefore able to convert UDP-Gal*p* and UDP-Gal*f in vivo*. UNGM has been shown to convert UDP-Gal*p* to UDP-Gal*f* at a ratio of 93:7 respectively, which is consistent the conversion ratio observed for ecUGM (Poulin *et al*., 2009; Zhang and Lui, 2000) (Figure 2.4).

A
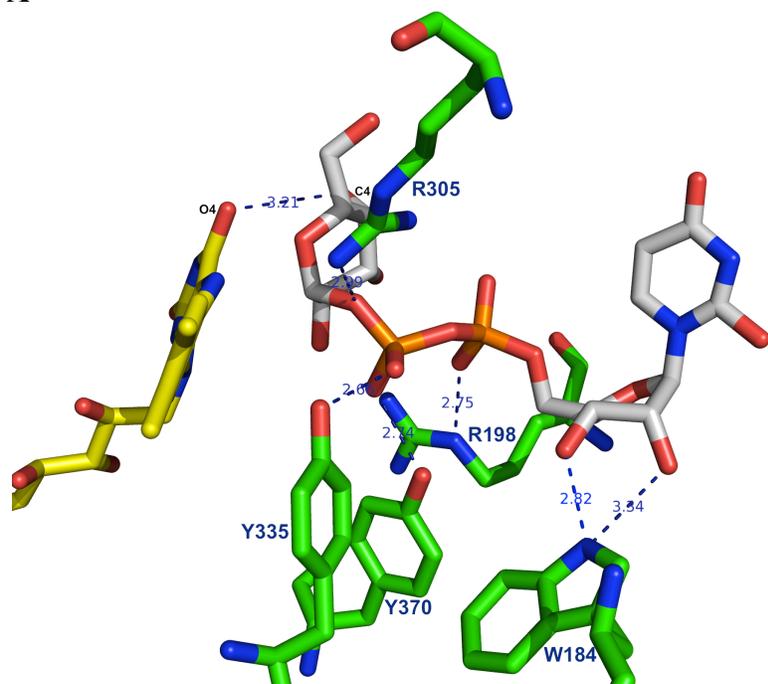
B
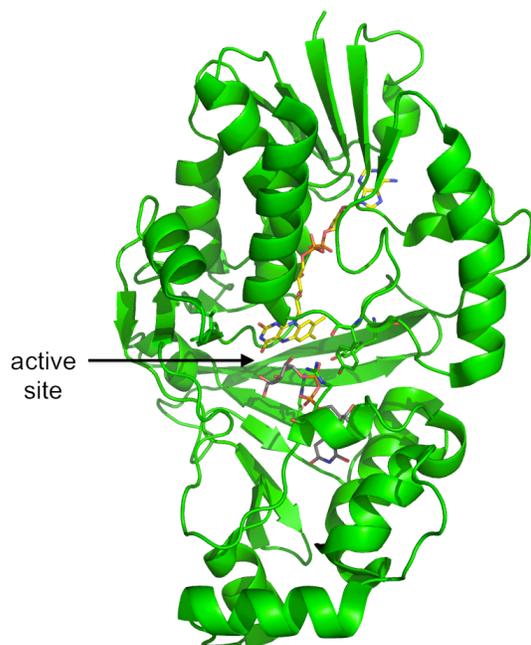
Figure 2.10: Substrate binding interactions of drUGM. **A-**The substrate UDP-Gal*p* (Carbons-gray) is stabilized by interactions with several conserved residues in the active site of drUGM (green) and the isoalloxazine ring of FADH⁻ (yellow). O4 in the isoalloxazine ring of the flavin cofactor is 3.2 Å from the C4 hydroxyl of UDP-Gal*p*. **B-** drUGM monomer, the active site is indicated by the black arrow. Figures generated in Pymol (Version 1.5.0.4 Schrödinger, LLC).

20

UNGM is the first known example of an enzyme able to catalyze the interconversion of UDP-GalpNAc and UDP-GalfNAc. Although the UGM homologues from *K. pneumoniae* (kpUGM) and ecUGM are highly related to UNGM, they are unable to catalyze the synthesis of UDP-GalfNAc sugars (Poulin *et al.*, 2009; Errey *et al.*, 2009). To our knowledge, there is no other enzyme that catalyzes the conversion of both of these substrates. Co-incubation studies of UNGM with UDP-Galf and UDP-GalfNAc indicated that the enzyme has a catalytic preference for UDP-Gal over UDP-GalNAc, with a ratio of 2.5:1 at equilibrium (Figure 2.11). The decreased rate of conversion of UDP-GalNAc may be due to additional bonding interactions with the active site residues. The stabilization of the acetamido group of GalNAc is crucial for the positioning of the sugar within a productive distance of the FAD cofactor. Mutagenesis of two non-conserved active site residues, Arg59 to His and Arg168 to Lys, reduced the ability of UNGM to catalyze the conversion of GalNAc residues (Poulin *et al.*, 2009). This result indicated that these residues are important for the specificity of the enzyme. It is worth noting that the conversion of both Gal and GalNAc substrates occurred with the R59H/R168K mutants, which indicated that other residues contribute to the specificity for GalNAc. Therefore, we conclude that the active site of UNGM is tailored specifically for this additional activity, and we hypothesize that the structural differences between UNGM and its UGM homologues will elucidate why this enzyme recognizes both substrates.
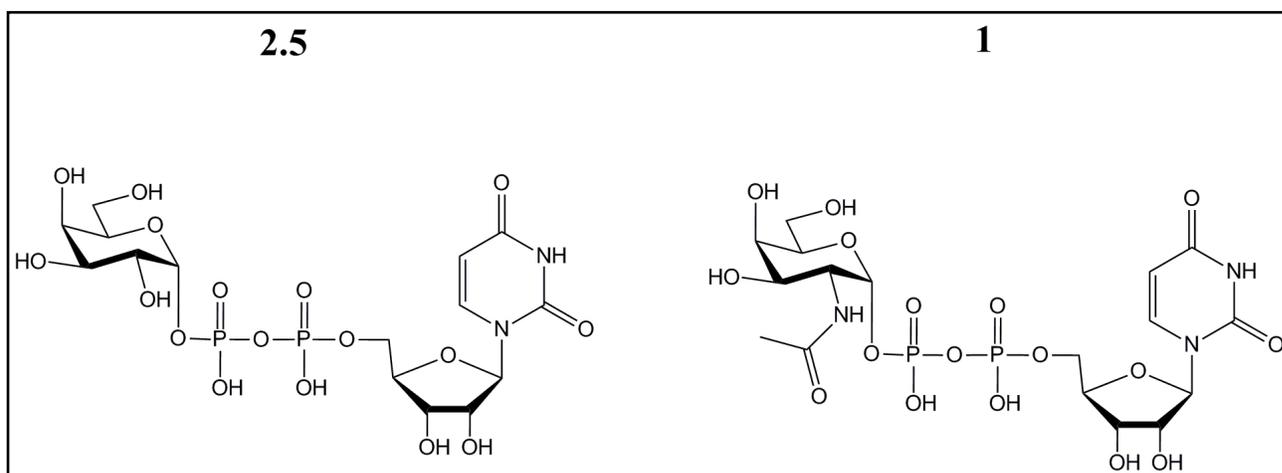
Figure 2.11: Catalytic preference of UNGM. At equilibrium, the ratio of UDP-Gal*p* to UDP-Gal*p*NAc is 2.5:1. UNGM is more catalytically active on UDP-Gal (left) than UDP-GalNAc (right). The figure was generated with ChemDraw (Mills, 2006).

CHAPTER 3: MATERIALS AND METHODS

3.1 Crystallographic Studies of UNGM

3.1.1 Protein Purification

C. jejuni 11168 UNGM (*cj1439c*) was cloned, expressed, and purified courtesy of Myles B. Poulin (Laboratory of Dr. Todd Lowary, Department of Chemistry, U of Alberta) (Appendix B). The purity of UNGM was assessed by SDS-PAGE.

3.1.2 Crystallization Screening

In order to discover an environment that is favorable for the crystallization of UNGM, it was necessary to screen a large number of different conditions. The crystallization of UNGM was initiated by performing screening trials using commercially available screening kits (Qiagen and Hampton Research), including the following suites (96 conditions in each suite): Classics, Classics Lite, Classics II, PEGs, PEGs II, Cryos, JCSG+, ComPAs, AmSO4, MPD, and PACT. The microbatch-under-oil technique was employed for screening because it requires a small volume of protein per experiment. The microbatch-under-oil plates contain 96 wells, and each well was treated as a separate experiment. In each well, 0.6 µL of protein solution to 0.6 µL of screening solution was mixed together, and the drop was subsequently covered with paraffin oil to slow the diffusion of the drop with the atmosphere (Figure 3.1). The diffusion of water through the paraffin oil allows for concentration of the reagents and protein to increase to a level that favors crystal formation.

Paraffin oil

Protein/crystallization solution
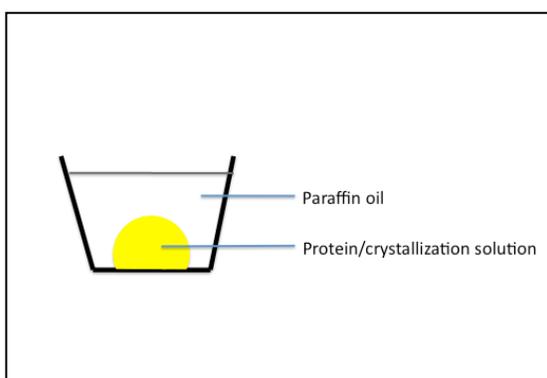
Figure 3.1: Microbatch under oil experimental design. The protein solution was mixed with the crystallization (or screening) solution in the well, and covered with paraffin oil. This technique was used for all of the crystallization screening experiments.

In order to achieve a UNGM complex structure with one of its substrates, the protein was incubated with 10 mM UDP-galactopyranose (UDP-Gal*p*) for 30 min prior to crystallization experimental setup. According to previous research in our lab, the presence of a reducing agent is normally required for the co-crystallization of UGM homologues with their respective substrates. Due to this observation, 10 mM of sodium dithionite was added to the protein sample prior to crystallization. Crystallization screening was performed as described above for the UNGM holoenzyme (Figure 3.1). Also, screening experiments were performed with an excess amount of UDP-Glc (5 to 20 mM) was added to 6.5 mg/mL of UNGM, and allowed to incubate for 30 min. The microbatch-under-oil technique was utilized for screening for suitable conditions for crystallization as previously described.

UDP has previously been shown to inhibit UGM, and is known to share the same binding mode as the natural substrate UDP-Gal*p* (Yuan *et al*. 2005, Karunan Partha *et al*., 2009). Therefore, co-crystallization of UNGM with UDP was attempted to determine if UDP binds the active site of UNGM as observed in the drUGM:UDP structure (PDB 3HE3). Crystallization screening was performed using commercially available kits (described previously) with the addition of 5 mM UDP and 10 mM sodium dithionite to reduce the FAD cofactor.

3.1.3 Crystal Optimization

Optimization of crystal conditions may improve the quality of crystal packing, crystal size, and crystal morphology, and therefore increase the quality of the diffraction data. Also, a large number of crystals may be required for testing different cryoprotectants and substrates. UNGM holoenzyme crystals identified during the screening experiments were optimized using a grid screen of the original condition that produced the crystals. The micro-batch under oil technique was employed for the grid screen. The grid screen varied the buffer pH (4.5, 5.0, 5.3, 5.6) and the precipitant polyethylene glycol 4000 (final concentration of 12 to 15%). For example, each crystallization drop had a final volume of 1.2 μL, and consisted of 0.6 μL of 6.8 mg/mL of UNGM in 20 mM Tris-HCl pH 7.6 mixed with 0.6 μL of condition #66 from the PEGs II Suite, containing 0.2 M Ammonium sulfate, 0.1 M tri-Sodium citrate pH 5.6, 25 % (w/v) Polyethylene glycol (PEG) 4000, covered with paraffin oil. Crystallization temperature is another variable that can be changed to improve the probability of crystal growth. Therefore,

parallel optimization experiments were set up at both room temperature and 4°C to increase the chances of crystal nucleation.

3.1.4 Microseeding Optimization

The conditions that produced crystals in the presence of UDP-Glc were optimized using the micro-seeding technique. Crystals were placed in a stabilization solution mimicking the mother liquor in which they crystallized, containing 0.2 M Ammonium sulfate, 0.1 M Sodium acetate-acetate pH 5.6, 30 % PEG 4000, 20 mM Tris-HCl pH 7.6, and 5 mM UDP-Glucose. Seed beads (Hampton Research) were added to the solution, vortexed for 1 min to crush the crystals into microseeds, and the resultant solution was diluted to 1:10, 1:50, and 1:100. Microseeding experiments were performed by mixing the following components in the well of a microbatch plate: 0.5 μL of 6.5 mg/mL UNGM, 0.5 μL of the screening condition, and 0.2 μL of the seeding solution. The drops were subsequently covered with paraffin oil.

3.1.5 Soaking Experiments and Cryoprotection

Prior to data collection, holoenzyme crystals were soaked for 5 min in a cryoprotectant of mother liquor containing 30 % glycerol, mounted on a cryoloop (Hampton Research), and subsequently flash-cooled in liquid nitrogen. The same technique was employed for other crystal conditions, unless otherwise noted. A simple method to produce a structure with bound substrate is to soak in the substrate into holoenzyme crystals. If the crystal packing is compatible with this method, such that the active site is solvent accessible and is not blocked by crystal contacts with other molecules, this method is likely to be successful for obtaining a complex structure. The substrate analog UDP-5-deoxy-Gal*f* (UDP-5d-Gal*f*) was used for holoenzyme crystal soaking experiments under the assumption that it may bind differently than UNGM's natural substrates (Figure 3.2). This analog is not an inhibitor of UNGM, however it has been observed to inhibit the synthesis of galactan chains by forming a "dead end" reaction product (Poulin *et al*., 2012). Crystals of UGM are yellow in color, due to the presence of oxidized flavin within the enzyme (Karunan Partha *et al*., 2009; van Straaten *et al*., 2012). Reduction of the flavin caused the dissipation of the yellow color to generate clear crystals. The crystals were reduced using 10 mM Sodium dithionite, followed by the addition of a cryoprotectant containing 30 % glycerol and 10 mM UDP-5d-Gal*f*. The crystals were allowed

to soak for 30 min, and did not dissolve or crack, and appeared to remain in a reduced form prior to freezing. In addition, co-crystals of UDP-Glc were soaked in a cryoprotectant containing 30 % glycerol and higher concentrations (5mM to 100 mM) of UDP-Glc, UDP-Gal*p*, UDP-Gal*f*NAc, and UDP-5d-Gal*f*. Sodium dithionite dissolved the crystals shortly after addition, and therefore was not used as a reducing agent for the co-crystals.

3.1.6 Diffraction and Data collection

All crystals were diffracted using synchrotron radiation at a wavelength of 0.9790 Å at the Canadian Macromolecular Crystallography Facility (CMCF-1) beamline (08ID-1) at the Canadian Light Source (Saskatoon, Saskatchewan, Canada). The crystal was kept at a temperature of 100 K in a stream of nitrogen gas during data collection. Diffracted x-rays were detected using a MAR 300CCD detector. Intensity data for the UNGM holoenzyme crystals was collected as 180 images with a 1s exposure over a 1° oscillation range, at a crystal-to-detector distance of 250 mm.

Intensity data for the UNGM crystals soaked in UDP-5d-Gal*f* were collected as 180 images over a 1° oscillation range with a 5 s exposure per image at a crystal-to-detector distance of 250 mm. Longer exposure to the synchrotron radiation was necessary due to the weak diffraction observed during screening at 1 s exposure. Intensity data for UNGM/UDP-Glc crystals from PEGs II #71 was collected as 180 images, with a 3 s exposure over a 1°
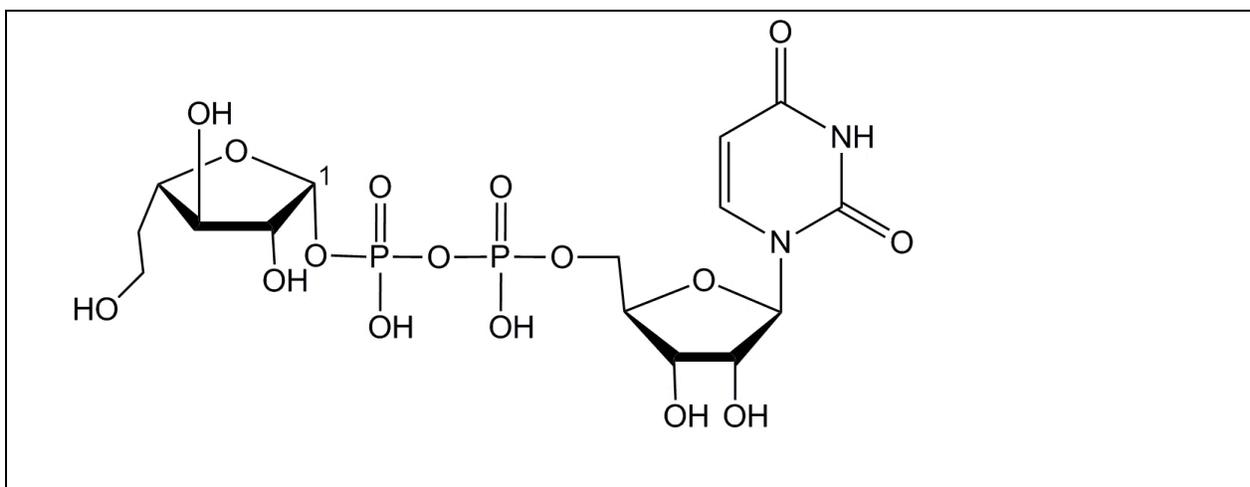
Figure 3.2: Substrate analog of UNGM. UDP-5-deoxy-Galactofuranose (UDP-5-deoxy-Gal*f*). Figure generated in ChemDraw (Mills, 2006).

oscillation range at a crystal-to-detector distance of 250 mm. Data for UNGM/UDP-Glc co-crystals from ComPAs #13 was collected as 180 images with a 5 s exposure over a 1° oscillation range at a crystal to detector distance of 250 mm. Lastly, UNGM/UDP-Gal*p* co-crystals were screened for diffraction by exposure to synchrotron radiation at a range of 1 to 10 s to determine if they were suitable for data collection.

3.1.7 Data Processing

Intensity data collected from the UNGM holoenzyme crystals were indexed, integrated and scaled using the XDS software package (Kabsch, 1993). Intensity data collected from the holoenzyme crystals soaked in UDP-5d-Gal*f* and the UNGM/UDP-Glc co-crystals from PEGs II #71 were indexed, integrated and scaled using d*TREK (Pflugrath, 1999). The data collected from the UNGM/UDP-Glc co-crystals were indexed, integrated and scaled using HKL2000 (Otwinowski and Minor, 1997). Additionally, the program Pointless (from CCP4 program suite) was used to determine the Laue group (Evans, 2006; Collaborative Computational Project Number 4, 1994). The diffraction data was analyzed for the presence of anisotropy using the Diffraction Anisotropy Server (Strong *et al*., 2006; http://services.mbi.ucla.edu/anisoscale/). Data quality was assessed using xtriage in Phenix (Afonine *et al*., 2005).

3.1.8 Structure Solution and Refinement

A molecular replacement solution for the UNGM holoenzyme was found using the atomic coordinates of UGM from *E. coli* (PDB entry 1I8T) which shares 59 % identity, using the program Molrep in MrBUMP in the CCP4 program suite (Vagin and Teplyakov, 1997; Keegan and Winn, 2007; Collaborative Computational Project, Number 4, 1994). Initial refinement was performed in Refmac5 (Murshudov *et al.*, 1997). Density fitting and addition of solvent (other than waters) was performed in the modeling software Coot (Emsley and Cowtan, 2004). Addition of the FAD cofactors, water molecules and the final refinement was performed in Phenix (Afonine *et al*., 2005). The data collected for the co-crystals and soaking experiments were solved by molecular replacement using the holoenzyme structure of UNGM using Molrep (Vagin and Teplyakov, 1997).

3.1.9 Validation of Structure

When the UNGM holoenzyme structure could no longer be improved using refinement and modeling, the final model was validated using Molprobity in Phenix (Murray *et al*., 2010; Afonine *et al*., 2005).

3.2 Docking Studies with GOLD

For the docking experiments, the search efficiency for the Genetic Algorithm (GA) was set to very flexible (200 %, unless otherwise stated), and is recommended for large highly flexible ligands such as UDP-Gal*p*. This setting delivers high predictive accuracy but is relatively slow due to the exhaustive number of operations performed. For UDP-Gal*p*, which has 8 rotatable bonds, we expect approximately 48,000 operations at a search efficiency of 100 %. Therefore, we can assume the number of operations for a search efficiency setting of 200 % would approximately equal 100,000.

For each experiment, the binding site was defined from the reference ligand UDP-Gal*p* from the structure of drUGM (PDB 3HDY). A search radius of 6 Å from the center of the ligand was chosen to encompass all of the proposed substrate-binding residues. Only those atoms specifically included in the binding site definition were considered during the docking experiments. Furthermore, the distance between N5 of the FAD cofactor and C1 of the sugar moiety of the substrate was constrained to 3.0 Å. Also, prior to docking experiments, the water molecules and ligand were removed from the model, and hydrogen atoms were added to the model to ensure proper protonation of the residues involved in substrate binding.

The scoring function CHEMPLP was used for initial docking experiments because it has been found to give the highest success rates for both pose prediction and virtual screening experiments against diverse validation test sets (Verdonk *et al*., 2003). Briefly, the CHEMPLP fitness score is calculated from the fitness score generated by Piecewise Linear Potential (PLP), which is used to model the steric complementarity between protein and ligand. Additionally, the Chemscore (total free energy change upon ligand binding) distance- and angle-dependent terms for hydrogen and metal bonding are subtracted from the PLP fitness score (Equation 3.1).

$$\text{fitness}_{\text{CHEMPLP}} = \text{fitness}_{\text{PLP}} - (f_{\text{CHEM-hb}} + f_{\text{CHEM-cho}} + f_{\text{CHEM-met}}) \qquad \text{(Equation 3.1)}$$

In addition, the scoring function Chemscore was used to validate the CHEMPLP solutions. The Chemscore fitness function estimates the total free energy change that occurs on ligand binding and was trained by regression against binding affinity data. As the fitness scores are dimensionless they cannot be used explicitly as values for binding energy or binding affinity. However, the $\Delta G_{binding}$ component of the score (ignoring the clash penalty and internal torsion terms) may provide a crude estimate of the binding affinity (Equation 3.2). The $\Delta G_{binding}$ estimates the total free energy change upon ligand binding. The "P" terms indicate the penalty terms, and "c" terms indicate torsion terms. If covalent constraints are included in the docking, then $c_{covalent}$ and $P_{covalent}$ terms are also included.

$$\textbf{Chemscore} = \Delta G_{binding} + P_{clash} + c_{internal}P_{internal} + (c_{covalent}P_{covalent} + P_{constraint}) \quad \text{(Equation 3.2)}$$

3.2.1 Generation of Docking Models

The structures of drUGM and UNGM were modified in Coot (Emsley and Cowtan, 2004). Structure idealization was performed in Refmac5 (Murshudov *et al*., 1997). The models were energy minimized using the Chiron Rapid Protein Energy Minimization server (http://dokhlab.unc.edu/tools/chiron), which uses short, discrete molecular dynamics simulations to remove steric clashes (Ramachandran *et al*., 2011).

3.2.2 Validation of Docking Models

The structure-idealized, energy minimized models were validated for Ramachandran and rotamer outliers using Molprobity (Chen *et al*., 2010). Flagged (outlier) residues were adjusted in Coot (Emsley and Cowtan, 2004). Geometric restraints were generated using the program Elbow in Phenix (Afonine *et al*., 2005).

3.2.3 Generation of Substrate Models for Docking

Models of UDP-Gal*f*, UDP-Gal*f*NAc, and UDP-Gal*p*NAc were built manually in the computational chemistry software Spartan (Hehre and Ohlinger, 2011). The Equilibrium Geometry was calculated using the molecular mechanics force field in Spartan. In this process, the bond lengths and angles are varied until the lowest energy is reached. This process was repeated several times to determine the lowest energy models.

3.3 Sequence Analysis

The sequence of *C. jejuni* 11168 UNGM (*Cj1439c*) was aligned with the sequences of the UGM homologues *C. jejuni* ssp. doylei, *E. coli*, *M. tuberculosis*, *K. pneumoniae, C. jejuni* ssp. 81116, *C. jejuni* ssp. ICDCCJ07001, *D. radiodurans* using clustalw (Thompson *et al.*, 1994). Secondary structure files were generated from the sequence alignment using DSSP (Kabsch and Sander, 1983).

CHAPTER 4: RESULTS

4.1 UNGM holoenzyme Structure

4.1.1 Crystallization of UNGM holoenzyme

The purity of a protein sample is essential for successful crystallization because the protein molecules pack in an ordered fashion during crystal formation, and contaminants can interfere with this process. The purity of UNGM in 20mM Tris HCl pH 7.6 was assessed by SDS-PAGE (Figure 4.1). The major component of the sample is approximately 43 kDa, which corresponds to the protein's calculated molecular weight of 43,445 Da. Minor impurities present in the sample were only visible in the highly concentrated samples of UNGM (lanes 2 and 3, Figure 4.1).

Therefore, the purity of the sample was deemed sufficient for crystallization trials. Furthermore, the theoretical isoelectric point (pI) of UNGM is 8.42, which is significantly higher than the pH of the storage buffer at 7.6. The lower pH was used for storage because proteins are more stable in solutions that differ from their isoelectric point. This is due to the charges that exist on the exterior of the protein which allow for interaction with the surrounding
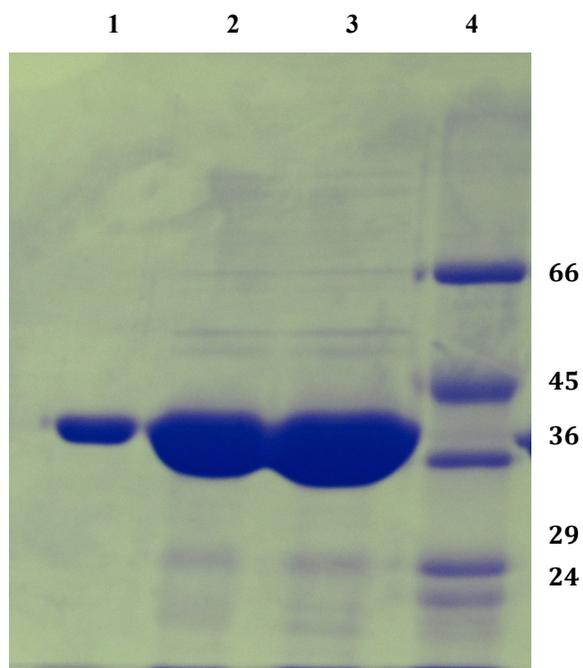
Figure 4.1: SDS-PAGE analysis of UNGM. Lanes 1, 2, 3 correspond to 1 μL, 5 μL, and 10 μL of concentrated UNGM (approximately 6 mg/mL). Lane 4 is the molecular weight (MW) marker.

water molecules. If the protein has no net charge (or is at its pI), the protein molecules are more likely to interact with each other and precipitate out of solution. The protein samples were stored at 193 K prior to use. The samples were inspected after thawing on ice, and appeared to have a yellow color and little or no precipitation depending on the aliquot.

The crystallization experiments were examined immediately after set up, to ensure that the crystallization drop remained mostly clear in appearance, with minimal protein precipitation. Screening kits that resulted in a high degree of precipitation initially did not yield any crystals and were not used for further screening. Diverse crystallization screening yielded several different crystal morphologies, including rod clusters, plate clusters, and needle (feathered-appearance) clusters. The yellow color of the crystals was indicative of the oxidation of the FAD cofactor. Therefore, the crystals were assumed to contain the inactive form of the enzyme (Sanders *et al.*, 2001). Since no FAD was added to the crystallization experiments, we can conclude that FAD co-purifies with UNGM. The best crystals were yellow with rod-shaped morphology, and grew within 24 h at 298 K with the dimensions 0.3 mm x 0.05 mm x 0.05 mm (Figure 4.2A). The crystals grew from the screening kit PEGs II #66, containing 0.2 M Ammonium sulfate, 0.1 M tri-Sodium citrate pH 5.6, 25 % (w/v) Polyethylene glycol (PEG) 4000 mixed in a 1:1 with 6.5 mg/mL UNGM in 20 mM Tris HCl pH 7.6.

4.1.2 Diffraction and Data Processing

The UNGM holoenzyme crystal from PEGs II #66 diffracted to 1.9 Å using synchrotron radiation (Figure 4.2B). The data was indexed, integrated, and scaled using the XDS software package (Kabsch, 1993). The Matthews coefficient had a value of 2.67 $Å^3$ Da^{-1} assuming two molecules are present in the asymmetric unit. The data were analyzed using the conditions for the systematic absence of reflections for the space group $P2_12_12_1$, which are h00: h = 2n + 1, 0k0: k = 2n + 1, 00l: l = 2n + 1. The probability of this space group based on the systematic absences was 0.898, with the next highest probability of 0.081 for the space group $P22_12_1$, suggesting that the crystal belonged to the orthorhombic space group $P2_12_12_1$ with unit cell parameters (Å) a = 48.4, b = 116.0, c = 165.4.

The data quality was assessed primarily by the value of $I/\sigma(I)$ (signal to noise ratio), equal to 2.12 in the highest resolution shell (1.9 Å to 1.95 Å). Intensity data must be
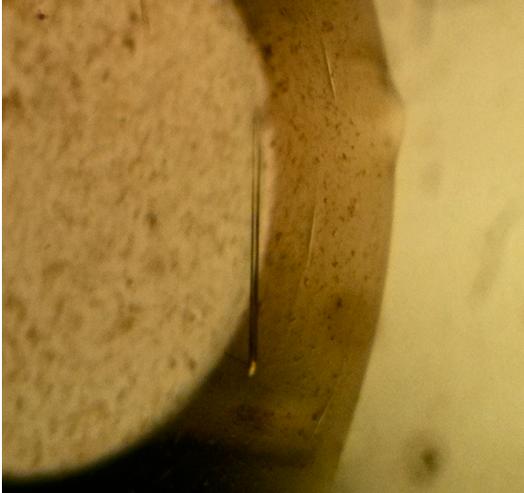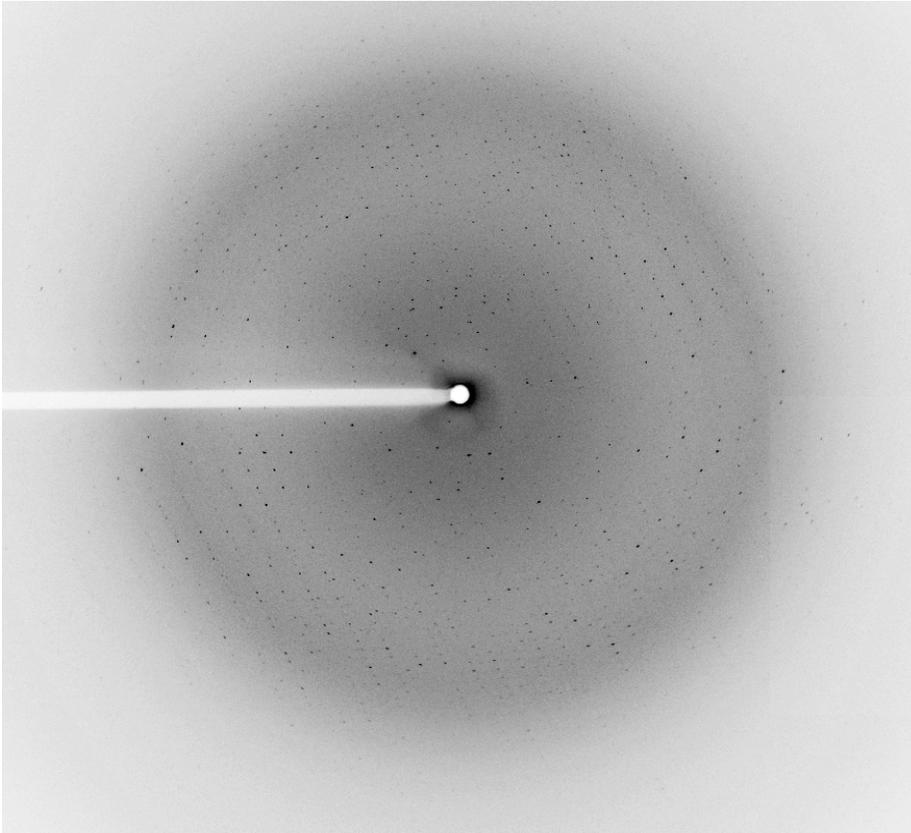
Figure 4.2: UNGM holoenzyme crystallization and diffraction. **A-** Crystals of UNGM were grown in PEGs II #66. **B-** Diffraction image of crystals grown from PEGs II #66.

distinguishable from the background noise that inherently arises from diffraction, therefore the average intensity divided by its error must be greater than 2.0 (Evans, 2006). Below this value, the data are likely too weak to be useful in structure determination. Therefore, the data at this resolution was of sufficient intensity for calculation of structure factors. Furthermore, the completeness of the data was 94.2 % (82.7 % in the highest resolution shell), which indicated that a high percentage of the total reflections present were measured. At this resolution, there are a large number of reflections (5397), high multiplicity (6.73). Furthermore, a decrease in completeness does not occur until a very high R-factor (R_{meas} or R_{merge}) is observed for the data (Figure 4.3A).

R-factors are indicators of data consistency, and R_{merge} is most commonly used for judging data quality (Equation 4.1; Diederichs and Karplus, 1997). The value of R_{merge} was large (0.169, 0.714 in the highest resolution shell). The data was also processed using the first 90 images for integration and scaling, however the lowered redundancy did not improve R_{merge} significantly (0.15, 0.70 in the highest resolution shell). In XDS, R_{meas} is more reliable than R_{merge} because it does not rely on the multiplicity (or redundancy) of the data (Equation 4.2). The value of R_{meas}, the multiplicity-weighted version of the R-factor, was equal to 16.2 % overall (77.1 % in the highest resolution shell). Although R_{meas} is large in the highest resolution shell, $I/\sigma(I)$ equaled 2.12, which indicated that there is a large amount of useable data present at this resolution.

$$R_{merge} = \Sigma_{hkl}\Sigma_i \,|I_i(hkl) - <I(hkl)>|\, /\Sigma_{hkl}\Sigma_i \,I_i(hkl) \qquad \text{(Equation 4.1)}$$

$$R_{meas} = \frac{\Sigma_h \sqrt{\frac{n_h}{n_h-1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h^{n_h}\sum_i I_{h,i}} \quad \text{with } \hat{I}_h = \frac{1}{n_h}\sum_i^{n_h} I_{h,i}$$

$$\text{(Equation 4.2)}$$

The data was also analyzed for the presence of anisotropy. Diffraction anisotropy, indicated as a dependence on direction in diffraction quality, is attributed to uniformity of crystal packing in one direction. The rod-like morphology of the crystals may diffract weakly through the smallest dimension (0.05 mm), and therefore anisotropic analysis was used to detect if the crystal defect was present. The data was subjected to ellipsoidal truncation using a

resolution limit of 1.9 Å in all three directions (x,y,z). Essentially no anisotropy was observed in the crystal based on the spread in values of the three principle components (the exponential scale factors used to correct for anisotropy), which had an anisotropic delta B equal to $5.36\ Å^2$. The value of the anisotropic delta B statistic (degree of anisotropy) indicates mild anisotropy above $10\ Å^2$, and strong anisotropy above $25\ Å^2$. Therefore, the data does not exhibit mild anisotropy and anisotropy should not contribute to a large R-factor.

A
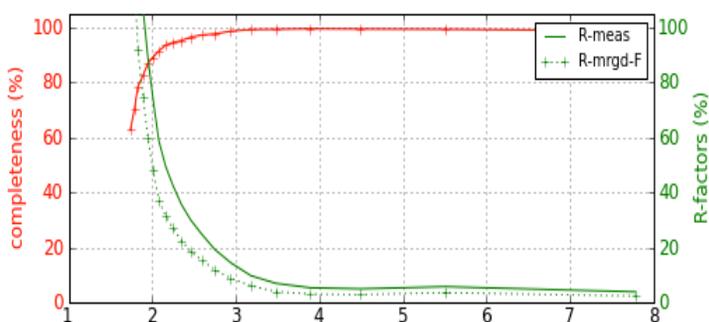
B
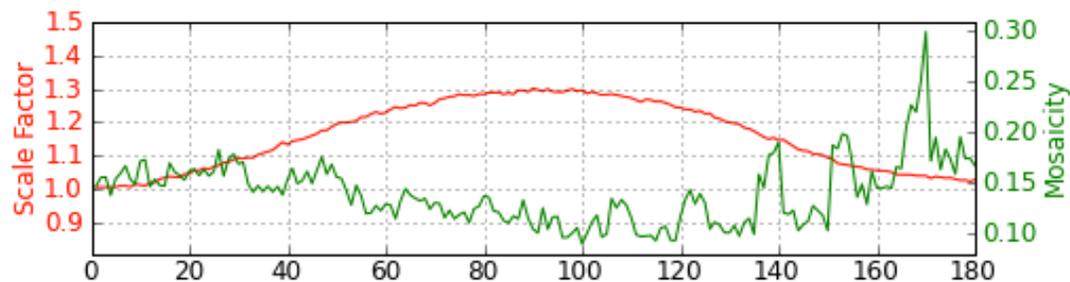
C
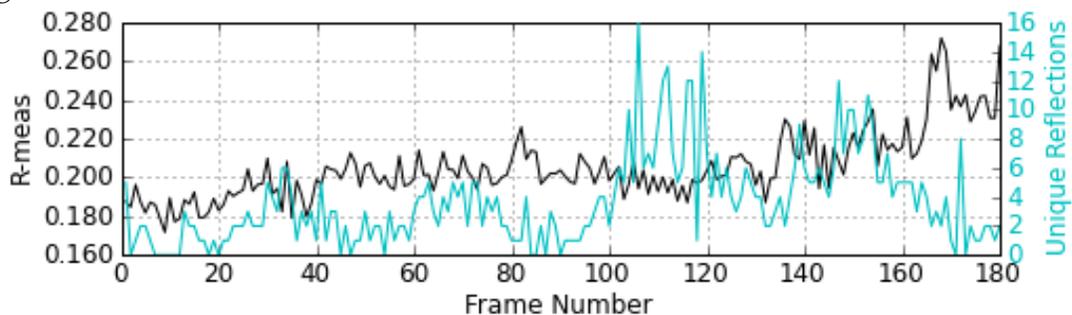
Figure 4.3: Data quality for the UNGM holoenzyme. A- Completeness of intensity data (left) and the R-factors (right) vs. shell. **B-** Mosaicity vs. the frame number. **C-** R_{meas} vs. frame number (bottom).

Due to the elevated R-factor, the data was re-processed in $P2_1$ to test for violation of space group symmetry and crystal twinning. Twinning is a common defect where the crystal is composed of distinct domains whose orientations differ, but are related in a specific way. The result is the overlapping of reflections from more than one lattice or crystal. Merohedral twinning occurs when the lattices of the different domains overlap in three dimensions. Hemihedral twinning, a specific type of merohedral twinning, involves two different domain orientations. Twinned crystals contain two or more domains, and the orientation of the domains are related by one or more symmetry elements that do not exist in the space group symmetry of the single crystal. In pseudo-merohedral twinning, higher apparent symmetry is observed in the diffraction pattern than what exists within the crystal. Pseudo-merohedral twinning occurs when the crystal metric corresponds to the higher symmetry system. The twinning operation is a symmetry element that is found in the higher-symmetry group but missing in the proper symmetry class of the crystal.

The data was indexed in monoclinic, and the unique axis b was chosen for all three possible permutations. The unit cell dimensions (Å) were a= 116.23, b= 48.47, c= 165.40, and β= 90.01°. The data was integrated with the mosaicity set at 0.29. The systematic absences in *hkl* reflections showed that the data contained a 2_1 screw axis, and followed the following conditions: k, l = 2n. The data was then re-indexed in $P2_1$ using dtcell. Finally, the data was scaled to a resolution of 2.0 Å. The value of R_{meas} did not improve significantly when the data was reprocessed in $P2_1$, with an overall value of 0.147 (0.604 in highest resolution shell). The intensity data (in $P2_1$) was analyzed using the L-test for twinning, which examines the relationship between local pairs of reflections (near each other in reciprocal space)(Padilla and Yeates, 2003). Acentric reflections have randomly distibruted atom phases, whereas for centric reflections the phase for atom pairs are related by the contribution of both atoms in the pair is equal to ϕ_c or $\phi_c + \pi$. Therefore, if there is twinning in the data, the data will fit the red curve (Figure 4.4A) that represents the theoretical acentric reflections for twinning. The data did not deviate significantly from the untwinned, theoretical acentric value (Figure 4.4A). Specifically, the Mean |L| = 0.494 (untwinned: 0.500, perfect twin: 0.375) and Mean L^2 = 0.327 (untwinned: 0.333, perfect twin: 0.200). Therefore, no twinning is expected in the data based on this test. The symmetry of the lattice and the intensity of the data suggested that the chosen space group was too low. One possible pseudo-merohedral twin operator (twin law: h,-k,-l), a two-fold

twinning operator on the *a*-axis was found within this symmetry. This is consistent with the condition that the angle *β* must ≈ 90° for the appearance of orthorhombic symmetry to exist in the monoclinic crystals. Twin fraction estimation was performed using the Britton analyses and the H-test, with values of 0.441 and 0.460 respectively. Britton analyses measure the fraction of reflections outside the allowed region given a certain twin fraction (Britton, 1972), and the H-test is used to detect meroherdral twinning (Yeates, 1988). These scores indicate that if twinning is present in this space group, it is nearly perfect. It is more likely that the symmetry of the processed data is incorrect, due to the presence of twin laws and the lack of suspected twinning in the intensity statistics. Additionally, structural refinement was unsuccessful in this

A
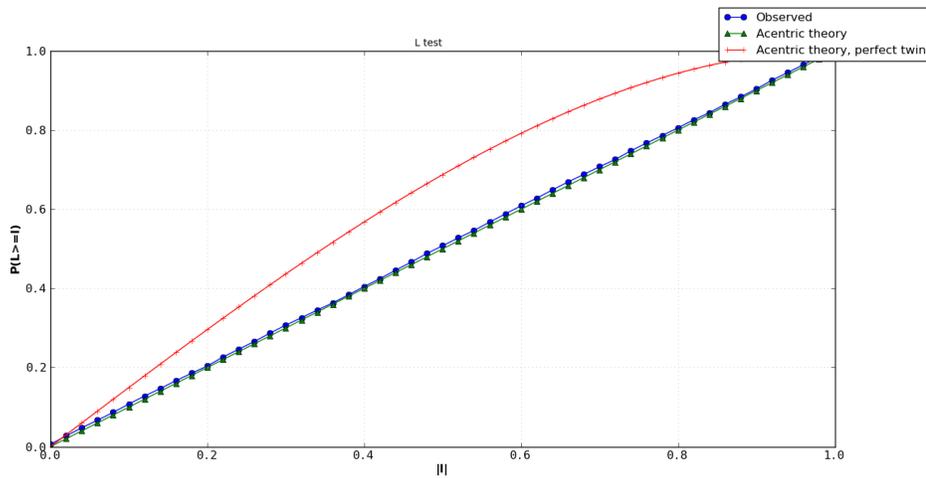
B
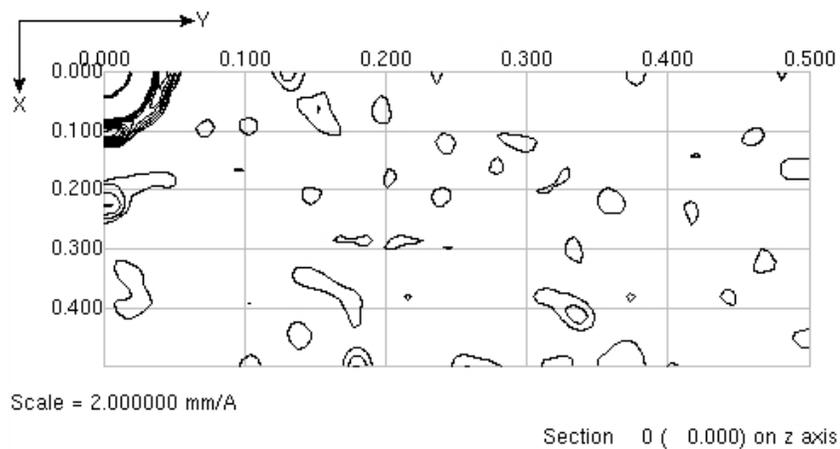
Figure 4.4: Twinning analysis for the UNGM holoenzyme. A- The L-test for twinning plot for intensity data (Padilla and Yeates, 2003). The data fits the untwinned theoretical value for acentric reflections. **B-** Patterson map for the UNGM holoenzyme in P2$_1$, no major non-origin peaks are observed.

space group, as the value of R_{free} remained above 0.43 despite using the same refinement strategy used for the structure in $P2_12_12_1$. Therefore, these findings suggest that the data belongs to the higher symmetry space group $P2_12_12_1$.

A large R-factor can also arise from crystal slippage, radiation damage, or bad absorption corrections. The presence of multiple lattices within the crystal will also result in an elevated R-factor. The diffraction images were visually inspected for the presence of an additional lattice and anisotropy, and did not show symptoms of either of these crystal defects. The reciprocal lattice did not appear to shift and there was no decrease in high-resolution reflections throughout data collection. Radiation damage may be a cause of the elevated R-factor, and R_{meas} was observed to increase for approximately the last 20 images from the data collection, which was also correlated to an increase in mosaicity of the diffraction in these images (Figure 4.3B and C). The elevation of these two parameters near the end of data collection strongly favors the hypothesis that radiation damage has occurred. If this were entirely true, we expect the R-factor to decrease with lowered multiplicity, however this was not observed. Alternatively, the radiation damage may occur very early on in the data collection, or freezing the crystals may also cause damage and increase in disorder. This damage may be prevented by further optimization of the cryoprotectant and reduction of exposure time to synchrotron radiation, or attenuation of the beam intensity. Patterson analysis was used to detect translational pseudosymmetry in the data, and did not contain any peaks other than the origin and its symmetry related peaks. Therefore, it is very unlikely that there is translational pseudosymmetry present (Figure 4.4B). Data-collection statistics are summarized in Table 4.1.

4.1.3 Structure Solution and Refinement

The structure of UNGM was solved by molecular replacement using the structure of UGM from *E. coli* (PDB 1I8T), which shares 59 % identity with UNGM. In order to determine the success of the molecular replacement, the model of UNGM was compared to the calculated electron density from the model phases. The $2|F_o|-|F_c|$ electron density map compares 2 x the observed amplitudes vs. the amplitudes calculated from the model. The $2|F_o|-|F_c|$ map was visually inspected for quality based on its presence of secondary structure, side chain and

Table 4.1: Data collection and refinement statistics for UNGM holoenzyme structure

Data Collection:	
Space group	P2$_1$2$_1$2$_1$
Unit-cell parameters (Å)	a =48.4 b=116.0 c=165.4
Wavelength (Å)	0.9790
Temperature (K)	100
Resolution range (Å)	47.5-1.90 (1.90- 1.95)
Matthews coefficient (Å^3 Da^{-1})	2.67
Solvent content (%)	54
No. of molecules in ASU	2
Molecular weight (Da)	43445 [368 amino acid residues]
No. of measured reflections	523944 (35448)
Total No. of unique reflections	74431 (5397)
Completeness (%)	94.2 (82.7)
Redundancy	7.04 (6.73)
R$_{meas}$	0.162 (0.771)
<I/σ(I)>	11.7 (2.12)
Refinement:	
R$_{work}$‡	0.1827
R$_{free}$§	0.2223
R.m.s.d. from ideal values	
Bond lengths (Å)	0.006
Bond angles (°)	0.975
Average *B* factor (Wilson Plot) (Å^2)	
Overall	23.70
Protein	21.02
Solvent	30.05
Waters	668
Ramachandran plot (%)	
Favored	98.2
Allowed	1.8

carbonyl bulges, as well as subunit boundaries. The C-alpha backbone of the model appeared to fit the $2|F_o|-|F_c|$ electron density map well at a contour level of approximately 1.0 sigma, with the exception of the N-terminal end of the protein, which was missing density for several residues. Initially, each residue in the model was manually checked and fitted to the $2|F_o|-|F_c|$ map in Coot. The regions of the model that deviated from the electron density were optimized using the real-space refinement and regularization functions. Real-space refinement maximizes the correlation between what is observed and the model-based density, whilst regularization corrects distorted bonds. In several areas, manual fitting was necessary to fit the model to the electron density. The second type of electron density map used for modeling is the $|F_o|-|F_c|$ map or difference map, which is calculated from the difference between the observed and calculated structure-factor amplitudes. These maps are used to identify regions of electron density not clarified by the model. As expected, positive electron density in the $|F_o|-|F_c|$ map (contoured to 3.0σ) was observed for FAD within the conserved FAD binding region of UNGM (Figure 4.5A and B).

Alternate conformations were observed for the active site residue Arg59 in one of the monomers of UNGM. Therefore, both conformations were modeled in the structure with occupancy of 0.5 assigned to each. Alternate conformations were not observed in the other monomer, although the conformation observed is in the proposed active conformation. Calculation of an omit map is useful to remove model phase memory and bias. Initially, the refined model was edited to remove the uncertain regions including the alternate conformations of Arg59, Glu285, and the C-terminal residues that exhibited high B-factors. Secondly, the model was subjected to simulated annealing (starting temperature of 1000 K), and the model phases recalculated in phenix.refine. Lastly, an omit map was calculated from the initial phases using Sfcheck in CCP4 (Vaguine *et al.*, 1999). Overall, the omit map fit the model very closely, which indicated that the model was correct. Density was observed in the omit map for the alternate conformation of Arg59 in monomer B only, therefore validating that both conformations exist in the structure and are not an artifact of model bias (Figure 4.5C). The C-
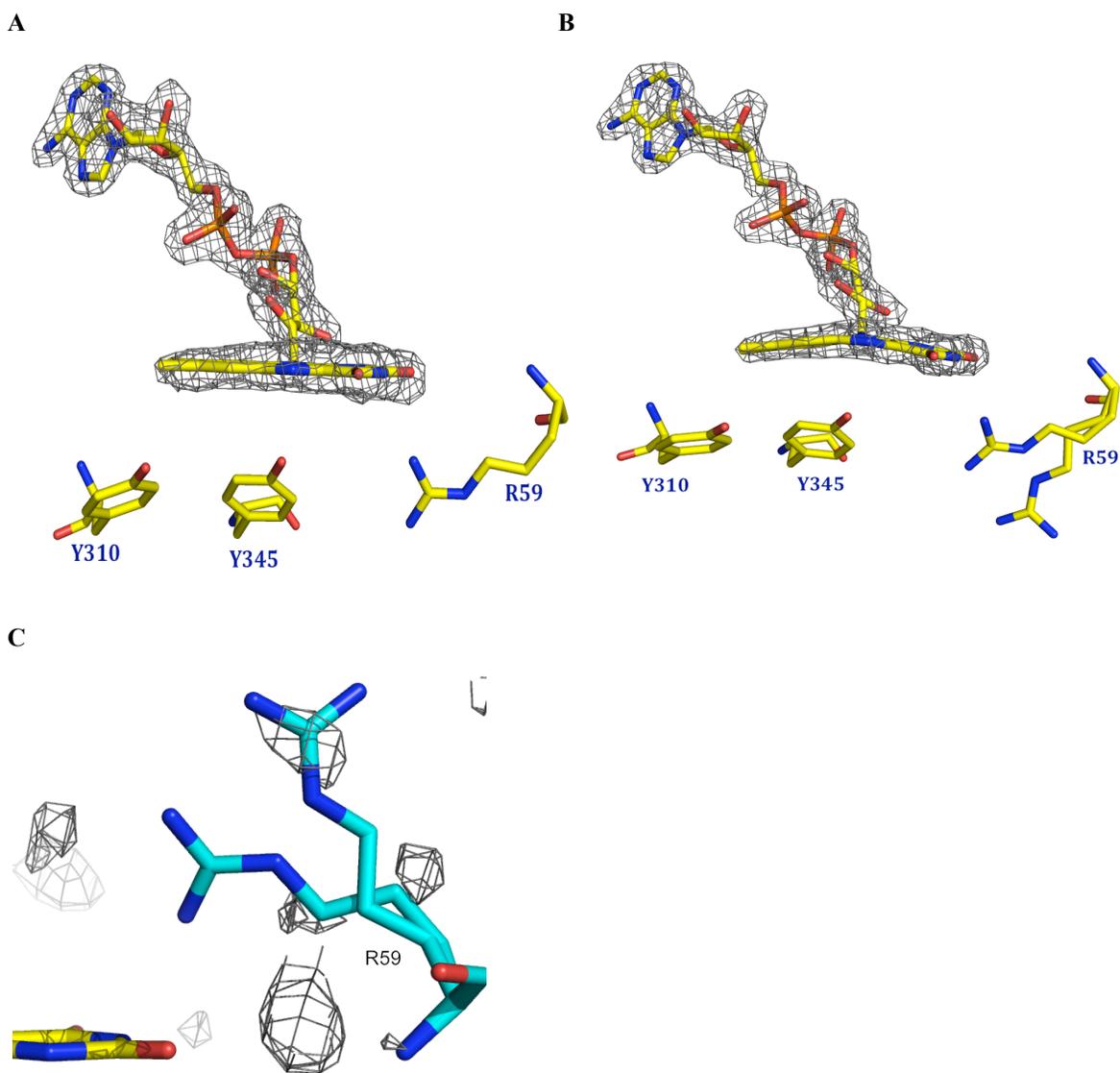
Figure 4.5: FAD conformations and alternate residue conformations in UNGM holoenzyme. A- The oxidized, planer FAD cofactor in monomer A of UNGM; $2|F_o|-|F_c|$ map shown in gray, contoured to 2.0 σ. **B-** The reduced, puckered FADH⁻ cofactor in monomer B of UNGM; the $2|F_o|-|F_c|$ map is shown in gray, contoured to 2.0 σ. Figure generated in Pymol (Version 1.5.0.4 Schrödinger, LLC). **C-** Observed density for the second conformation of Arg59 in monomer B of UNGM; $|F_o|-|F_c|$ omit map (gray) was contoured to 3.0 σ.

terminus of the model (residues 364-366) was deficient in electron density, and was removed from the model.

Further refinement of the structure was performed using phenix.refine to improve the initial phases by decreasing errors in the model. The following parameters were selected initially for refinement: xyz coordinates, rigid body, occupancies, and individual B factors. B factors for all atoms were refined individually as isotropic at this resolution. Also, simulated annealing was performed (starting temperature of 1000°, 50 steps) to find the energy minimum for the model. Importantly, no NCS restraints were used during the refinement due to significant differences between each monomer of the UNGM homodimer. After initial refinement in phenix.refine, R_{work} dropped from 0.3104 (R_{free} = 0.3064) to 0.2414 (R_{free} = 0.2777). The Ramachandran and rotamer outliers were checked with the structure validation program Molprobity (Chen *et al*., 2010), which relies heavily on the optimized hydrogen placement and all-atom contact analysis, as well as covalent-geometry and torsion-angle criteria. The identified outliers were modified accordingly using "Auto Fit Rotamer" or "Edit Chi Angles" in Coot. Furthermore, absences and defects present in the model were identified using the $|F_o|-|F_c|$ map.

The next step in the structural refinement of UNGM was to fit the flavin adenine dinucleotide cofactors into each monomer. LigandFit in Phenix was used to generate geometric constraints for FAD fit the electron density observed in the $|F_o|-|F_c|$ difference map. In monomer A, the cofactor FAD is in the planar oxidized form, which had an overall correlation with difference map of 0.84 (0.7 or better indicates an excellent fit), whereas the reduced conformation (FADH⁻) had a correlation of 0.82 (Figure 4.5A). In monomer B, the flavin cofactor is in the reduced puckered conformation (FADH⁻), with an overall correlation with the $|F_o|-|F_c|$ map of 0.84, and FAD in the oxidized form had a correlation of 0.83 (Figure 4.5B). Therefore, the conformation of the cofactor with the highest correlation to the difference map was fitted to the model and used for further refinement. After visual inspection in Coot, the flavin cofactors were then merged with the model of UNGM, and the model was subjected to another round of refinement in Phenix. The same parameters were refined, with the exception of a lower starting temperature of 500° for simulated annealing resulted in lower R_{work} and R_{free} values of 0.2242 and 0.2588 respectively. For further refinements of the structure, no rigid

body refinement or simulated annealing was used. Additional un-modeled density in the $|F_o|$-$|F_c|$ map was resolved using automatic water picking in phenix.refine using a 3.0 sigma cutoff in the $|F_o|$-$|F_c|$ map. Other ordered solvent molecules were added manually in Coot using the "Difference Map Peaks" function.

Chi1-Chi2 Plots were used to identify outliers, which were manually adjusted. The Isotropic B Factor (mean) was 23.66, and individual residues with a B-factor higher than 60.0 were manually checked for their atomic positions. As expected, all residues exhibiting very high B-factors were on the periphery of the molecule and were solvent accessible. Therefore, we expect a higher degree of flexibility in these residues and a higher value for the B-factor. The final B-factors were analyzed using Baverage in the CCP4 software suite (Collaborative Computational Project 4, 1994). The final R_{free} and R_{work} were 0.2223 and 0.1827 respectively. The significant decrease in R_{free} during the refinement of the model validated that the model and the phases derived from the original data shared a higher correlation with each other after refinement than prior to refinement. Final refinement statistics are summarized in Table 4.1.

4.1.4 Validation of Structure

Once the structure could no longer be improved by refinement and modeling, it was analyzed for Ramachandran and rotamer outliers using Molprobity and Comprehensive Validation in Phenix (Chen *et al*., 2010; Afonine *et al*., 2005). The structure contained no Ramachandran outliers, with 98.2 % of residues in the Ramachandran favored region (Figure 4.6). Also, a small number of rotamer outliers (0.8 %) remained in the structure, however the outliers appeared to fit the electron density well, and were not involved in steric clashes with other residues. Validation statistics are summarized in Table 4.1.
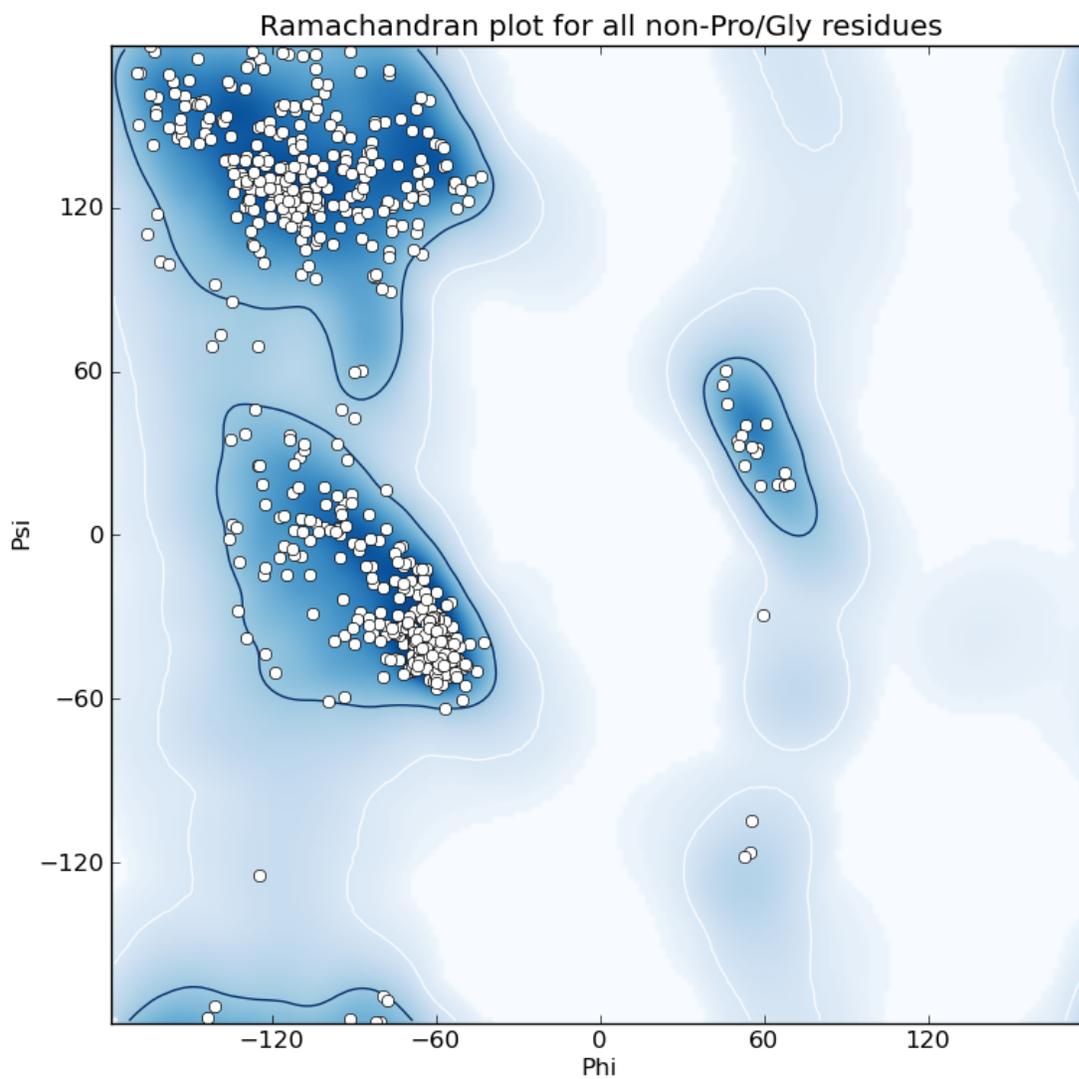
Figure 4.6: Validation of holoenzyme structure of UNGM. Ramachandran plot for the final refinement of the holoenzyme structure of UNGM. The dark blue region is the most favored region in the Ramachandran plot (Ramakrishnan and Ramachandran, 1965). The light blue region is the additional allowed region. The disallowed region is white. All the residues were within the allowed regions of the plot.

4.2 Structural Analysis of UNGM holoenzyme

4.2.1 Crystal Structure of UNGM holoenzyme

All UGMs share a conserved overall fold, including UNGM (Figure 4.7A). UNGM forms a homodimer in its active form and belongs to the α/β class of proteins (Figure 4.7B). The tertiary structure of UNGM consists of 3 major domains. Domain 1 is the FAD-binding site, which contains a αβα Rossman fold. Domain 2 is a 5-helix bundle and domain 3 is a six-stranded antiparallel β-sheet. In the holoenzyme of UNGM there are 4 helices in domain 2, although we expect another helix is formed in mobile loop 2 upon interaction with substrate, as was observed in the substrate-bound structure of kpUGM, but not in the holoenzyme structure (Gruber *et al*., 2009a and 2009b). The substrate binding cleft is located adjacent to the isoalloxazine ring of FAD, with mobile loops 1 and 2 near the opening of the cleft. In other UGMs, these loops move upon binding of substrate within the cleft (Karunan Partha *et al*., 2009; Chad *et al*., 2007; Yuan *et al*., 2008; Yao *et al*., 2009; Gruber *et al*., 2009a). Due to the high degree of conservation within the active site, we expect that UNGM shares a similar binding mode with the other UGMs. The superposition of the substrate UDP-Gal*p* (from kpUGM:UDP-Gal*p* structure; PDB 3INT) with the active site of UNGM indicated that this enzyme likely binds the substrate via conserved residues previously identified in UGM homologues (Figure 4.8A).

Each monomer of UNGM contains two mobile loops (shown in yellow) near the active site of the protein (Figures 4.7C and 4.8C). The mobile loops undergo significant conformational changes upon binding of substrate within the active site, and elevated temperature factors in the loops indicated that there is increased flexibility in these regions. Mobile loop 1 (residues 123-129) moves toward the active site upon substrate binding, which allows for movement of the α5 helix inward to adopt the closed conformation of the enzyme (Figures 4.7C and 4.8B). The length of this loop is shorter than the same loop observed in drUGM by 2 residues; however we believe that it serves the same function in UNGM. Mobile loop 1 does not contain secondary structure, which is consistent with other structures of UGM. Previous crystallographic studies of UGM have identified that mobile loop 2 (residues 156-186) moves inward toward the substrate binding cleft upon interaction with substrate (Karunan Partha *et al*., 2009; Gruber *et al*., 2009a; Gruber *et al*., 2009b). Residues in mobile loop 2
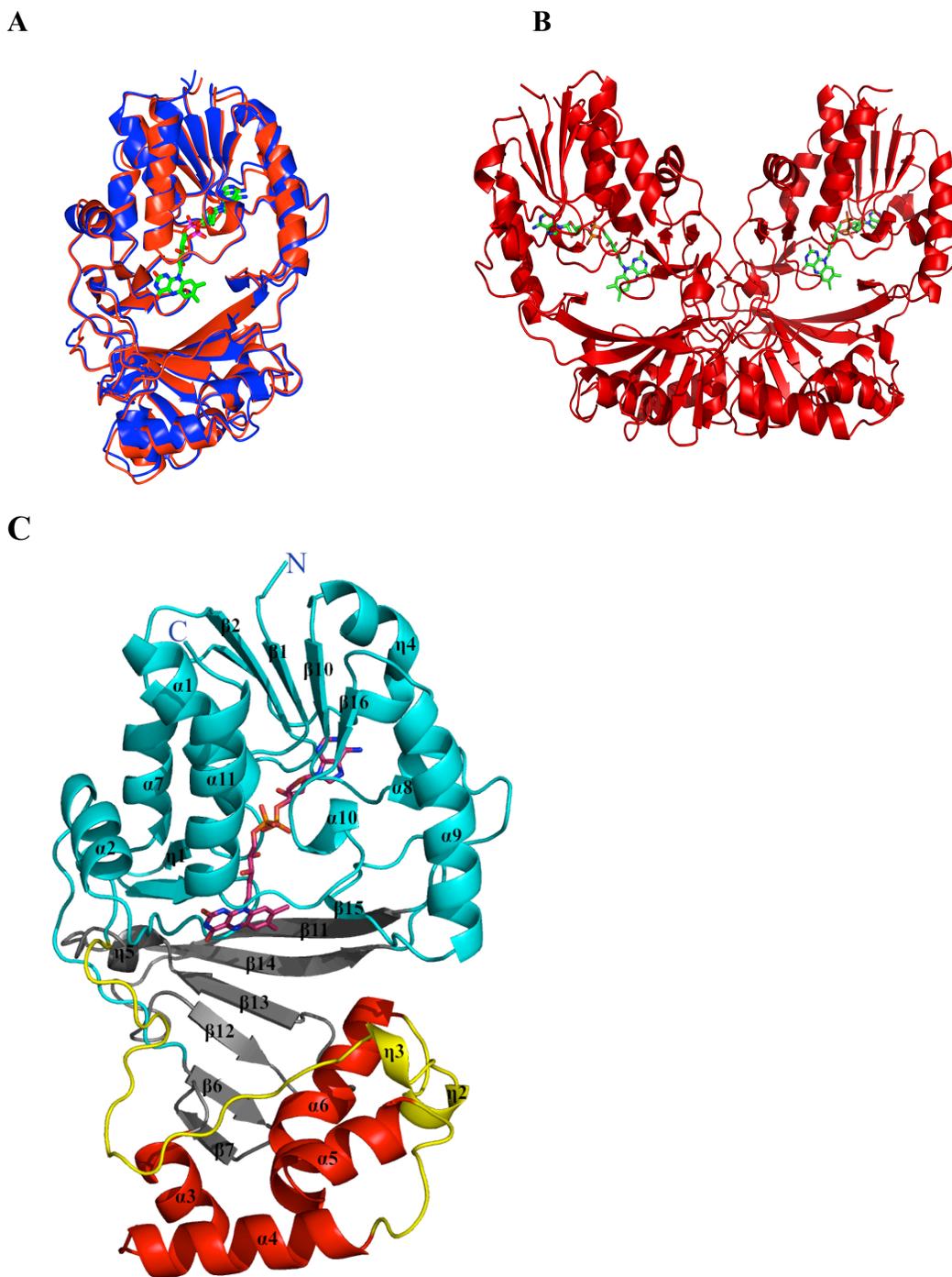
Figure 4.7: Crystal structure of UDP-*N*-acetylgalactopyranose mutase holoenzyme. A- Ribbon diagram representation of a monomer of UNGM (red) was superimposed over the structure of ecUGM (blue; PDB code 1I8T) using Secondary Structure Matching (SSM) superposition (Krissinel and Henrick, 2004); FAD cofactor is rendered as sticks in green. **B-** The asymmetric unit of the UNGM holoenzyme structure, which is a homodimer. **C-** Domains 1, 2 and 3 of UNGM monomer are shown in cyan, red, and gray respectively, and the mobile loops are shown in yellow. The FAD cofactor is shown in pink. Secondary structures are labeled in black. Figures generated with QtMg and Pymol (Potterton *et al*., 2002).
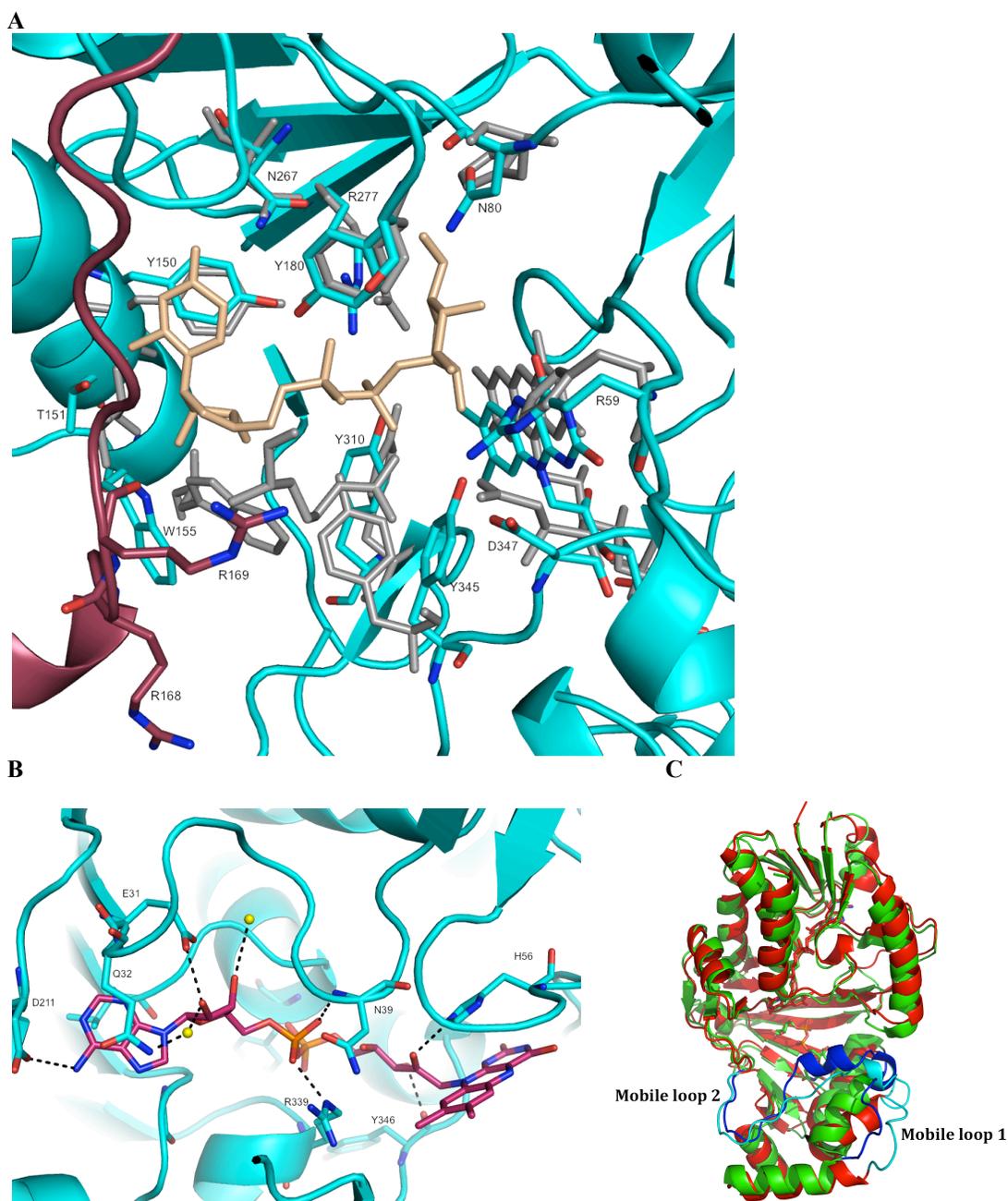
Figure 4.8: The binding mode of the UNGM. A- UDP-Gal*p* from drUGM (wheat) is within favorable distances for stabilization in the active site of UNGM (cyan). Residues within a distance of 4.0 Å to the substrate are shown in cyan for UNGM; gray for drUGM; mobile loop 2 in UNGM shown in raspberry. The structure of UNGM was superimposed with the drUGM:UDP-Gal*p* (3HDY) using Least Squares Fit (LSQ superposition) in Coot (Emsley *et al.*, 2010). **B-** FAD-binding domain of UNGM, the cofactor flavin adenine dinucleotide (magenta-C, red-O, blue-N, orange-P) is non-covalently bound by the αβα Rossman fold. Water molecules are shown as yellow spheres. **C-** The mobile loop arrangement of UNGM. The UNGM holoenzyme tertiary structure (red) and drUGM reduced structure (green) was superposed with Secondary structure matching (SSM) (Krissinel and Henrick, 2004). Mobile loops are highlighted in cyan for UNGM, and blue for drUGM.

participate in the stabilization of the substrate in the substrate binding cleft. Specifically, a conserved Arginine, Arg169 in UNGM, forms salt-bridge interactions with the α and β phosphates of UDP. We speculate that this loop should adopt the same conformation observed in the structure of drUGM upon substrate binding (Figure 4.8). Mobile loop 2 does not contain a region of α-helical secondary structure as observed in the substrate-bound structures of drUGM and kpUGM, although it does contain a 3-residue helical turn (164-166), annotated as a 3_{10} helix (η3) in Figure 4.7C.

4.2.2 Sequence analysis

The sequence of *C. jejuni* 11168 UNGM (*Cj1439c*) was aligned with the sequences of the UGM homologues *C. jejuni* ssp. doylei, *E. coli*, *M. tuberculosis*, *K. pneumoniae, C. jejuni* ssp. 81116, *C. jejuni* ssp. ICDCCJ07001, *D. radiodurans* using Clustalw (Thompson *et al*., 1994) (Figure 4.9). UNGM shares a high degree of conservation with UGM homologues within the active site (indicated by black asterisks in Figure 4.9). Mobile loop 2 (156-186 in UNGM) also shares a high degree of conservation with other UGMs, indicating that this loop shares a similar function in UNGM. Mobile loop 1 (123-129 in UNGM) does not appear to share any conservation with other UGMs, which is consistent with the lack of conservation and secondary structure in this region. Arg169 in UNGM is conserved in the same position on the mobile loop in every known UGM. This high degree of conservation indicates the importance of Arg169 in the function of UGM. The specificity of UNGM for both UDP-Gal and UDP-GalNAc sugars is believed to be due to the contributions of several active site residues (Table 4.2). Mutagenic studies of Arg168 to lysine have shown that this residue conferred specificity for UDP-GalNAc over UDP-Gal (Poulin *et al*., 2009). Arginine is found in this position in only *C. jejuni* ssp. 11168 and doylei, whilst lysine and alanine are more common. Mutagenesis of another active site arginine (59) has shown that this residue also contributes to the dual specificity of UNGM (Poulin *et al*., 2009). Upon comparison with its UGM homologues, histidine is found in the conserved position in the all of the UGM homologues, whilst arginine is found exclusively in the *C. jejuni* ssp. 11186 and doylei. Because *C. jejuni* doylei shares the unique residues Arg168 and Arg59, we hypothesize that the *glf* gene from this species will also encode an enzyme with bifunctional activity for UDP-Gal and UDP-GalNAc. In addition, the
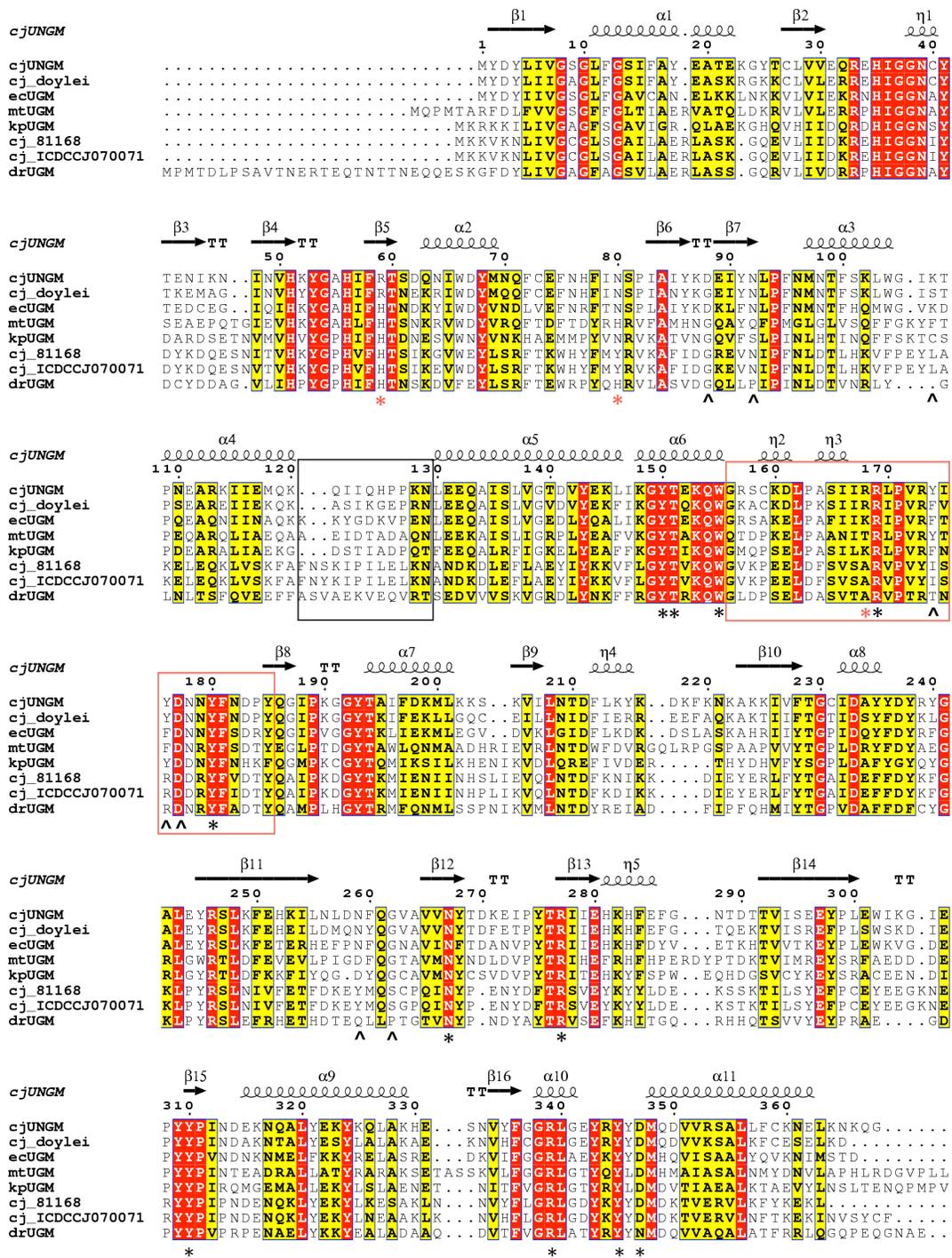
Figure 4.9: **Sequence alignment of** *C. jejuni* **11168 UNGM with UGM homologues** (top to bottom): *C. jejuni* ssp. doylei, *E. coli*, *M. tuberculosis*, *K. pneumoniae, C. jejuni* ssp. 81116, *C. jejuni* ssp. ICDCCJ07001, and *D. radiodurans.* Mobile loops 1 & 2 are indicated by black and red boxes respectively. Conserved active site residues annotated with black asterisk, non-conserved putative active site residues are annotated with red asterisk. Dimer contacts in UNGM indicated by ^. The figure was generated using EsPript (Gouet *et al.* 1999).

top 100 UGM homologues were aligned with UNGM, and arginine was absent in this position in every species other than *C. jejuni* (not shown).

Table 4.2: Candidate residues for specificity of UNGM.

Amino acid residue	Active site location	Conserved (Y/N)
Arg59	sugar-binding region	N
Asp347	sugar-binding region	Y
Arg168	diphosphate-binding region	N

4.2.3 FAD-binding region

The crystals of UNGM were yellow in color, which indicated the presence of an oxidized flavin compound. As expected, density for FAD was observed within the FAD binding region of UNGM (Figure 4.5A & B). The FAD molecule is stabilized by interactions with highly conserved residues in the αβα Rossman fold (Figure 4.8B). A pocket containing the conserved residues Gln32, Val7, Gly230, and Phe212 binds the adenine moiety of FAD. Hydrogen bonds are formed between the backbone nitrogen atoms of Gln32 and Phe212. We also observed a hydrogen bond between N6 of adenine and the side chain of Asp211. This type of bond was observed in the structure of drUGM, but not in the structure of kpUGM. The ribosyl moiety is stabilized by hydrogen bonding between Glu31 and the C2 and C3 hydroxyl groups. Additionally, 2 conserved waters also interact with these hydroxyl groups. The negatively charged diphosphate region is stabilized by interactions with the conserved Arg339, as well as the backbone nitrogen atoms of Phe12 and Asn39. The ribitol moiety is stabilized by hydrogen bonds between the backbone oxygen of Tyr346, and the side chain of His56.

The isoalloxazine ring is sandwiched between His56 and Tyr345. The same arrangement is observed in the structure of kpUGM in complex with UDP-Gal*p*, however the conserved tyrosine (Tyr370) in drUGM/UDP-Gal*p* complex is not involved in stabilization of the flavin. Rather, this tyrosine has been found to stabilize the α-phosphate of UDP-Gal*p*. We are unable to confirm if the conserved Tyr345 also has a role in this stabilization of the diphosphate of UDP, however we conclude that this is a possibility due to its position in the crystal structure. As observed in other oxidized structures, N5 of FAD is within hydrogen

bonding distance (2.9 Å) with the carbonyl group of Ala55 (Pro59 in kpUGM, Pro84 in drUGM). However, we do not expect a hydrogen bond to form because N5 would have to be protonated, which is a rare event. We also observed that O2 of FAD forms a hydrogen bond with the backbone nitrogen of Met348, which is conserved in the other structures.

The oxidized form of FAD was observed in monomer A of UNGM, which had a planar conformation in the isoalloxazine ring (Figure 4.5A). In monomer B, the reduced form of the flavin cofactor, FADH- was observed (Figure 4.5B). The isoalloxazine ring of FAD must be in the reduced or puckered conformation for the enzyme to be active. This pucker is due to N5 of FADH⁻ adopting sp³ hybridization, whereas N5 in FAD is sp² hybridized and forms a planar ring system. Therefore, the FAD cofactor in the UNGM holoenzyme structure is in the inactive conformation in monomer A, and in the catalytically active conformation in monomer B.

4.2.4 Uridine-binding region

The active site of UNGM encompasses the isoalloxazine ring of FAD and the substrate-binding region. The active site consisted mainly of conserved residues known to be involved in the binding and stabilization of substrate in UGM homologues. The UGM homologue from *D. radiodurans* shares 36 % identity with UNGM, and was used for comparative structural analysis due to the availability of crystal structures in complex with UDP-Gal*p*, UDP, and UDP-Glucose. The active site of UGM can be divided into 3 major regions: the sugar-binding region, the diphosphate-binding region, and the uridine-binding region.

The uridine binding region of UNGM was found to maintain a high degree of conservation with other structures of UGM. The uridine moiety is stabilized through hydrogen bonding with the pocket containing Trp155, Thr151, Tyr150, and Asn267 (Figure 4.8A & 4.10A). Previous modeling studies have identified a conserved tryptophan (Trp155 in UNGM) that forms π-edge stacking interactions with the uridine moiety (Yuan *et al.*, 2005). Most recently, structural analysis of this residue has identified the formation of hydrogen bonds between the indole nitrogen of tryptophan and the C2 and C3 hydroxyl groups of the ribose of uridine (Karunan Partha *et al.*, 2009). Kinetic studies have identified that this residue was critical for the binding of substrate, with a decrease of approximately 10⁶ in binding affinity (Chad *et al.*, 2007). In the position were ribose is expected to bind, Gln154 is pointed up

towards the substrate binding site in the pocket, which is consistent with other oxidized holoenzyme structures (Sanders *et al*., 2001; Gruber *et al*., 2009a). In contrast, this residue is expected to bend away from the binding pocket when the ribose is bound. We predict that there are π-stacking interactions between the conserved Tyr150 and the uracil ring, as there are no other aromatic rings present in this region to the active site as seen in other UGM structures. Furthermore, we expect that the conserved residues Thr151 and Asn267 also form hydrogen bonds with the uracil as previously observed in homologous structures (Gruber *et al*., 2009a; Karunan Partha *et al*., 2009).

4.2.5 Diphosphate binding region

UNGM shares a highly conserved diphosphate binding region, and we expect that the conserved residues bind to the diphosphate region of the substrate(s) in the previously observed manner (Figure 4.10B). UGMs contain a mobile loop (2) region that must be in a closed conformation for catalytic activity of the enzyme (Figure 4.11B). In the holoenzyme structure, the positions of these residues are in different conformations than what is seen in the substrate-bound structures of kpUGM and drUGM (drUGM is shown in Figure 4.8A). The loop contains a conserved Arginine, Arg169 (Arg198 in drUGM), which has previously been shown to be essential for mutase activity. Specifically, in kpUGM, the conserved arginine (Arg174) was mutated to alanine, which resulted in an inactive enzyme (Chad *et al*., 2007). Structurally, this arginine stabilizes the negatively charged diphosphate backbone of the sugar nucleotide substrate via the formation of a salt-bridge (Karunan Partha *et al*., 2009; Gruber *et al*., 2009a). Upon comparison of UNGM with other known structures of UGM, we have identified that Arg169 is likely to be the conserved residue essential for catalysis.

Previous work by our collaborators targeted the adjacent, non-conserved arginine, Arg168 for mutagenesis and kinetic analysis. The R168K mutants had a less pronounced effect on catalysis than expected and a 2-fold increase in selectivity for UDP-Gal*f* over UDP-Gal*f*NAc (Poulin *et al*., 2009). Also the K_m value for UDP-Gal*f*NAc increased, but was unchanged for UDP-Gal*f*, which supports that Arg168 has a role in the binding and stabilization of Gal*f*NAc. In contrast, the eukaryotic UGM from *Aspergillus fumigatus* was mutated at the conserved arginine (182) to lysine, which caused a large decrease in the catalytic activity of the enzyme, but interestingly increased the binding affinity for substrate (van Straatan et al., 2012). In the
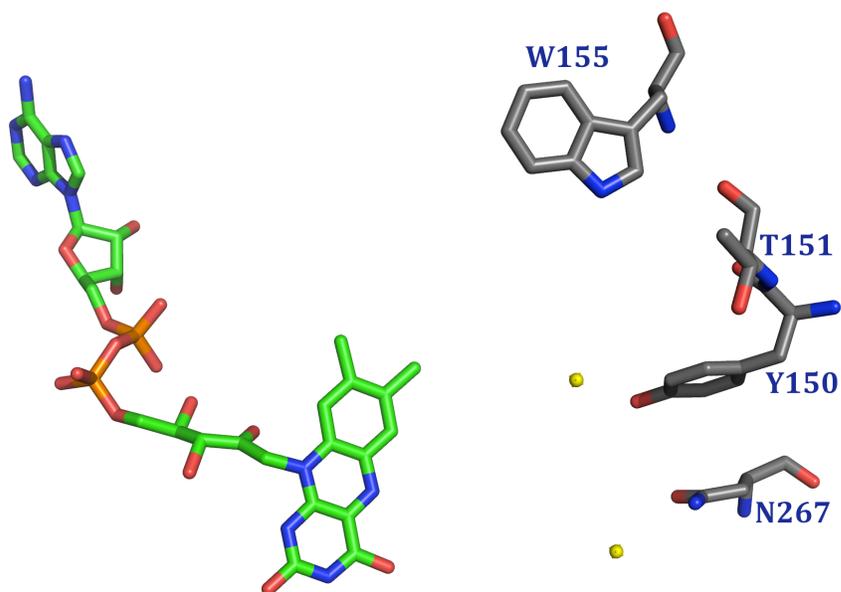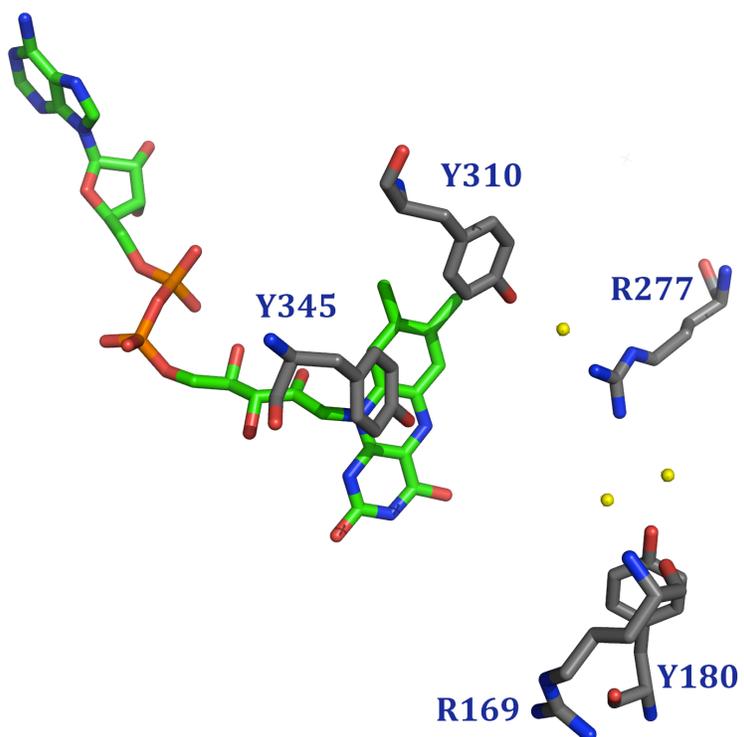
Figure 4.10: The uridine diphosphate binding region of UNGM. A- The proposed uridine binding region of UNGM. **B-** The proposed diphosphate binding region of UNGM. Amino acid residues (gray-C, red-O, blue-N); FAD (green); and bonded waters (yellow).

structure of afUGM R182K mutant (PDB 3UKK), there was a lack of density for the galactose moiety, which suggested that Arg182 is important for positioning of the sugar. Together, these studies indicate that lysine successfully stabilizes the binding of substrate in the active site, but arginine is important for orienting the sugar moiety. This role is not apparent in the holoenzyme structure of UNGM, as Arg168 adopts an open conformation, pointing away from the active site. In ecUGM, Lys169 occupies the same position as Arg168 from UNGM when the loop is in the open conformation. In the closed loop conformation of ecUGM, Lys169 is pointed toward the active site whereas the conserved arginine, Arg170 (Arg169 in UNGM), is pointing away from the substrate binding cleft (Figure 4.11C). Superposition of ecUGM and UNGM has shown that Arg168 could adopt the same conformation as Lys169 in the closed conformation, which occupies the putative acetamido binding region adjacent to Arg59. Therefore, Arg168 may function to stabilize the conformation of Arg59, which has been found to confer specificity for UDP-Gal*f*NAc, or directly interact with the acetamido group of the sugar. A second highly conserved arginine, Arg277 (Arg305 in drUGM), forms salt-bridges with the phosphate groups in a similar fashion to Arg169 (Figure 4.8A & 4.10B). In the homolog *Klebsiella pneumoniae* UGM (PDB code 2B17), site-directed mutagenesis of Arg280 (Arg277 in UNGM) to Ala rendered the enzyme inactive, which indicated that it is essential for substrate binding or stabilization (Chad *et al*. 2007). Additionally, UNGM contains several conserved tyrosine residues, Tyr345 (Tyr370 in *D. radiodurans*), Tyr310 (Tyr335 in *D. radiodurans*), and Tyr180 (Tyr209 in *D. radiodurans*), which have been previously shown to be involved in the stabilization of α and β phosphates through hydrogen bonding (Chad *et al*., 2007, Karunan Partha *et al*, 2009).

4.2.6 Sugar binding region

The sugar moiety of UDP-Gal*p* has been shown bind several conserved residues present in the sugar-binding region of the active site. In drUGM, stabilization occurs primarily through water-mediated hydrogen bonds with these residues. In UNGM, the residues Arg59, Asn80, Arg339, and Asp347 are believed to form the aforementioned hydrogen bonds with the sugar moiety (Figure 4.12A). Furthermore, the isoalloxazine ring of the flavin is believed to stabilize the sugar moiety at the C4 hydroxyl via a hydrogen bond with O4 (of the flavin). Mechanistically, we expect the formation of a FAD-Gal or FAD-GalNAc adduct.
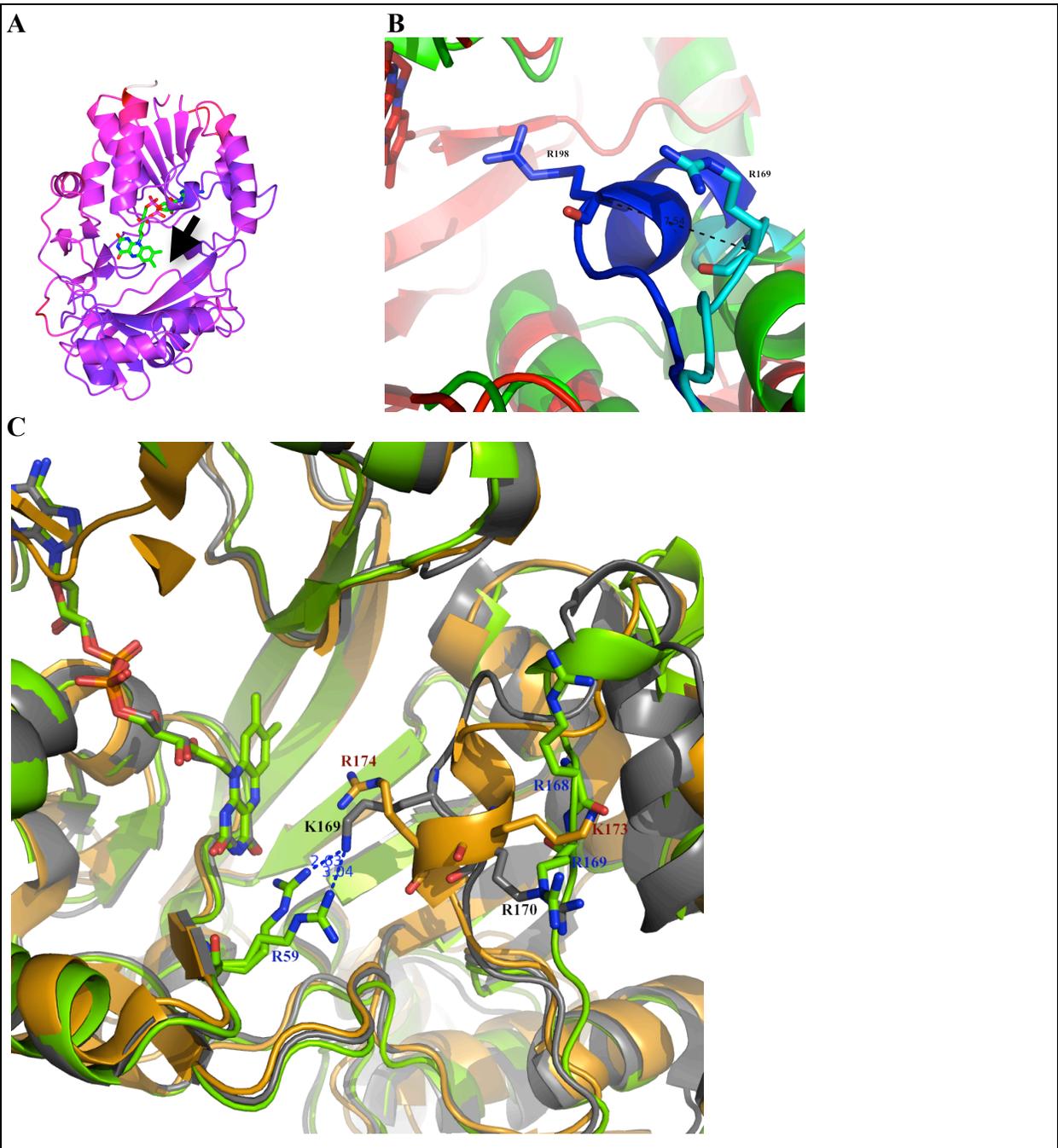
Figure 4.11: Mobile loop 2 conformations of UNGM. A- UNGM temperature factor representation. Regions with high temperature factors are shown in red, low temperature factors shown in blue. Mobile loop 2 indicated by black arrow. **B-** The mobile loop 2 of UNGM (cyan) is approximately 7 Å from the position of the conserved loop in drUGM reduced structure (navy) (PDB 3HDY). **C-** The putative acetamido binding region. In ecUGM (gray; black label) mobile loop 2 in a non-productive closed conformation; Lys169 (R168 in UNGM; K173 in kpUGM) occupies putative acetamido binding region near Arg59 (shown in 2 alternate conformations) in UNGM (green; blue label). KpUGM (orange; red label) is in the productive closed conformation with the conserved R174 (R169 in UNGM; R170 in ecUGM) facing the active site.

The bifunctional activity of UNGM is thought to depend heavily on the interaction of the *N*-acetylated sugar moieties with Arg59. The putative acetamido binding region is thought to contain two non-conserved residues, Arg59 and Arg168 (Figure 4.11C). Arg59 (His88 in drUGM) is speculated to stabilize the binding of the acetamido moiety of UDP-GalNAc (Figure 4.12). In monomer B, Arg59 exists in two conformations; one faces the active site in the sugar-binding region, and the other points away from the active site. This monomer also contains the reduced form of the flavin cofactor (Figure 4.5A). When Arg59 faces the active site, the guanidinium group is 2.9 Å from N3 of the isoalloxazine ring. In the other monomer, the oxidized isoalloxazine ring (N3) is 3.1 Å from Arg59. In both monomers, the flavin appears to stabilize the active conformation of Arg59. Arginine is a bulkier residue than histidine, and it is able to project further into the sugar-binding site. Furthermore, site-directed mutagenesis of Arg59 to His resulted a 3-fold increase in k_{cat} for UDP-Gal*f* and a 4-fold decrease in k_{cat} for UDP-GalNAc (Poulin *et al*., 2009). The value of K_m did not significantly change between these substrates when compared to the changes in k_{cat}. These results indicate that Arg59 is an essential residue for stabilization of *N*-acetylated moieties. Although Arg59 is important for this specificity, the R59H mutant was still able to recognize and catalyze the inter-conversion between UDP-Gal*p*NAc and UDP-Gal*f*NAc, and therefore other residues or structural changes play a role.

Other differences were observed when we compared the sugar-binding region of drUGM with UNGM. The residue Asp347 (Asn372 in drUGM) is adjacent to Arg59, and forms a hydrogen bond with the positively-charged Arg59, forcing it into its alternate conformation facing towards the active site (Figure 4.12B). In the holoenzyme structure, the distance between the carboxylic acid of Asp347 and the guanidinium group of Arg59 is 3.5 Å, which is an acceptable distance for such an interaction. Also, a water molecule is within hydrogen-bonding distance (3.1 Å) of the guanidinium group, which may also stabilize this conformation. This observation, when coupled with the putative acetamido moiety stabilization by Arg168 from mobile loop 2, is consistent with a synergistic role for these amino acids based on mutagenic studies (Figure 4.11C) (Poulin *et al*., 2009).
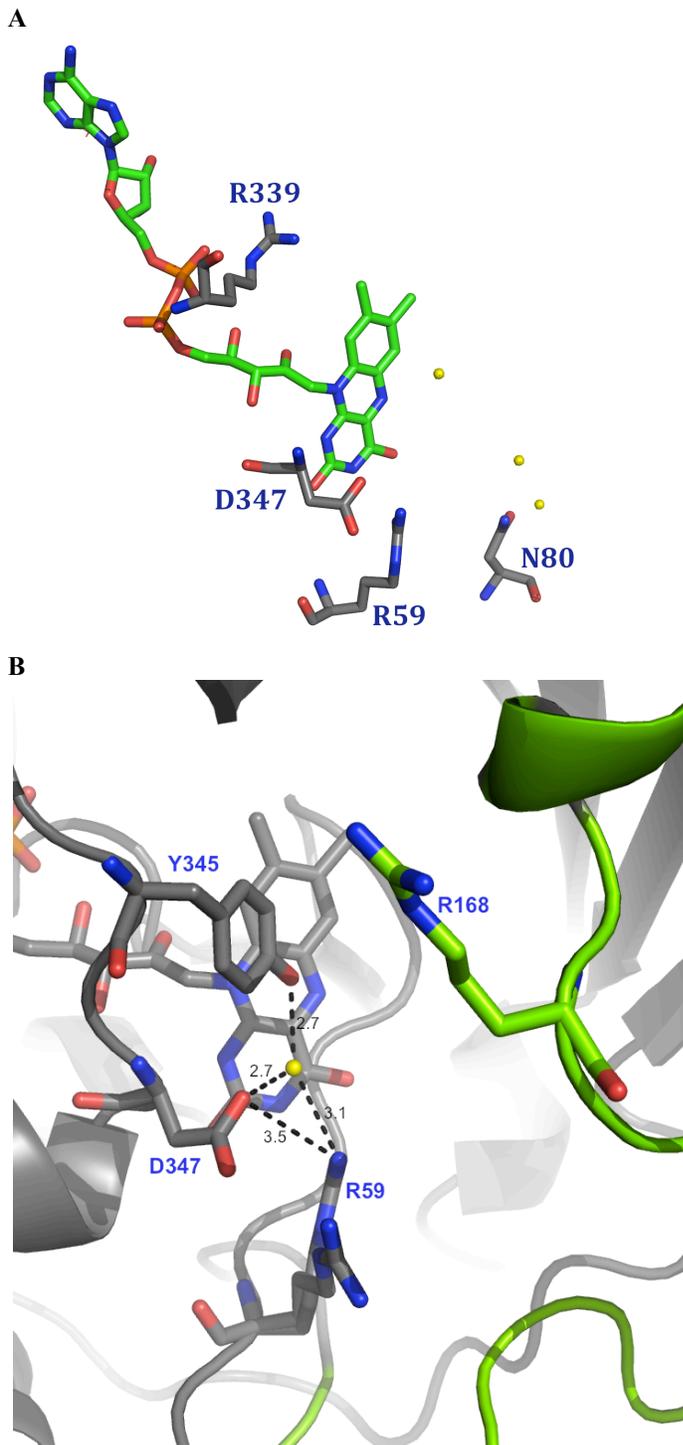
Figure 4.12: The sugar binding region of UNGM. **A-** The sugar binding residues of UNGM; residues are highlighted in gray-C, red-O, blue-N; FAD in green; bonded waters in yellow. **B-** The acetamido binding region. One conformation of Arg59 (facing the sugar binding site) is stabilized by hydrogen bonding with Asp347 and a water molecule. Tyr345 is also within hydrogen-bonding distance with this water molecule. Residues are highlighted in gray, mobile loop 2 shown in green.

4.2.7 Crystal contacts

The crystal packing between asymmetric units of UNGM is due to intermolecular and inter-dimer contacts. These interactions are important for the stabilization of the crystal packing, which allows for order within the crystal. The quaternary structure of UNGM is reliant on the interactions between the following residues at the dimer interface: Tyr176 with Gly262, Asn259 with Tyr174, Lys107 with Asp88, Asp178 with its symmetry-related residue and Asn92 with its symmetry-related residue (Figure 4.9). Along with these interactions, water molecules and several solvent molecules assist with stabilization of this domain (discussed in Section 4.2.8).

Intermolecular contacts are observed between residues on the external surface of the protein molecules (Figure 4.13). The exterior surface charge contains two large regions of positively-charged residues, as seen in the active site cleft and the upper left region of Fig. 4.13A. The positive charge of the active site is beneficial for binding of the negatively charged diphosphate region of the substrate. Intermolecular contacts occur the between the conserved arginine (Arg169) of mobile loop 2 and Asp290 at a distance of 3.0 Å, forming a salt-bridge (Figure 4.13C). In contrast, Arg169 in the other monomer does not make this contact; Asp290 is 6.8 Å away, which is too large a distance for salt-bridge formation. Therefore, we conclude that mobile loop 2 is in different positions within each monomer.

The other region of positive-charge on the exterior (left upper region) appears to function in forming intermolecular crystal contacts (Figure 4.13A). The intermolecular contact region largely occurs within the range of residues 215-224 in both monomers. Tyr215 forms a hydrogen bond with the backbone oxygen of Trp104 at a distance of 2.8 Å. Tyr215 in the second monomer forms a salt-bridge with Asp161, and a distance of 2.5 Å separates the two residues. In the other monomer Asp161 also forms a dipole interaction with Asn333 a distance of 2.4 Å (Figure 4.13B). The residue Lys224 forms a salt-bridge with Glu152 at a distance of 3.0 Å. Lys224 and Lys160 both form hydrogen bonds with a sulfate ion at this interface.
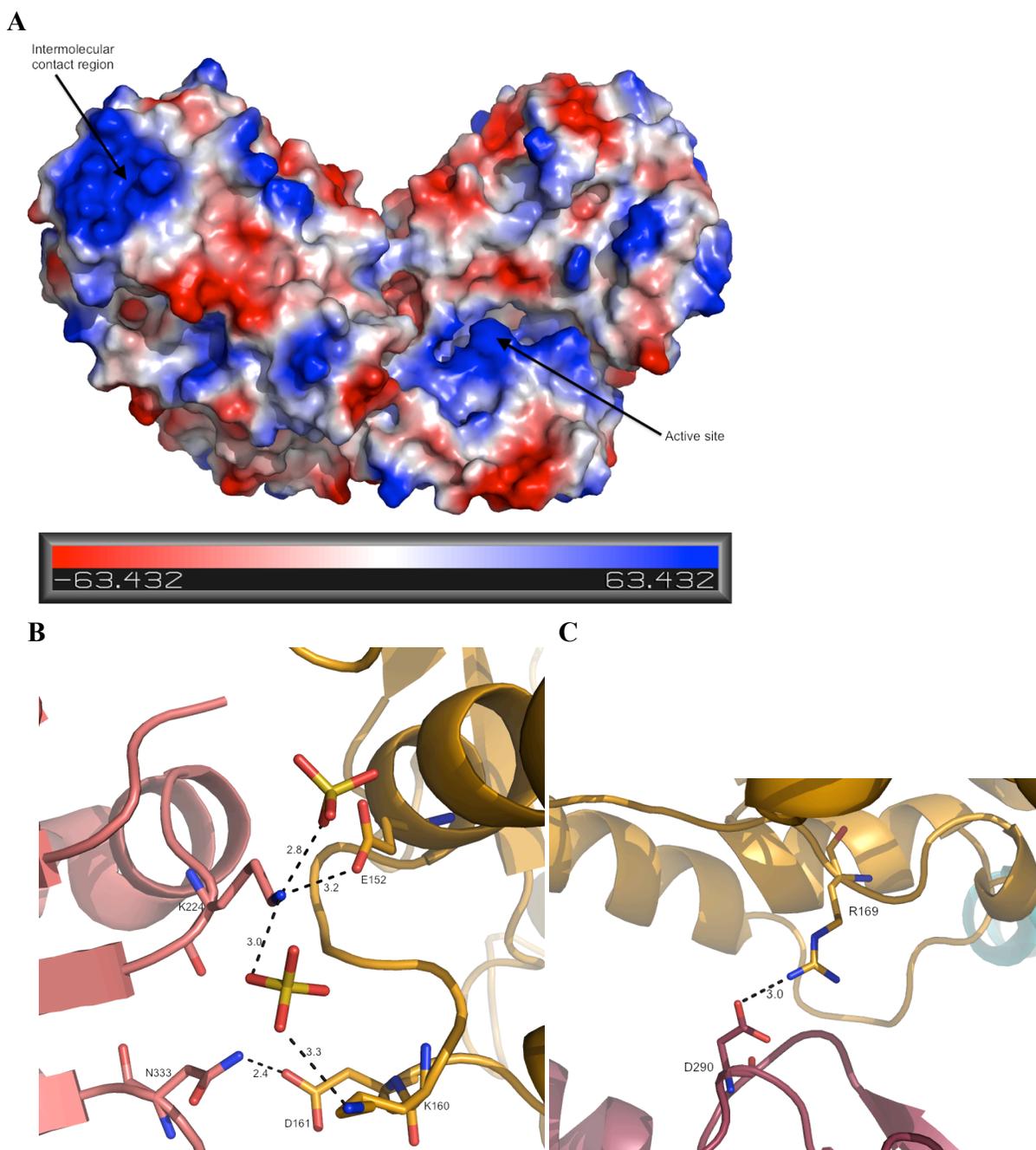
Figure 4.13: The surface view and intermolecular contacts of UNGM holoenzyme. A- The exterior surface charge of UNGM has near equal positive (blue) and negative (red) charges visible on the protein exterior. The active site of UNGM is largely positive, which promotes the binding of the negatively-charged substrate. Figure generated with Pymol. **B-** Two sulfate ions are found at a intermolecular interface. **C-** Mobile loop 2 (Arg169) forms a salt-bridge with Asp290 of another molecule.

4.2.8 Solvent molecules

Within the active site, there are several water molecules that appear to form hydrogen bonds with the conserved residues in place of a substrate. Within the sugar binding region, Arg59 and Asn80 both coordinate water molecules which are bound in place of the sugar moiety (Figure 4.12B). The water bonded to Arg59 is common to other UGM structures, although the interaction is with Histidine in every other known structure. Also, Asn80 commonly bonds to a water molecule in UGM holoenzyme structures, however this interaction does not occur in the UDP-Gal*p* complex structures of kpUGM and drUGM (PDB codes 3INT and 3HDY). The diphosphate binding region also contains several bonded waters with the conserved tyrosine residues Tyr180, Tyr310, and Tyr345, as well as Arg277 (Figure 4.14A). Tyr180 forms a hydrogen bond with a water molecule, and this interaction is observed in the structures of ecUGM holoenzyme and drUGM:UDP-Gal*p*. Tyr310 binds to a water molecule instead of the β-phosphate as observed in UGM complex structures. Also, this interaction was not observed in any of the holoenzyme structures. Tyr345 interacts with a water molecule, which is also bound by Arg59 and Asp347 (Figure 4.12B). This interaction was not observed in any other the other UGM structures. Arg277 is bonded to a water molecule, which is also coordinated by Tyr310 (Figure 4.14A). Waters are present in this region in the holoenzyme structures and in the kpUGM:UDP-Gal*p* structure, however none have a water molecule that interacts with both residues. Within the uridine binding region, Tyr150 forms a hydrogen bond with a water molecule in every known UGM structure. Asn267 bonds a water molecule in place of the uracil ring of UDP in the holoenzyme structures (Figure 4.14B).

Several glycerol molecules were observed within the structure of UNGM. Glycerol is commonly used as a cryoprotectant, and has been found to stabilize protein structures in disordered regions and enhance diffraction (Sousa *et al*., 1990; Rould *et al*., 1991). The addition of glycerol should stabilize the crystals by preventing ice nucleation within the crystal. The cryoprotectant contained 15 % glycerol, which likely displaced water or other solvent molecules present during crystallization. Two glycerol molecules were observed at the solvent-protein interface, and do not appear to aid in crystal packing. At the dimer interface, there are two glycerol molecules that interact with the ε-amino group of Lys102 in both monomers at a distance of 3.2 Å (Figure 4.14C).

Also present in the structure are several ammonium ions that all appear to form hydrogen bonds with water molecules. All of the ammonium ions are present at the solvent-protein interface, with the exception of one molecule within the active site of monomer A, which interacts with Arg277 and Asn80 through hydrogen bonding. Furthermore, 2 sulfate ions are present at the intermolecular interface of two monomers (Figure 4.14D). The sulfate ions form hydrogen bonds with Lys224 in both structures, and this interaction contributes to the stabilization of crystal packing.

4.2.9 Differences between monomers

Significant differences were observed between monomer A and B of the UNGM homodimer using the Non-crystallographic Symmetry (NCS) differences function in Coot (Emsley and Cowtan, 2004). Therefore, NCS restraints were not used during the refinement of the model, and each monomer was treated as a separate entity. The sugar-binding residue Arg59 existed in two conformations in monomer B, whilst in monomer A, Arg59 exists in a single conformation. Differences were also observed in the mobile loops, which are known to affect crystal packing in this structure. Mobile Loop 1 deviated between monomer A and monomer B by a distance of 3.1 Å. Specifically, Gln121 of mobile loop 1 of monomer A is 3.8 Å from Asn110 of the adjacent monomer B. However, Gln121 in monomer B is 3.1 Å from Lys114 of the adjacent monomer A. Therefore, mobile loop 1 is in different positions within each monomer of the dimer. Mobile loop 2 differed in the positions of monomer A and B by a distance of 2.6 Å. The differences in the mobile loops are consistent with the elevated B factors of the residues, indicating that there is a higher degree of flexibility in these regions. The loops are important for crystal contacts, and are therefore we do not expect substrate to bind successfully in a crystal form that restricts movement of the mobile loops. The range of residues 215-224 differed by 3.6 Å between monomer A and B. This region was found to be very important in the formation of crystal contacts. Therefore, we expect that packing of adjacent molecules altered the arrangement of this region.
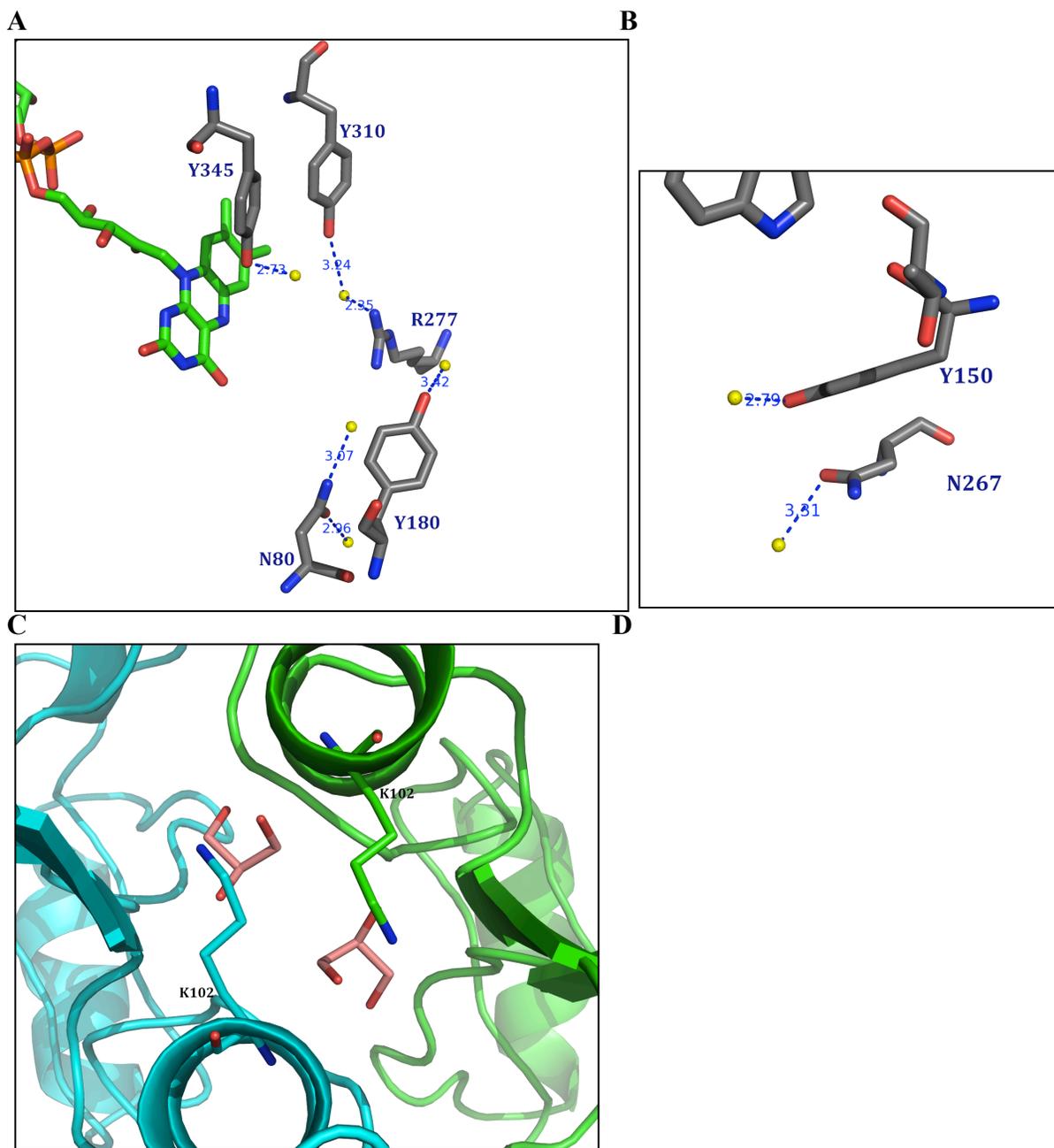
Figure 4.14: Solvent molecules present in the UNGM holoenzyme structure. A- Bonded waters (yellow) in the sugar and diphosphate binding region. **B-** Bonded waters (yellow) found in the uridine binding region. **C-** Two glycerol molecules are bound at the dimer interface. Chain A and B are shown in green and cyan respectively, glycerol molecules are highlighted in pink. **D-** Two sulfate ions are found at the intermolecular interface which function to stabilize the crystal packing.

4.3 Structure of UNGM soaked in UDP-5-deoxy-Gal*f*

4.3.1 Crystallization and soaking experiments

The majority of the holoenzyme crystals were not compatible with soaking experiments. Dissolution and cracking was observed in most of the crystals, which may indicate that the binding of substrate caused movement within the crystal and weakened the crystal packing. Another hypothesis is that the crystals were unstable prior to soaking experiments, and that any change in buffer concentration or environment would also produce the same results. The best crystals were grown in condition #67 from PEGs II suite containing 0.1 M Sodium acetate, 0.2 M Lithium sulfate, 0.1 M HEPES pH 7.5, 25 % (w/v) PEG 4000 (Figure 4.15A). These plate crystals did not dissolve after the addition of the substrate analog UDP-5d-Gal*f* and sodium dithionite (Figure 4.15E).
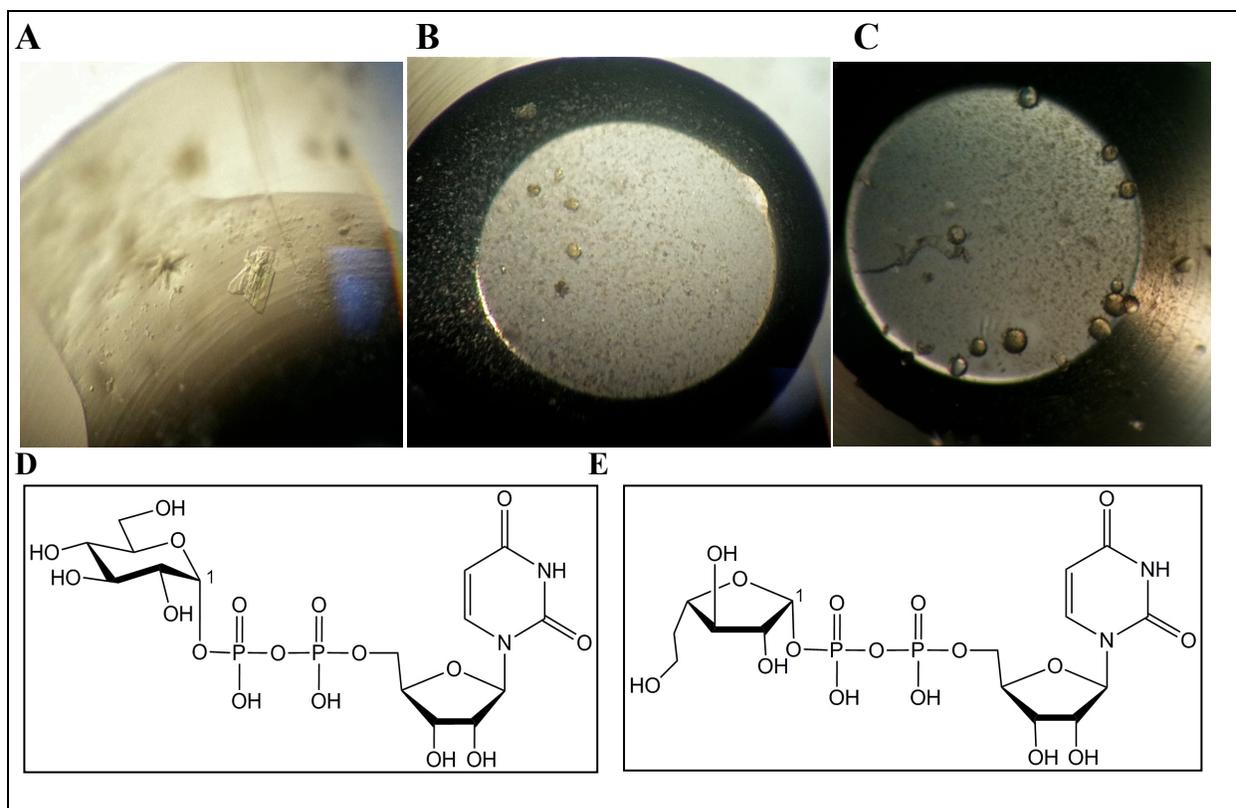
Figure 4.15: Crystal morphologies for UNGM and UNGM:UDP-Gal*p*. **A-** Plate-like crystals of UNGM grown in #67 from PEGs II suite, used for soaking-in experiments. **B-** Co-crystallization of UNGM with UDP-Gal*p*, their yellow color is indicative of flavin bound within UNGM. **C-** Quasi-crystals of UNGM co-crystallized with UDP-Gal*p*, grown from #11 of ComPAS suite. **D-** Structure of UDP-Glucose (UDP-Glc). **E-** UDP-5-deoxy-galactofuranose (UDP-5-deoxy-Gal*f*). Figures generated in ChemDraw (Mills, 2006).

4.3.2 Diffraction and Data Processing

The crystals were diffracted to a resolution of 3.3 Å. The data belonged to the space group $P2_1$ with the unit cell parameters (Å): a = 74.53, b = 48.2, c = 122.2, β = 95.2°. The data was indexed, integrated and scaled using d*TREK to a resolution of 3.3 Å (Pflugrath, 1999). Data processing was initiated using the "Find" function at sigma cutoff of 3, which gave a sufficient number of reflections to be indexed. After indexing, the Bravais lattice type with the lowest least-square residual (%) and the highest symmetry was primitive monoclinic. The data was refined to fit most of the reflections, which appeared to be successful with most of the selections were on the intense reflections and a low number of red circles observed (indicative of error in the choice of point group). The integration of the data was successful with the majority of the diffraction sitting within the predicted lattice for the space group, and the unit cell parameters were similar to the values observed prior to integration. The data was scaled to 3.3 Å based on the statistics in the highest resolution shell. The R_{merge} was 33.4 % and $I/\sigma(I)$ of 2.21 with 91.8 % completeness in the highest resolution shell. Data collection statistics are summarized in Table 4.3.

4.3.3 Structure Solution and Refinement

The structure was solved by molecular replacement using the holoenzyme structure of UNGM using Molrep (Vagin and Teplyakov, 1997). The model appeared to correlate to the electron density despite the low resolution of the data. Refinement of the model was performed using phenix.refine (Afonine, 2005) on the following parameters: group B-factors, torsion angle restraints, secondary structure restraints, and simulated annealing. In Coot, real-space refinement and regularization was used for modeling poorly fit regions. The starting values of R_{work} and R_{free} were 0.5112 and 0.5102 respectively. The FAD cofactor was manually modeled to fit the $|F_o|-|F_c|$ map. Preliminary refinement resulted in R factor values of 0.2795 (R_{work}) and 0.3376 (R_{free}). Refinement statistics are summarized in Table 4.3. The structure was not further refined due to the low resolution of the data. The C-α backbone of the model was observed to fit the electron density in the $2|F_o|-|F_c|$ map; however density was missing for several side chains. The r.m.s bonds and angles remained high after initial refinement, and the model would need to be manually re-built to generate more acceptable values. This structure is not publishable at this resolution and quality. However, the solution was used to distinguish if

substrate was bound within the active site. Upon analysis of the structure, we discovered that there was no density observed for UDP-5d-Gal*f* within the substrate-binding site (Figure 4.16). We expect a region of positive density (in the $|F_o|$-$|F_c|$ map) directly adjacent to the isoalloxazine ring of FAD if the substrate were bound within the substrate-binding region. The difference map was contoured to 1.0 σ to show the absence of density for the substrate. Density in the $|F_o|$-$|F_c|$ for this region corresponded to water molecules only. This result indicated that soaking experiments alone were likely insufficient to produce a structure in the active substrate-bound form.
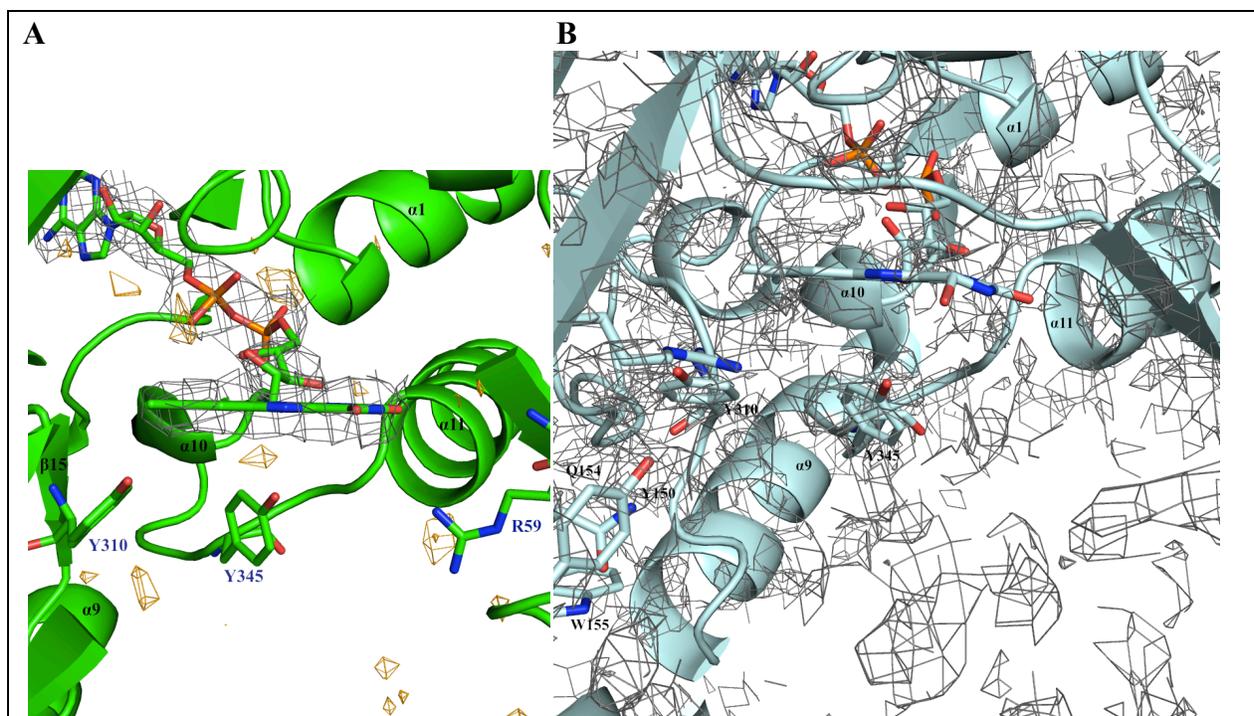
Figure 4.16: The active site of UNGM crystals soaked in UDP-5d-Gal*f*. **A-** The $2|F_o|$-$|F_c|$ map (grey; contoured to 1.0 σ) harbored density for the FAD cofactor in the oxidized conformation in both monomers. The $|F_o|$-$|F_c|$ map is shown in orange, contoured to 3.0 σ. **B-** The $|F_o|$-$|F_c|$ map does not show density for UDP-5d-Gal*f*. If substrate analog was bound to the substrate binding site (highlighted by residues in stick representation), a large region of positive density is expected in the $|F_o|$-$|F_c|$ map (in gray; contoured to 1.0 σ), however no electron density in this region is large enough to encompass this molecule.

Table 4.3: Data collection and refinement statistics for UNGM with UDP-5-deoxy-Gal*f*

Data Collection:

Space group	P2$_1$
Unit-cell parameters (Å)	a = 74.53 b = 48.2 c =122.2 β = 95.2°
Wavelength (Å)	0.9790
Temperature (K)	100
Resolution range (Å)	30-3.3 (3.42-3.30)
Matthews coefficient (A^3 Da^{-1})	2.34
Solvent content (%)	47
No. of molecules in ASU	2
No. of measured reflections	24524
Total No. of unique reflections	11721
Completeness (%)	87.7 (91.8)
Redundancy	2.09 (2.07)
R$_{merge}$†	0.177 (0.334)
$<I/\sigma(I)>$	3.3 (2.1)
Refinement:	
R$_{work}$‡	0.2795
R$_{free}$§	0.3376
RMSD bond lengths	0.025
RMSD bond angles	2.269

†$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where I$_i$(*hkl*) are the individual intensities of the *i*th observation of reflection *hkl* and $<I(hkl)>$ is the average intensity of reflection *hkl* with summation over all data. ‡ $R_{work} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}|| /\Sigma_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. § R$_{free}$ is equivalent to the R_{work} but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.

4.4 Co-crystallization of UNGM

4.4.1 Co-crystallization and Diffraction of UNGM with UDP-Gal*p*

Co-crystallization screening of UNGM with UDP-Gal*p* yielded crystals in several conditions, including Classics Suite #96, containing 12 % (w/v) PEG 20000, 0.1 M MES pH 6.5, mixed in a 1:1 ratio with 6 mg/mL UNGM in 20 mM Tris HCl pH 7.6, 10 mM UDP-Gal*p*, 10 mM Sodium dithionite (Figure 4.15B). Quasi-like crystals were also observed in ComPAs #11, containing 20 % PEG 8000, 0.1 M MES pH 6.5, and 0.2 M Magnesium acetate (Figure 4.15C). Crystals were also yellow in color; however the partial oxidation of the crystals is expected to occur over time as sodium dithionite loses its ability to reduce FAD. The crystals were dyed with IZIT protein dye (Hampton Research) to confirm that the crystals were most likely not salt. The crystals were dyed successfully, indicating that solvent channels existed within the crystals, which is a phenomenon observed in protein crystals and not salt crystals.

The crystals were screened for diffraction with synchrotron radiation at the Canadian Macromolecular Crystallography Facility (CMCF-1) beamline (08ID-1) at the Canadian Light Source (Saskatoon, Saskatchewan, Canada). Unfortunately, co-crystallization of UNGM with UDP-Gal*p* did not yield diffraction quality crystals; we did not observe any diffraction using synchrotron radiation. The yellow color of the crystals is indicative of UNGM, but the crystals observed were not well ordered as indicated by the round morphology. Furthermore, the round morphology was observed in several conditions, which contained different crystallization buffers, salts, and precipitants. Optimization of these conditions did not improve the morphology or diffraction of the crystals. Therefore, the presence of UDP-Gal*p* within the active site of these crystals is unknown.

4.4.2 Structure of UNGM PEGs II #71, crystallized with UDP-Glucose

4.4.2.1. Crystallization

Although UDP-Glc is not a substrate for UNGM, it very closely resembles the natural substrates UDP-Gal*p* and UDP-GalNAc (Figure 4.17C). Binding affinity data for kpUGM (which shares 40 % identity with UNGM) indicated that UDP-Glucose (UDP-Glc) had a K_d of 750 ± 250 μM, which is only 2-fold weaker binding when compared to UDP-Gal*p* ($K_d = 290 \pm$

40 µM for the oxidized enzyme and K_d = 52 µM for the reduced), as these molecules differ at the C4 hydroxyl group (Gruber *et al.*, 2009a). Furthermore, K_M can be used as an approximation of K_d (discussed in Section 4.11). UNGM approximately has a binding affinity of 40 µM for UDP-Gal*p* in the reduced enzyme, which is nearly identical to the K_d of kpUGM (52 µM) (Poulin *et al.*, 2009). Therefore, UDP-Glc should serve as a ligand for UNGM. Structures of UGM in complex with UDP and UDP-Gal*p* have shown that the majority of substrate binding occurs via interactions with the UDP moiety of the substrate and the UDP-binding region (Karunan Partha *et al.*, 2009). As UGM and UNGM share a highly conserved UDP-binding region, the sugar nucleotide should also bind UNGM with a similar affinity. Co-crystallization of UNGM with UDP-Glc in absolute configuration (UDP-D-glucose) was attempted without the addition of sodium dithionite. Complex crystals of kpUGM with an excess of UDP-Glc (5 mM to 50 mM) were observed to grow without a requirement for sodium dithionite (Gruber *et al.*, 2009a). Additionally, the crystals were soaked in a high concentration of UDP-Gal*p* prior to freezing, which yielded a structure with UDP-Gal*p* bound within the active site (Gruber *et al.*, 2009b). We hypothesized that a crystallization condition for UNGM in complex with a substrate would also allow for other substrates to bind and co-crystallize in the same environment. Crystallization screens were checked the daily after set up, and after 5 d, pale yellow rod clusters were observed in the PEGs II #71 (0.2 M Ammonium sulfate, 0.1 M Sodium acetate-acetate pH 5.6, 30 % PEG 4000) (Figure 4.17A). Importantly, crystal growth was only observed in this condition when UDP-Glucose was also present (Figure 4.15D).

4.4.2.2. Diffraction and Data Processing

The best-diffracting crystals from the condition PEGs II #71 were soaked in a cryoprotectant containing 30 % glycerol and 10 mM UDP-Glc for 30 min prior to flash-cooling in liquid nitrogen. The data was indexed, integrated and scaled as previously described in section 4.3.2, to a resolution of 3.29 Å. The intensity of the data was at this resolution was sufficient for solving the structure. The data belonged to the space group $P2_1$, with unit cell parameters (Å) of a = 73.7, b = 48.0, c = 119.8, β = 96.5°. The Matthews coefficient of 2.42 Å^3 Da^{-1} indicated 2 molecules per asymmetric unit. The structure was primarily used to determine if this crystal contained UDP-glucose in the active site, and is not at a high enough resolution to publish.
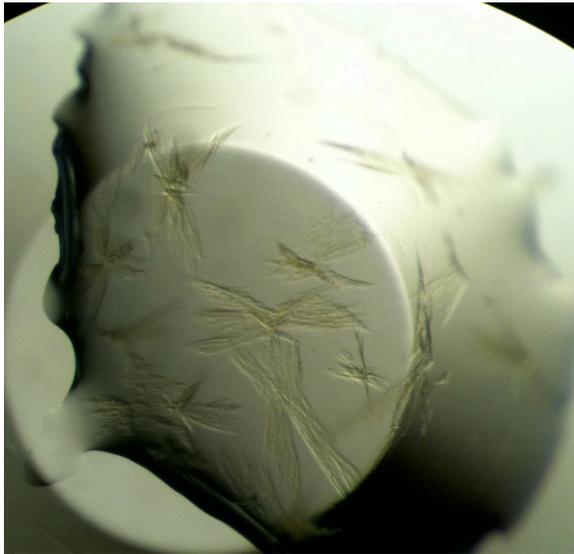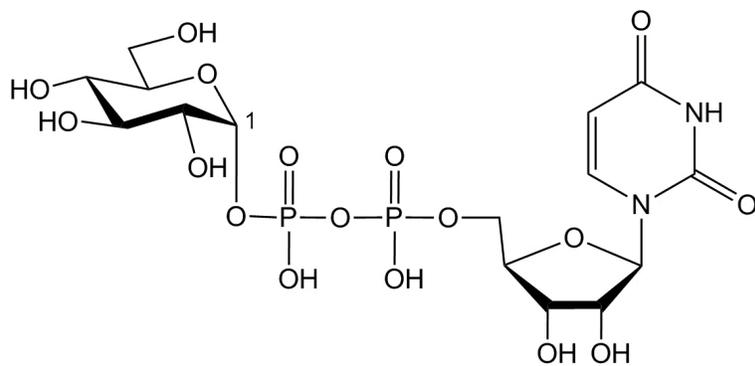
Figure 4.17: Co-crystals of UNGM grown in the presence of UDP-Glc. Crystals grown in PEGs II #71 (**A**) and ComPAs #13 (**B**) both yielded data sets. **C**- UDP-Glucose

4.4.2.3 Structure Solution and Refinement

A molecular replacement solution was found using the holoenzyme UNGM structure as a model. Refinement was performed on the following parameters: group B-factors, rigid body refinement, torsion angle refinement and real-space refinement. The starting values for R_{work} and R_{free} were 0.4521 and 0.4423 respectively. The final $R_{work} = 0.2244$ and $R_{free} = 0.2942$. Refinement statistics are summarized in Table 4.4. The $|F_o|$-$|F_c|$ map (contoured to 3.0 σ), identified electron density for the FAD cofactor but did not show the presence of UDP-Glc in either monomer of UNGM (Figure 4.18B). Density in the $|F_o|$-$|F_c|$ map (3.0 σ) for several water molecules were bonded by active site residues in place of a substrate; this was consistent with the other holoenzyme structures of UNGM (Figure 4.18A). Tyr345 and Tyr310 make hydrogen bonds with the water molecules in absence of a substrate.
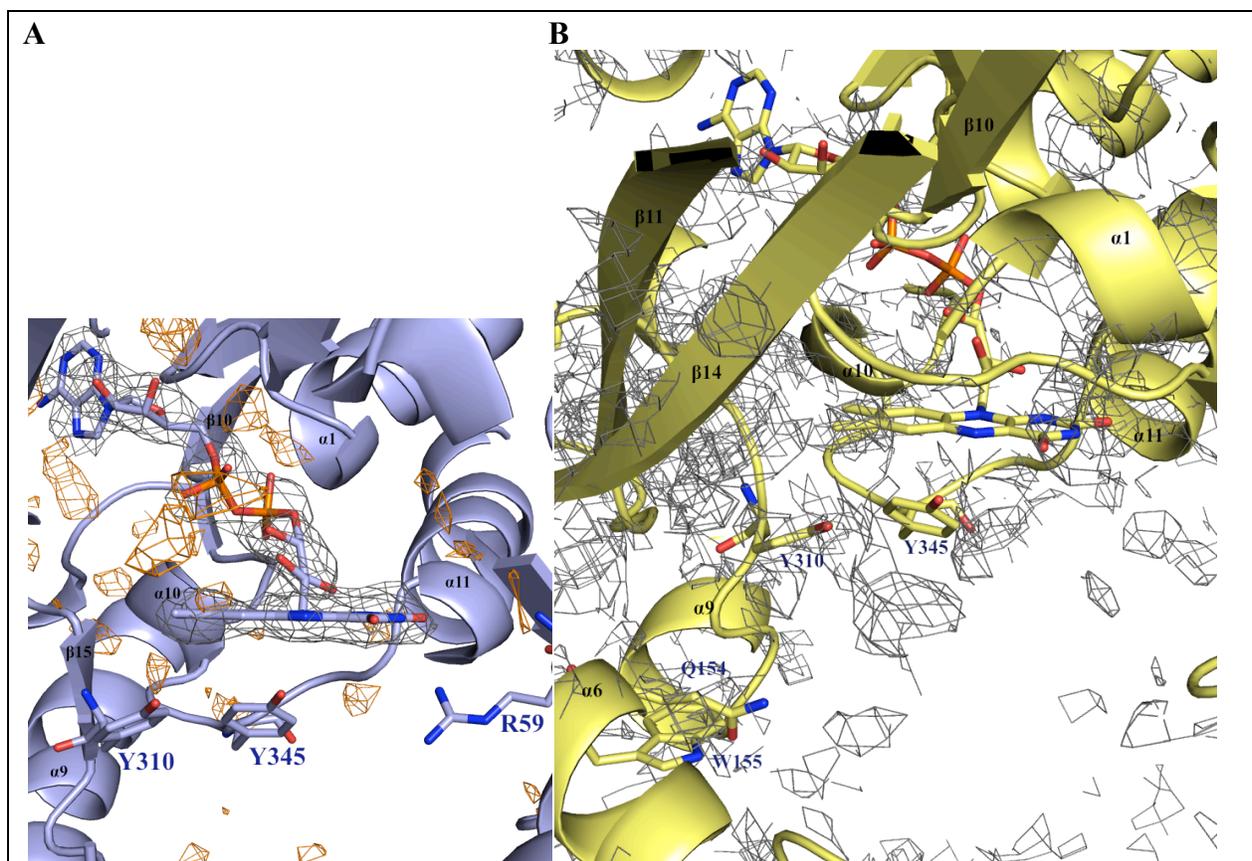
Figure 4.18: Active site of UNGM:UDP-Glc co-crystals from PEGs II #71. A- Density for the oxidized FAD cofactor was observed in the $2|F_o|$-$|F_c|$ map (contoured to 1.0 σ (gray) and the $|F_o|$-$|F_c|$ map was contoured to 3.0 σ (orange). **B-** Electron density was not observed for UDP-glucose in the $|F_o|$-$|F_c|$ map (gray), contoured to 1.0 σ.

Table 4.4: Data collection and refinement statistics for UNGM with UDP-Glc (PEGs II #71).

Data Collection:

Space group	P2$_1$
Unit-cell parameters (Å)	a = 73.7 b = 48.0 c = 119.8 β = 96.5°
Wavelength (Å)	0.9790
Temperature (K)	100
Resolution range (Å)	44.5-3.29 (3.41-3.29)
Matthews coefficient (Å^3 Da^{-1})	2.42
Solvent content (%)	49
No. of molecules in ASU	2
No. of measured reflections	48369
Total No. of unique reflections	12297
Completeness (%)	99.8 (100)
Redundancy	3.07 (3.23)
R$_{merge}$ †	0.25 (0.45)
<I/σ(I)>	7.15 (4.29)
Refinement:	
R$_{work}$‡	0.2876
R$_{free}$§	0.3382
R.m.s.d. from ideal values	
Bond lengths (Å)	0.025
Bond angles (°)	1.254

†$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where I$_i$(hkl) are the individual intensities of the ith observation of reflection *hkl* and <I(*hkl*)> is the average intensity of reflection *hkl* with summation over all data. ‡ $R_{work} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}||/\Sigma_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. § R$_{free}$ is equivalent to the R_{work} but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.

4.4.3. Structure of UNGM ComPAs #13, crystallized with UDP-Glucose

4.4.3.1 Crystallization and Soaking Experiments

Crystallization screens produced yellow plate cluster crystals that grew within 24 h, with the largest dimension of 0.25 mm. The crystals grew from the solution of 20 mM Tris pH 7.6 and ComPAs suite #13 which included: 0.2 M Ammonium sulfate, 0.1 M MES sodium salt pH 6.5, 22 % (w/v) PEG 8000 (Figure 4.17B). Crystal growth was only observed in this condition when UDP-Glucose was also present in the crystallization buffer. Microseeding produced the best crystals, grown with the 1:10 and 1:50 seeding solutions. This technique generated larger crystals and better diffraction than previously observed for this condition. The crystals dissolved after addition of sodium dithionite; therefore this technique was not used to reduce the crystals. The soaking experiments with UDP-Gal*p*, UDP-5d-Gal*f*, and UDP-Gal*f*NAc did not dissolve the crystals.

4.4.3.2 Diffraction and Data Processing

All of the crystals were diffracted using synchrotron radiation, however only UDP-Glc crystals were of sufficient quality for data collection. All other substrates appeared to decrease the overall resolution of diffraction as well as an increase in the mosaicity of the diffraction, which indicated an increase in disorder within the crystals. Importantly, this observation allows for the assumption that the substrate binds within the crystals, and the mobile loop movement is likely a cause of disorder in the crystal packing.

The best crystals were grown in the condition ComPAs #13 were soaked in a cryoprotectant containing a final concentration of 5 mM UDP-Glc prior to diffraction (Figure 4.19A). The intensity data were processed using HKL2000 (Otwinowski and Minor, 1997). The reflections were indexed at an appropriate sigma level such that the majority of reflections were selected. After indexing, the Bravais lattice was determined to be primitive orthorhombic, P222. The Chi^2 values were examined for each reflection and diffraction images with values higher than 1.0 were excluded from further refinement (Figure 4.19B). Therefore, 120 images were used for refinement and integration, with the mosaicity kept constant at a value of 1.3. The data was scaled to a resolution of 2.40 Å. The highest resolution shell (2.49-2.40) had an I/σ(I) of 3.22, which is sufficient to distinguish the data from the background noise. The

highest resolution shell had a large value for R_{merge} (47.1 %), and remained high overall (0.159). R_{merge} did not a dramatically decrease from the previously processed data (56 %), where all 180 images were included. The large value of R_{merge} was consistent with data sets from other crystals of UNGM. The data belonged to the space group $P2_12_12_1$ with the unit cell parameters a = 47.8 b = 119.3 c = 147.3. The Matthews coefficient was 2.47 $Å^3$ Da^{-1} with 2 molecules present in the asymmetric unit. Data collection statistics are summarized in Table 4.5.

4.4.3.3 Structure Solving and Refinement

A molecular replacement solution was found using the UNGM holoenzyme structure. The electron density was observed to fit the generated model well, with the crystal packing between monomers consistent with what was observed with the UNGM holoenzyme structure in the same space group $P2_12_12_1$. Initially, the R_{work} and R_{free} equaled 0.4881 and 0.4983 respectively. The initial refinement was performed on the following parameters: xyz coordinates, individual B-factors, occupancies, real-space refinement, and rigid body refinement. Simulated annealing was performed at a starting temperature of 3000 for 50 cycles to a final temperature of 300. Additionally, the manual building and regularization of disordered regions was performed in Coot. Final refinement was performed on xyz coordinates, real-space refinement, occupancies, and individual B-factors. The final values for R_{work} = 0.2232 and R_{free} = 0.2813. Refinement statistics are summarized in Table 4.5. Within the active site, the $|F_o|-|F_c|$ map (contoured to 3.0 σ) clearly showed density for the FAD cofactor, however there was no density observed for UDP-Glc as we had expected (Figure 4.19C & D). Once again, only water molecules were present in this region. Water molecules were not added to the model due to the lower resolution of the structure in comparison to the UNGM holoenzyme.
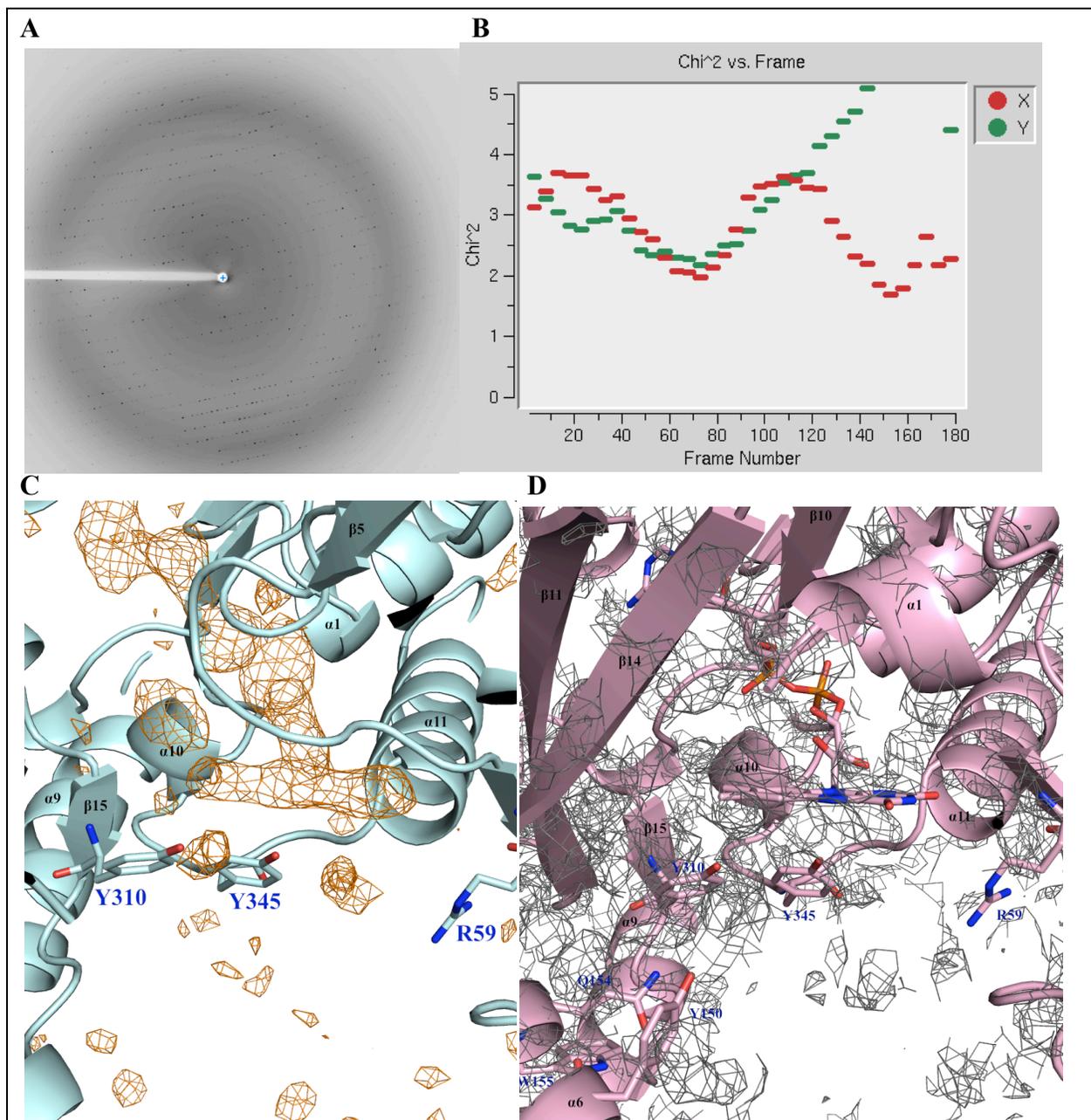
Figure 4.19: **Diffraction of UNGM:UDP-Glc co-crystals from ComPAs #13.** **A-** Diffraction image of ComPAs #13 crystal. **B-** Chi^2 vs. Frame number for the ComPAs #13 data set. **C-** Active site of UDP-Glc co-crystallization in ComPAs #13. The $|F_o|-|F_c|$ map (orange, contoured to 3.0 σ) shows density for the oxidized FAD cofactor. **D-** The $|F_o|-|F_c|$ map (in gray) does not show electron density for UDP-Glucose in the substrate binding site (highlighted by residues in stick representation).

Table 4.5: Data collection and refinement statistics for UNGM with UDP-Glc (ComPAs #13).

Data Collection:

Space group	P2$_1$2$_1$2$_1$
Unit-cell parameters (Å)	a = 47.8 b = 119.3 c = 147.3
Wavelength (Å)	0.9795
Temperature (K)	100
Resolution range (Å)	50.00-2.40 (2.49-2.40)
Matthews coefficient (Å^3 Da^{-1})	2.47
Solvent content (%)	50
No. of molecules in ASU	2
Molecular weight (Da)	43445 [368 amino acid residues]
No. of measured reflections	295240
Total No. of unique reflections	43321
Completeness (%)	99.7 (100)
Redundancy	3.6 (3.4)
R$_{merge}$†	0.159 (0.471)
<I/σ(I)>	6.1 (3.22)
Refinement:	
R$_{work}$‡	0.2233
R$_{free}$§	0.2813
RMSD bond lengths	0.009
RMSD bond angles	1.21

†$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where I$_i$(*hkl*) are the individual intensities of the *i*th observation of reflection *hkl* and <I(*hkl*)> is the average intensity of reflection *hkl* with summation over all data. ‡ $R_{work} = \Sigma_{hkl} ||F_{obs}| - |F_{calc}||/\Sigma_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. § R$_{free}$ is equivalent to the R_{work} but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.

4.4.4 Co-crystallization with UDP

Crystals were observed in the condition JCSG #62 (3.15 M Ammonium sulfate, 0.1 M tri-Na-citrate pH 5.0) within 12 d (Figure 4.20). The crystals did not have a defined morphology, however they were believed to be protein due to their yellow color and ability to uptake IZIT dye. Unfortunately, the crystals were not reproducible upon optimization of this condition. Therefore, these crystals were not diffracted to confirm if they were indeed comprised of protein.

4.5 Docking Studies in GOLD

Due to the lack of a UNGM structure in complex with one of its substrates, we initiated docking studies using the program GOLD (Verdonk *et al.*, 2003). GOLD (Genetic Optimization for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites. The genetic algorithm (GA) takes a population of potential solutions that are set up at random. Each member of the population is encoded as a chromosome, which contains information about the mapping of ligand hydrogen-bond atoms onto protein model hydrogen-bond atoms, mapping of hydrophobic points on the ligand onto protein hydrophobic points, and the conformation around flexible ligand bonds and protein OH groups. Each chromosome is assigned a fitness score based on its predicted binding affinity and the chromosomes within the population are ranked according to their fitness score. At each step, a point mutation may occur in a chromosome, or two chromosomes may mate (combine) to give

Figure 4.20: Crystals of UNGM with UDP.

a child. The parent chromosomes that are fitter members of the population (ligand dockings with good fitness scores) are selected for the optimized final population. The fitness scoring functions are helpful to predict the biological activity of ligands.

4.6 Generation of docking models

The holoenzyme structure of UNGM was modified to resemble the closed (active) conformation observed in the drUGM reduced in complex with UDP-Gal$p$ structure (PDB 3HDY). The mobile loop 2 region of UNGM (residues 156-186) was modeled to resemble the arrangement observed in drUGM (Figure 2.6B). This region is known to be essential for proper binding of substrates; specifically the conserved Arginine (Arg169 in UNGM) is oriented towards the substrate-binding site and makes an electrostatic bond with the α-phosphate group of UDP (Karunan Partha *et al*., 2009). In order to maintain the most realistic model of UNGM, no other active site residues were altered prior to further experiments. This model will be referred to as UNGM oxidized model A, or UNGM$_{ox}$ A. In addition, a second model for UNGM was modified such that Arg168 was modeled to mimic the conformation of Lys169 in the closed loop structure of ecUGM (Figure 4.10C). The model will be referred to as UNGM oxidized model B, or UNGM$_{ox}$ B.

It is essential that the docking model be in the lowest free energy state in order to mimic the state observed *in vitro* for the protein. Proteins exist in the lowest kinetically available state, and a low free energy state is favored. Due to the significant modification to the mobile loop of the oxidized holoenzyme structure, it was essential to optimize the energy minimum of the model. Energy minimization was attempted with Chiron (Ramachandran *et al*., 2011). The energy clash score or Van der Waals repulsion energy is calculated using CHARMM non-bonded parameters (Brooks *et al*., 1983). The initial energy clash score for this model was 964 kcal/mol, which decreased to a final value of 221 kcal/mol. Although the energy clash score decreased during this experiment, this type of energy minimization was unsuccessful because the mobile loop moved to the inactive open conformation. The binding of the substrate stabilizes the mobile loop, and it can be assumed that without this interaction the closed loop is in a high-energy state, and that the loop in its open conformation is in the lowest energy state.

In addition, the drUGM structure in complex with UDP-Gal$p$ was also energy minimized with Chiron to determine whether the mobile loop would also move to its open conformation. We did not observe a change in the energy clash score throughout the minimization, which maintained a value of 87.92 kcal/mol. This result was consistent with the lack of mobile loop translocation. With this knowledge, we expected UNGM$_{ox}$ A with UDP-Gal$p$ artificially docked into the substrate-binding site would also remain in the closed conformation after energy minimization. Initially, the energy clash score had a value of 1015 kcal/mol, which decreased significantly during minimization, with a final value of 217 kcal/mol. However, despite the presence of UDP-Gal$p$, this model did not remain in the desired closed conformation.

Therefore, the closed-conformation UNGM$_{ox}$ A was energy minimized using structure idealization in Refmac5 (Murshudov *et al*., 1997). Structure idealization is used to optimize the bond geometry, which improves the energy minimum for the model. This program also requires a geometry restraint file for the FAD cofactor, which was generated using the program Elbow in Phenix. After structure idealization, the model was visually inspected to confirm that the mobile loop remained in the closed conformation and that the modified residues remained in the same positions within the substrate binding site. The structure-idealized model was subjected to further minimization with Chiron, which did not result in a change in the energy clash score. Both the initial and final energy clash scores shared a value of 32.426 kcal/mol, however the mobile loop was observed to move to the open conformation in this model. Therefore, the structure-idealized model was used for further analysis. The UNGM$_{ox}$ B was also subjected to structure idealization and energy minimization as previously described for UNGM$_{ox}$ A. The energy clash score for the UNGM oxidized model B was identical to model A (32.42 kcal/mol).

The structure-idealized UNGM$_{ox}$ A was edited to contain the reduced form of flavin adenine dinucleotide, FADH$^-$ in the FAD-binding region. The UNGM reduced model was subjected to energy minimization in Chiron, which resulted in no change in the value of the energy clash score (103 kcal/mol). The sole presence of FADH$^-$ in the active site resulted in the stabilization of the mobile loop in the closed conformation. This observation conflicts with

previous results, which pointed to the necessity of the UDP moiety of the substrate(s) to stabilize the closed conformation.

The structure of drUGM:UDP-Gal*p* (PDB 3HDY) was used as a comparative docking model for the docking studies. For consistency, the drUGM monomer used for UNGM modeling was also used for the docking studies. Mobile loop 2 is in the closed conformation, and was not modified. The structure of drUGM (oxidized FAD, in complex with UDP-Gal*p*) was subjected to energy minimization to ensure that the model was consistent with the UNGM model (discussed in Section 4.5). Initially, the UDP-Gal*p* substrate was removed from the structure and subjected to structure idealization. After structure idealization, the structure was inspected to ensure the mobile loop remained in the closed conformation and the key substrate-binding residues were in the same positions within the active site. The structure was then subjected to further energy minimization in Chiron. As previously observed, the energy clash score shared the same value at the beginning and end of minimization, 92.53 kcal/mol. Importantly, the mobile loop remained in the closed conformation without the presence of UDP-Gal*p* bound in the active site. No significant changes were observed in the Chiron energy minimized structure of drUGM, and therefore the structure-idealized version of drUGM was used for further analysis. This model will be referred to as the drUGM oxidized model (drUGM$_{ox}$).

The energy minimized, oxidized drUGM model was edited by removing the FAD cofactor and the cofactor was replaced with FADH$^-$. Further structure idealization resulted in the loss of the reduced conformation of FAD, and therefore we did not subject the flavin cofactor to structure idealization. This model will henceforth be referred to as the drUGM reduced model (drUGM$_{red}$).

4.7 Validation of Docking Models

The structure-idealized drUGM docking models were validated using Molprobity in Phenix (Chen *et al*., 2010) (Table 4.6). There were no Ramachandran outliers observed in the structure. Rotamer outliers were present in the structure outside of the active site, and therefore were deemed to have no effect on the docking experiments. Bond geometry and angles

analysis (in Molprobity) was also utilized for the validation of the UNGM docking models. Ramachandran and rotamer outliers were identified, as well as unfavorable contacts that occurred during the modification of the holoenzyme structure. The residues flagged as outliers were examined for their effect of substrate binding and manually adjusted. The remaining outliers (Table 4.6) were in areas of the protein that are not suspected to have an effect on substrate-binding, and therefore were not refined any further.

Table 4.6: Final validation of docking models.

Model	UNGM$_{ox}$ A & B	UNGM$_{red}$	drUGM (oxidized & reduced)
Ramachandran outliers	1.6 %	1.6 %	0.0 %
Rotamer outliers	1.2 %	1.2 %	2.9 %
Ramachandran most favored	94.1 %	94.1 %	96.6 %

4.8 Generation of Substrate Models for docking

Models of UDP-Gal*f*, UDP-Gal*f*NAc, and UDP-Gal*p*NAc were built manually in the computational chemistry software Spartan (Hehre and Ohlinger, 2011) (Figure 4.21). It is not known if any of the substrates carry a charge when they bind to the active site environment within UNGM. Therefore, each substrate was assigned a neutral charge, anion or dianion on the phosphate region prior to energy minimization. The models were energy minimized in prior to docking experiments to ensure they were in the lowest energy state. The final energies of the models are listed in Table 4.7.

Table 4.7: Final energies of substrate models after minimization.

Substrate model	Energy (kJ/mol)
UDP-Gal*f*	-271.3
UDP-Gal*f*NAc	-1099
UDP-Gal*p*NAc	-722.9
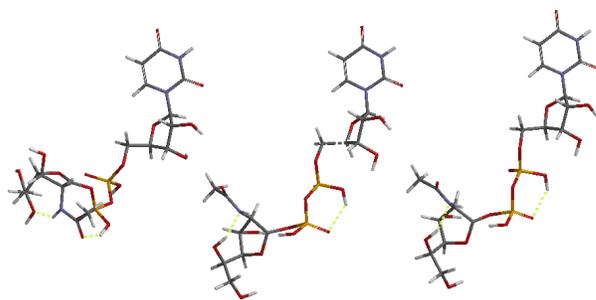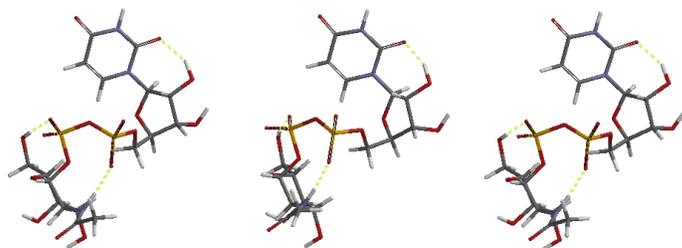
Figure 4.21: Substrate models for docking studies. A- UDP-Gal*f*NAc models for docking experiments; Neutral, anion, and dianion-charged structures left to right. **B-** UDP-Gal*f* models; Neutral, anion, and dianion-charged structures left to right. **C-** UDP-Gal*p*NAc models; Neutral, anion, and dianion-charged structures left to right. Models were generated and energy minimized with the molecular modeling program Spartan (Hehre and Ohlinger, 2011).

4.9 Docking experiments with UGM from *Deinococcus radiodurans*

Water molecules and the ligand UDP-Gal*p* were removed from the structure, and hydrogen atoms were added prior to docking experiments. Unless otherwise noted, the parameters for the docking experiments were exactly the same. The GA search efficiency was set to 200 %, with a search radius of 6 Å around the ligand to incorporate the essential amino acid residues for substrate binding, and a distance constraint of 3.0 Å was placed between the anomeric carbon of the sugar and N5 of the FAD cofactor. The active site residues were held rigid during the experiments, to retain the substrate binding conformations observed in the complex structure. For consistency, the same parameters were used for the oxidized and reduced drUGM models.

4.9.1 drUGM oxidized model

4.9.1.1 UDP-Gal*p*

As a control docking experiment, the structure of drUGM was used to dock UDP-Gal*p* into the active site. The top docking solution had a fitness score (Chemscore) of 17.0 and near perfect superposition with the UDP-Gal*p* observed in the crystal structure, which is annotated as the reference ligand (Figure 4.22A). The distance between N5 of the isoalloxazine ring of FAD is approximately 3.1 Å from C1 of galactopyranose. This observation is consistent with previous studies on the enzyme mechanism, in which a FAD-galactose adduct forms between N5 (FAD) and C1 (galactose) (Karunan Partha *et al*., 2009). The observed distances between these atoms were within the range of 2.8 to 3.3 Å. Therefore the ligand was successfully docked in a productive binding mode. The parameters used for this experiment were used for all further docking experiments. The docking studies with drUGM$_{ox}$ are summarized in Table 4.8.

4.9.1.2 UDP-Gal*f*

The natural substrate UDP-Gal*f* (with a neutral charge) was docked into drUGM$_{ox}$, with a Chemscore fitness score of 13.4. Even though a distance constraint of 3.0 Å between the anomeric carbon and N5 of FAD was used during the calculations, all of the solutions resulted in an interatomic distance of 4 Å or greater. The top docking solution was bound within the

uridine and diphosphate regions in a similar mode as UDP-Gal*p*; however unexpectedly, the sugar is too far for the formation of the FAD-Gal*f* adduct (Figure 4.22B).

Table 4.8: Docking studies with drUGM

Model	Substrate	Chemscore fitness score	Interatomic distance between C1 (sugar) and N5 (flavin) (Å)	Distance between acetamido moiety of sugar and δ^1N of His88	Productive binding mode (Y/N)
drUGM$_{ox}$					
	UDP-Gal*p*	17.0	3.1	N/A	Y
	UDP-Gal*f*	13.4	4.1	N/A	N
	UDP-Gal*p*NAc	10.2	3.1	2.8 (8.1 for ChemPLP)	Y
	UDP-Gal*f*NAc	7.8	3.6	5.0	Y
drUGM$_{red}$					
	UDP-Gal*p*	21.0	2.9	N/A	Y
	UDP-Gal*f*	21.0	4.2	N/A	N
	UDP-Gal*p*NAc	9.5	3.4	6.8	Y
	UDP-Gal*f*NAc	4.8	3.8	5.1	Y
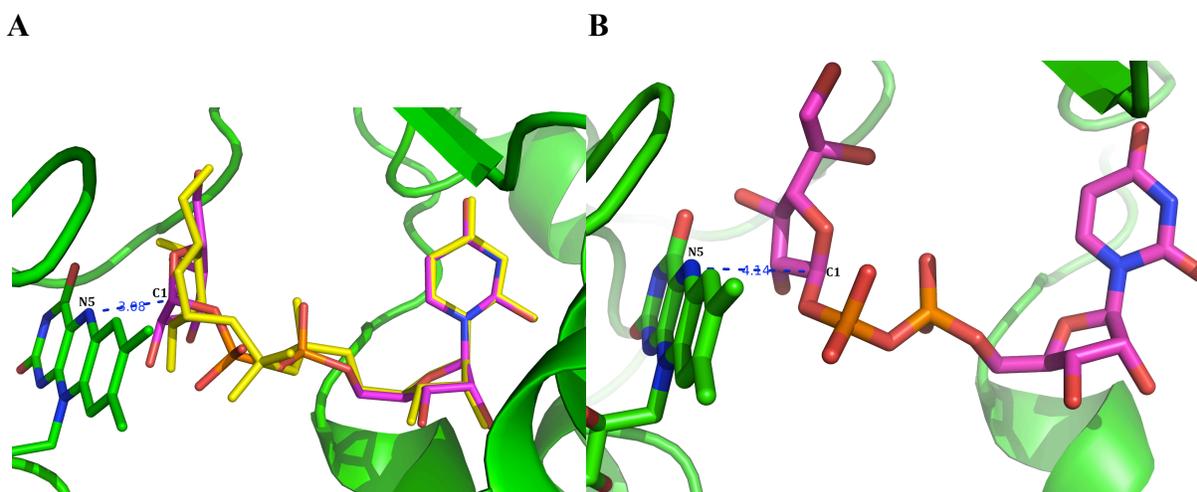
Figure 4.22: Control docking experiments for UDP-Gal*p* and UDP-Gal*f* into drUGM$_{ox}$. **A**-The top docking solution for UDP-Gal*p* is shown in magenta, drUGM$_{ox}$ and FAD in green, reference ligand UDP-Gal*p* in yellow. **B**- UDP-Gal*f* is shown in magenta and drUGM$_{ox}$ in green.

4.9.1.3 UDP-Gal*p*NAc

The top docking solution had a Chemscore fitness score of 10.2, generated with UDP-Gal*p*NAc dianion model. Although UDP-Gal*p*NAc is not a substrate of drUGM, the docking of this molecule into the substrate-binding site unexpectedly yielded a productive binding mode (Figure 4.23). The distance between C1 of Gal*p*NAc and N5 of FAD is approximately 3.1 Å, within the range for the formation of an FAD-galactose adduct. Also, the distance between His88 and the acetamido group of Gal*p*NAc was 2.8 Å, which indicated the formation of a hydrogen bond. This result was unexpected because drUGM does not recognize this substrate and we do not expect it to bind in a productive way. Additionally, the docking was rescored using CHEMPLP, with the top solution fitness score of 91 (not shown). Importantly, the distance between C1 (Gal*p*NAc) and N5 (FAD) was significantly larger (3.8 Å). The acetamido group of the sugar was oriented away from His88 at a distance of 8.1 Å, which is not favorable for hydrogen bond formation.
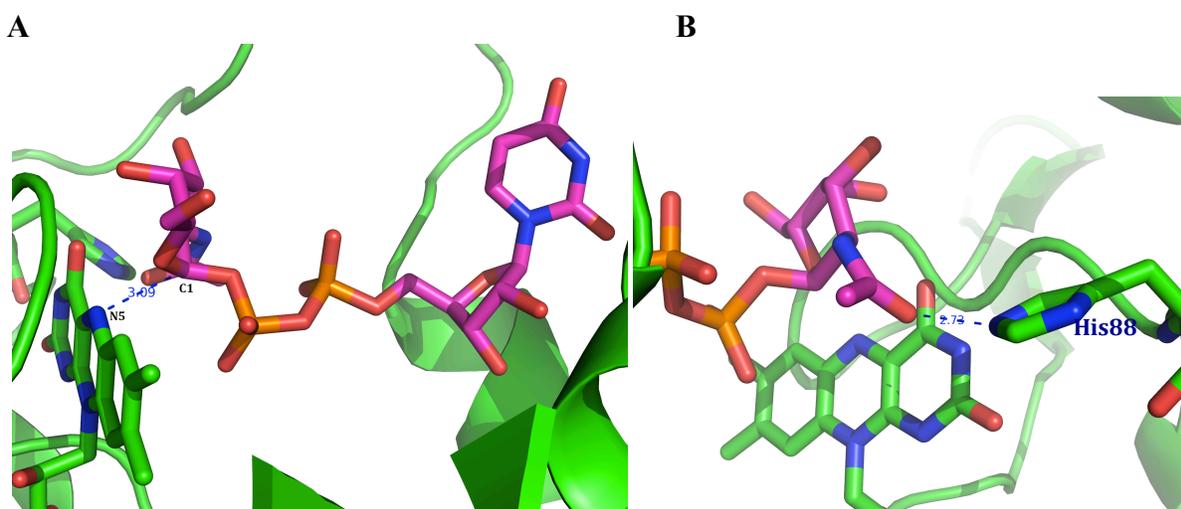
Figure 4.23: The top docking solution for UDP-Gal*p*NAc into the drUGM$_{ox}$. A- Top docking solution for UDP-Gal*p*NAc (magenta) binds the active site of drUGM$_{ox}$ (green); the isoalloxazine ring of FAD shown in green. **B-** The carbonyl oxygen of the acetamido group of UDP-Gal*p*NAc is 2.73 Å from δ^1N of His88 in drUGM (180° rotation from A).

4.9.1.4 UDP-Gal*f*NAc

The structure of drUGM was also used for UDP-Gal*f*NAc docking experiments. The top solution had a fitness score of 7.8, and was generated using UDP-Gal*f*NAc with a neutral charge (Figure 4.24). The uridine and diphosphate regions fit in a similar fashion to UDP-Gal*p*, however the furanose ring is shifted further away from FAD. The docking pose of UDP-Gal*f*NAc into the drUGM oxidized model was deemed to be productive. The distance between C1 of galactofuranose and N5 of FAD is approximately 3.6 Å, which is a greater distance than observed for UDP-Gal*p*NAc (Figure 4.24A). Importantly, the distance from the carbonyl of the acetamido group (Gal*f*NAc) and δ^1N of His88 is approximately 5.0 Å (Figure 4.24B). This distance is not compatible for forming a hydrogen bond, and therefore we expect a decrease in the binding stability of UDP-Gal*f*NAc into drUGM.

4.9.2 drUGM reduced model

For consistency, the energy minimized drUGM oxidized model was edited by removing the FAD cofactor, and replaced with the reduced FADH⁻ cofactor in the same position. The reduction of the cofactor is expected to improve the efficiency of the docking experiments, due to the catalytic dependence on reduction of the cofactor. For the natural substrates of UDP-Gal*p* and UDP-Gal*f*, the fitness score improved when the flavin cofactor was in the reduced
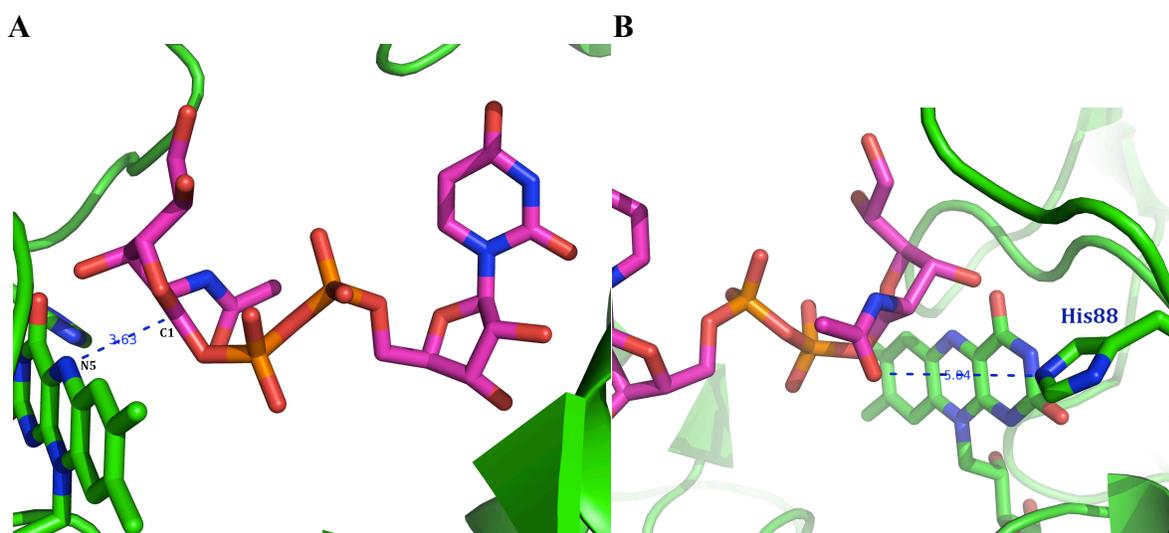
Figure 4.24: The top docking solution for UDP-Gal*f*NAc into drUGM$_{ox}$. A- The best docking solution for UDP-Gal*f*NAc (magenta) into drUGM$_{ox}$ (green); FAD cofactor shown in green. **B-** The carbonyl oxygen of the acetamido group of UDP-Gal*f*NAc (magenta) is 5.04 Å from the δ^1N of His88 in drUGM (180° rotation from A).

form, which is consistent with previous requirements of cofactor reduction for catalysis and efficient substrate binding. For UDP-Gal*p*, the interatomic distance between C1 of Gal*p* and N5 of the reduced flavin was closer than observed in the oxidized model, which indicated that cofactor reduction promoted this interaction. The docking studies with this model are summarized in Table 4.8.

The ligands UDP-Gal*p*NAc and UDP-Gal*f*NAc are not natural substrates of drUGM, and both scored lower for the reduced model than the oxidized model. This result is opposite of what we expect if the ligand binds more favorably to reduced structures. The docking pose of the top solution for UDP-Gal*p*NAc was significantly worse than the oxidized model (Figure 4.23), with larger distances between both the acetamido group and His88 (6.8 Å), and the sugar and cofactor (3.4 Å) (Figure 4.25). These results indicate that GalNAc sugars do not dock consistently between the oxidized and reduced models, and that the reduction of the cofactor does not improve the binding of these ligands.
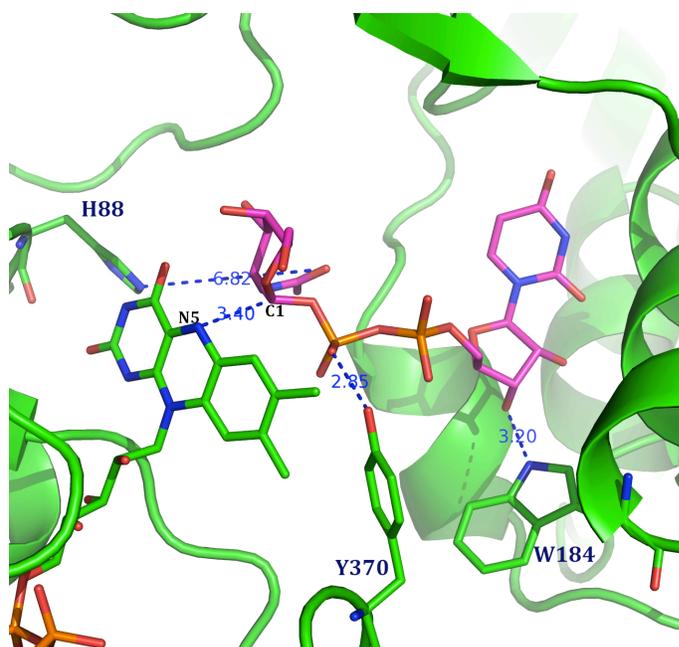
Figure 4.25: The top docking solution for UDP-Gal*p*NAc into drUGM$_{red}$. UDP-Gal*p*NAc is shown in magenta, drUGM$_{red}$ is shown in green. The uridine and diphosphate regions are bound as expected from the drUGM:UDP structure, however the sugar region is not stabilized by His88 in this model.

4.10 Docking studies with UNGM

Docking experiments were performed using both the UNGM oxidized and the UNGM reduced models. The same parameters used for drUGM docking experiments were also used for the UNGM docking experiments. Docking results for the UNGM models are summarized in Table 4.9.

Table 4.9: Docking studies with UNGM

Model	Substrate	Fitness score	Interatomic distance between C1 (sugar) and N5 (flavin) (Å)	Distance between acetamido moiety of sugar and Arg59	Productive binding mode (Y/N)
$UNGM_{ox}$ A					
	UDP-Gal$p$	11.4	3.3	N/A	Y
	UDP-Gal$f$	8.9	3.1	N/A	Y
	UDP-Gal$p$NAc	18.3	3.6	3.0	Y
	UDP-Gal$f$NAc	11.2	3.8	2.8	Y
$UNGM_{ox}$ B					
	UDP-Gal$p$	7.22	3.4	N/A	Y
	UDP-Gal$f$	11.6	3.2	N/A	Y
	UDP-Gal$p$NAc	16.5	3.0	3.0	Y
	UDP-Gal$f$NAc	17.1	3.6	2.9	Y
$UNGM_{red}$					
	UDP-Gal$p$	15.1	3.3	N/A	Y
	UDP-Gal$f$	9.9	3.8	N/A	Y
	UDP-Gal$p$NAc	20.0	3.6	3.0	Y
	UDP-Gal$f$NAc	13.5	3.6	2.7	Y

4.10.1 UNGM oxidized model A ($UNGM_{ox}$ A)

The best docking solutions for the $UNGM_{ox}$ A model were obtained using a modified UNGM structure in mobile loop 2 (residues 156-186) as well as the active site residues Trp155 and Tyr345 (Figure 4.26). Arg169 on the mobile loop was modeled to face towards the substrate binding region and facilitate favorable interactions with the diphosphate backbone of UDP. The conserved active site Trp155 is thought to stabilize the binding of the uracil portion via hydrogen bonding interactions between the indole nitrogen of Trp155 and the C2 and C3 hydroxyl groups of ribose. We saw a drastic improvement in the docking of the uridine moiety

after Trp155 was adjusted to the observed position in drUGM (Trp184). Tyr345 is one of 3 conserved tyrosine residues previously shown to stabilize the binding of the diphosphate backbone through hydrogen bonding (Chad *et al.*, 2007). The position of Tyr345 was adjusted to mimic Tyr370 in drUGM to improve the binding of the diphosphate region (Figure 4.26).

4.10.1.1 UDP-Gal*p*

The best solution for the UNGM$_{ox}$ A with its natural substrate UDP-Gal*p* had a Chemscore fitness score of 11.4 (Figure 4.27A). The distance between C1 of Gal*p* and N5 of FAD is 3.3 Å, which is a productive binding mode for catalysis. The uridine and diphosphate regions of the substrate were stabilized by interactions with Trp155 and Tyr345 respectively. Arg169 appears to further stabilize the β-phosphate of UDP.

4.10.1.2 UDP-Gal*f*

The top docking solution for UDP-Gal*f* (with a neutral charge) into the UNGM$_{ox}$ A had a Chemscore fitness score of 8.9. The docking pose was deemed to be in a productive binding mode, with a distance of 3.1 Å between N5 of FAD and C1 of galactofuranose (Figure 4.27B).
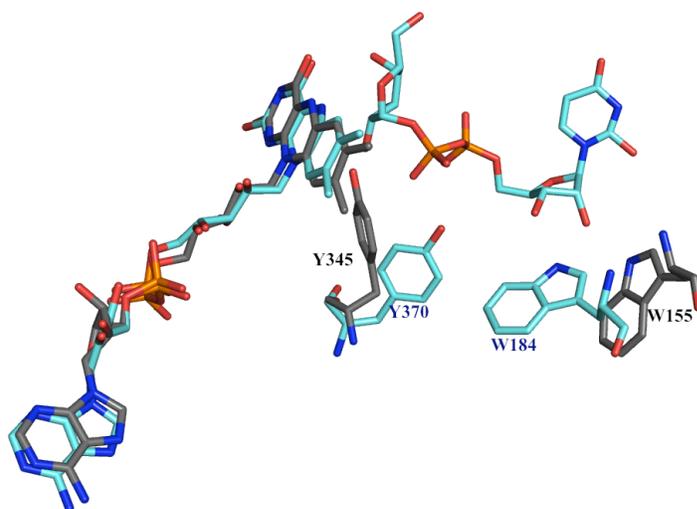
Figure 4.26: The superposed active sites of UNGM and drUGM. UNGM (gray; black label) and drUGM (cyan; blue label) are highly similar, however when substrate is bound in the active site, movement is observed in the residues Tyr370 and Trp184 in drUGM; UDP-Gal*p* highlighted in cyan. The structures were superposed using LSQ superposition in Coot, which performs alignment using the residues selected from both models (FAD, Tyr370 and Trp184) (Emsley *et al.*, 2010).
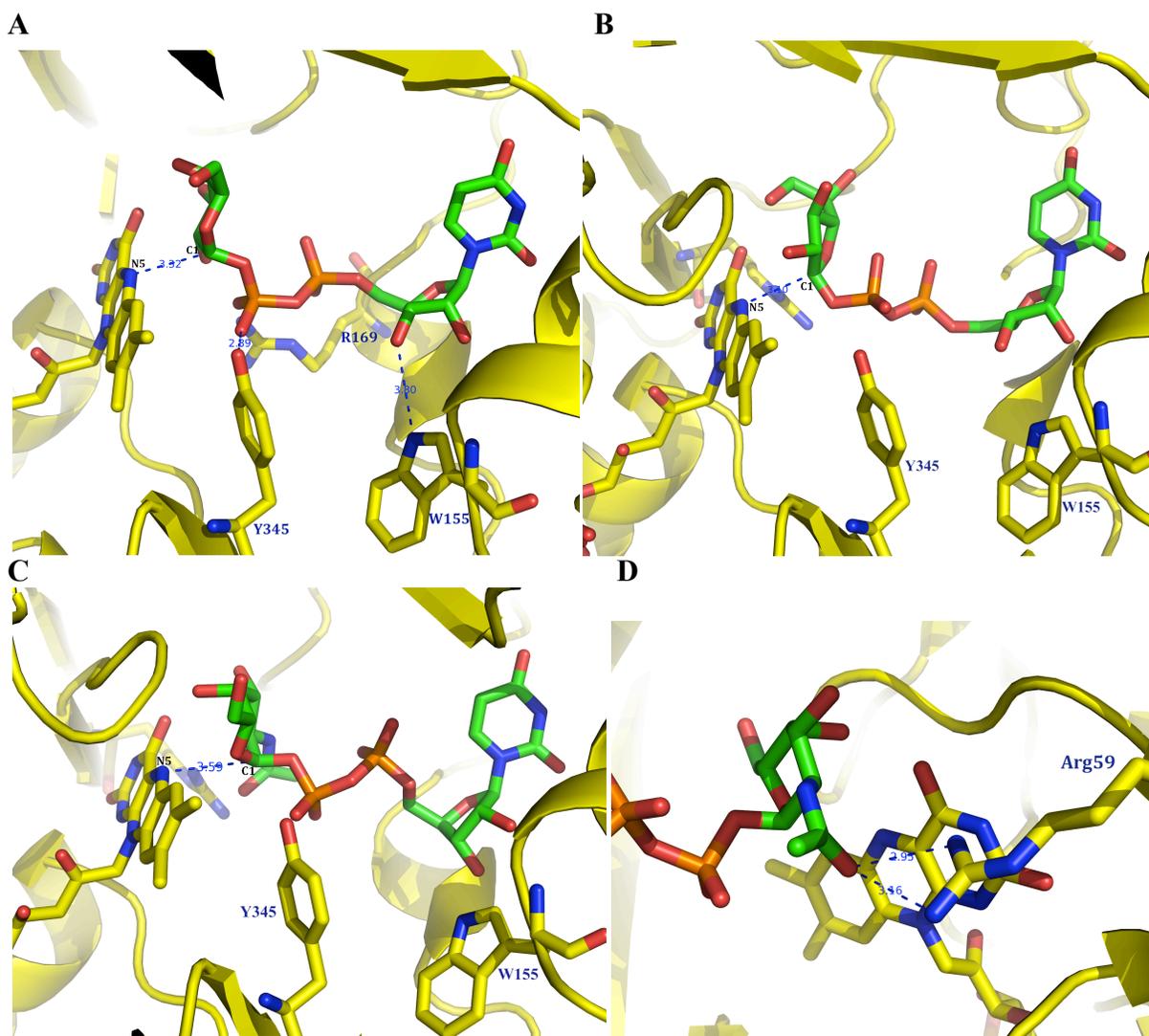
Figure 4.27: The top docking solutions for UNGM_{ox} A. A- UDP-Gal*p* (green) was bound to the UNGM oxidized model A (yellow) in a favorable conformation; FAD shown in yellow. Tyr345, Trp155, and Arg169 are labeled in navy. **B-** The top docking solution for UDP-Gal*f* into UNGM_{ox} A. UDP-Gal*f* (green) is docked in a favorable conformation in the UNGM oxidized model A (yellow); FAD cofactor shown in yellow. Tyr345 and Trp155 are highlighted in yellow. **C-** UDP-Gal*p*NAc (green) was docked into UNGM_{ox} A (yellow), and the substrate formed favorable interactions with the conserved active site residues. **D-** Hydrogen bond formation is likely to occur between the carbonyl oxygen of the acetamido group of UDP-Gal*p*NAc and Arg59 (180° rotation from **C**).

4.10.1.3 UDP-GalpNAc

The top docking solution for UDP-GalpNAc into UNGM was found using the dianion charged model for UDP-GalpNAc (Figure 4.27C). The best results were generated with the GA flexibility set to 100 %. The Chemscore fitness score of 18.3 indicated a high degree of fit with the active site of UNGM, with the next solution at a fitness score of 13.0. UNGM is known to actively catalyze the interconversion of UDP-GalpNAc to UDP-GalfNAc, so we expect that both of these substrates will dock with a high fidelity. The distance between the sugar and FAD was approximately 3.6 Å, a favorable distance for stabilization of the sugar. Also, the uridine and diphosphate regions are bonded to the active site in a highly conserved manner. In the sugar-binding region of UNGM, Arg59 is thought to stabilize the binding of UDP-GalpNAc via its guanidinium group. The carbonyl oxygen of the acetamido group of UDP-GalpNAc is approximately 3 Å from the guanidinium group of Arg59 (Figure 4.27D), which is a favorable distance for the formation of a hydrogen bond.

4.10.1.4 UDP-GalfNAc

The natural substrate UDP-GalfNAc (neutral charge) was docked into $UNGM_{ox}$ A with a Chemscore fitness score of 11.2 (Figure 4.28). The GA parameter for these experiments was
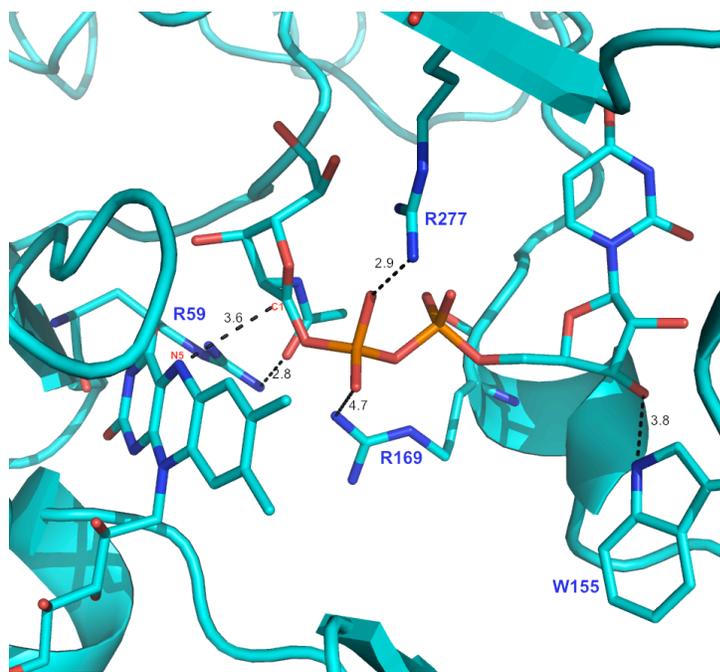
Figure 4.28: The top docking solution for UDP-GalfNAc into $UNGM_{ox}$ A. UDP-GalfNAc (C-cyan, O-red, P-orange, N- blue) occupies the active site of $UNGM_{ox}$ A (cyan) in a favorable conformation. The acetamido group of UDP-GalfNAc forms a hydrogen bond with Arg59.

lowered in flexibility to 100 %, due to poor results at higher flexibility settings (200 %). The distance between the anomeric carbon and N5 of FAD was 3.8 Å, which is within of the range of a favorable bonding distance, and therefore the substrate is in a productive binding mode. The uridine and diphosphate regions of the molecule adhere to previously observed binding modes. Importantly, the carbonyl oxygen of the acetamido of UDP-Gal*f*NAc is coordinated by N3 and N4 of the guanidinium group of Arg59 at a distance of approximately 2.8 Å (Figure 4.28). These observations are comparable to the docking of UDP-Gal*p*NAc into UNGM.

4.10.2 UNGM oxidized model B (UNGM$_{ox}$ B)

Mobile loop 2 of UNGM$_{ox}$ A was modified to position Arg168 for interaction with the substrate instead of Arg169, designated as UNGM oxidized model B (UNGM$_{ox}$ B). If Arg168 participates in the stabilization of the acetamido group of the sugars, we expect that the Chemscore fitness will be larger for the *N*-acetylgalactosamine sugars when compared to galactose sugars. The same settings were used for these docking experiments as previously described for UNGM$_{ox}$ A. In addition, Arg168 side chain flexibility was allowed for the 34 known rotamers of arginine during the docking experiments for this model. If the position of this residue is incorrect in the model, increasing the flexibility will allow for multiple conformations, and thereby increase the probability of favorable interactions.

4.10.2.1 UDP-Gal*p*

The best docking solution for UDP-Gal*p* in UNGM$_{ox}$ B corresponded to a Chemscore fitness score of 7.22 (Figure 4.29A). The substrate was observed to bind the enzyme in a near identical pose to UNGM$_{ox}$ A (not shown). The distance between C1 of UDP-Gal*p* and N5 of FAD was 3.4 Å, and the ribose of UDP is stabilized by hydrogen bonding with Trp155. Significantly, Arg168 functions to stabilize the binding of the diphosphate region, as this residue is approximately 4 Å away.

4.10.2.2 UDP-Gal*f*

The best docking solution for UDP-Gal*f* into UNGM$_{ox}$ B had a Chemscore fitness score of 11.6 (not shown). The sugar and FAD were within a favorable distance for catalysis (3.2 Å), and the substrate docked in a near identical pose as UDP-Gal*p* in UNGM$_{ox}$ B (Figure 4.29A).
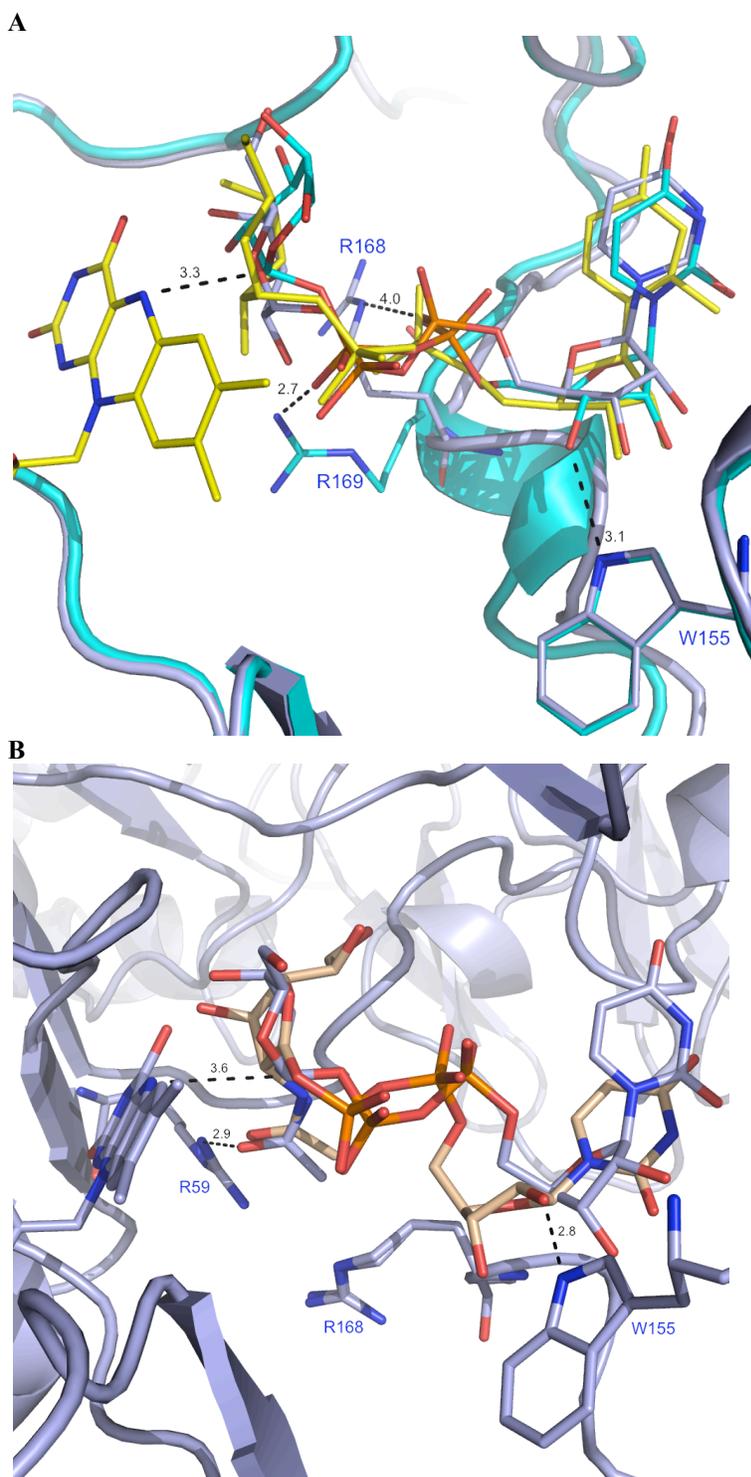
Figure 4.29: The top docking solutions in UNGM$_{ox}$ B. A- UDP-Gal$p$ in UNGM$_{ox}$ B (light blue) binds in near identical fashion as seen in drUGM (yellow) and UNGM$_{ox}$ A (cyan). Arg168 in UNGM$_{ox}$ B is within range for stabilization of the diphosphate region. **B-** UNGM$_{ox}$ B stabilizes UDP-Gal$f$NAc (C-wheat) and UDP-Gal$p$NAc (C-light blue) in a highly conserved manner, with the acetamido group within hydrogen-bonding distance with Arg59 in both dockings. The substrates are stabilized in the active site by: Arg59, Trp155, and Arg168.

4.10.2.3 UDP-GalpNAc

The best docking solution for UDP-GalpNAc into UNGM$_{ox}$ B had a fitness score of 16.5, and the pose of the substrate was deemed to be productive (Figure 4.29B). The distance between Arg59 and the carbonyl oxygen of the acetamido group is approximately 3.0 Å, which is a favorable distance for stabilization. The guanidinium group of Arg168 is 5.2 Å from the carbonyl oxygen of the acetamido group of the sugar, which indicated that Arg168 does not directly interact with the sugar moiety in this experiment. Rather, Arg168 may stabilize the binding the diphosphate region. Trp155 is 4.1 Å from the C3' hydroxyl of ribose, and is within the range for stabilization through a hydrogen bonding interaction.

4.10.2.4 UDP-GalfNAc

The best docking solution for UDP-GalfNAc (neutral charged) into the UNGM$_{ox}$ B had a fitness score of 17.1, and was also in a productive binding mode. The docking of the sugar portion of UDP-GalfNAc was highly similar to GalpNAc in the same model (Figure 4.32B). A shift is observed in the uridine region in UDP-GalfNAc, however the ribose moiety is stabilized by W155 at a distance of approximately 2.8 Å.

4.10.3 UNGM reduced model (UNGM$_{red}$)

In order to test the significance of cofactor reduction, the docking experiments with the UNGM reduced model were performed using the exact same parameters used for UNGM$_{ox}$ A. Also, UNGM$_{ox}$ A was used for comparative analysis because this model is identical to UNGM$_{red}$, with the exception of the cofactor conformation. The substrates UDP-Galp and UDP-Galf both had larger fitness scores in the reduced model for UNGM than in UNGM$_{ox}$ A, and therefore it can be assumed that the reduction of the cofactor improved the efficiency of binding. The substrates UDP-GalpNAc and UDP-GalfNAc where both found to bind in a productive binding pose. Furthermore, the acetamido group of both of these substrates was within the range for hydrogen bond formation with Arg59. These results indicate that the reduction of the isoalloxazine ring favors more stable binding of substrates in this model.

4.11 Estimation of Binding Affinity through $\Delta G_{binding}$ and K_M

Prediction of a protein's binding affinity for a ligand using docking programs has proven to be extremely difficult. Currently, no methods are available to accurately predict the energetics of protein-ligand binding. The docking programs must search the space of possible poses and conformations to find the energy minimum of the protein-ligand complex. The fitness scores calculated in GOLD are dimensionless, and therefore cannot be used explicitly as values for binding energy or binding affinity. The free energy of binding, $\Delta G_{binding}$ or the binding affinity constant, K, is a desirable term to calculate the binding affinity of a ligand-protein complex. The affinity can be quantitated through the binding free energy as:

$$\Delta G_{binding} = -RT\ln K_{eq} \qquad \text{(Equation 4.3)}$$

where K_{eq} = equilibrium constant
T = temperature in Kelvin
R = gas constant = 8.314 Jmol^{-1}

The $\Delta G_{binding}$ term is commonly estimated by the contributions of hydrophobic effect, conformational entropy, and hydrogen bonding (Murphy, 1999). The $\Delta G_{binding}$ (free energy of binding) component of the Chemscore function may provide a crude estimate of the binding affinity (Equation 3.2). Each component of $\Delta G_{binding}$ (Equation 4.4) is the product of a term dependent on the value of a physical contribution to free energy, and a scale factor determined by regression. Physical contributions considered in these calculations include hydrogen bonding, metal-binding, lipophilic interactions, and bond rotation.

$$\textbf{Chemscore} = \Delta G_{binding} + P_{clash} + c_{internal}P_{internal} + (c_{covalent}P_{covalent} + P_{constraint}) \qquad \text{(Equation 3.2)}$$

$$\Delta G_{binding} = \Delta G_o + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot} \qquad \text{(Equation 4.4)}$$

The correlation of $\Delta G_{binding}$ was found to correlate to binding affinity data approximately 60% of the time, with an average error of 2.5 kcal/mol (Verdonk *et al.*, 2003). The terms that comprise the Chemscore are defined in Section 3.2. Therefore, the clash penalty and internal torsion terms (P and c terms respectively) are ignored in Equation 3.2, such that only the $\Delta G_{binding}$ component is used for correlation with binding affinity data. A negative value for $\Delta G_{binding}$ indicates a more favorable docking solution and a larger fitness score. All of the

$\Delta G_{binding}$ terms were negative; indicating that all of the ligands tested formed favorable interactions with the protein (Table 4.10).

Table 4.10: The $\Delta G_{binding}$ terms from Chemscore fitness scores.

Docking Model	Substrate	$\Delta G_{binding}$
UNGM oxidized model A	UDP-Gal*p*	-20.98
	UDP-Gal*f*	-25.45
	UDP-Gal*p*NAc	-27.57
	UDP-Gal*f*NAc	-21.99
UNGM oxidized model B	UDP-Gal*p*	-14.39
	UDP-Gal*f*	-17.87
	UDP-Gal*p*NAc	-26.09
	UDP-Gal*f*NAc	-26.94
UNGM reduced	UDP-Gal*p*	-21.59
	UDP-Gal*f*	-18.98
	UDP-Gal*p*NAc	-28.35
	UDP-Gal*f*NAc	-23.28
drUGM oxidized	UDP-Gal*p*	-25.56
	UDP-Gal*f*	-28.39
	UDP-Gal*p*NAc	-27.57
	UDP-Gal*f*NAc	-21.92
drUGM reduced	UDP-Gal*p*	-31.52
	UDP-Gal*f*	-29.97
	UDP-Gal*p*NAc	-28.63
	UDP-Gal*f*NAc	-21.19

The substrate constant (K_s) is a measure of the affinity of the enzyme for the substrate. K_s is also known as the dissociation constant or binding affinity (K_d), the concentration at which the enzyme active site has a probability of half occupancy. Therefore, a low value for K_d indicates that the enzyme-substrate complex (ES) is stabilized through energetically favorable interactions, and the binding affinity is high. The change in free energy upon the binding of ligand, or $\Delta G_{binding}$, is expected to be negative if favorable interactions exist between the enzyme and ligand. The Michaelis constant (K_M) is equal to the substrate concentration at which the reaction rate is half its maximum value, and can be used as an *approximation* of K_d when:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P$$

Where k_1 = rate of the forward reaction
k_{-1} = rate of the reverse reaction
k_2 = rate of product formation
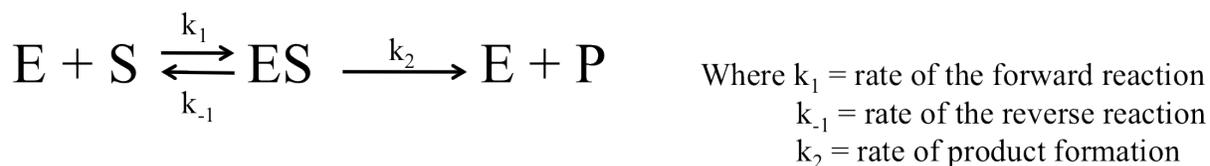
$$K_M = \frac{k_{-1} + k_2}{k_1} \qquad \text{(Equation 4.5)}$$

If $k_{-1} \gg k_2$, then:

$$K_M \approx \frac{k_{-1}}{k_1} = K_d \qquad \text{(Equation 4.6)}$$

When the rate constant of the reverse reaction (k_{-1}) is much larger than the rate constant of product formation (k_2), we assume that K_M is approximately equal to the dissociation constant (K_d). If K_M is approximately equal to the binding affinity (K_d), a low K_M value should result when energetically favorable interactions exist between the enzyme and substrate (ES) or the binding affinity is high. The docking score ($\Delta G_{binding}$ component) was compared with the experimental K_M value, which should be low if the free energy of binding is high. The data used for correlating kinetic data and $\Delta G_{binding}$ are summarized in Table 4.11.

Kinetic parameters for UNGM were calculated by monitoring the conversion of UDP-Gal*f* to the product UDP-Gal*p* using HPLC (Poulin *et al.*, 2009). The same technique was employed for kinetic analysis of UDP-Gal*f*NAc. The value of K_M for UDP-Gal*f* was 45 μM, which was correlated to a $\Delta G_{binding}$ = -18.98 in UNGM$_{red}$ and $\Delta G_{binding}$ = -25.45 in UNGM$_{ox}$ A (where R169 is the catalytic arginine in mobile loop 2). This model did not follow the expected correlation, as the lowest value for K_M (UDP-Gal*f*NAc = 40 μM) did not correlate to the largest change in $\Delta G_{binding}$ (UDP-Gal*f* = 25.45 μM). In UNGM$_{ox}$ B, the values of K_M and $\Delta G_{binding}$ for

UDP-Gal$f$ and UDP-Gal$f$NAc followed the expected correlation, as UDP-Gal$f$NAc was found to have the better K_M value and the larger free energy change.

The kinetic parameters for drUGM were calculated in the same manner as described earlier (Karunan Partha *et al*., 2009). The value of K_M for UDP-Gal$f$ was 55 μM, which correlated to a $\Delta G_{binding}$ = -29.97 for drUGM$_{red}$, and $\Delta G_{binding}$ = -28.39 for drUGM$_{ox}$. The value of K_M for UDP-Gal$f$ is higher for UNGM when compared to drUGM, therefore we expect a larger value for $\Delta G_{binding}$ in UNGM. Unexpectedly, the $\Delta G_{binding}$ for drUGM was more negative when compared to the $\Delta G_{binding}$ for UNGM. Although UNGM has been shown to bind UDP-Gal$f$ with higher affinity (a lower value for K_M) than drUGM, this was not evident in the results of our docking experiments. Lastly, since drUGM is not catalytically active on UDP-Gal$f$NAc, we do not have kinetic parameters to compare with the $\Delta G_{binding}$ derived from the docking experiments with this enzyme.

Table 4.11: Correlation of K_M with $\Delta G_{binding}$

Model Used	Substrate	$\Delta G_{binding}$	K_M (μM)
UNGM$_{ox}$ A	UDP-Gal$f$	-25.45	45 ± 3
	UDP-Gal$f$NAc	-21.99	40 ± 6
UNGM$_{ox}$ B	UDP-Gal$f$	-17.87	45 ± 3
	UDP-Gal$f$NAc	-26.94	40 ± 6
UNGM$_{red}$	UDP-Gal$f$	-18.98	45 ± 3
	UDP-Gal$f$NAc	-23.28	40 ± 6
drUGM$_{ox}$	UDP-Gal$f$	-28.39	55 ± 7
drUGM$_{red}$	UDP-Gal$f$	-29.97	55 ± 7

CHAPTER 5: DISCUSSION

5.1 UNGM holoenzyme structure

The UNGM holoenzyme shares the conserved overall fold and secondary structural elements of all other known structures of UGM. UNGM forms a homodimer, which is the only known type of quaternary structure that UGM adopts, as seen in the crystal structures of ecUGM, kpUGM, and drUGM. UNGM also shares the conserved mobile loop regions, which have previously been found to be essential for substrate binding. The substrate binding cleft is located adjacent to the isoalloxazine ring of FAD, with mobile loops 1 and 2 near the opening of the cleft. Both the oxidized and reduced forms of flavin adenine dinucleotide were observed in the structure of UNGM. Monomer A of UNGM holoenzyme structure contains FAD (in the inactive conformation), and monomer B contains the catalytically active conformation FADH⁻. The reduction of the flavin does not appear to affect the conformation of the mobile loops, as the loops are in the open conformation in both monomers. Therefore, binding of substrate is required for this enzyme to adopt the closed conformation. The active site consisted mainly of conserved residues known to be involved in the binding and stabilization of substrate in UGM homologues. We have proposed that the majority of interactions with the substrate occur via hydrogen bonding with these residues.

Several active site residues differed from other structures of UGM. Firstly, Arg59 in the sugar binding region has been found confer specificity for UDP-Gal*f*NAc. The interaction between the acetamido moiety of UDP-GalNAc and the guanidinium group of Arg59 (commonly a histidine residue) is speculated to stabilize the binding of these sugars within the sugar binding site of UNGM. As arginine is a bulkier residue than histidine, and it is able to project further into the sugar-binding site and may form a more stable interaction with N-acetylated sugars. Secondly, Arg168 (commonly a lysine residue) on mobile loop 2 is known to confer specificity for GalNAc sugars, and is believed to occupy the putative acetamido binding region adjacent to Arg59 in the sugar-binding region. Therefore, this residue may function to stabilize the active site conformation of Arg59 or directly interact with the acetamido group of the sugar. Alternatively, this residue may also stabilize the diphosphate backbone of UDP in the same manner as the conserved arginine, 169.

Several differences were observed between monomers A and B of the UNGM holoenzyme. Alternate conformations were observed for Arg59 and Glu285 in monomer B only. In addition, monomer B contained the reduced form of the flavin cofactor. Differences were also observed in the mobile loops, which are known to affect crystal packing in this structure. Therefore, we expect that packing of adjacent molecules altered the arrangement of these residues.

5.2 Co-crystallized structures of UNGM

The lack of any diffraction from the yellow co-crystals of UDP-Gal*p* indicated that the crystals were not well-ordered and most likely not composed of salt from the crystallization condition. Unstable packing within the crystals could be due to the lack of a reducing environment, which would result in oxidation of the FAD. This would decrease the affinity for the substrate, allowing it to be released from the active site.

In the highest resolution UNGM structure (holoenzyme), Arg169 forms a salt-bridge with Asp290 in one of the monomers. This interaction restricts the movement of the mobile loop 2, and would prevent the interaction with substrate. In the lower resolution structures, these interactions are not directly observed, however deemed possible within a range of 4.5 to 6 Å between Arg169 and Asp290. Within the UNGM inhibitor-soaked (UDP-5-deoxy-Gal*f*) structure, there was no density for the inhibitor observed within the active site. Also, there were no intermolecular interactions observed with mobile loop 2 in either monomer with the adjacent molecules, therefore we expect that the soaking experiments should have worked within these crystals. However, we did observe hydrogen-bond formation between Tyr215 and Asp161, as previously observed in the UNGM holoenzyme structure. Tyr86 from monomer A forms a hydrogen bond with the backbone oxygen of Gly24 of adjacent monomer B and vice versa.

The structure of UNGM crystallized in the presence of UDP-Glc (grown in the condition PEGs II #71) did not harbor any density for the substrate analog as expected (Figure 4.18). Also, there were no interactions observed between mobile loop 2 and the adjacent monomers. Once again, this result infers that the binding of the substrate is possible in the

active site. We assume that UDP-glucose participates in the crystallization of the protein in ways different from active site binding. Although this condition crystallized in the space group P2$_1$, the hydrogen bond formation between Tyr215 and Asp161 was observed in this structure. Crystal packing occurred in a different manner than observed in the holoenzyme structure as expected. Hydrogen bonding occurred with Tyr240 and the backbone oxygen of Lys249 (both monomers), and Tyr86 (B) with backbone oxygen of Gly24 (adjacent monomer A). Also, density was observed for a glycerol molecule near Lys102 (in both monomers) as previously observed in the holoenzyme structure. No other structures of UGM homologues have been found to contain glycerol molecules at the dimer interface.

The structure of UNGM crystallized in the presence of UDP-Glc (grown in the condition ComPAs #13) did not contain any density for the sugar nucleotide in the substrate binding site (Figure 4.19). Once again, there were no interactions with mobile loop 2 from either monomer with the symmetry-related molecules. Therefore, the ability of the substrate to access the substrate binding region of the active site was not hindered by the crystallographic packing. Hydrogen bonding interactions were observed between the residues Tyr215 and Asp161, Lys214 with the backbone oxygen of Asp161, and Lys249 with the backbone oxygen of Tyr240. Hydrogen bonding interactions were observed between Tyr215 and Asp161 in every crystal structure of UNGM regardless of space group. Therefore, mutagenesis of one these residues to a non-polar residue would prevent the formation of the hydrogen bond which appears to stabilize the crystal-packing in every structure. This destabilization may promote the formation of different interactions between molecules, which may favor the substrate-bound, closed-conformation of the structure.

Importantly, all of the structures that have substrate bound (drUGM and kpUGM) contain an α-helix within mobile loop 2, which confers additional rigidity to the region and may be important for stabilization of the complex through crystal contacts. In the holoenzyme structure of kpUGM, no secondary structure is observed in this loop, and therefore the secondary structure occurs upon interaction with substrate (Gruber *et. al*. 2009b). We hypothesize that the formation of an α-helix in mobile loop 2 is important for obtaining a crystal of UGM in complex with substrate.

5.3 Quality of intensity data from UNGM crystals

Commonly, the data is truncated before the R_{merge} reaches 0.6 to 0.8 to ensure that the data is of good quality (Equation 4.1; Karplus and Diederichs, 2012). The R_{merge} value is not a very good measure of data quality as it only measures the inconsistency between observations and takes no account of the improvement in the merged intensity by averaging many observations (R_{meas})(Evans, 2006).

$$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i I_i(hkl) \qquad \text{(Equation 4.1)}$$

At high resolution, R_{merge} diverges towards infinity. This is rationalized by the formula for R_{merge}, in which the average net intensity approaches zero, while the numerator is dominated by background noise and is constant (Karplus and Diederichs, 2012). Furthermore, R_{merge} and R-factors from refinement are commonly considered to follow the same criterion for truncation. Specifically, the data are considered not useful past a value of 0.6, however, there is no valid basis for this practice.

The data sets collected from all UNGM crystals exhibited high R_{merge} values, with a range of 0.159 to 0.25 overall and 0.334 to 0.714 in the highest resolution shell. As R_{merge} tends to increase with increasing multiplicity of the data processed. R_{meas} for the same data was 0.162 (0.771). Data collected from UNGM crystals share a high R-factor, with is consistent with the majority of published UGM structures (Table 5.1).

As a test, the data was processed with lower multiplicity (between 90 to 120 images). For example, the data set collected from UNGM co-crystallized with UDP-Glc (ComPAs #13) was processed with 120 images, which resulted in a large R_{merge} value of 0.159 overall (0.471 in highest resolution shell) despite the lower multiplicity. Repetition of this technique for the other data sets did not significantly improve the value of R_{merge} overall. In addition, high multiplicity is always desirable for a data set, as it provides an estimate of the accuracy that can be anticipated from the data collected (Berger, 2007). Ignoring a large portion of the data (in order to obtain a lower R_{merge}) lowers the level of accuracy we can expect from the data, and is not desirable for a high-resolution structure. Upon analysis of the data collected from the UNGM holoenzyme crystal, radiation damage or crystal slippage is likely the cause of the elevated R-factor, as R_{meas} (the multiplicity weighted version of the R-factor) was observed to

increase near the end of the data collection, which was also correlated to an increase in mosaicity of the diffraction in these frames. If the crystal is exposed to more radiation, we expect the mosaicity will also increase if radiation damage occurs. Furthermore, as the data did not exhibit symptoms of anisotropy, the crystal orientation is not likely the cause of this increase in mosaicity. The data sets discussed in this report exhibited an $I/\sigma(I)$ value above 2.0, high completeness, a large number of reflections, and high multiplicity, which indicated that there was a large amount of useable data present at the resolution(s) chosen for data truncation. Therefore, an elevated R_{merge} appears to be an inherent problem for these data sets, and should not be taken as the sole indicator of data quality due to the unreliability of this statistic. The majority of published UGM structures also have a high R_{merge} (Table 5.1).

Table 5.1: Merging R-factors of UGM homologues.

UGM homologue	R_{merge} or R_{sym}
AfUGM	10.7 (58.7)
AfUGM: UDP-Gal*p* (reduced)	11.4 (54.9)
AfUGM R327K:UDP-Gal*p*	15.5 (77.3)
AfUGM R237A: UDP-Gal*p*	23.8 (53.5)
DrUGM$_{ox}$	11.6 (68.9)
DrUGM$_{red}$	18.0 (61.0)
DrUGM:UDP	17.2 (57.8)
DrUGM:UMP	19.2 (54.6)
DrUGM:UDP-Glc	16.7 (59.9)
DrUGM:UDP-Gal*p*	18.9 (61.6)
MtUGM	10.2 (61.4)

5.4 Docking Studies

5.4.1 Comparison of binding modes for UNGM and UGM

Previous mechanistic studies have elucidated the requirement of the substrate to be within a close distance to the isoalloxazine ring system of FAD for catalysis. Importantly, the docking studies in both UNGM and drUGM consistently found that the sugar moiety of the substrate was adjacent to the isoalloxazine ring, although it was not always within distance for the formation of an adduct between to the two. The docking solutions were judged using two criteria: 1) Ligand position, and 2) fitness function comparison.

Previous work has shown that the majority of stabilization of the ligand occurs through interactions with the UDP moiety, which was present in every ligand tested in these studies (Karunan Partha *et al*., 2009a). Specifically, the UDP moiety of the substrate was consistently docked within bonding distance of the conserved tryptophan (Trp155 in UNGM; Trp184 in drUGM), which is known to stabilize the ribosyl moiety of uridine. In the control experiment, UDP-Gal*p* was removed from the substrate binding site of drUGM:UDP-Gal*p* structure, and successfully docked back into the same position. Therefore, we expected that other similar substrates should also bind the active site in a conserved manner.

The binding mode for UDP-Gal*p* is expected to be nearly identical in UNGM and drUGM, as both enzymes are known to catalyze the conversion of UDP-Gal*p* to UDP-Gal*f*. UDP-Gal*p* was docked in near identical positions in both enzymes. The substrate UDP-Gal*f* was also expected to dock in a similar mode in both enzymes, however it did not dock in a productive binding mode in drUGM, although the docking mode was productive in the model of UNGM.

In drUGM, the docking of UDP-Gal*p*NAc yielded inconsistent results. This ligand is not a substrate of drUGM, however the docking studies with Chemscore revealed that it was bound in a productive binding mode. The UDP moiety is stabilized by interactions with the residues Trp184, Arg198, and Tyr370, and the acetamido moiety of the sugar is within bonding distance of the conserved histidine (His88). CHEMPLP was used as a second scoring function and also generated a productive binding mode. However, the acetamido moiety of Gal*p*NAc was not oriented in a favorable position for stabilization by the conserved histidine (His88) in

102

the sugar binding region. Therefore, the active site of drUGM does not always favor the binding of the acetamido moiety, and does not consistently bind this substrate favorably.

In comparison, the docking of UDP-Gal*p*NAc into UNGM was favorable for the formation of the FAD-galactose adduct as well as the stabilization of the acetamido moiety by Arg59. This result was consistent for both scoring methods utilized in this study. UNGM is catalytically active on UDP-Gal*p*NAc, and therefore these results are compatible with kinetic analysis that has shown Arg59 is important for the recognition of GalNAc sugars. We can conclude that UNGM binds UDP-Gal*p*NAc more consistently than drUGM binds this substrate. The docking of UDP-Gal*f*NAc was attempted in both UNGM and drUGM. We expected that drUGM would not accommodate this substrate in a productive binding mode. Although the docking of both substrates was deemed to be productive, the acetamido group of Gal*f*NAc was not within an acceptable distance for stabilization by His88 in drUGM. In UNGM, Arg59 formed a hydrogen bond with the acetamido moiety of Gal*f*NAc. Therefore, these results favor that this ligand binds more favorably to UNGM than drUGM, which is consistent with the enzyme's ability to recognize and convert UDP-Gal*f*NAc as a substrate.

In both enzymes, the oxidized and reduced models were not significantly different in docking position of the substrates. However, the fitness scores were generally higher for the reduced FADH⁻ models and corresponded to a larger $\Delta G_{binding}$ (Table 4.9). In drUGM$_{red}$, the natural substrates UDP-Gal*p* and UDP-Gal*f* produced higher fitness scores in comparison with GalNAc sugars, which is consistent with drUGM's inability to convert GalNAc sugars. In addition, docking studies using the crystal structure of UNGM produced significantly improved results when compared to previous homology modeling experiments in Autodock 4.0, using a model based on ecUGM (Poulin *et al*., 2009; Morris *et al*., 1998). Specifically, the docking of UDP-Gal*f*NAc in the modified UNGM structure was within a productive distance for catalysis and in an ideal orientation for stabilization, whereas the previous study identified that this substrate was bound in an inactive, non-productive conformation. Additionally, the docking results for UDP-Gal*f* were consistent between our studies in GOLD and the previous docking studies in Autodock.

5.4.2 Comparison of UNGM oxidized models A and B

The binding modes for the UNGM oxidized models A and B were nearly identical, with only slight variations observed for the distance between the sugar and FAD cofactor (within 0.5 Å of each other). The results from docking studies with the UNGM oxidized model B were consistent with our hypothesis that the GalNAc sugars would generate better fitness scores when compared to the substrates without an acetamido moiety. Specifically, UDP-Gal*p*NAc and UDP-Gal*f*NAc had significantly larger fitness scores when compared to the scores for UDP-Gal*p* and UDP-Gal*f*. This result favors that Arg168 (in model B) has an additional role in the stabilization of *N*-acetylated sugars. This finding is consistent with the kinetic results from R168K mutants, in which the K_M value for UDP-Gal*f*NAc increased, but was unchanged for UDP-Gal*f*. Also, the lysine mutant showed a 2-fold increase in selectivity for UDP-Gal*f* over UDP-Gal*f*NAc. These results support that Arg168 has an important role in the binding and stabilization of Gal*f*NAc. In model A, the fitness scores did not vary significantly between the GalNAc and Gal sugar nucleotides. Therefore, Arg169 in model A does not favorably bind one substrate over another; rather it was observed to stabilize the diphosphate region of all of the substrates.

5.5 Dual specificity of UNGM

The lack of a UNGM structure in complex with substrate has prevented the direct analysis the interactions between the protein and ligand. Specifically, the structural roles of Arg59 and Arg168 are of significant interest due to their effect on the enzyme's ability to catalyze UDP-GalNAc sugar nucleotides. The results of docking studies with UNGM have shown that both of these residues aid in the stabilization of the substrate(s). Arg59 is within hydrogen bonding distance of the carbonyl oxygen of the acetamido moiety of GalNAc, which is consistent with a role for stabilization. Previously, it has been hypothesized that this stabilization is key for the preventing the formation of an oxazoline intermediate (Poulin *et al*., 2009) (Figure 5.1). Arg168 also contributes to the specificity of UNGM, and likely plays a role in stabilization of GalNAc sugars. Docking studies have indicated that Arg168 plays a role in stabilization of the diphosphate region of the substrates. The long side chain of Arg168 may also play a role in the positioning of the sugar, which is has also been proposed for Arg182 in

the structure of afUGM R182K mutant (van Straaten *et al.*, 2012). Therefore, the dual specificity of UNGM is likely due to the concerted contributions of several residues, Arg59 and Arg168, amongst others not yet identified, in the stabilization of the acetamido moiety of GalNAc.

A

B
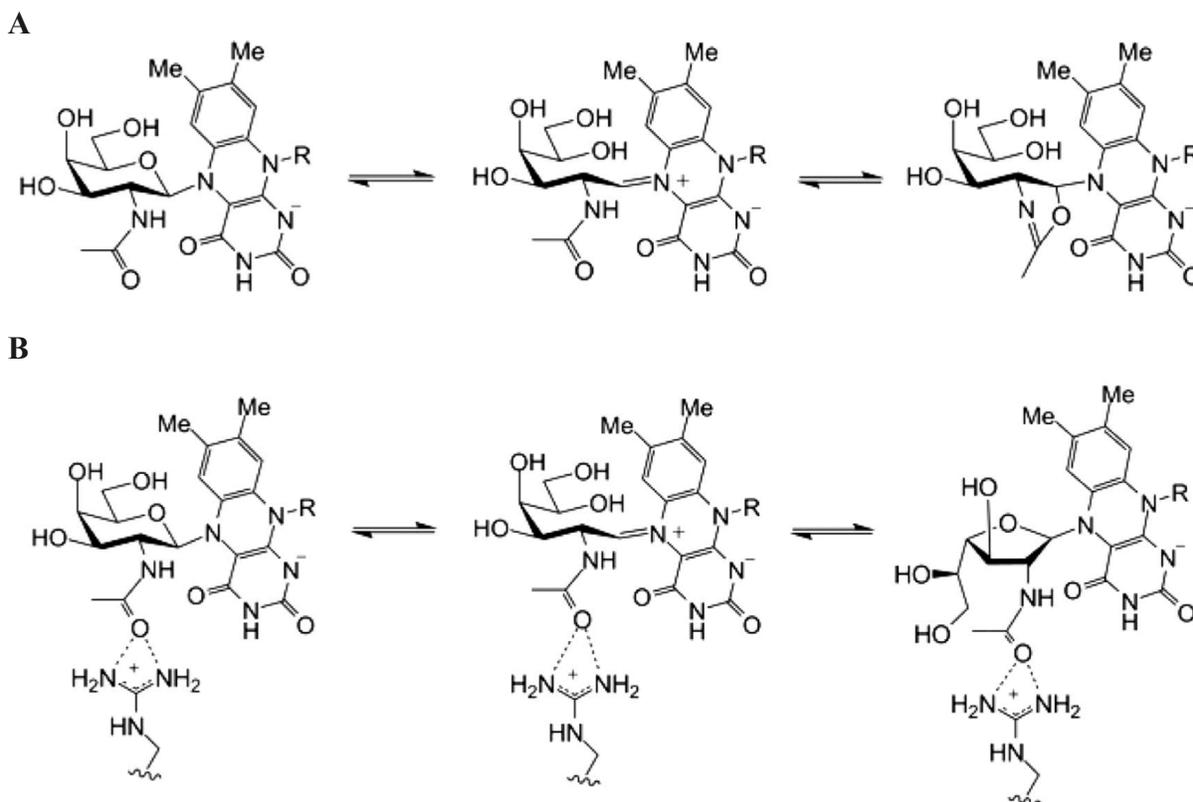
Figure 5.1: Formation of oxazoline intermediate. A- The occurrence of an unproductive intramolecular reaction within the acetamido group of GalNAc can form an FAD-bound oxazoline intermediate, which prevents conversion to GalfNAc. **B-** Arg59 forms a hydrogen bond with the acetamido group of GalNAc that would prevent formation of an oxazoline intermediate and allow for conversion of Gal*p*NAc. This research was originally published in the Journal of Biological Chemistry (Poulin *et al.*, 2009). © The American Society for Biochemistry and Molecular Biology.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

The crystal structure of UDP-*N*-acetylgalactopyranose mutase holoenzyme has shown the cofactor FAD bound to each monomer, which is essential for the catalytic activity of this enzyme. Furthermore, the FAD cofactor was in the oxidized or inactive conformation in one of the monomers, and in the catalytically active reduced conformation in the other. The mobile loop arrangement was in an open conformation for both monomers, which was consistent with other holoenzyme structures of UGM. Due to the high degree of conservation within the UNGM active site, we are confident in the roles of the conserved residues described in this research. Furthermore in the holoenzyme, water molecules form hydrogen bonds with many of the proposed catalytic residues, indicating that the environment of the active site is favorable for substrate binding. The roles of the active site residues in UNGM are based on comparisons with complex structures of UGM, but we do not have a structure of UNGM in complex with any of its known substrates to confirm these claims crystallographically.

Co-crystallization of UNGM was attempted using UDP-Glc, UDP-Gal*p*, and UDP to capture the substrate within the active site. Even though the co-crystallization method has recently been successful for drUGM and kpUGM, previous work on the UGM homologous structure with the highest sequence identity (59 %), ecUGM, was also unsuccessful in co-crystallization and soaking experiments. This type of enzyme has been notoriously difficult to co-crystallize, although the holoenzyme has been crystallized in several UGM homologues. The movement of the loop regions is a possible hindrance to co-crystallization, which can disrupt crystal packing in the presence of substrate. The conserved arginine (R198 in drUGM) in mobile loop 2 is primarily responsible for the stabilization of the closed conformation by forming a salt-bridge with the α-phosphate of the substrate and also by cation-π interaction with a conserved tyrosine (Tyr370 in drUGM) (Karunan Partha *et al*., 2009). All known UGM's contain the aforementioned conserved residues that stabilize the closed conformation; however all homologues do not form stable structures in complex with substrate (Figure 4.9). In the holoenzyme UNGM structure, the conserved arginine forms a salt bridge with Asp290 of a symmetry-related molecule, promoting the open conformation. Due to the low resolution of the co-crystallized structures, the formation of this salt-bridge interaction is not reliably confirmed;

however these residues are in close proximity to each other. If this interaction does occur, a stable closed conformation would not be favored in any of the structures of UNGM. Our results show that the mobile loop 2 in UNGM does not form a stable complex with substrate in crystalline form. The mobile loops in UNGM remained in an open conformation despite the presence excess substrate and a reducing agent to stabilize substrate binding during crystallization. In order to co-crystallize UGM, our lab has studied drUGM, which shares 36 % identity to UNGM, the lowest identity structure discussed in this study. Therefore, the exterior surface of drUGM may be more favorable for intermolecular contacts that stabilize the closed loop conformation, which would allow for the trapping of substrate within the active site.

The interactions observed in the protein-ligand complex of each docking were favorable for stabilization, which should correlate to a high affinity for the substrate. All of the substrates tested appear to bind in the conserved UDP binding region. The reliability of protein-ligand docking is a major challenge, and it is not apparent that binding affinity can be accurately determined using this method. The docking models were not found to accurately predict the binding affinity of UNGM, as the experimental value of K_M did not appear to agree with the docking scores ($\Delta G_{binding}$). Specifically, the binding affinity estimations for UDP-Gal$f$ in drUGM and UNGM did not fit the approximation we have proposed. These are estimations, and UGMs from different species may correlate better using this approximation, or a more accurate closed conformation model of UNGM may also improve these results.

Better binding interactions, as evidenced by the larger fitness scores (and a larger $\Delta G_{binding}$), were observed for the drUGM models, which is likely due to subtle differences in the conformation of the active site residues. The model for drUGM is expected to be more accurate than the UNGM model(s), as the structure of drUGM is in complex with UDP-Gal$p$ and UNGM was not in complex with any substrate. The interactions with substrate have been found to cause conformational changes to active site residues in drUGM, and therefore it is expected that similar changes occur in UNGM. The *N*-acetylated sugars appeared to bind drUGM and UNGM in a highly similar manner, as both enzymes bind the UDP moiety of the substrate in a conserved binding pocket. The sugar moiety of GalNAc was more consistently docked in a productive binding mode in UNGM, although drUGM also accommodates this sugar. Based on these results, we cannot conclude why drUGM does not also catalyze the

conversion of GalNAc. The bifunctional activity (or dual specificity) of UNGM is likely due to the concerted contributions of several residues, Arg59 and Arg168, amongst others not yet identified, in the stabilization of the acetamido moiety of GalNAc. The holoenzyme of UNGM is expected to differ significantly from the reduced, substrate-bound form. Major changes in the loop conformation are expected upon interaction with the substrate, as well as conformational changes in the active site residues.

Modeling studies identified that Tyr345 and Trp155 are likely to move toward the active site, as the re-positioning of these residues (to mimic drUGM) was essential for consistent docking results. Arg169 was observed to stabilize the diphosphate region of every substrate in both drUGM and UNGM, which is in accordance with crystallographic studies. Our results favor that Arg168 contributes to the specificity of UNGM, and likely plays a role in stabilization of GalNAc sugars. Arg168 appears to stabilize the diphosphate region, as previously proposed for the conserved arginine, Arg169. We conclude that both of these residues are capable of substrate stabilization, but only Arg168 appears to favor GalNAc binding. These studies also favor that Arg59 in UNGM has a role in the stabilization of the acetamido moiety in both the pyranose and furanose forms of GalNAc.

6.2 Future Research

We did not obtain a structure of UNGM in complex with a ligand, and therefore site-directed mutagenesis of the residues Tyr215 and Asp161, which are involved in crystal stabilization through hydrogen bonding, could change the packing of the protein molecules such that a complex structure would be favored. This is more likely to occur if the enzyme crystallizes in a different space group. This technique has improved the crystal packing interactions in other proteins, although it has not been attempted in UGM homologues. Further crystallization screening may prove successful, using different screening kits, additives or ligands.

If further docking studies were endeavored, a closed loop crystal structure of UNGM would be desirable as little adjustment would be necessary in the modeling process, which would reduce the errors in the model. Attainment of a closed loop structure of UNGM may

occur by further crystallization screening, or the use of other reducing agents for flavin reduction. If a better model is available, docking experiments with an extensive library of inhibitors will lead to further characterization of the active site. Identification of inhibitors that bind UNGM and utilize the nucleophilicity of the flavin N5 are of particular interest. These inhibitors should impair capsular polysaccharide biosynthetic pathway, and therefore act as antibacterial agents.

The structure of UNGM may be used for designing and screening inhibitors that are recognized in a similar way to UDP-Gal*p*NAc, which is common to the CPS, *N*-linked glycans, and the LOS. The inhibition of UNGM is not expected to directly prevent the *N*-linked modification of proteins or LOS biosynthesis. However, inhibitors or substrate analogs for UNGM may be incorporated into the *N*-linked glycan structures, which would prevent their proper assembly. Although the exact roles of these structures are poorly understood, there is evidence that they are also crucial for virulence in this organism. The disruption of the biosynthetic processes that require GalNAc will likely hinder the ability of this pathogen to invade host cells.

References:

Afonine, P.V., Grosse-Kunstleve, R.W., and Adams, P.D. (2005). The Phenix Refinement Framework. CCP4 Newsl. *42*, contribution 8, 1-7.

Allos, B.M. (2001). *Campylobacter jejuni* infections: Update on emerging issues and trends. Clin. Infect. Dis. *32,* 1201-1206.

Allos, B.M., and Blaser, M.J. (2009). *Campylobacter jejuni* and Related Species. In Principles and Practice of Infectious Diseases 7th ed. Mandell, G.L., Bennett, J.E., and Dolin, R. eds. (Philadelphia: Elsevier) pp. 2793-2802.

Bacon, D.J., Szymanski, C.M., Burr, D.H., Silver, R.P., Alm, R.A., and Guerry, P. (2001). A phase variable capsule is involved in virulence of *Campylobacter jejuni* 81-176. Mol. Microbiol. *40*, 769–777.

Barlow, J.N., Girvin, M.E., and Blanchard, J.S. (1999). Positional isotope exchange catalyzed by UDP-galactopyranose mutase. J. Am. Chem. Soc. *121*, 6968-6969.

Basso, L.A., Pereira da Silva, L.H., Fett-Neto, A.G., Azevedo Jr., W.F., Moreira, I.S., Palma, M.S., Calixto, J.B., Filho, S.A., dos Santos, R.R., Soares, M.B.P., and Santos, D.S. (2005). The use of biodiversity as a source of new chemical entities against defined molecular targets for treatment of malaria, tuberculosis, and T-cell mediated diseases. Mem. Inst. Oswaldo. Cruz. *100*, 475-506.

Beis, K., Srikannathasan, V., Lui, H., Fullerton, S.W., Bamford, V.A., Sanders, D.A.R., Whitfield, C., McNeil, M.R., and Naismith, J.H. (2005). Crystal Structures of *Mycobacteria tuberculosis* and *Klebsiella pneumoniae* UDP-Galactopyranose Mutase in the Oxidized State and *Klebsiella pneumoniae* UDP-Galactopyranose Mutase in the (Active) Reduced State. J. Mol. Biol. *348*, 971-982.

Berger, J.M. (2007). Looking Past the Pictures: Evaluating X-Ray crystal Structure Papers. In Evaluating Techniques in Biochemical Reasearch, D. Zuk, ed. (Cambridge, MA: Cell Press), http://www.cell-press.com/misc/page?page=ETBR, pp. 7-12.

Bernard, M., and Latge, J.-P. (2001). *Aspergillus fumigatus* Cell Wall: Composition and

Biosynthesis. Med. Mycol. *39* (Suppl. I), 9-17.

Besra, G.S., Khoo, K.H., McNeil, M.R., Dell, A., Morris, H.R., and Brennan, P.J. (1995). A new interpretation of the structure of the mycolyl-arabinogalactan complex of *Mycobacterium tuberculosis* as revealed through characterization of oligoglycosylalditol fragments by fast-atom bombardment mass spectrometry and 1H nuclear magnetic resonance spectroscopy. Biochemistry *34*, 4257-4266.

Britton, D. (1972). Estimation of twinning parameter for twins with exactly superimposed reciprocal lattices. Acta Crystallogr. A. *28*, 296-297.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaninathan, S., and Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J. Comp. Chem. *4*, 187-217.

Chad, J.M., Sarathy, K.P., Gruber, T.D., Addala, E., Kiessling, L.L., and Sanders, D.A.R. (2007). Site-directed mutagenesis of UDP-galactopyranose mutase reveals a critical role for the active-site, conserved arginine residues. Biochemistry *46*, 6723– 6732.

Chen, V.B., Arendall, B.W.III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). Molprobity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D. *66*, 12-21.

Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallogr. D. *50*, 760-763.

Corbett, D., and Roberts, I.S. (2009). Genetics and Regulation of Bacterial Polysaccharide Expression in Human Pathogenic Bacteria. In Bacterial Polysaccharides: Current Innovations and Future Trends. Matthias Ullrich, ed. (Norwich, UK: Horizon Scientific Press), pp. 69-86.

Diederichs, K., and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat. Struct. Biol. *4*, 269-275.

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

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D. *66*, 486-501.

Errey, J.C., Mann, M.C., Fairhurst, S.A., Hill, L., McNeil, M.R., Naismith, J.H., Percy, J.M., Whitfield, C., and Field, R.A. (2009). Sugar Nucleotide Recognition by *Klebsiella pneumoniae* UDP-D-Galactopyranose Mutase: Fluorinated Substrates, Kinetics and Equilibria. Org. Biomol. Chem. *7*, 1009-1016.

Evans, P. (2006). Scaling & assessment of data quality. Acta Crystallogr. D. *62*, 82-82.

Fauchere, J.L., Rosenau, A., Veron, M., Moyen, E.N., Richard, S., and Pfister, A. (1986). Association with HeLa cells of *Campylobacter jejuni* and *Campylobacter coli* isolated from human feces. Infect. Immun. *54*, 283-287.

Feng, L., Senchenkova, S.N., Yang, J., Shashkov, A.S., Tao, J., Guo, H., Cheng, J., Ren, Y., Knirel, Y.A., Reeves, P.R., and Wang, L. (2004). Synthesis of the heteropolysaccharide O antigen of *Escherichia coli* O52 requires an ABC transporter: Structural and genetic evidence. J. Bacteriol. *186*, 4510-4519.

Friedman, J., Neimann, J., Wegener, H.C., and Tauxe, R.V. (2000). Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations. Campylobacter, 2nd Ed. I. Nachimkin, M.J. Blaser, eds. (Washington, DC: ASM Press), pp. 121-138.

Fullerton, S.W.B., Daft, S., Sanders, D.A.R., Ingledew, W.J., Whitfield, C., Chapman, S.K., and Naismith, J.H. (2003). Potentiometric analysis of UDP-galactopyranose mutase: Stabilization of the flavosemiquinone by substrate. Biochemistry *42*, 2104-2109.

Ghisla, S., and Massey, V. (1989). Mechanisms of flavoprotein-catalyzed reactions. Eur. J. Biochem. *181*, 1-17.

Gilbert, M., Brisson, J.R., Karwaski, M.F., Michniewicz, J., Cunningham A.M., Wu, Y., Young, N.M., and Wakarchuk, W.W. (2000). Biosynthesis of Ganglioside Mimics in *Campylobacter jejuni* OH4384. J. Biol. Chem. *275*, 3896-3906.

Gilbert, M., Karwaski, M.F., Bernatchez, S., Young, M.N., Taboada, E., *et al*. (2002). The

genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, Campylobacter jejuni. Biosynthesis of sialylated ganglioside mimics in the core oligosaccharides. J. Biol. Chem. *277*, 327-337.

Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. Bioinformatics *15,* 305-308.

Gruber, T.D., Borrock, J.M., Westler, W.M., Forest, K.T., and Kiessling, L.L. (2009a). Ligand Binding and Substrate Discrimination by UDP-Galactopyranose Mutase. J. Mol. Biol. *391*, 327-340.

Gruber, T.D., Westler, W.M., Kiessling, L.L., and Forest, K.T. (2009b). X-ray crystallography reveals a reduced substrate complex of UDP-galactopyranose mutase poised for covalent catalysis by flavin. Biochemistry *48*, 9171-9173.

Guan, S., Clarke, A.J., and Whitfield, C. (2001). Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of *Klebsiella pneumoniae*. J. Bacteriol. *182*, 3318-3327.

Hehre, W., and Ohlinger, S. (2011). Spartan'10 Tutorial and User's Guide. Irvine, CA: Wavefunction, Inc.

Höltje, J.V. (1998). Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. Microbiol. Mol. Biol. Rev. *62*, 181-203.

Huang, Z., Zhang, Q., and Lui, H. (2003). Reconstitution of UDP-galactopyranose mutase with 1-deaza-FAD and 5-deaza-FAD: analysis and mechanistic implications. Bioorg. Chem. *31*, 494-502.

Itoh, K., Huang, Z., and Lui, H. (2007). Synthesis and analysis of substrate analogues for UDP-galactopyranose mutase: Implication for an oxocarbenium ion intermediate in the catalytic mechanism. Org. Lett. *9*, 879-882.

Jacobs, B.C., Endtz, H., van der Meche, F.G., Hazenberg, M.P., Achtereekte, H.A., and van Doorn, P.A. (1995). Serum anti-GQ1b IgG antibodies recognize surface epitopes on *Campylobacter jejuni* from patients with Miller Fisher syndrome. Ann. Neurol. *37*, 260–264.

Jann, K., and Jann, B. (1997). Capsules of *Escherichia coli*. In *Escherichia coli* Mechanisms of Virulence, M. Sussman, ed. (Cambridge, The Press Syndicate of the University of Cambridge), pp. 113-144.

Jannsson, P.E. (1999). Endotoxin in Health and Disease. Brade, H., Opal, S.M., Vogel, S.N., Morrison, D.C., eds. (New York/Basel: Marcel Dekker), pp. 155-78. In Endotoxin in Health and Disease. Brade, H., Opal, S.M., Vogel, S.N., Morrison, D.C., eds. (New York/Basel: Marcel Dekker), pp. 155-78.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795-800.

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

Kaldor, J., and Speed, B.R. (1984). Guillain-Barré syndrome and *Campylobacter jejuni*: a serological study. Br. Med. J. *288*, 1867–1870.

Karplus, P.A., and Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. Science *336*, 1030-1033.

Karunan Partha, S., Sadeghi-Khomami, A., Slowski, K., Kotake, T., Thomas, N.R., Jakeman, D.L., and Sanders, D.A.R. (2010). Chemoenzymatic synthesis inhibition studies, and X-ray crystallographic analysis of the phospho analog of the UDP-Galp as an inhibitor and mechanistic probe for UDP-galactopyranose mutase. J. Mol. Biol. *403*, 578-590.

Karunan Partha, S., van Straaten, K.E., and Sanders, D.A.R. (2009). Structural Basis of Substrate Binding to UDP-galactopyranose Mutase: Crystal Structures in the Reduced and Oxidized State Complexed with UDP-galactopyranose and UDP. J. Mol. Biol. *394*, 864-877.

Keegan, R.M., and Winn, M.D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. Acta Crystallogr. D. *63*, 447-457.

Knirel, Y.A, and Kochetkov, N.K. (1994). Structure of lipopolysaccharides from gram-negative bacteria. III. Structure of O-specific polysaccharides. Biochemistry –USSR+. *59*, 1784-1851.

Kremer, L., Dover, L.G., Morehouse, C., Hitchin, P., Everett, M., Morris, H.R., Dell, A., Brennan, P.J., McNeill, M.R., Flaherty, C., Duncan, K., and Bersa, G.S. (2001). Galactan biosynthesis in *Mycobacterium tuberculosis*: Identification of a bifunctional UDP-Galactofuranosyltransferase. J. Biol. Chem. *276*, 26430-26440.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D. *60*, 2256-2268*.*

de Lederkremer, R.M., and Colli, W. (1995). Galactofuranose-containing Glycoconjugates in Trypanosomatids. Glycobiology *5*, 547-552.

Mansoorabadi, S.O., Thibodeaux, C.J., and Liu, H. (2007). The Diverse Roles of Flavin Coenzymes - Nature's Most Versatile Thespians. J. Org. Chem. *72*, 6329-6342.

Massey, V. (2000). The chemical and biological versatility of riboflavin. Biochem. Soc. Trans. *28*, 283-296.

McConville, M.J., Homans, S.W., Thomas-Oates, J.E., Dell, A., and Bacic, A. (1990). Structures of the Glycoinositolphospolipids from *Leishmania major*. A Family of Novel Galactofurnanose-containing Glycolipids. J. Biol. Chem. *265*, 7385-7394.

Mikuŝová, K., Yagi, T., Stern, R., McNeil, M.R., Besra, G.S., Crick, D.C., and Brennan, P.J. (2000). Biosynthesis of the galactan component of the mycobacterial cell wall. J. Biol. Chem. *275*, 33890-33897.

Mills, N. (2006). ChemDraw Ultra 10.0. J. Am. Chem. Soc. *128*, 13649–13650.

Mølbak, K., and Havelaar, A. (2008). Burden of illness of campylobacteriosis and sequelae. Campylobacter, 3rd Ed. I. Nachamkin, C. Szymanski, and M.J. Blaser eds., (Washington, DC: ASM Press), pp. 151-162.

Monteio, M.A., Baqar, S., Hall, E.R., Chen, Y. H., Porter, C.K., Bentzel, D.E., Applebee, L., and Guerry, P. (2009). Capsule Polysaccharide Conjugate Vaccine against Diarrheal Disease Caused by *Campylobacter jejuni*. Infect. Immun. *77*, 1128-1136.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olsen, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. *19*, 1639-1662.

Murphy, K.P. (1999). Predicting Binding Energetics from Structure: Looking Beyond ΔG°. Med. Res. Rev. *19,* 333-339.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood method. Acta Crystallogr. D. *53*, 240-255.

Nassau, P.M., Martin, S.L., Brown, R.E., Weston, A., Monsey, D., McNeil, M.R., and Duncan, K. (1996). Galactofuranose Biosynthesis in Escherichia coli K-12: Identification and Cloning of UDP-Galactopyranose Mutase. J. Bacteriol. *178*, 1047-1052.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology. Volume 276: Macromolecular Crystallography, part A, C.W. Carter, Jr. *et al*., eds., (New York: Academic Press), pp. 307-326.

Padilla, J.E., and Yeates, T.O. (2003). A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. Acta Crystallogr. D. *59*, 1124-1130.

Parker, C.T., Miller, W.G., Horn, S.T., and Lastovica, A.J. (2007). Common genomic features of *Campylobacter jejuni* subsp. doylei strains distinguish them from *C. jejuni* subsp. jejuni. BMC Microbiol. *7*, 1-9.

Peltier, P., Euzen, R., Daniellou, R., Nugier-Chauvin, C., and Ferrières, V. (2008). Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. Carbohydr. Res. *343*, 1897-1923.

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. Acta Crystallogr. D. *55*, 1718-1725.

Pitout, J.D., Hanson, N.D., Church, D.L., and Laupland, K.B. (2004). Population-based laboratory surveillance for *Escherichia coli*-producing extended-spectrum beta-lactamases: importance of community isolates with blaCTX-M genes. Clin. Infect. Dis. *38*, 1736-1741.

Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002). The CCP4 molecular-graphics project. Acta Crystallogr. D. *58*, 1955-1957.

Poulin, M.B., Nothaft, H., Hug, I., Feldman, M.F., Szymanski, C.M., and Lowary, T.L. (2009). Characterization of a Bifunctional Pyranose-Furanose Mutase from *Camplyobacter jejuni* 11168. J. Biol. Chem. *285*, 493-501.

Poulin, M.B., Zhou, R., and Lowary, T.L. (2012). Synthetic UDP-galactofuranose analogs reveal critical enzyme-substrate interactions in GlfT2-catalyzed mycobacterial galactan assembly. Org. Biomol. Chem. *10*, 4074-4087.

Raetz, C.R., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. Annu. Rev. Biochem. *71*, 635-700.

Ramachandran, S., Kota, P., Ding, F., and Dokholyan, N.V. (2011). Automated minimization of steric clashes in protein structures. Proteins *79*, 261-270.

Ramakrishnan, C., and Ramachandran, G.N. (1965). Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformation for a pair of peptide units. Biophys. J. *5*, 909-933.

Roberts, I.S. (1996). The Biochemistry and genetics of capsular polysaccharide production in bacteria. Annu. Rev. Microbiol. *50*, 285-315.

Rould, M.A., Perone, J.J., and Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Nature *352*, 213-218.

Sanders, D.A.R, Staines, A.G., McMahon, S.A., McNeil, M.R., Whitfield, C., and Naismith, J.H. (2001). UDP-galactopyranose mutase has a novel structure and mechanism. Nat. Struct. Biol. *8*, 858-863.

Sarvas, M., and Nikaido, H. (1971). Biosynthesis of T1 antigen in *Salmonella*: origin of D-galactofuranose and D-ribofuranose residues. J. Bacteriol. *105*, 1063-1072.

Scheffers, D.-J., and Pinho, M.G. (2005). Bacterial Cell Wall Synthesis: New Insights from Localization Studies. Microbiol. Mol. Biol. Rev. *69*, 585-607.

Soltero-Higgin, M., Carlson, E.E., Gruber, T.D., and Kiessling, L.L. (2004). A unique catalytic mechanism for UDP-galactopyranose mutase. Nat. Struct. Mol. Biol. *11*, 539-543.

Sousa, R., Lafer, E.M., and Wang, B.C. (1990). Preparation of Crystals of T7 RNA Polymerase Suitable for High-Resolution X‑ray Structure Analysis. J. Cryst. Growth, *110*, 237-246.

Splain, R.A., and Kiessling, L.L. (2010). Synthesis of galactofuranose-based acceptor substrates for the study of the carbohydrate polymerase GlfT2. Bioorg. Med. Chem. *18*, 3753-3759.

St. Michael, F., Szymanski, C. M., Li, J., Chan, K.H., Khieu, N.H., Larocque, S., Wararchuk W.W., Brisson, J.R., and Monteiro, M.A. (2002). The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. Eur. J. Biochem. *269*, 5119-5136.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA. *103*, 8060-8065.

Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., and Guerry, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol. Microbiol. *32*, 1022–1030.

Tappe, D., Schulze, M.H., Oesterlein, A., Abele-Horn, M., Baron, S., Durchhloz, D., Langen, H.-J., Jany, B., and Schoen, C. (2012). Molecular Detection of *Campylobacter jejuni* as a Cause of Culture-Negative Spondylodiscitis. J. Clin. Microbiol. *50*, 1499-1500.

The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673-80.

Valtonen, V.V., Sarvas, M., and Mäkelä, P.H. (1971). The Effect of T1 Antigen on the Virulence of Salmonella Strains for Mice. J. Gen. Microbiol. *69*, 99-106.

Vagin, A.A., and Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. J. Appl. Crystallogr. *30*, 1022-1025.

Vaguine, A.A., Richelle, J., and Wodak, S.J. (1999). SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and their agreement with atomic model. Acta Crystallogr. D. *55*, 191-205.

van Straaten, K.E., Routier, F.H., Sanders, D.A.R. (2012). Structural Insight into the Unique Substrate Binding Mechanism and Flavin Redox State of UDP-galactopyranose Mutase from *Aspergillus fumigatus*. J. Biol. Chem. *287*, 10780-10790.

Vaughan-Shaw, P.G.L., Rees, J.R., White, D., and Burgess, P. (2010). *Campylobacter jejuni* cholecystitis: a rare but significant clinical entity. BMJ Cases Rep. http://dx.crossref.org/10.1136%2Fbcr.10.2009.2365.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D. (2003). Improved Protein-Ligand Docking using GOLD. Proteins *52*, 609-623.

Yeates, T.O. (1988). Simple statistics for intensity data from twinned specimens. Acta Crystallogr. A. *44*, 142-144.

Young, K.T., Davis, L.M., and DiRita, V.J. (2007). *Campylobacter jejuni*: molecular biology and pathogenesis. Nature Rev. Microbiol. *5*, 665-79.

Yuan, Y., Wen, X., Sanders, D.A.R., and Pinto, M.B. (2005). Exploring the Mechanism of Binding of UDP-Galactopyranose Mutase by STD-NMR Spectroscopy and Molecular Modeling. Biochemistry *44*, 14080-14089.

Weston, A., Stern, R.J., Lee, R.E., Nassau, P.M., Monsey, D., Martin, S.L., Scherman, M.S., Besra, G.S., Duncan, K., and McNeill, M.R. (1998). Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. Tuber. Lung Dis. *78*, 123-131.

Whitfield, C. (2006). Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. Annu. Rev. Biochem. *75*, 39-68.

Zhang, Q., Lui, H. (2000). Studies of UDP-D-Galactopyranose Mutase from *Escherichia coli*, An Unusual Role of Reduced FAD in Catalysis. J. Am. Chem. Soc. *122*, 9065-9070.

Appendix A:

Overview:

      The objective of this study is to solve the structure of IolG4 from *Lactobacillus plantarum* by X-ray crystallography, which will provide a platform for elucidating the catalytic mechanism and substrate specificity of the enzyme. The significance of the proposed research is to understand the involvement of the incomplete *iol* operon in MI metabolism of *L. plantarum*. The specific goals of the project are to crystallize highly purified IolG4, solve the structure of the protein, and elucidate its role in the metabolism of inositol-related compounds.

A.1 Literature Review

      The genome of *Lactobacillus plantarum* WCSF1 is the largest known amongst lactic acid bacteria (LAB) as well as the first *Lactobacillus* species to be sequenced (Kleerbezum, *et al.*, 2003). *L. plantarum* has been found to inhabit many different environmental niches, including the gastrointestinal tract (GI) of humans (Ahrne *et al.*, 1998). Although its function in the GI is still unknown, research has pointed out that the presence of certain strains of *L. plantarum* are beneficial and mutalisitic (Moran *et al.*, 2000; Wels, 2007). For example, *L. plantarum* 299v has been tested for its probiotic benefits and is currently marketed as a probiotic supplement (Cunningham-Rundles *et al.*, 2000). Probiotic benefits include the increase in cytokine production, enhancement of phagocytosis of pathogens, and the decrease in symptoms of Irritable Bowel Syndrome (Kleerbezum *et al.*, 2003; Niedzielin *et al.*, 2001). Nicotinamide adenine dinucleotide (phosphate) $(NAD(P)^+)$-dependent dehydrogenases catalyze the oxidation of the substrate via the transfer of a hydride ion from the substrate to the nicotinamide ring of $NAD(P)^+$(Weiner and Hurley, 2005). Inositol dehydrogenase (IDH; EC 1.1.1.18) catalyzes the NAD^+-dependent oxidation of *myo*-inositol to *scyllo*-inosose (Figure A.1). *Myo*-Inositol (MI) is an important carbon source for many bacteria due to its abundance in the surrounding environment.

      Like many symbionts that have developed a niche in a host, *L. plantarum* WCFS1 has lost many of the genes required for complete catabolism of *myo*-inositol. This mechanism of gene reduction is commonly found in symbionts that depend on their host for nourishment

(Moran *et al.*, 2000). Metabolism of MI is not commonly seen in lactic acid bacteria (LAB). In most cases, the genes involved in MI utilization are absent altogether in *lactobacillales* that have been sequenced (Yebra *et al.*, 2007). *L. plantarum* has genes encoding for 4 *iolG*'s, *iolE*, and two putative sugar permeases, while it lacks the rest of the genes required for MI catabolism, including *iolA, iolB, iolC, iolD, iolJ, iolK,* and *iolR* (Figure A.2). Determination of the function of IolG4 may be key to understanding why *L. plantarum* has an incomplete MI catabolic pathway. *Lactobacillus plantarum* strain WCFS1 possesses four *iolG* genes coding for putative IDH proteins, including IolG4 (van Straaten *et al.*, 2010). Each of these 4 *iolG* genes represents a subgroup for IDH. Classification of the subgroups is based on the geometry of the active site and the proposed residues for substrate binding. Each subgroup of IDH is differentiated by 6 sequence motifs (Figure A.3).

myo-inositol   *scyllo*-inosose

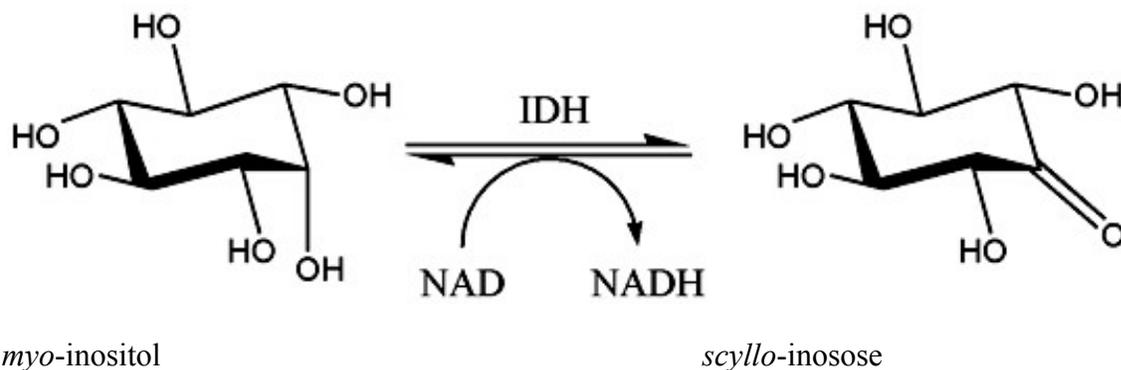**Figure A.1: First step of *myo*-inositol catabolism**. IDH stereoselectively oxidizes carbon 2 of *myo*-inositol to produce *scyllo*-inosose, which is further oxidized by other enzymes in the *myo*-inositol catabolic pathway. This reaction is NAD$^+$ dependent, and results in the production of NADH. Figure generated in Chemdraw.
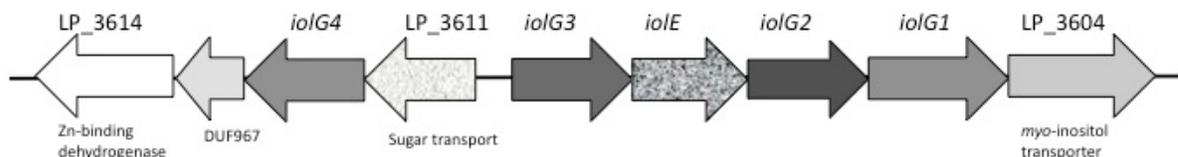

Figure A.2: The incomplete *iol* operon in *Lactobacillus plantarum*.

IolG4 is differentiated from the other subgroups largely by consensus motif IV where the active site is present (van Straaten *et al.*, 2010). Currently, there are no known structures of the subgroup to which IolG4 belongs. Motif IV of subgroup 4 (to which IolG4 belongs) contains a higher number of hydrophobic residues when compared to the other 3 subgroups of IDH (Figure A.4). Another notable difference is the presence of phenylalanine instead of methionine (subgroups 2 and 3), and threonine (subgroup 1). This phenylalanine residue (likely 176 in IolG4) appears to be in a favorable position to form stacking interactions with the substrate. IolG4 is annotated as a *myo*-inositol 2-dehydrogenase, however there is no detectable IDH activity when assayed kinetically (unpublished results, Hari Aamudalapalli). IolG4 consists of 341 amino acid residues with a molecular weight of 38,379 Da, however there is little experimental evidence to suggest which reaction IolG4 catalyzes. IolG4 is a paralog of IDH, which is a gene that is related by duplication within a genome. Paralogous genes can

Subgroup	Motif I	Motif II	Motif III
1	$Gx_2GxGx_2Gx_2H$	Hx_6Ax_9CEKP	QxGFMRx(Y/F)D
2	$(G/A)x_2GLGRLGx_2H$	$Px_2FHx_{13}VFxEKP$	$Fx_2GFMRR(Y/F)D$
3	$(G/A)x_2GxGRIGx_2H$	$Hx_{15}CEKP$	GFNRR(Y/F)D
4	GxIGLGRxGxMH	$Hx_{15}CEKP$	G/A)FNRRXD

Subgroup	Motif IV		Motif V
1	$HxNx_6Yx_7$	$(D/E)(T/S)x_2HEx(D/N)$	$Vx_3(Y/F)GY(D/Q)(V/I)x_2(E/D)$
2	$RxY(G/S)xDPx_7Fx_2FAx_4SGGxFxDMx_2HDxD$		$Rx_3HGY(Q/H)(V/I)ExE$
3	$SRDPx_6Yx_3$	$SGGx_3DMxLHDxD$	$Rx_3YGYDQRxE$
4	$SRDPx_3PHx_2IxRIGGLxFDFTMHDFD$		$Rx_3YGYDQRVE$

Subgroup	Motif VI
1	$(W/F)x_2RFx_2AY$
2	$FxERFx_2A(Y/F)$
3	$FFx_2RYx_3(Y/F)$
4	$Fx_2RYx_2A(Y/F)$

Figure A.3: Six consensus motifs defining the active and cofactor-binding sites of the four subgroups of IDH. Classification of subgroups is largely based on motif IV, the catalytic motif. Motifs I and II together comprise the Rossman Fold, involved in NAD binding. The other motifs (III, V, and VI) are likely involved in substrate binding. Residues in red are proposed substrate-binding residues, and residues in blue are thought to compose the catalytic triad. (Figure courtesy of Karin van Straaten)

evolve new functions or retain related functions to the known gene (Koonin, 2005). It is unknown whether the multiple *iol*G genes have redundant functions or differ in substrate affinities toward inositol isomers or inositol-related molecules.

A.2 Cloning and Expression of *iol*G4

Cloning and transformation of *iol*G4 was performed by Hari Aamudalapalli in the laboratory of Dr. David R. J. Palmer, Department of Chemistry at the University of Saskatchewan. *E. coli* BL21-Gold (DE3) competent cells harboring the *iol*G4-pET28b expression construct for over-expression of N-terminal His$_6$-tagged IolG4 were cultured at 310 K and 250 rev min^{-1} in Terrific Broth medium containing 50 μg/mL ampicillin. Expression was induced by the addition of 1 mM IPTG once the OD$_{600}$ reached 0.3. The cells were grown for 4 h until the OD$_{600}$ had reached approximately 1.4.

A.3 Protein Purification

The expression and purification of IolG4 was optimized from the protocol for IDH from *Bacillus subtilis* (van Straaten *et al.,* 2008). The cells were harvested and suspended in lysis buffer (buffer A) containing 25 mM HEPES pH 7.5, 0.3 M NaCl. The addition of 2 mM lysozyme, 1 mM AEBSF and catalytic amounts of DNAse was followed by sonication to rupture the cells. The cell debris was removed by centrifugation and the IolG4-containing supernatant was filtered using a 0.45 μm Millipore filter and loaded onto a Ni-loaded Protino

Figure A.4: Sequence alignment of motif IV of IDH-subgroup 4, including IolG4 (LACPL) aligned with the 3 highest sequence identity matches. Motif IV is thought to encompass residues 152 to 188 in IolG4. Residues that are identical are highlighted in red, and residues that are considered to be similar are in red within blue boxes (Figure adapted from van Straaten *et al*., 2010).

Ni-IDA column (Macherey-Nagel) pre-equilibrated with buffer A. After several successive washes with buffer A, the bound His$_6$-tagged IolG4 was eluted with a linear gradient of 2.5 mM to 250 mM imidazole using a elution buffer containing 25 mM HEPES pH 7.5, 0.3 M NaCl, 250 mM imidazole. Fractions collected during the observed peak in UV absorbance were subsequently concentrated using Vivaspin 20 (10,000 MW cutoff) concentrators. Fractions were also analyzed by SDS-PAGE to determine which contained the highest concentration of IolG4. The fractions were then pooled and dialyzed against 25 mM HEPES pH 7.5, 2 mM tris(2-carboxyethyl)phosphine (TCEP).

A.4 Characterization of IolG4

A.4.1. Dynamic Light Scattering

Dynamic Light Scattering (DLS) is used to determine the dispersity of a solution containing macromolecular components (Koppel, 1972). Polydispersity is indicative of the aggregation of molecules, which can interfere with intermolecular protein contacts during crystal growth. Monodisperse solutions are composed of homogeneous molecules in both weight and shape, which is very desirable for crystallization. DLS was used to determine the stability of IolG4 in 25 mM Tris pH 8.2 buffer. The sample of IolG4 was found to be largely polydisperse, which indicated a high degree of protein aggregation and instability in this buffer.

A.4.2 pH- and buffer-dependent stability analysis

The stability of IolG4 at different pH(s) was assessed using the JBScreen pH-2D (Jena Bioscience) which included 15 different buffering systems with a pH range of 4 to 10. The screen solution was mixed with IolG4 in a ratio of 1:1 and allowed to equilibrate under oil. The best buffers observed (e.g. the least amount of precipitation) were within the pH range of 7.0 to 7.7. HEPES and sodium dihydrogen phosphate buffers were chosen as the most suitable buffers for further dialysis experiments. IolG4 was then dialyzed out of 25 mM Tris pH 8.2 into several different buffers (Table A.1) using dialysis buttons (Hampton Research). DLS analysis of the samples confirmed that IolG4 was the most stable in 50 mM HEPES pH 7.5, 2

mM TCEP, with 99.8% of the mass analyzed found to have a radius of 4.75 nm with a polydispersity of 11.2%.

Table A.1: List of buffers for IolG4 micro-dialysis trials.

Buffers for Dialysis
50 mM HEPES pH 7.0
50 mM HEPES pH 7.5, 2 mM TCEP
50 mM HEPES pH 7.5, 0.15 M NaCl, 2 mM TCEP
50 mM Na$_2$H$_2$PO$_4$ pH 7.5
50 mM Na$_2$H$_2$PO$_4$ pH 7.5, 0.15 M NaCl

The purity of the dialyzed IolG4 sample was assessed by SDS-PAGE (Figure A.5) after concentration to 11 mg/mL, which is a suitable concentration for crystallization. The overall purity of IolG4 appeared to improve when using HEPES buffer instead of Tris buffer. The clarity of the sample improved from a yellow color to colorless, as well a decrease in the visible high MW contaminants as seen in lanes 5 (HEPES) and 6 (Tris) in Figure A.5. This could be due to the increased affinity of the His$_6$-tagged protein for the Ni-IDA resin in the presence of HEPES, which would decrease non-specific binding of contaminant proteins. In addition, the use of Tris buffer is not recommended with the Ni-IDA resin because it can coordinate with Ni^{+2} ions, causing a decrease in capacity.

A.4.3 Spectrophotometric analysis

IolG4 is annotated as an IDH and contains a Rossman fold, which is commonly known to bind the cofactor NAD(H). A scan of the protein sample from 200 nm to 400 nm was performed using the Cary 50 Bio UV-visible spectrophotometer to determine if there was a visible peak(s) for the cofactor bound to the enzyme (Figure A.6). This experiment did not conclusively determine if NAD(H) is present during the purification of IolG4, and further work using HPLC should elucidate whether NAD(H) is the cofactor for the enzyme.

A.4.4 High Performance Liquid Chromatography (HPLC) analysis

In order to determine if the NAD(H) is present during the purification of IolG4, we subjected the supernatant of the denatured IolG4 sample to analysis with HPLC. Initially, the protein buffer was exchanged to 20 mM Potassium phosphate pH 6.8 using a Vivaspin 20 (10,000 MW cutoff) concentrator. The protein was denatured in ice cold 20% Trichloroacetic acid (TCA) and allowed to incubate for 10 min. The sample was subjected to centrifugation at 10,000 x g for 2 min to remove the denatured protein. The cleared supernatant was filtered with 0.22 μm Millipore filter prior to loading on a Zorbax SB-C18 reverse phase (non-polar) column. The sample was eluted using acetonitrile, a non-polar solvent that is able to elute compounds that the non-polar stationary phase of the column. The theoretical concentration of NAD bound to IolG4 was calculated to equal 52 μM (Equation A.1). Therefore, standards for NAD and NADH were made to a final concentration of approximately 50 μM.
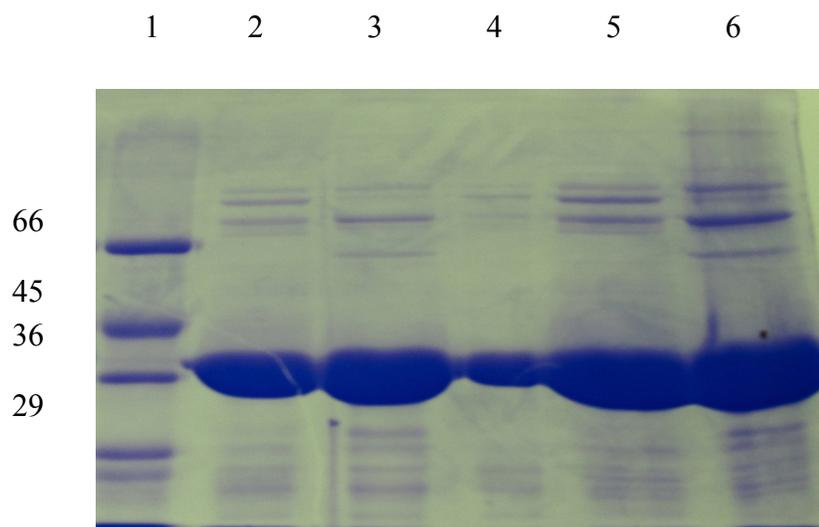

Figure A.5: SDS-PAGE of IolG4: Lane 1- MW marker, Lane 2- 5 μL of IolG4 in 25mM Tris pH 8.2, Lane 3- 5 μL IolG4 in 25mM HEPES pH 7.5, 2 mM TCEP, Lane 4- 1μL IolG4 (in HEPES), Lanes 5 and 6- 10 μL of IolG4 in HEPES and Tris respectively. All samples were approximately 10 mg/mL.

Equation A.1: <u>concentration of enzyme</u>

molar mass of 1 subunit

Standards for NAD (peak 1) and NADH (peak 2) were ran as controls to determine the retention time of the cofactors, at 9 and 11 min respectively (Figure A.7B). The supernatant of IolG4 in did not appear to contain NAD or NADH, as evidenced by the absence of a peak for either cofactor in the chromatogram (Figure A.7A). However, the 2 peaks that are observed at approximately 3 and 6 min were confirmed to be contaminants from TCA treatment (not shown). In addition, the same procedure was performed using ice cold 70% ethanol as the denaturant. The supernatant did not contain the large peaks at 3 and 6 min observed in the presence of TCA, however there were no peaks that correlated to NAD or NADH in the sample (Figure A.7B). Therefore, these results did not conclusively show the presence of NAD(H) during the purification of IolG4.
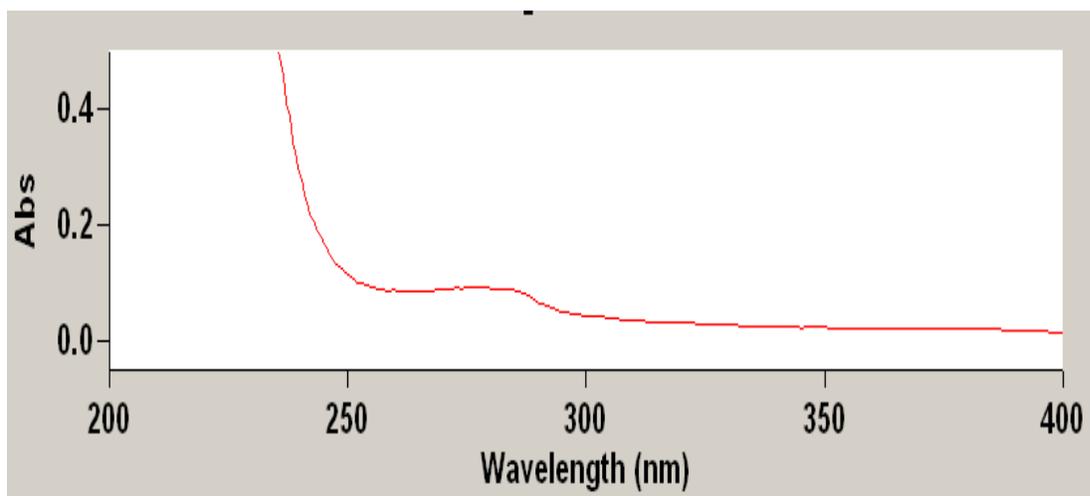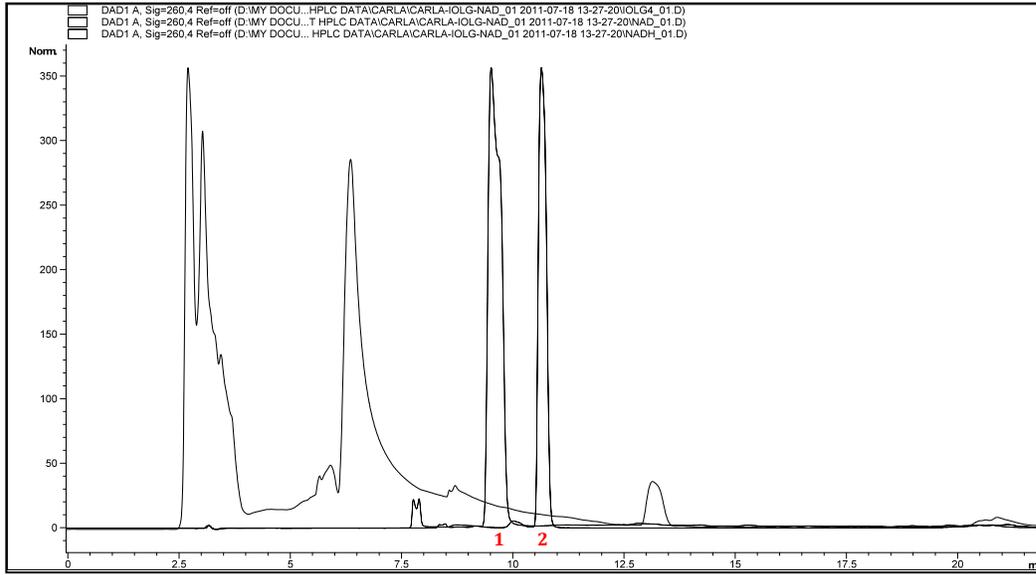

Figure A.6: At a concentration of 0.1 mg/mL IolG4, there was no definite peak in absorbance seen at 260 nm (the absorbance peak for NAD^+), nor at 340 nm (the absorbance peak for NADH).
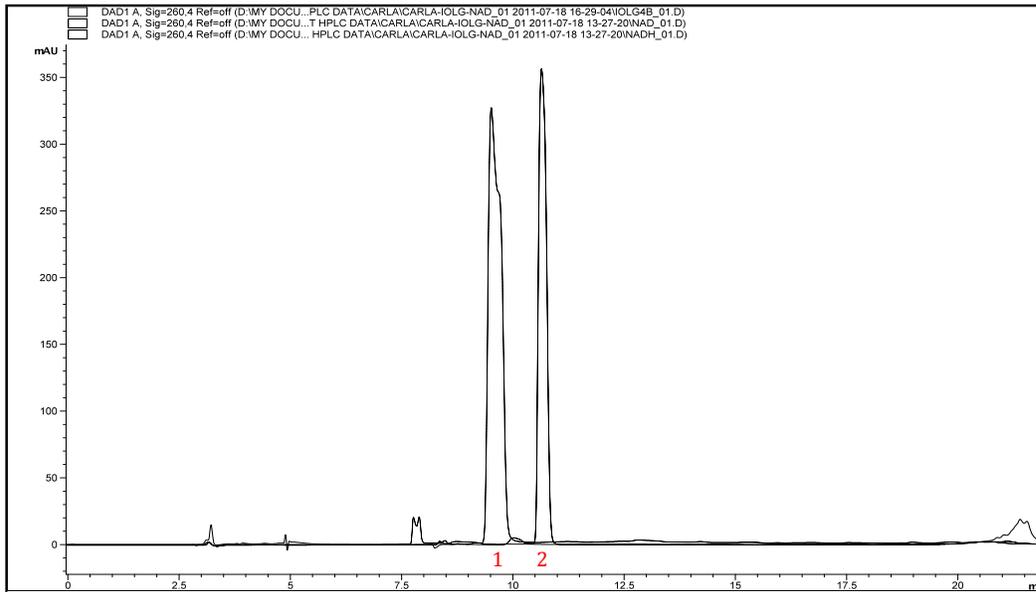
A

B

Figure A.7: A- Overlay of the chromatograms for the supernatant from IolG4 denatured in TCA and the cofactor standards: NAD (peak 1) and NADH (peak 2) indicated in red. B- Overlay of IolG4 denatured in ethanol and the cofactor standards (red).

A.5 Crystallization of IolG4

Broad crystallization screening (as described in Section 3.1.2) using of purified IolG4 in 25 mM Tris pH 8.2 was endeavored. Determination of crystallization conditions for IolG4 was attempted using the microbatch-under-oil technique, through utilization of commercial screening kits at 4°C and room temperature. The crystallization plates were dispensed using an Oryx6 Crystallization Robot (Douglas Instruments). Use of a robotic crystallization setup allows for higher precision, smaller volume of protein sample required per plate, smaller drop volumes, and prevention of pipetting errors. Each drop contained 0.5 µL of 10 mg/mL IolG4 in 50 mM HEPES pH 7.5, 2 mM TCEP and 0.5 µL of the screening solution, covered with paraffin oil. Several conditions yielded crystals, including #74 from Classics (0.2 M CaCl$_2$, 0.1 M 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) Na pH 7.5, 28% (v/v) PEG 400) and #3 from ComPAS suite (0.2 M Na acetate, 0.1 M MES pH 6.5, 2.0 M NaCl) (Figure A.8). The crystals were diffracted at the CMCF-1 beamline (08ID-1) at the Canadian Light Source (Grochulski *et al*., 2011). All of the crystals were confirmed to consist of salt, and therefore the growth of protein crystals was unsuccessful for the all of the crystallization experiments. Broad crystallization screening was also performed with the addition of 10 mM NAD$^+$ to the IolG4 protein sample prior to experimental setup. If NAD$^+$ is a cofactor of IolG4, it is expected to stabilize the protein. However, no crystallization conditions yielded protein crystals despite the presence of NAD$^+$.
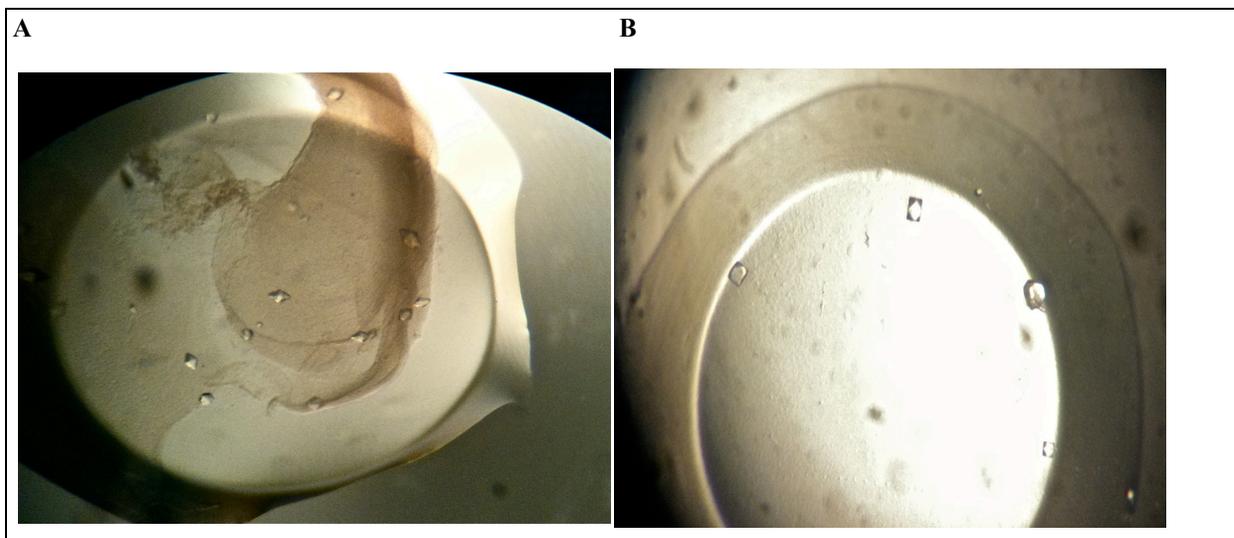

Figure A.8: A- Crystals observed in condition #74 from Classics. B- Condition #3 from ComPAS suite.

A.6 Conclusions and Future Work

The function of IolG4 remains largely elusive after extensive protein characterization, as well as kinetic analyses (as performed by our collaborators in Dr. D. R. J. Palmer's laboratory). There are no known substrates of this protein, and there is no detectable NAD cofactor present during the purification. Also, crystallization of this protein was unsuccessful after extensive screenings, which was surprising due to the high promiscuity of IDH homologues for different crystallization conditions (van Straatan *et al*., 2010).

We hypothesize that IolG4 is a paralog of IDH, which is a gene that is related by duplication within a genome. This paralog was likely created by gene duplication. The function of *iolG4*'s gene product has yet to be determined, therefore it may be a pseudogene, a presumed non-functional paralog of a functional protein-coding gene generated by gene duplication or retrotransposition. The transcription of pseudogenes can produce non-coding RNAs (ncRNA) that have been shown to regulate gene clusters by silencing the paralogous mRNAs *in trans*. Also, in the absence of an *iolR* transcriptional repressor, the silencing of this operon could be controlled by this alternate mechanism. If this hypothesis is true, it may explain why the iolG4 does not appear to recognize any inositol-related compounds.

Future Work

Further optimization of the purification protocol including additional gel filtration of the protein sample should improve purity. Growth of suitable protein crystals should be attempted and the subsequent collection of diffraction data will lead to structure solution. Several structures in the PDB are suitable for molecular replacement on the basis of their shared sequence identity (Table A.2). Structural studies will allow for the analysis of the enzymatic mechanism, substrate specificity, and chemical properties of the active site. Molecular modeling of the active site may elucidate the possible substrates for IolG4, while kinetic assays can determine which substrate results in the highest enzymatic activity. The products formed during these assays may also be analyzed by HPLC. Once a probable substrate is discovered, the substrate(s) can be used for co-crystallization screening for conditions. The structural

details of the active site, as well as the interactions between specific residues and the substrate(s) can then be assessed.

Table A.2: Highest-ranking sequence identity structures with IolG4 in the Protein Data Bank.

PDB code	Structure	Identity (%)
3EZY	probable dehydrogenase from *T. maritima*	45
3EUW	*myo*-inositol dehydrogenase from *C. glutamicum*	34
3CEA	IolG1, *myo*-inositol dehydrogenase from *L. plantarum*	26
3EC7	putative dehydrogenase from *S. typhimurim* LT2	28

References:

Ahrne, S., Nobaek, S., Jeppsson, B., Adlerberth, I., Wold, A.E. and Molin, G. (1998). The normal *Lactobacillus* flora of healthy human rectal and oral mucosa. J. Appl. Microbiol. *85,* 88-94.

Cunningham-Rundles, S., Ahrne, S., Bengmark, S., Johann-Liang, R., Marshall, F., Metakis, L., Califano, C., Dunn, A.M., Grassey, C., and Hinds, G. (2000). Probiotics and immune response. Am. J. Gastroenterol. *95,* S22-25.

Grochulshi, P., Fodje, M.N., Gorin, J., Labiuk, S.L., and Berg, R. (2011). Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. J. Synchrotron Rad. *18,* (Pt 4), 681-684.

Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D., Kuipers, O.P., Leer, R., Tarchini, R., Peters, S.A., Sandbrink, H.M., and Fiers, M.W. (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc. Natl. Acad. Sci. USA, *100,* 1990-1995.

Koonin, E.V. (2005). Orthologs, Paralogs, and Evolutionary Genomics. Annu. Rev. Genet. *39,* 309-338.

Koppel, D.E. (1972). Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. J. Chem. Phys. *57,* 4814.

Moran, N.A., and Wernegreen, J.J. (2000). Lifestyle evolution of symbiotic bacteria: insights from genomics. Trends Ecol. Evol. *15,* 321-326.

Niedzielin, K., Kordecki, H. and Birkenfeld, B. (2001). A controlled, double-blind, randomized study on the efficacy of *Lactobacillus plantarum* 299V in patients with irritable bowel syndrome. Eur. J. Gastroenterol. Hepatol. *13,* 1143-1147.

van Straaten, K.E., Zheng, H., Palmer, D.R.J., and Sanders, D.A.R. (2010). Structural Investigation of myo-inositol Dehydrogenase from *Bacillus subtilis*: Implications for Catalytic Mechanism and Inositol Dehydrogenase Subfamily Classification. Biochem. J. *432,* 237-247.

van Straaten, K.E., Hoffort, A., Palmer, D.R.J., and Sanders, D.A.R. (2008). Purification, crystallization and preliminary X-ray analysis of inositol dehydrogenase (IDH) from *Bacillus subtilis*. Acta. Cryst. *F64*, 98-101.

Yebra, M.J., Zuniga, M., Beaufils, S., Perez-Maritinez, G., Deutscher, J., and Monedero, V. (2007). Identification of a Gene Cluster Enabling *Lactobacillus casei* BL23 to Utilize *myo*-Inositol. Appl. Environ. Microb. *73*, 3850-3858.

Weiner, H., and Hurley, T.D. (2005). NADP$^+$ Binding to Dehydrogenases. Enc. Life Sci. (available online).

Wels, M.W.W. (2007). Unravelling the regulatory network of *Lactobacillus plantarum* WCFS1. PhD thesis, Wageningen Unversity and Research Centre, The Netherlands.