A STRUCTURAL AND BIOCHEMICAL INVESTIGATION OF THE BIOSYNTHESIS OF DEOXYSUGARS FROM PATHOGENIC BACTERIA

by

Ari J. Salinger

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN — MADISON

2017

Date of oral examination: 12/12/16

The dissertation is approved by the following members of the Final Oral Committee:

Hazel M. Holden, Professor, Biochemistry
Ivan Rayment, Professor, Biochemistry
Michael G. Thomas, Professor, Bacteriology
Samuel E. Butcher, Professor, Biochemistry
Aaron A. Hoskins, Assistant Professor, Biochemistry

ABSTRACT

Carbohydrates represent the most abundant family of biomolecules on earth where they function in a variety of biological processes. For example, they are the majority component of plant cell walls, and are critical in cell adhesion, the immune response, antibiotic activity, and microbial pathogenicity. This is all in addition to the integral role carbohydrates play in energy storage for advanced life forms. Oftentimes, hexose sugars are deoxygenated at one or more carbon positions, and these variations can result in quite unusual looking products. Some of the most fascinating and diverse sugars found in nature are attached to the exterior part of Gram-negative bacteria on their lipopolysaccharide O-antigens. The sugar moieties decorating the O-antigen have been linked to the pathogenicity of the bacteria that contain them. These carbohydrates are so diverse and specialized, they are even used to classify serogroups of bacteria within a strain.

Among the deoxysugars appending the exterior of the lipopolysaccharide, the most common are of the 6-deoxyhexose family. Many of the biosynthetic pathways which yield these sugars begin by attaching glucose-1-phosphate with a nucleotide-monophosphate. The (d)NDP-hexoses can then be deoxygenated at C-2′, C-3′, C-4′, and/or C-6′. These modifications, in addition to possible dehydration reactions, methylations, formylations, aminations, epimerizations, and acylations, create the

plethora of diverse pyranose rings found in O-antigens. The sugars, therefore, afford the exterior of the bacteria unique local characteristics such as altered acidity, chemical stability/reactivity, and hydrophobicity.

Chapter 2 of this dissertation discusses the enzymes responsible for the third step of the biosynthesis of either 3-acetamido-3,6-dideoxy-D-glucose or 3-acetamido-3,6-dideoxy-D-galactose. Whereas both ketoisomerase enzymes utilize the same substrate, the stereochemistries of their products differ around C-4′. In this chapter, FdtA and QdtA are explored to elucidate the manner in which they control the stereochemistry of their products. Various site-directed mutants of each enzyme were constructed, and their dTDP-sugar products were analyzed via nuclear magnetic resonance (NMR) spectroscopy. In addition, their kinetic parameters were measured and a novel structure of mutant QdtA was solved and analyzed. This investigation demonstrated that it is possible to convert a "galacto" enzyme into a "gluco" enzyme, and vice versa.

In Chapter 3, the biosynthesis of an extraordinarily rare CDP-linked sugar is discussed. WbcA is a 3′-monoepimerase enzyme expressed by a specific strain of *Yersinia enterocolitica*, and while its catalytic mechanism seemed straight-forward based on structural analysis of homologues, this enzyme was different. Exploration of the primary sequence showed that much of the active site architecture of this enzyme is conserved, however, a cysteine had replaced a tyrosine and was the proposed catalytic acid. The work presented in Chapter 3 reveals the three-dimensional structure of WbcA. Site

directed mutagenesis alongside this informative model reveals that another tyrosine from WbcA swings into the active site to act as an acid. This refutes the theory that the predicted cysteine does not play an integral catalytic role.

Chapter 4 details the characterization of an *N*-acyltransferase from *Acinetobacter nosocomialis*. This sugar-modifying enzyme has previously been described as a putative member of the larger family of GCN5-like *N*-acyltransferase enzymes (GNAT). However, it had never been biochemically studied. In Chapter 4, the first documented ternary structure of a bacterial, sugar-modifying GNAT family member with bound sugars and CoA products is presented. Further enzymatic analysis through mutagenesis experiments evidences the existence of a, "catalysis via approximation," reaction mechanism, which necessitates neither an active site base, nor a specific water-mediated proton transfer. Furthermore, this chapter documents the promiscuity of this enzyme in regards to both the sugar substrate and CoA substrate.

Chapter 5 describes a unique outreach program that was established in 2009. Referred to as Project CRYSTAL (Colleagues Researching with Young Scientists Teaching And Learning), this program paired local middle school students, who were eager to learn advanced scientific techniques, with graduate student mentors. The middle school students were brought into the lab weekly and walked through a graduate-level research project which spanned the nine-month school year. As a mentor, I, alongside my colleagues, were responsible for organizing weekly lectures, experiments, and helping

the young scientists prepare final presentations for the Madison Middle School Science Symposium. In its most recent year, we developed a curriculum in the hopes of disseminating the program to other willing labs around the country.

ACKNOWLEDGEMENTS

In December 2012, I placed my books on my shelves.

I unpacked my baseball and toys. I cranked Christmas music noise.

While the weather outside seemed drab, I was happy to have found my lab.

Rachel welcomed me first. Her fast pace kept her immersed.

She was a truly great role model with her only speed "full-throttle."

I helped her solve protein EvaA- the first of its kind. Hooray!

Matthew was there as well. A cool kid that I met at Grinnell.

Most important there on floor two, was Jim Thoden and his undergrads too.

Jim is so incredibly smart. I'm having trouble with where to start.

Into this lab, he pours his whole heart. He makes crystallography a true work of art.

Always willing to give, not taking. He is a trained chemist with a knack for baking.

From getting the stupid pumps to drip, to looking at pics from his recent trip

To giving advice on cryo conditions, to sharing his political positions

To being ok that I broke a rotor, to talking about his car's sweet motor

Jim is the best ally to acquire because his love for his work ceases to tire

Another person I cannot forget, is the friendly Haley from scenic Wilmette

Haley joined me in year two. A welcome addition through-and-through

Although Haley might seem quiet, she's a true friend and regular riot
It's been so fun to watch her grow, and her humor is best when her blood sugar's low
Because of her presence, Haley I must thank. Also, a shout out to her pet dog, Hank.

And now to go right to the source- I wouldn't be here without Hazel, of course!
She welcomed me in and let me explore. She made work a joy and had an open door
She let me be me, which is all I need. She allowed me to teach, to preform, and succeed.
When things got tough, she stuck up for me. When I got results, she was filled with glee.
Empathy is never something she'd lack. She's so open minded and sharp as a tack.
I'm so happy I got to be part of her team. It was here I found my science self-esteem.
I can't wait to see what more she will write. With such good leadership, it's infinite.

To my committee, you all never fell short. Thank you all for all your support.
For his great questions, Aaron was key. He is so darn smart, it's a little scary!
Michael's expertise is amazing to behold. His brilliant ideas I've put in the fold.
Sam, although you are a late addition, since rotations you have been a contribution
Ivan- aside from your jams, which are delish, and all of the most wonderful fish
Your experience and advice have been integral. Your whit and demeanor indispensable
I couldn't have done this without your aid, with all the comradery you have displayed.

There are so many people who've helped as well- like Kimberly, Kath, Beth and Danielle.
Like Daniel and Shirly (from Israel, you know!). Like the Rostons and Ora and Shlomo.
Oh Duh! I need to thank my Mom and my Dad! Who gave me confidence I never had.
There's Karli and Jeremy and Maia too. Who remind me I'm me so, of course, thank you!
I need to thank my tequila and lime. And the best advice I get from my dog, Enzyme.
To Jordan and Jordan and Simon and Phil- you guys are the best and are really a pill.
To Molly and Anna and Leah and Yoni. If friends were a musical, you'd get the Tony.
To the Chicago Cubs- What a surprise! We actually did it this year, guys!
To MBI, my comedy team- I would have died without y'all. Y'all are a dream.

And then, to the one who changed my life, thank you to Ali, my brilliant wife.
If it wasn't for you, where would I be? I'd have eaten nothing but doughnuts and coffee.
I'd be bored, dirty and unmotivated. You keep me excited, engaged, fresh, and elated!
You anchor my crazy and keep me afloat. The best sides of me you seem to promote.

And finally, thank you reader for taking the time, to get through all these silly rhymes
I hope my research is engaging to you. I am proud of my work. I hope you are too.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES AND SCHEMES

CHAPTER 1:

INTRODUCTION TO UNUSUAL SUGARS AND THEIR BIOSYNTHESES

1.1 CARBOHYDRATES

Biological molecules are classed into four categories: nucleotides, amino acids, lipids, and carbohydrates. Of these, carbohydrates constitute the largest portion (~50%) of the Earth's biomass. Carbohydrates are involved in a plethora of biological processes and are necessary for all domains of life. For example, carbohydrates are essential for energy acquisition and storage— a ubiquitous process which is often one of the first pathways taught to beginning biology students. Glucose, a monosaccharide, represents the end of photosynthesis and the beginning of cellular respiration. Glucose, and its simple derivatives, can also assemble into structural components inside and outside of a cell such as cellulose, chitin, and the glycan components of cartilage (2).

Despite the biological importance and abundance of carbohydrates, many of their structural and functional roles are overlooked. Sugars play a necessary role in cell adhesion, blood clotting, plant cell wall structure, the immune response, fertilization, antibiotic activity, and microbial pathogenicity. Because of the diversity of sugars, the most miniscule of differences are quite impactful. An example of this lies in human ABO blood typing. Red blood cells, or erythrocytes, have chains of carbohydrates decorating their exterior. These chains are also known as antigenic determinants. They are extremely specific, and blood type is identified by the presence of one of three determinants on the surface of erythrocytes (2-4). The simplest of these determinants is the H antigen, which is composed of three carbohydrate units: galactose, *N*-acetylglucosamine, and a

deoxygenated, 6-carbon sugar called fucose. Individuals with an additional *N*-acetylglucosamine residue at the non-reducing end of the antigenic unit have type A blood, while those with type B contain a galactose instead (2, 5, 6).

Although these differences seem insignificant when examining the overall structure of an erythrocyte, the clinical consequences spell life or death for an individual. These cell-surface interactions are a carefully orchestrated part of the biological world, and it is often a result of the interactions between carbohydrates and their substrates.

1.2 UNUSUAL SUGARS

While the function of common sugars, such as glucose, fructose, or mannose, are discussed in most life-science classes, their more complex and fascinating counterparts are skipped over. It turns out that successive modifications to these common sugars create a massive library of multifarious possibilities for final sugar products. These modified sugars typically have one of two common precursors. Many pathways begin with a glucose-1-phosphate sugar, while others commence with a fructose-1-phosphate sugar. These precursors are then activated by the attachment of a nucleotide monophosphate to the already existing phosphate on C-1′. The nucleotidyltransferase, which carries out this reaction, positions the phosphate appended to the sugar in such a way that it is able to attack the α-phosphate of the (d)NTP. Following this addition, the sugars are deoxygenated at one or more positions to become deoxysugars. These deoxysugars have more labile protons and can, therefore, undergo more complex chemical reactions such as aminations, epimerizations, methylations, reductions, and metal-dependent redox reactions [Figure 1.1]. The final sugar product will always have different chemical properties from its precursor— it may have different hydrophobic properties, stability, reactivity, and/or electrochemical variability. The specificity of these unusual deoxysugars can be pivotal in the survival of the organism that makes them (3, 4, 7-9).

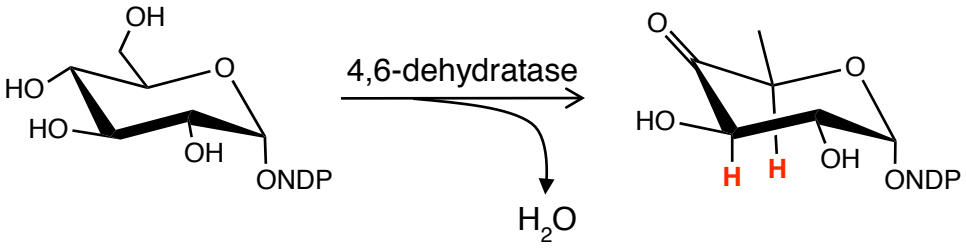


Figure 1.1. Depiction of labile protons on a deoxysugar. After having undergone a reaction with a 4,6-dehydratase, nucleotide-linked sugars become far more reactive. The ketone on C-4′ allows for electron delocalization around parts of the pyranose ring. This, in turn, can lower the pK_a of the hydrogens on C-3′ and C-5′ (red) by many orders of magnitude. In addition, C-4′ itself becomes electrophilic.

Many prokaryotic cells expend a great deal of cellular energy on sugar modification and diversification, and have evolved to use these sugars in a wide variety of manners. For instance, many unusual deoxysugars are incorporated into the assorted antiobiotic and antifungal secondary metabolites created by soil-dwelling species of *Streptomyces* and *Saccharopolyspora*. These small molecules, which have many potential human health implications, are synthesized by the bacteria in order to carve out a niche for them in their respective ecosystems (4, 7).

One class of secondary metabolites produced by various prokaryotes is the class of macrolide antibiotics. These molecules are based around a 14, 15, or 16-membered polyketide ring to which one or more deoxysugars are added. These sugars are integral to the pharmacology of the metabolites such that even the most minor alteration could yield a completely ineffective compound (3, 4, 7). Erythromycin is one such example of these sorts of antibiotics [Figure 1.2]. It is isolated from *Saccharopolyspora erythraea* and is commonly prescribed to humans for respiratory infections (10). Note how it has two deoxysugars appended to the macrolide core: desosamine and cladinose. In the absence of either methyl group on desosamine's amino modification, the antibiotic properties are entirely eliminated (11, 12).

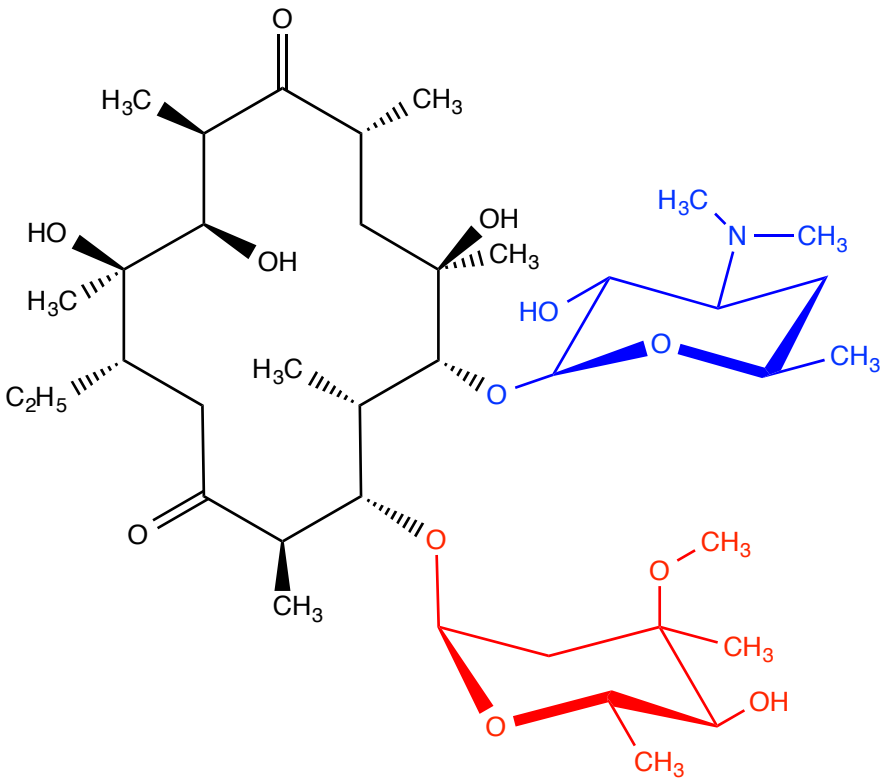


Figure 1.2. Erythromycin. Erythromycin is a macrolide antibiotic produced by the soil-dwelling bacteria, *Saccharopolyspora erythraea*. It is often prescribed for the treatment of human respiratory infections and human sexually-transmitted bacterial infections (1). The polyketide core structure is appended with two sugars, desosamine (blue) and cladinose (red).

Aside from antibiotics, some of the most diverse and unique sugars on Earth are found in the O-antigen of Gram-negative bacteria. Other sugars like 6-deoxy-D-gulose and 3-[(R)-3-hydroxybutanoylamino]-3,6-dideoxy-D-galactose can be found on this distal layer of the lipopolysaccharide. Over 100 different modified carbohydrates have been identified on various bacterial O-antigens. These sugars and their arrangements on bacteria are so specific, that they are the foundation for bacterial serotyping (13, 14).

1.3 THE LIPOPOLYSACCHARIDE

Gram-negative bacteria differ from Gram-positive bacteria in a number of ways. The most discerning of which is that Gram-negative bacteria have a second cell-membrane. Anchored into the membrane furthest from the center of the cell is the lipopolysaccharide, or LPS, which is estimated to occupy nearly 75% of the total bacterial surface area. As the name implies, this most distal layer of the bacteria is comprised of a lipid-based, polysaccharide chain (14, 15).

There are three distinct regions of the LPS: the lipid A, the core oligosaccharide, and the repeating O-specific unit [Figure 1.3]. The lipid A layer of the LPS anchors the rest of the polysaccharide chain into the outer membrane. Unlike other membrane-bound moieties, the lipid A is not connected through a glycerol. In fact, it is linked to the outer membrane via a glucosamine-based phospholipid. Commonly, lipid A is comprised of two glucosamine sugars connected by an α(1→6) linkage. Each hydroxyl group is appended with a fatty acid chain bestowing hydrophobic properties to adhere this unit to the outer membrane. These fatty acids are β-hydroxy hydrocarbons which vary between 10 and 16 carbons in length (13, 14).

The core oligosaccharide is comprised of a relatively small chain of sugars beginning with one to three acidic heptoses, 3-deoxy-D-mannooctulosonic acid (KDO). The KDO sugars can be modified via phosphorylation or with the addition of an ethanolamine. Further from the cell membrane are more seven-carbon pyranose rings

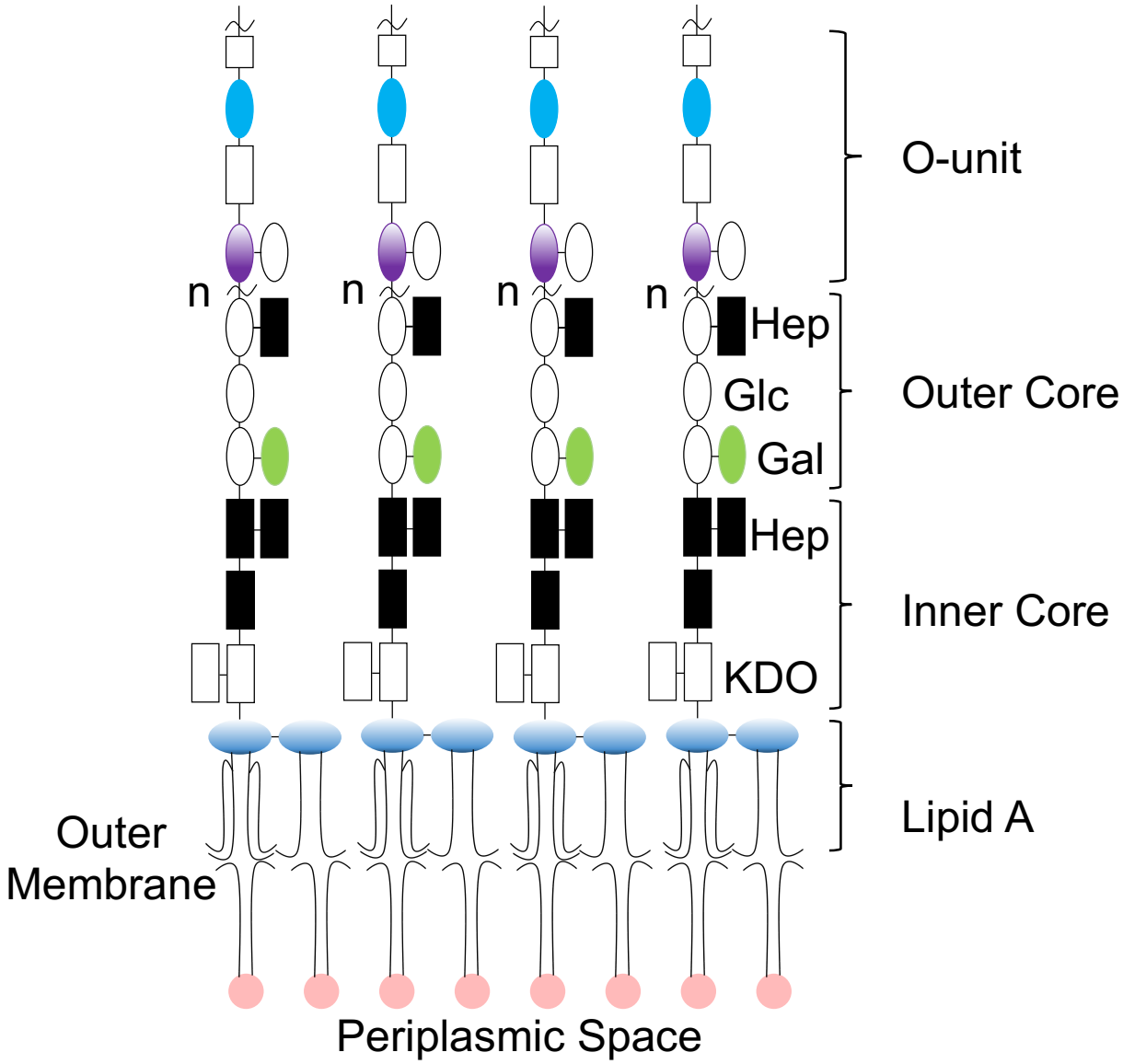


Figure 1.3. Schematic representation of the lipopolysaccharide. Specifically, this schematic is a depiction of the lipopolysaccharide from *E. coli* K12. The LPS, from bottom to top, consists of the lipid A, inner and outer core oligosaccharides, and the O-unit. KDO: 3-deoxy-D-manno-octulosinic acid, Hep: heptose, Gal: galactose, Glc: glucose.

that can be phosphorylated as well. These primary, seven-carbon sugars are building blocks of the inner core polysaccharide. The outer core is made up of common hexose sugars such as D-glucose, D-mannose, D-galactose, among others (14).

The most distal layer of the LPS is the O-unit, or O-antigen. Built from a repeating unit of two to six hexose sugars, the O-antigen is the most solvent exposed part of a Gram-negative bacterial cell. It is in this layer where some of the most diverse sugars in nature can be found. While there does exist many common sugars like glucose or galactose in the O-antigen, it is also home to an array of complex, modified deoxysugars.

Genes that code for the proteins necessary for the biosynthesis of O-antigen sugar units are believed to be constitutively expressed (8, 13, 14). The repeating sets of hexoses are compiled on a membrane-bound carrier protein, undecaprenyl lipid carrier (UndP). The assembly is carried out by glycosyltransferases that are either soluble or peripherally associated with the inner cell membrane. The export of the repeating units is accomplished by one of three different pathways: the Wzy-dependent pathway, the ABC-transporter dependent pathway, or the synthetase dependent pathway. Although fairly different, all three pathways end in the same ligation step via WaaL ligase (14).

The variations in the O-antigen define the serological specificity of an organism. For example, for *Escherichia coli* alone, there are over 170 O serotypes (13). While it is unclear why sophisticated deoxysugars exist on this layer, studies of *Salmonella*

demonstrate that serogroups with exotic deoxygenated sugars are more pathogenic than serogroups that contain only common sugars (16). Furthermore, animals infected with one specific serogroup of *Salmonella,* with unique sugars in the O-antigen, were then susceptible to other serogroups with different sugars after having gained resistance to the previous pathogen (17, 18).

In fact, the structure of an antibody interacting with an O-unit from *Salmonella* serogroup B was solved in 1994, giving molecular details to the relationship between the O-antigen and the immune response. Interestingly, the structure reveals that the unusual deoxysugar, abequose, is buried deeply in the antibody globule domain [Figure 1.4]. However, the common sugars remain more solvent exposed. This evidence supports the hypothesis that the presence of deoxysugars in an O-antigen better equips an organism to evade the host response until an antibody specific to that serogroup is generated (19).

Laboratory strains of bacteria can be grown in the absence of an O-antigen. These strains are termed as having a "rough phenotype." Strains of bacteria with a fully intact O-antigen are termed as having a "smooth phenotype." While bacteria with a rough phenotype can be cultured under laboratory conditions, the important coat of polysaccharides from the O-antigen is necessary for these strains of bacteria to resist environmental stress, host immune response, and antibiotics (20). An example of this sort of finding comes from work done locating virulence factors from *Yersinia sp.* Mutations in the dideoxysugar synthesis pathway yielded a strain of bacteria which no longer

plagued its host (16, 21, 22). Furthermore, the presence and length of the O-antigen in *Shigella flexneri* was shown to be essential for the invasiveness and inter- and intracellular spread of the bacteria (23-25).

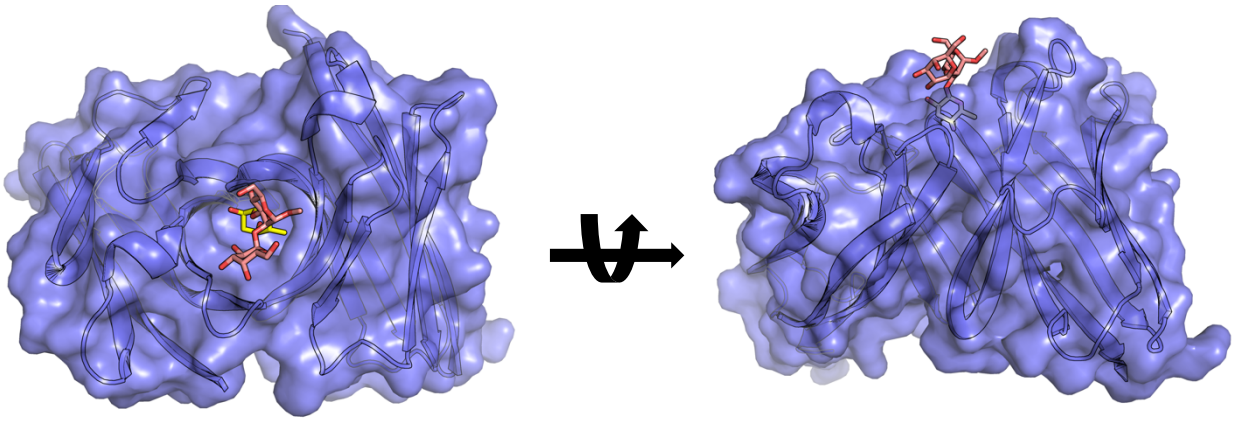


Figure 1.4. Structure of a single-chain antibody variable domain binding to trisaccharide antigen. This model depicts the globular domain of a human antibody in conjunction with an O-unit from *Salmonella* serogroup B: α-D-Gal(1→2)[α-D-abequose(1→3)]α-D-Man*p*1→OMe. The unusual deoxysugar, abequose, is colored yellow. The model is deposited as 1MFA in the Protein Data Bank.

1.4 SUGAR 3,4-KETOISOMERASES

Some of the most common deoxysugars found in the O-antigen are derived from 3-amino-3,6-dideoxyhexoses. The amino nitrogens on these sugars are susceptible to further modification by acetylation and methylation. While the molecular formula of these sugars are constant, the stereochemistry can differ about carbon C-4′ (26-28).

The biosynthesis of these 3-amino-3,6-dideoxyhexose sugars begins with the common precursor, glucose-1-phosphate. The sugar is acted upon a nucleotidyltransferase RmlA (E.C. 2.7.7.24) yielding dTDP-glucose. Studies of RmlA began in the mid 1960's when scientists realized the key role it played in bacterial virulence. The original purification and enzymatic studies were performed on the enzyme from *Pseudomonas aeruginosa.* Interestingly, the enzyme, aside from having a very strong affinity for the substrates (apparent K_m is $5x10^{-6}$ mM for glucose-1-phosphate and $4x10^{-6}$ mM for dTTP), also demonstrates product-induced inhibition. This means that the final sugar in the pathway, in this case dTDP-rhamnose, can alter the activity of the enzyme (29). The appendage of a dTMP to the sugar-phosphate designates the carbohydrate for the sugar-modification pathway as opposed to the energy manufacturing pathway (4, 13).

The next enzyme in the pathway is RmlB. This enzyme is a 4,6-dehydratase (E.C. 4.2.1.46) which acts as an oxidoreductase transforming dTDP-glucose into dTDP-4-keto-6-deoxy-D-glucose. In order for this reaction to take place, the enzyme requires an NAD^+

cofactor (13). The reaction occurs in three distinct steps: 1) dTDP-glucose is oxidized to a 4-keto sugar, 2) water is eliminated from C-5′ and C-6′ yielding a glucoseen intermediate, 3) and finally, reduction with a hydride produces the final sugar (30) [Figure 1.5].

RmlB is a member of the vast short chain dehydrogenase/reductase (SDR) superfamily. This family of enzymes carries out a wide range of reactions—isomerizations, oxidoreductions, dehydrations, decarboxylations, and more. The active site not only necessitates some form of NAD(P)(H), but it also requires a catalytic diad of tyrosine and lysine (31, 32). In the case of sugar modification, the resulting keto product is very reactive which allows for further chemistry.

To biosynthesize the 3-amino sugar, the next step in the pathway is a keto-isomerization. The fascinating question of the isomerization reaction is what will the stereochemistry be about C-4′ after the reaction is complete [Figure 1.6]? This question became evident when it was found that both the equatorial quinovose ("gluco") configuration and the axial fucose ("galacto") configuration were found in various bacterial species (27, 28). The enzymes responsible for these reactions are QdtA and FdtA respectively. Sugars that undergo this sort of reaction can be found on the O-antigen of *Campylobacter coli*- a bacteria strongly associated human infection. Whereas the O-antigen from *C. coli* contains a final 3-amino sugar in the quinovose conformation, certain *Bordetella* strains (which can cause human reparatory infections) have a 3-amino fucose derivative in their O-antigen (33, 34).

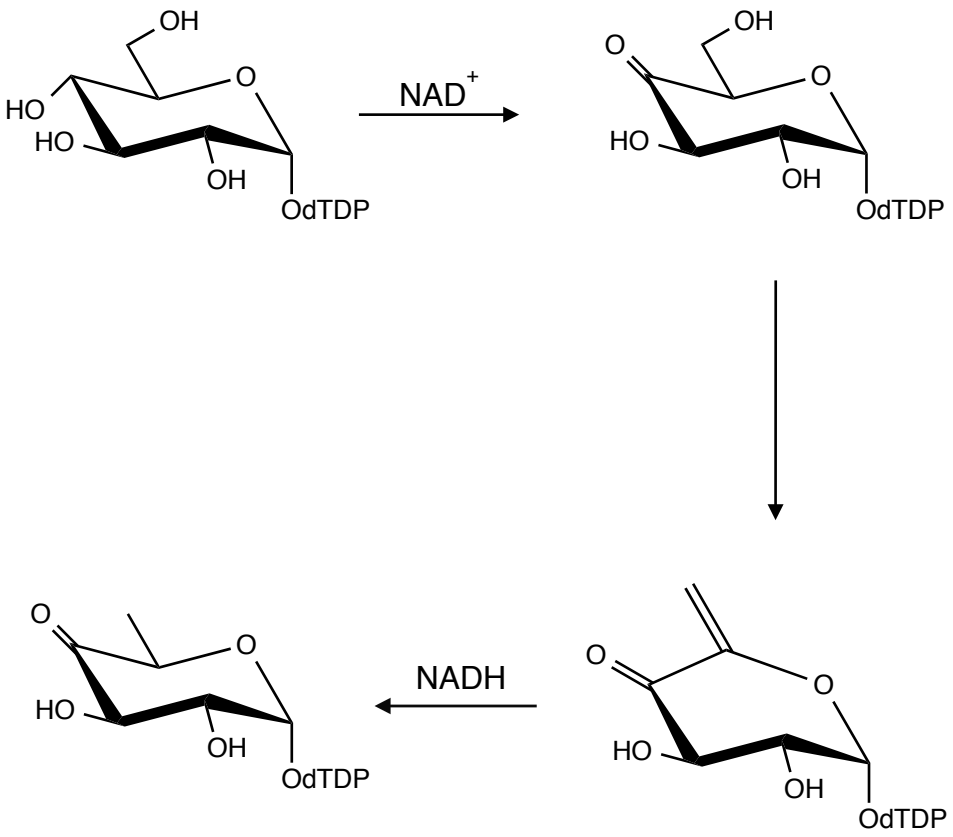


Figure 1.5. Reaction of RmlB, a 4,6-dehydratase. RmlB acts upon nucleotide-linked sugars, in this case, a dTDP linked glucose. RmlB utilizes an NAD^+ cofactor to carry out its chemistry in the three steps show here.

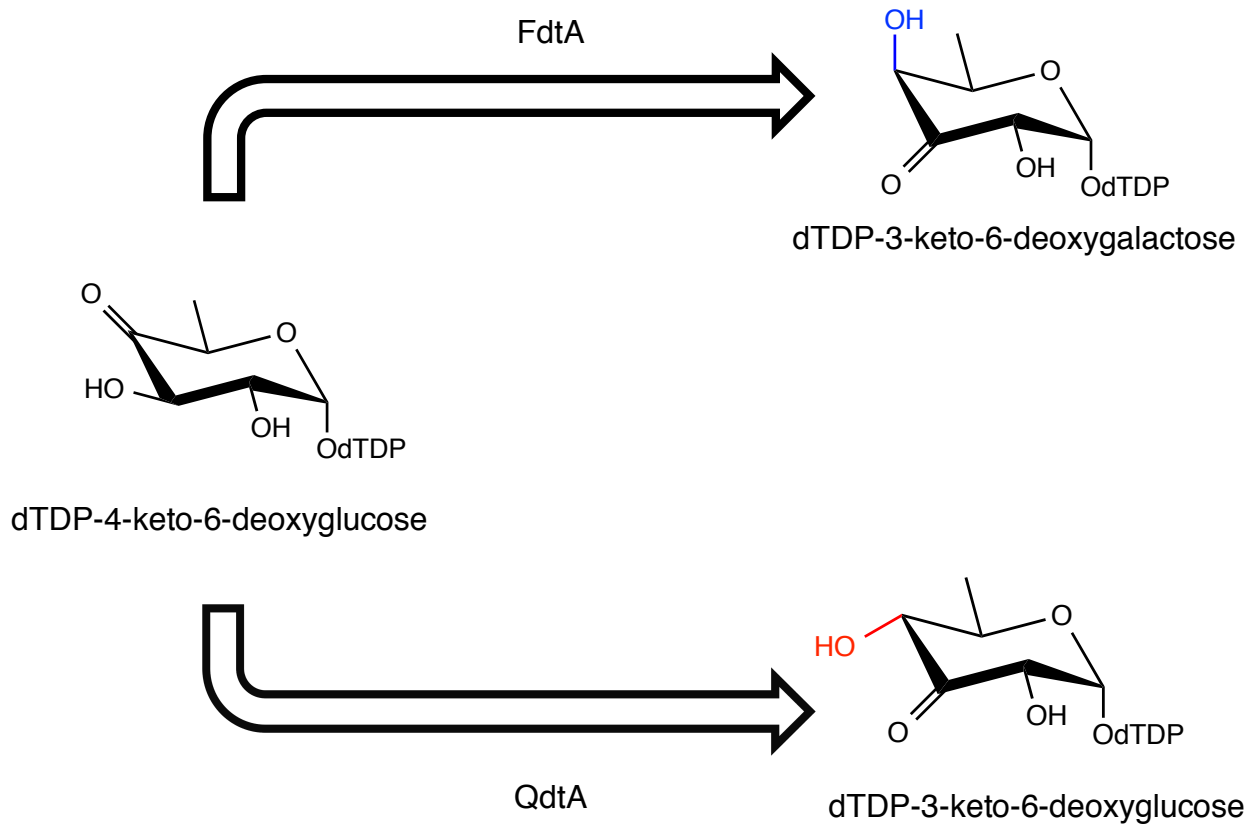


Figure 1.6. Diagram distinguishing the enzymes FdtA and QdtA. Both FdtA and QdtA accept dTDP-4-keto-6-deoxyglucose. However, the geometry of their products varies. QdtA retains the stereochemistry of glucose around C-4′ (red). FdtA inverts the stereochemistry (blue).

As of 2016, there were over 1,200 enzymes annotated as dTDP-3,4-ketoisomerases in the UNIPROT database. They are predicted members of the cupin family of proteins, and are dimers in solution. They come from a wide variety of bacteria and are involved in diverse processes beyond O-antigen biosynthesis. They modify sugars destined for the S-layer of Gram-positive bacteria or sugars destined to be appended to macrolide antibiotics (27, 28, 35). In addition, there are known bifunctional enzymes with active dTDP-3,4-ketoisomerase domains. These are most likely a result of gene duplication which results in having the ketoisomerase on the same polypeptide chain as another sugar-modifying enzyme (36). These diverse variations on the theme of the ketoisomerization make these enzymes interesting to study.

In Chapter 2, two ketoisomerases are studied in detail. Through thorough mechanistic and structural exploration, light is shed upon the question of how an enzyme may or may not change the geometry of its sugar substrate. This chapter culminates in the engineering of a stereochemistry-altering ketoisomerase from a non-altering enzyme and then vis versa.

1.5 3-MONOEPIMERASES

As previously stated, a nucleotide-linked sugar which has been modified by a 4,6-dehydratase is chemically reactive due to the newly acidic hydrogens on the carbons adjacent to the ketone. As chemists know, ketones exist in equilibrium with their enol conformation. While in biological conditions, the ketone form is favored, through a simple acid-base mechanism, an enzyme can take advantage of this chemistry, and a sugar can be epimerized about C-3′ and/or C-5′. These epimerizations are just another way that nature can add variations to sugars.

Predicted to work on CDP- and dTDP-linked sugars, the 3,5-epimerases are relatively small proteins with molecular masses of around 22 kDa (while GDP-linked sugars may also go through a similar epimerization, the enzymes responsible are from the SDR superfamily) (37). Although small, di-epimerases are responsible for shuffling four protons around to complete their reactions, and they do so without any cofactors. Sugar epimerases are unique to prokaryotic cells and are, therefore, potential drug targets. For example, their deletion renders *Streptococcus mutans* unable to initiate or maintain infection (38). Similarly, *Vibrio cholerae* missing their sugar-epimerase carry a phenotype where they are defective in colonization (39).

Interestingly, these cupin family sugar-epimerases exist as both di-epimerases, as well as monoepimerases. In the circumstance of a monoepimerase, the sugar will be epimerized solely around either C-3′ or C-5′. Thus is the case in the biosynthesis of

ascarylose, wherein the enzyme AscE acts on a CDP-linked sugar as a 5′-monoepimerase (40) [Figure 1.7]. Ascarylose is a sugar destined for the O-antigen of *Yersinia pseudotuberculosis* VA. An example of a 3′-monoepimerase can be found in the biosynthetic pathway for dTDP-6-deoxy-D-allose. There are four enzymes necessary for *Streptomyces bikiniensis*, a soil dwelling bacteria, to produce dTDP-6-deoxy-D-allose. This sugar is eventually appended to the antibiotic, chalcomycin. This macrolide antibiotic has strong antibiotic activity towards *Staphylococcus aureus* and *Streptococcus pyogenes* (41). ChmJ is the third enzyme in the dTDP-6-deoxy-D-allose biosynthesis gene cluster and it, indeed, is a 3′-monoepimerase. It was biochemically characterized and found to utilize an active-site tyrosine as a catalytic acid, and a histidine as a catalytic base. It is believed that small changes in active site architecture, particularly around the acidic tyrosine, dictate whether the enzyme will act on C-3′ and/or C-5′ (37, 42).

In Chapter 3, an unusual CDP-sugar monoepimerase, WbcA, is explored. WbcA's primary amino acid sequence deviates from the consensus of hundreds of annotated monoepimerases. Our studies show that this enzyme, while there are different peptides predicted to be in the active site, has an overall fold and mechanism which are analogous to other members of the cupin family of epimerases.

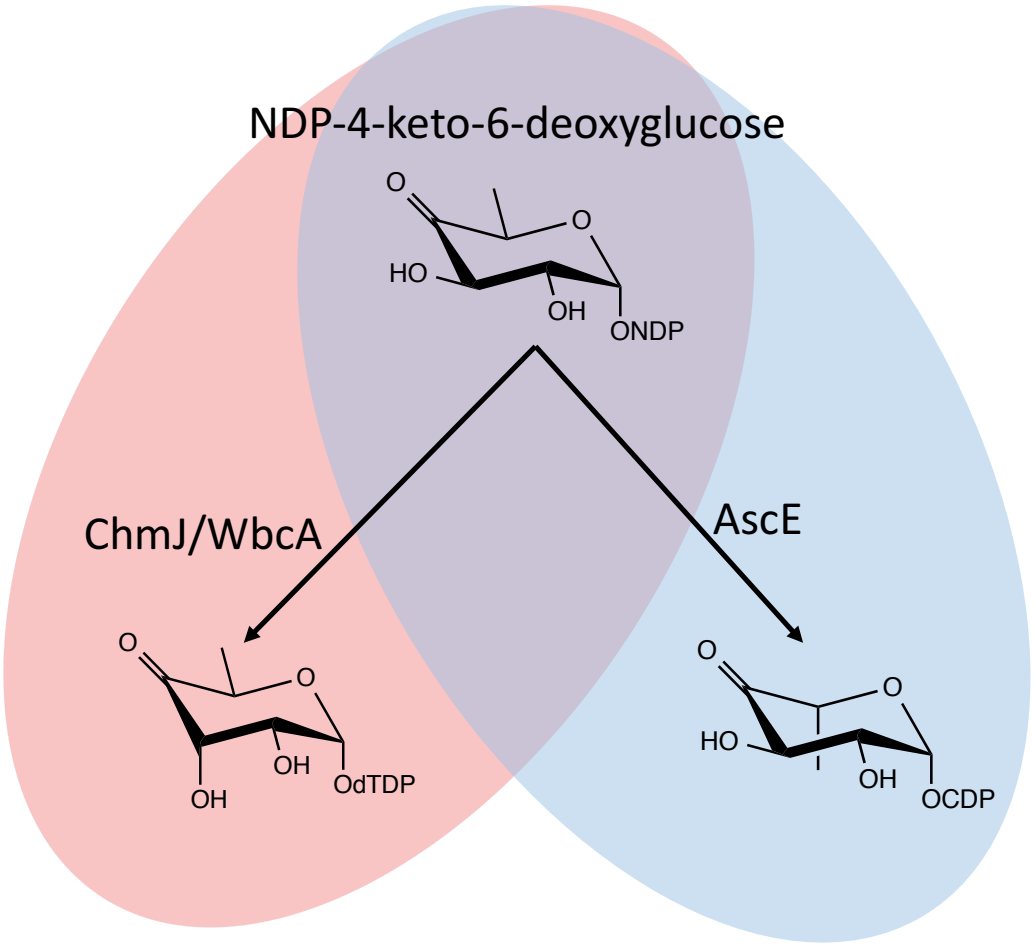


Figure 1.7. 3′-monoepimerase vs. 5′-monoepimerase. AscE acts solely to epimerize the group on C-5′. In contrast, ChmJ and WbcA both act upon the C-3′ hydroxyl group and flip the orientation from axial to equatorial.

1.5 *N*-ACYLTRANSFERASES

As discussed in Section 1.3, many nucleotide-linked deoxysugars undergo a ketoisomerization to yield a 3-keto intermediate. It is not just that carbons adjacent to the ketone that are chemically active, but the ketone itself is quite labile. A common enzymatic alteration is that of amination at either the C-3′ or C-4′ ketone. The resultant deoxyamino sugars can be found in a variety of different bacteria, fungi, and plants (43).

Regardless of the position, whether C-3′ or C-4′, the amino group is transferred to the carbohydrate scaffold via a pyridoxyl 5′-phosphate (PLP) dependent aminotransferase (43, 44). These enzymes utilize glutamine as a co-substrate and eliminate α-ketoglutarate alongside the desired sugar. One of the best-characterized examples of a deoxysugar-aminotransferase is QdtB from *Thermoanaerobacterium thermosaccharolyticum* [Figure 1.8A]. These studies confirmed that this enzyme can accommodate both dTDP-3-keto-D-4,6-dideoxygalactose and dTDP-3-keto-4,6-D-dideoxyglucose as substrates. The amino group is exclusively attached equatorially to the sugar scaffold. This sort of reaction also takes place in the biosynthesis of desosamine sugars from various bacterial strains. The enzyme DesV from the pathway modifies C-3′, while DesI acts as a C-4′ aminotransferase (45-47).

After the amination reaction, once again, the sugar can be further modified: both methyl and acyl groups are frequently found appended to sugar amines. The biological

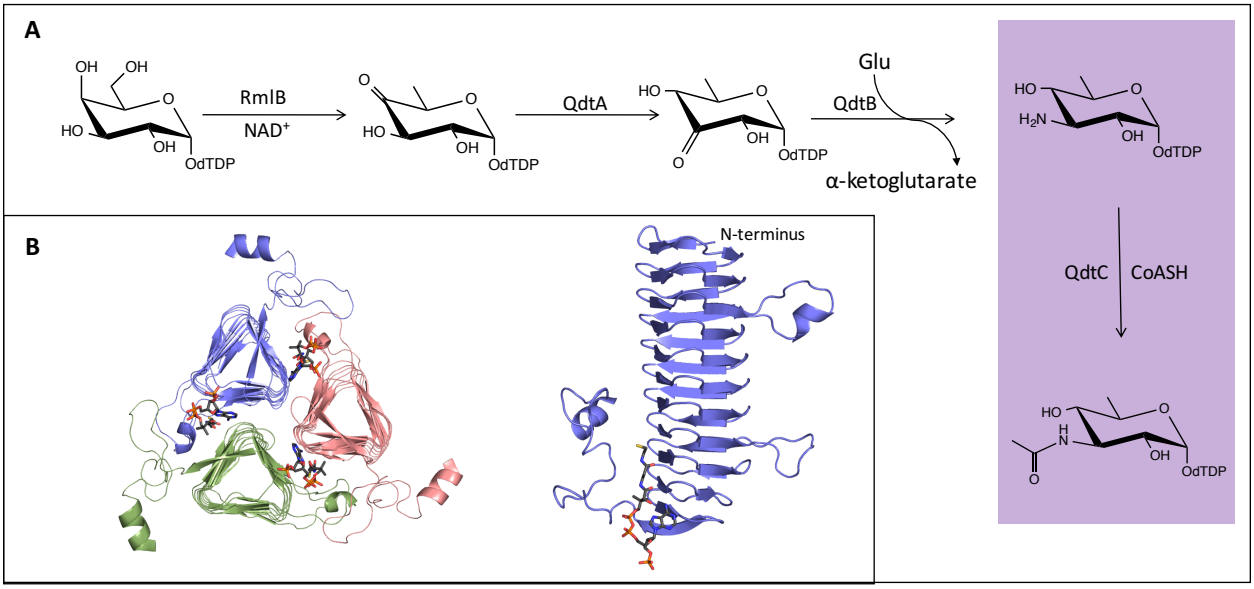


Figure 1.8. A deoxysugar metabolic pathway from *Thermoanaerobacterium thermosaccharolyticum*. A) The complete pathway in the biosynthesis of an acetyl quinovosamine including the amino transferase QdtB and the acyltransferase QdtC. B) The crystal structure of QdtC, a left-handed β-helix protein in its active trimeric assembly. It is also portrayed as a monomer to better illustrate the protein fold.

importance of acetylation emerged in the last decade. This is due to more research suggesting that the signaling role of acyl Coenzyme A (CoA) rivals the importance of phosphorylations (48-50). In fact, it is evidenced that the acyltransfer reaction has a higher energy potential than the phosphoryltransfer reaction (51).

There are two families of enzymes responsible for acetylating nucleotide-linked amino-deoxysugars: the left-handed β-helix family and the GCN5-related *N*-acetyltransferase (GNAT) superfamily. Members of the left-handed β-helix family were first discovered and characterized by Christian Raetz and Steven Roderick (52). They noted that these enzymes had a characteristic tandem-repeat of a six-residue sequence motif, which they named the "isoleucine patch," but has since become known as a "hexapeptide repeat." The pattern observed was that at every sixth position, there was an aliphatic residue- usually Ile, Leu, or Val. In the acyltransfer protein they were studying, LpxA, the repeat occurred 28 times. The fold's name emphasizes that when these chains are properly oriented, they remarkably form regular, triangular helices with left-handed crossover connections [Figure 1.8B]. The structure orients the hydrophobic side chains towards the interior of the β-helix stacks. These enzymes operate as trimers where the active site lies in the groove between monomers.

The GNAT family of enzymes was first characterized 20 year ago (53), however the first member of the family which was described in the literature dates back to 1965 (54). The founding member of the GCN5 *N*-acetyltransferases is a histone

acetyltransferase from *Saccharomyces cerevisiae* (53). As of January 2016, there were well over 100,000 members of the GNAT superfamily. They have been annotated in the genomes of archaea, bacteria, and eukaryotes. While their primary sequences vary significantly, a pyrophosphate binding loop remains semiconserved between GNATs (R/Q-X-X-G-X-A/G) (49, 50, 55).

There are a vast number of potential acyl groups that can be transferred to the sugar substrate. For example, QdtC in an acyltransferase which follows the amino transferase QdtB in the sugar modification pathway from *T. thermosaccharolyticum* [Figure 1.8] (56). QdtC is a member of the left-handed β-helix superfamily and catalyzes the transfer of a single acetyl group from acetyl-CoA to a 3′-amino dTDP-linked deoxysugar. Another example is PerB which attaches an acetyl group from acetyl-CoA to a 4′-amino GDP-linked deoxysugar destined for the O-antigen of *E. coli* O157:H7 (57). While both QdtC and PerB transfer a relatively small acetyl appendage, another left-handed β-helix superfamily member, AntD from *Bacillus anthracis*, transfers a much larger 3-hydroxy-3-methylbutyryl group to a 4′-amino dTDP-linked deoxysugar (58).

These various acyl groups can alter the local chemistry of the sugar quite dramatically. Adding just a simple acetyl group not only gives the sugar another hydrophobic patch, but the amide oxygen can accept two new hydrogen bonds. These variations add another level of complexity to the final sugar product that the bacteria produce.

Until recently, GNAT enzymes which accept nucleotide-linked amino sugars as substrates had not been biochemically characterized. Chapter 4 shows evidence that these enzymes may not require an active-site base to catalyze the acyltransfer reaction. Furthermore, this chapter reveals for the first time the active site architecture of a GNAT enzymes in a ternary complex with the product sugar and a Coenzyme A.

1.6 PROJECT CRYSTAL

A 2012 study by the Program for International Student Assessment (PISA), which was published by the Pew Research Center, ranked The United States of America as 25th among participating countries for science education by evaluating the basic science knowledge of 15-year-old students. As the United States fights to remain competitive at the forefront of science, administrators seem to forget that they must nurture the love of discovery early on in education. Research has confirmed that if educators do not capture their students' interest and enthusiasm for science by the age of 13, the odds of kindling an appreciation for the topic later on are very low. In a collaboration between the Holden Laboratory in the Biochemistry department at the University of Wisconsin-Madison and Spring Harbor Middle School, a pioneering program aimed at middle school science education has been developed and implemented.

Starting in 2009, Project CRYSTAL (Colleagues Researching with Young Scientists Teaching And Learning) was established to bridge the gap between science education and applicable scientific exploration. Since its conception, it has evolved each year to become a better program. In general, four students from Spring Harbor Middle School are chosen from a pool of applicants. These students come to the laboratory for an afternoon, once a week, every week that school is in session. In the laboratory, graduate school student mentors walk the middle school students through a graduate-level research project which begins with isolating genomic DNA and culminates in the

purification and crystallization of a target protein. The final step is to build a three-dimensional model of the crystallized protein. Additionally, the students who have participated have successfully presented their research at the Madison Middle School Science Symposium. Through these means, Project CRYSTAL aims to instill a love and a fascination of science into the lives of young students in the hopes that it continues long into their chosen path of work.

Apart from laboratory work, project CRYSTAL is actively looking into other means to reach as many middle school students as possible. By disseminating custom reading materials, lectures, and experimental equipment designed for middle school education, the Project CRYSTAL team has broadened its outreach to Marquette University and to Indiana University-Indianapolis.

This unique program was in operation for six continuous years and has educated 22 wonderful students. Not only does this program open the eyes of young researchers to the applications of scientific discovery, but it also gives the mentors the opportunity of explaining science in new and exciting ways. More information on the accomplishments of Project CRYSTAL can be found in Chapter 6.

1.8 REFERENCES

1. Hamilton, RJ (2016) Tarascon Pocket Parhacopoeia (2016) 24-26, 63, 77-78, 91, 96-97, 154, 184.

2. Nelson DL, Cox MM, & Lehninger AL (2013) *Lehninger principles of biochemistry* (W.H. Freeman, New York).

3. Holden HM, Cook PD, & Thoden JB (2010) Biosynthetic enzymes of unusual microbial sugars. *Curr. Opin. Struct. Biol.* 20(5):543-550.

4. Thibodeaux CJ, Melancon CE, & Liu HW (2008) Natural-Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chem.-Int. Edit.* 47(51):9814-9859.

5. Lowe JB (1993) Red Cell Membrane and Red Cell Antigens 7: The Blood Group-Specific Human Glycosyltransferases. *Baillière's Clinical Haematology* 6(2):465-492.

6. Voet D, Voet JG, & Pratt CW (2013) *Fundamentals of biochemistry : life at the molecular level* (Wiley, Hoboken, NJ).

7. Weymouth-Wilson AC (1997) The role of carbohydrates in biologically active natural products. *Natural Product Reports* 14(2):99-110.

8. Trefzer A, Salas JA, & Bechthold A (1999) Genes and enzymes involved in deoxysugar biosynthesis in bacteria. *Natural Product Reports* 16(3):283-299.

9. Blankenfeldt W, Asuncion M, Lam JS, & Naismith JH (2000) The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidylyltransferase (RmlA). *Embo J.* 19(24):6652-6663.

10. Oliynyk M, *et al.* (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat Biotech* 25(4):447-453.

11. Oh TJ, Mo SJ, Yoon YJ, & Sohng JK (2007) Discovery and molecular engineering of sugar-containing natural product biosynthetic pathways in actinomycetes. *J. Microbiol. Biotechnol.* 17(12):1909-1921.

12. Tang L & McDaniel R (2001) Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity. *Chem. Biol.* 8(6):547-555.

13. Knirel YA, Valvano MA, & Valvano MA (2011) Bacterial Lipopolysaccharides: Structure, Chemical Synthesis, Biogenesis and Interaction with Host Cells. (Springer Vienna, Vienna).

14. Raetz CRH & Whitfield C (2002) Lipopolysaccharide Endotoxins. *Annual Review of Biochemistry* 71:635-700.

15. Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, & Rick PD (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology* 4(12):495-503.

16. Lüderitz O, Staub AM, & Westphal O (1966) Immunochemistry of O and R antigens of *Salmonella* and related *Enterobacteriaceae*. *Bacteriological Reviews* 30(1):192-255.

17. Kingsley RA & Bäumler AJ (2000) Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm: MicroReview. *Molecular Microbiology* 36(5):1006.

18. Lindberg AA, Segall T, Weintraub A, & Stocker BA (1993) Antibody response and protection against challenge in mice vaccinated intraperitoneally with a live aroA O4-O9 hybrid *Salmonella* dublin strain. *Infection and Immunity* 61(4):1211-1221.

19. Zdanov A, Li Y, Bundle DR, Deng, SJ, MacKenzie CR, Narang SA, Young NM, & Cygler M (1994) Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. *Proceedings of the National Academy of Sciences of the United States of America* 91(14):6423-6427.

20. Rittig MG, Kaufmann A, Robins A, Shaw B, Sprenger H, Gemsa D, Foulogne V, Rouot B, & Dornand J (2003) Smooth and rough lipopolysaccharide phenotypes of *Brucella* induce different intracellular trafficking and cytokine/chemokine release in human monocytes. *Journal of Leukocyte Biology* 74(6):1045-1055.

21. Darwin AJ & Miller VL (1999) Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. *Molecular Microbiology* 32(1):51-62.

22. Skurnik M, Venho R, Bengoechea J-A, & Moriyón I (1999) The lipopolysaccharide outer core of *Yersinia enterocolitica* serotype O:3 is required for virulence and plays a role in outer membrane integrity. *Molecular Microbiology* 31(5):1443-1462.

23. Morona R, Daniels C, & Van Den Bosch L (2003) Genetic modulation of *Shigella flexneri* 2a lipopolysaccharide O antigen modal chain length reveals that it has been optimized for virulence. *Microbiology-(UK)* 149:925-939.

24. Sandlin RC, Lampel KA, Keasler SP, Goldberg MB, Stolzer AL, & Maurelli AT (1995) Avirulence of rough mutants of *Shigella flexneri*: requirement of O antigen for correct unipolar localization of IcsA in the bacterial outer membrane. *Infection and Immunity* 63(1):229-237.

25. VanDenBosch L, Manning PA, & Morona R (1997) Regulation of O-antigen chain length is required for *Shigella flexneri* virulence. *Molecular Microbiology* 23(4):765-775.

26. Melancon CE, Hong L, White JA, Liu YN, & Liu HW (2007) Characterization of TDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase from the D-mycaminose biosynthetic pathway of *Streptomyces fradiae*: In vitro activity and substrate specificity studies. *Biochemistry* 46(2):577-590.

27. Pfoestl A, Hofinger A, Kosma P, & Messner P (2003) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91(T). *J. Biol. Chem.* 278(29):26410-26417.

28. Pfostl A, Zayni S, Hofinger A, Kosma P, Schäffer C, & Messner P (2008) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose. *Biochem. J.* 410:187-194.

29. Melo A & Glaser L (1965) Nucleotide Specificity And Feedback Control Of Thymidine Diphosphate D-Glucose Pyrophosphorylase. *J. Biol. Chem.* 240(1):398-&.

30. He XM & Liu H-w (2002) Formation of Unusual Sugars: Mechanistic Studies and Biosynthetic Applications. *Annual Review of Biochemistry* 71(1):701-754.

31. Jörnvall H, Persson B, Krook M, Atrian S, Gonzàlez-Duarte R, Jeffery J, & Ghosh D (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 34(18):6003-6013.

32. Gerratana B, Cleland WW, & Frey PA (2001) Mechanistic Roles of Thr134, Tyr160, and Lys 164 in the Reaction Catalyzed by dTDP-Glucose 4,6-Dehydratase. *Biochemistry* 40(31):9187-9195.

33. Vinogradov E & Caroff M (2005) Structure of the *Bordetella trematum* LPS O-chain subunit. *FEBS Lett.* 579(1):18-24.

34. Aspinall GO, McDonald AG, Pang H, Kurjanczyk LA, & Penner JL (1993) Lipopolysaccharide Of *Campylobacter coli* serotype O30 - Fractionation And Structure Of Liberated Core Oligosaccharide. *J. Biol. Chem.* 268(9):6263-6268.

35. Thoden JB & Holden HM (2014) The molecular architecture of QdtA, a sugar 3,4-ketoisomerase from *Thermoanaerobacterium thermosaccharolyticum*. *Protein Sci.* 23(6):683-692.

36. Chantigian DP, Thoden JB, & Holden HM (2013) Structural and Biochemical Characterization of a Bifunctional Ketoisomerase/N-Acetyltransferase from *Shewanella denitrificans*. *Biochemistry* 52(46):8374-8385.

37. Dong C, Major LL, Srikannathasan V, Errey JC, Giraud MF, Lam JS, Graninger M, Messner P, McNeil MR, Field RA, Whitfield C, & Naismith JH (2007) RmlC, a C3′ and C5′ Carbohydrate Epimerase, Appears to Operate via an Intermediate with an Unusual Twist Boat Conformation. *Journal of Molecular Biology* 365(1):146-159.

38. Yamashita Y, Tomihisa K, Nakano Y, Shimazaki Y, Oho T, & Koga T. (1999) Recombination between gtfB and gtfC Is Required for Survival of a dTDP-Rhamnose Synthesis-Deficient Mutant of *Streptococcus* mutans in the Presence of Sucrose. *Infection and Immunity* 67(7):3693-3697.

39. Chiang SL & Mekalanos JJ (1999) rfb Mutations in *Vibrio cholerae* Do Not Affect Surface Production of Toxin-Coregulated Pili but Still Inhibit Intestinal Colonization. *Infection and Immunity* 67(2):976-980.

40. Thorson JS, Lo SF, Ploux O, He X, & Liu HW (1994) Studies of the biosynthesis of 3,6-dideoxyhexoses: molecular cloning and characterization of the asc (ascarylose) region from *Yersinia pseudotuberculosis* serogroup VA. *Journal of Bacteriology* 176(17):5483-5493.

41. Coffey GL, Anderson LE, Douros JD, Erlanderson AL, Fisher MW, Hans RJ, Pittillo RF, Vogler DK, Weston KS, & Ehrlich J (1963) Chalcomycin, A New Antibiotic: Biological Studies. *Canadian Journal of Microbiology* 9(5):665-669.

42. Allard STM, Giraud MF, & Naismith JH (2001) Epimerases: structure, function and mechanism. *Cellular and Molecular Life Sciences CMLS* 58(11):1650-1665.

43. Nedal A & Zotchev SB (2004) Biosynthesis of deoxyaminosugars in antibiotic-producing bacteria. *Applied Microbiology and Biotechnology* 64(1):7-15.

44. Timmons SC & Thorson JS (2008) Increasing carbohydrate diversity via amine oxidation: Aminosugar, hydroxyaminosugar, nitrososugar and nitrosugar biosynthesis in bacteria. *Current opinion in chemical biology* 12(3):297-305.

45. Burgie ES & Holden HM (2007) Molecular Architecture of DesI: A Key Enzyme in the Biosynthesis of Desosamine. *Biochemistry* 46(31):8999-9006.

46. Burgie ES, Thoden JB, & Holden HM (2007) Molecular architecture of DesV from *Streptomyces venezuelae*: A PLP-dependent transaminase involved in the biosynthesis of the unusual sugar desosamine. *Protein Science : A Publication of the Protein Society* 16(5):887-896.

47. Thoden JB, Schäffer C, Messner P, & Holden HM (2009) Structural Analysis of QdtB: An Aminotransferase Required for the Biosynthesis of dTDP-3-Acetamido-3,6-Dideoxy-α-D-Glucose. *Biochemistry* 48(7):1553-1561.

48. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, & Mann M (2009) Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* 325(5942):834-840.

49. Favrot L, Blanchard JS, & Vergnolle O (2016) Bacterial GCN5-Related *N*-Acetyltransferases: From Resistance to Regulation. *Biochemistry* 55(7):989-1002.

50. Vetting MW, *et al.* (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Archives of Biochemistry and Biophysics* 433(1):212-226.

51. Cleland WW & Hengge AC (1995) Mechanisms of phosphoryl and acyl transfer. *Faseb J.* 9(15):1585-1594.

52. Raetz CRH & Roderick SL (1995) A Left-Handed Parallel β Helix in the Structure of UDP-*N*-Acetylglucosamine Acyltransferase. *Science* 270(5238):997-1000.

53. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, & Allis CD (1996) Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation. *Cell* 84(6):843-851.

54. Okamoto S & Suzuki Y (1965) Chloramphenicol-, Dihydrostreptomycin-, and Kanamycin-Inactivating Enzymes from Multiple Drug-Resistant Escherichia coli Carrying Episome /`R/'. *Nature* 208(5017):1301-1303.

55. Neuwald AF & Landsman D (1997) GCN5-related histone *N*-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends in Biochemical Sciences* 22(5):154-155.

56. Thoden JB, Cook PD, Schäffer C, Messner P, & Holden HM (2009) Structural and Functional Studies of QdtC: an *N*-Acetyltransferase Required for the Biosynthesis of dTDP-3-Acetamido-3,6-Dideoxy-α-D-Glucose. *Biochemistry* 48(12):2699-2709.

57. Thoden JB, Reinhardt LA, Cook PD, Menden P, Cleland WW, & Holden HM (2012) Catalytic Mechanism of Perosamine *N*-Acetyltransferase Revealed by High-Resolution X-ray Crystallographic Studies and Kinetic Analyses. *Biochemistry* 51(16):3433-3444.

58. Kubiak RL & Holden HM (2012) Structural Studies of AntD: An *N*-Acyltransferase Involved in the Biosynthesis of D-Anthrose. *Biochemistry* 51(4):867-878.

CHAPTER 2:

THE BACTERIAL SUGAR 3,4-KETOISOMERASES: STRUCTURAL INSIGHT INTO PRODUCT STEREOCHEMISTRY

This work was published as: Thoden, J.B., Vinogradov, E., Gilbert, M., Salinger, A.J., and Holden, H.M. (2015) The Bacterial Sugar 3,4-Ketoisomerases: Structural Insight into Product Stereochemistry, *Biochemistry 54, 4495-4506.*

2.1 ABSTRACT

3-acetamido-3,6-dideoxy-D-galactose (Fuc3NAc) and 3-acetamido-3,6-dideoxy-D-glucose (Qui3NAc) are unusual sugars found on the lipopolysaccharides of Gram-negative bacteria and on the S-layers of Gram-positive bacteria. The 3,4-ketoisomerases, referred to as FdtA and QdtA, catalyze the third steps in the respective biosynthetic pathways for these sugars. Whereas both enzymes utilize the same substrate, the stereochemistries of their products are different. Specifically, the hydroxyl groups at the hexose C-4' positions assume the "galactose" and "glucose" configurations in the FdtA and QdtA products, respectively. In 2007 we reported the structure of the apoform of FdtA from *Aneurinibacillus thermoaerophilus*, which was followed in 2014 by the X-ray analysis of QdtA from *Thermoanaerobacterium thermosaccharolyticum* as a binary complex. Both of these enzymes belong to the cupin superfamily. Here we report a combined structural and enzymological study to explore the manner in which these enzymes control the stereochemistry of their products. Various site-directed mutant proteins of each enzyme were constructed and their dTDP-sugar products analyzed by NMR spectroscopy. In addition, the kinetic parameters for these protein variants were measured, and the structure of one, namely the QdtA Y17R/R97H double mutant form, was determined to 2.3-Å resolution. Finally, in an attempt to obtain a model of FdtA with a bound dTDP-linked sugar, the 3,4-ketoisomerase domain of a bifunctional enzyme from *Shewanella denitrificans* was cloned, purified, and crystallized in the

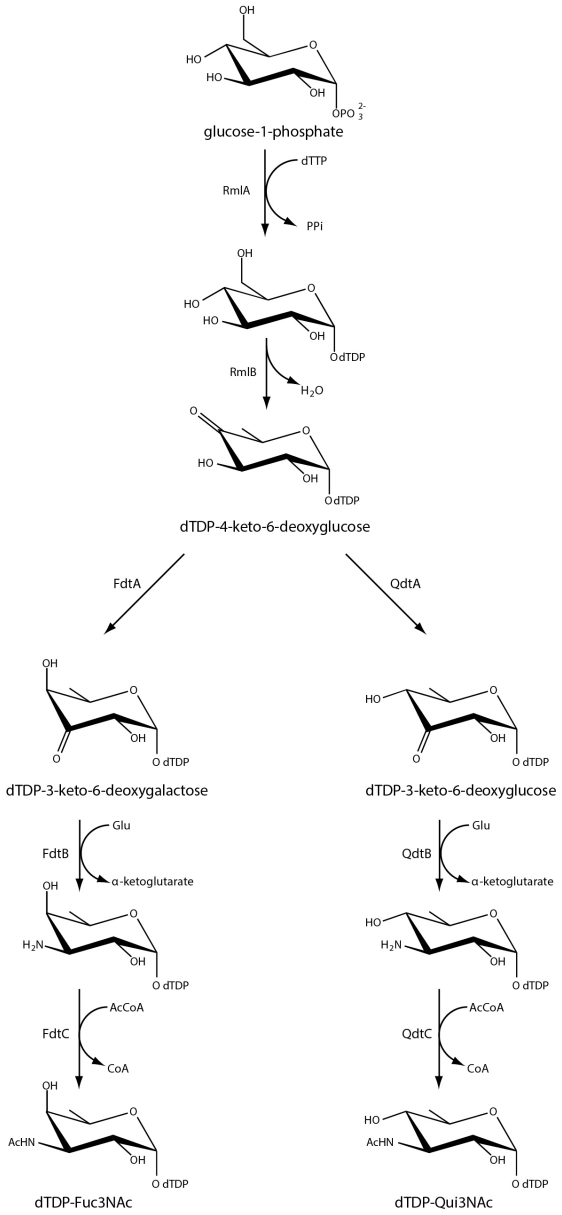
presence of a dTDP-linked sugar analogue. Taken together, the results from this investigation demonstrate that it is possible to convert a “galacto” enzyme into a “gluco” enzyme and vice versa.

2.2 INTRODUCTION

The lipopolysaccharide is a complex glycoconjugate found on the outer membranes of Gram negative bacteria. Conceptually it can be envisioned in terms of three specific components: the lipid A, the core polysaccharide, and the O-antigen. It is the O-antigen, which consists of repeating sugar units, that varies significantly from species to species. Within recent years, it has become increasingly apparent that the O-antigen contributes to bacterial virulence. For example, inactivation of the first gene in the O-antigen cluster in *Francisella tularensis*, the causative agent of rabbit fever, results in severe attenuation of virulence (1). It has also been demonstrated that the O-antigen is essential for full virulence of *Yersinia enterocolitica* O:8, and its absence affects the expression of other *Yersinia* virulence factors (2). As another example, an O-antigen gene cluster has been shown to have a major role in the virulence of the *Escherichia coli* meningitis clone O45:K1:H7 (3).

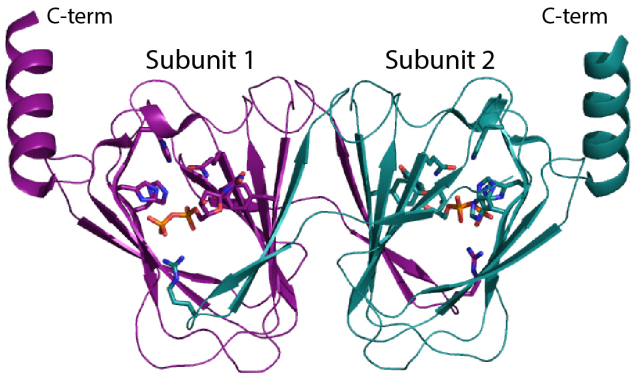
The wide variations observed in the O-antigens of Gram negative bacteria arise from the identities of the sugars present and the order and linkages in which they are attached to one another. Some of the sugars are quite unusual di- and trideoxysugars that contain additional amino, methyl, acetyl, acyl, and formyl groups, amongst others. Two of them, 3-acetamido-3,6-dideoxy-D-galactose (Fuc3NAc) and 3-acetamido-3,6-dideoxy-D-glucose (Qui3NAc) have been the focus of our research attention for the last several years. The biosynthetic pathways for these novel carbohydrates, as outlined in

Scheme 2.1, were first described by the Messner laboratory in 2003 and 2008, respectively (4, 5). As can be seen, both pathways initiate with the attachment of glucose-1-phosphate to dTMP and the subsequent dehydration of dTDP-glucose to yield dTDP-4-keto-6-deoxyglucose. The 3,4-ketoisomerases, referred to as FdtA and QdtA, catalyze the third steps. Whereas both enzymes utilize the same substrate, the stereochemistries of their products are different. Specifically, the hydroxyl groups at the hexose C-4' positions assume the "galactose" and "glucose" configurations in the FdtA and QdtA products, respectively. For the sake of clarity, FdtA and QdtA, are referred to, respectively, as "galacto" or "gluco" enzymes with reference to the orientation about the hexose C-4' carbon.

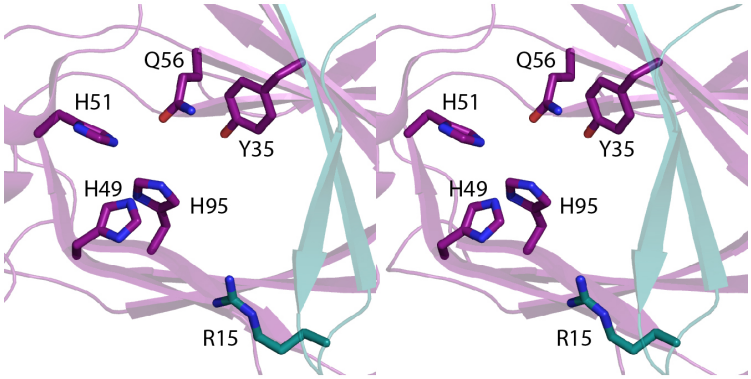


Scheme 2.1 dTDP-3NAc sugar biosynthetic pathways. After appending glucose-1-phosphate with a thyamine nucleotide monophosphate, the resulting sugar is dehydrated about C-4' and C-6' by RmlA. The ketoisomerases in question then act upon the sugar. When FdtA acts upon the substrate, it yields a sugar in the "galacto" configuration about C-4' (left). Whereas when the keto-sugar substrate is acted upon by QdtA, the resulting sugar is in the "gluco" configuration about C-4' (right)

The first structure of a sugar 3,4-ketoisomerase to be reported from our laboratory was that of FdtA from *Aneurinibacillus thermoaerophilus*, a "galacto" enzyme (6). Crystals of the enzyme were grown in the presence of dTDP. Each subunit of the dimeric enzyme was shown to adopt a classical cupin-type fold as can be seen in Figure 1a. A cluster of three histidine residues was observed lining the active site, two of which, His 49 and His 51, were known to be absolutely conserved amongst the 3,4-ketoisomerases [Figure 2.1b]. Subsequent site-directed mutagenesis experiments and activity assays suggested a possible catalytic mechanism for FdtA as outlined in Scheme 2.2. Accordingly, His 49 abstracts the proton from the sugar C-3' and shuttles it to the sugar C-4' on the same side of the pyranosyl ring. His 51 functions in the mechanism by shuttling a proton between the C-3' and C-4' oxygens. The combined action of His 49 and His 51 results in inversion of the stereochemistry about C-4'. Whether the mechanism is step wise, as proposed in Scheme 2.2, or concerted is still unknown. Unfortunately, it was never possible to solve the structure of FdtA in the presence of its substrate or a substrate analog, and thus the proposed mechanism relied solely on hypothetical modeling.

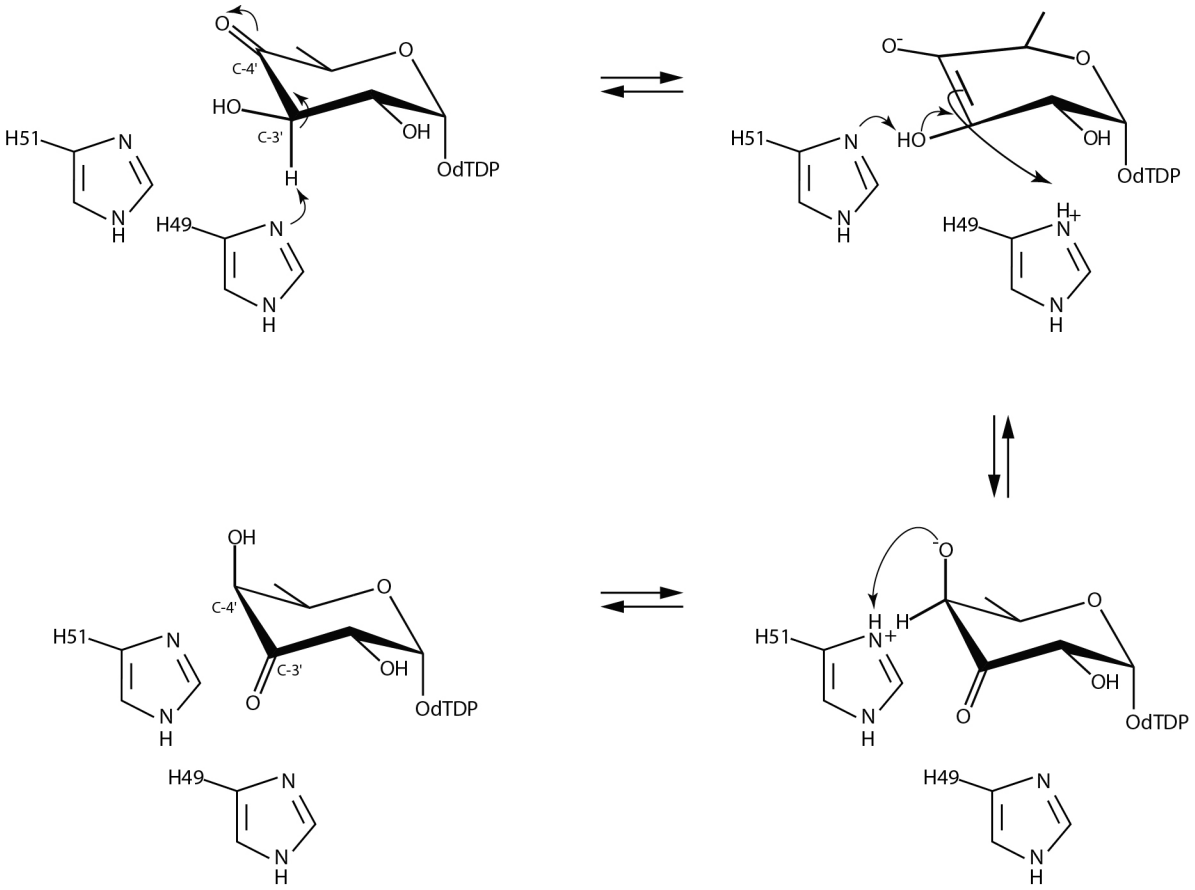


(a)



(b)

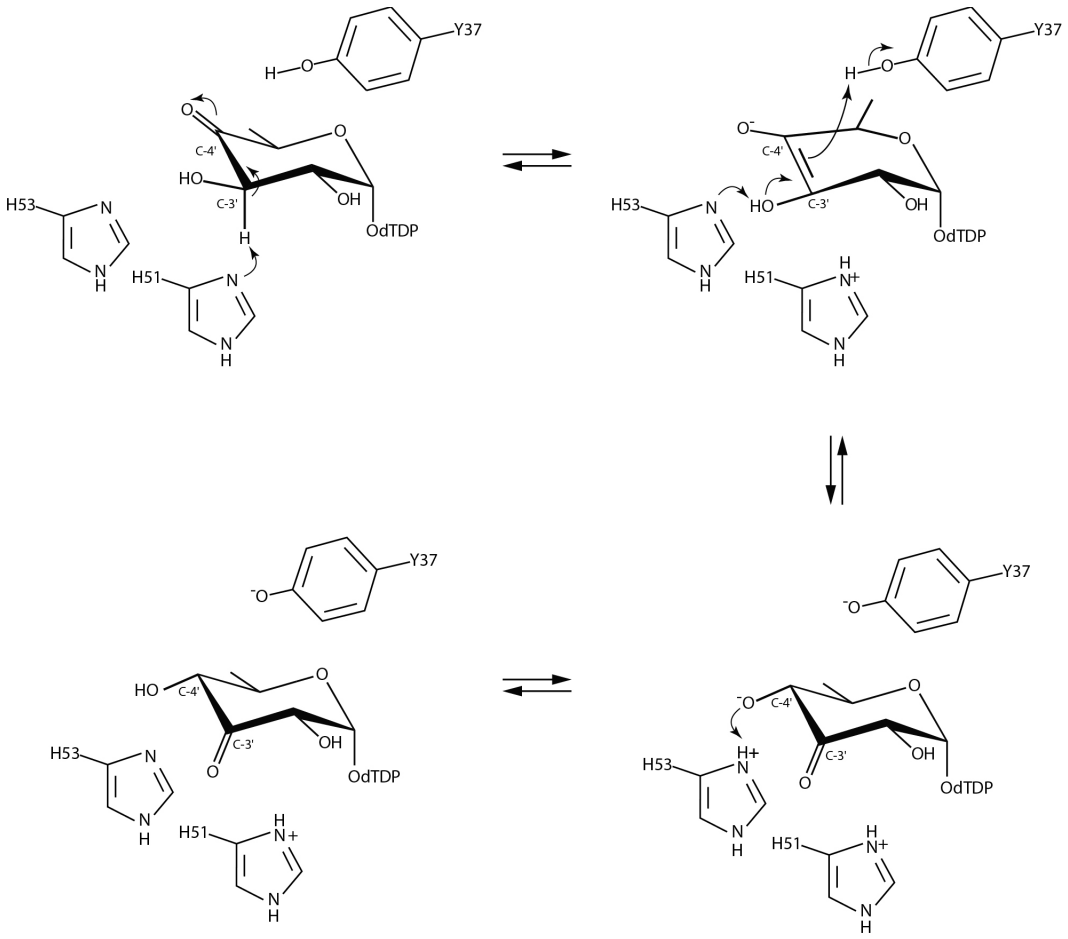
Figure 2.1 The structure of FdtA from *A. thermoaerophilus* (PDB code 2PA7). A ribbon representation of the FdtA dimer is shown in (a) with subunits 1 and 2 displayed in dark violet and dark teal, respectively. The enzyme belongs to the cupin superfamily (18). A close-up view of the active site for subunit 1 is shown in (b). Due to classical domain swapping, Arg 15 is contributed by subunit 2. This figure and Figures 2, 4, 5, 6, and 7 were created with PyMOL (19).



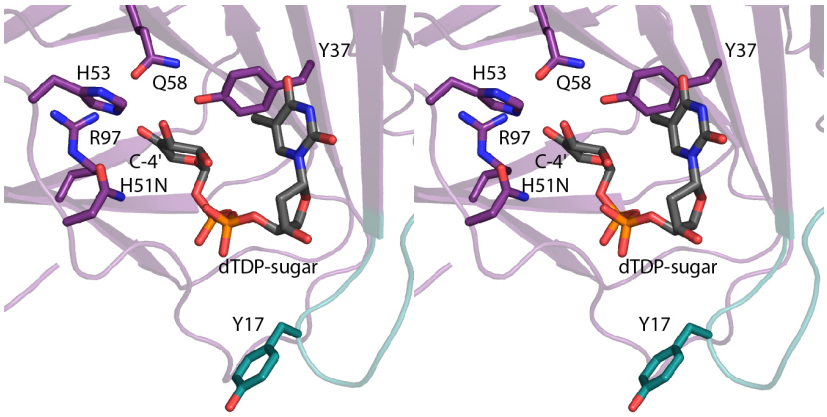
Scheme 2.2 Hypothetical mechanism of 3,4-ketoisomerase FdtA. After mutagenesis experiments, it was proposed that the enzyme utilizes 2 catalytic histidines.

In 2014 we reported the crystal structure of QdtA from *Thermoanaerobacterium thermosaccharolyticum,* which is a "gluco" enzyme (7). Given that QdtA and FdtA demonstrate an amino acid sequence identity of 55%, it was not surprising that QdtA adopted a similar cupin-type architecture. Importantly, however, it was possible to trap the dTDP-4-keto-6-deoxyglucose substrate into the QdtA active site by utilizing the catalytically inert H51N site-directed mutant variant for crystallizations. This structure thus represented the first model of a 3,4-ketoisomerase in complex with a dTDP-sugar ligand [Figure 2.1a] and provided new insight into a "gluco" enzyme. A catalytic mechanism was proposed as highlighted in Scheme 2.3. In the case of QdtA, it was suggested that His 51 abstracts the C-3' proton leading to the formation of an enolate intermediate and the side chain of Tyr 37 functions as an active site acid by protonating the C-4' carbon thereby leading to retention of the "glucose" configuration of the product. The role of His 53 was proposed to function as a proton shuttle between the C-3' hydroxyl and the C-4' keto group. This mechanism was previously proposed for the 3,4-ketoisomerase from *Streptomyces fradie*, which is also a "gluco" enzyme (8). A crystal structure of the *S. fradi*e enzyme has not yet been reported.

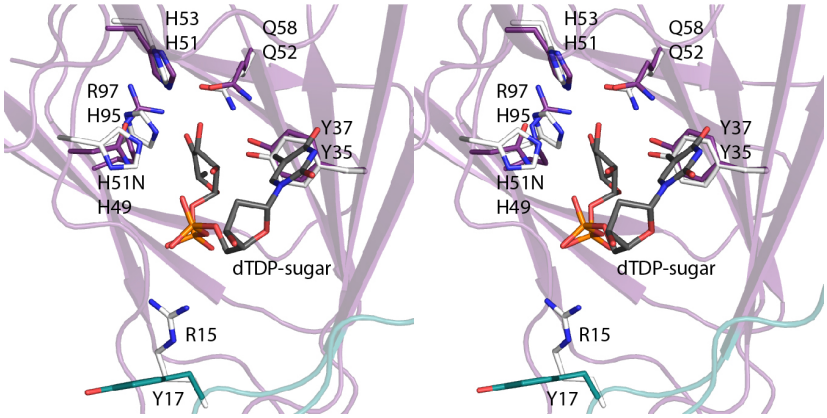
A comparison of the FdtA and QdtA active sites [Figure 2.2b] is especially intriguing given that the two histidines and the tyrosine are conserved. The question then arises as to why FdtA functions as a "galacto" enzyme even though there is a tyrosine in the correct position to function as an active site acid. There are two notable



Scheme 2.3 Hypothetical mechanism of 3,4-ketoisomerase QdtA. In addition to the catalytic, histidine dyad, it is hypothesized that Y37 acts as the catalytic base responsible for the final "gluco" configuration.



(a)



(b)

Figure 2.2 Active site of QdtA from *T. thermosaccharolyticum* (PDB code 4O9G). A close-up view of the QdtA active site with dTDP-4-keto-6-deoxyglucose bound is displayed in (a). A superposition of the QdtA and FdtA active sites is presented in (b). Those residues belonging to QdtA or FdtA are colored in purple (and teal) or white, respectively. The top and bottom amino acid labels refer to QdtA and FdtA, respectively.

changes between the active sites of FdtA and QdtA, however. In FdtA, there is an arginine at position 15, which is replaced by a tyrosine in QdtA [Figure 2.2b]. Additionally, in FdtA there is a histidine at position 95 that is replaced with an arginine in QdtA. To test the roles of these residues in determining the stereochemical outcome of the product, four site-directed mutant proteins of each enzyme were constructed and their dTDP-sugar products analyzed by NMR spectroscopy. In addition, the kinetic parameters for all eight site-directed mutant proteins were measured, and the three-dimensional structure of the QdtA Y17R/R97H double mutant protein was determined. Finally, in an attempt to obtain a model of FdtA with a bound dTDP-linked sugar, the 3,4-ketoisomerase domain of a bifunctional enzyme from *Shewanella denitrificans* was cloned, purified, and crystallized in the presence of a dTDP-linked sugar analogue. The structure was solved and refined to a nominal resolution of 1.7 Å. Here, we describe the results from this combined biochemical and structural investigation and demonstrate that it is, indeed, possible via site-directed mutagenesis to convert a "gluco" enzyme into a "galacto" enzyme and vice versa.

2.3 MATERIALS AND METHODS

2.3.1 Synthesis and Purification of the dTDP-Sugar Ligands and Substrates

dTDP-4,6-dideoxy-4-formamido-glucose was prepared as previously described (9). dTDP-4-keto-6-deoxyglucose was prepared by mixing 25 mM dTDP-glucose, 100 mM HEPES (pH 7.5), and 5 mg/ml *E. coli* RmlB (4,6-dehydratase) to a final volume of 1 mL. The reaction was incubated at 37° C for 7 hours, and RmlB was subsequently removed by filtration with an Ultracel 10 kDa NMWL centrifugal filter (Millipore). Analysis of the reaction mixture via HPLC demonstrated complete conversion of the dTDP-glucose to dTDP-4-keto-6-deoxyglucose. The dTDP-4-keto-6-deoxyglucose was used without further purification.

2.3.2 Site-Directed Mutagenesis

All site-directed mutant proteins of FdtA and QdtA were generated via the QuikChange method of Stratagene. The protein variants were expressed and purified as described for the wild-type enzymes (6, 7).

2.3.3 Cloning of the 3,4-Ketoisomerase Domain from *S. denitrificans FdtD*

The following primers were designed to amplify that portion of the *fdtD* gene encoding the 3,4-ketoisomerse domain:

5'-*CATATG*AGCAATAATTTCGAATTAAAATTGCAGATGTCAAAGGTTAAAGGCG-3'

(NdeI site in italics) and

5'-*CTCGAG*TTAGTTTTGTCTCATTTGTTTAAAAGTTGAATAATCACGAATGTAATCATC-3'

(XhoI site in italics).

The purified PCR product was subsequently A-tailed and ligated into the pGEM-T vector (Promega), which was used to transform *E. coli* DH5α cells for subsequent sequencing. Plasmids with the correct sequence were digested, and the gene encoding the 3,4-ketoisomerase domain ligated into the pET28t vector generating a construct for protein expression with an N-terminal His_6-tag and a TEV cleavage site (10).

2.3.4 Protein Expression and Purification

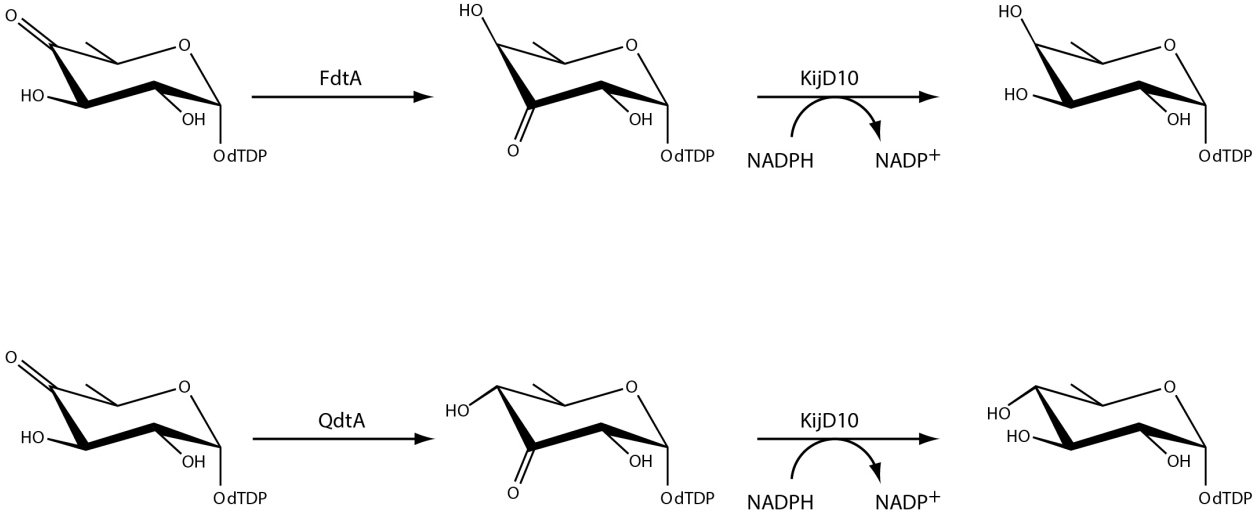
The pET28t plasmid was used to transform Rosetta2(DE3) *E. coli* cells (Novagen). The cultures were grown in lysogeny broth supplemented with kanamycin and chloramphenicol at 37° C with shaking until an optical density of 0.8 was reached at 600 nm. The flasks were cooled in an ice bath, and the cells were induced with 1 mM IPTG and allowed to express protein at 16° C for 24 hours.

The cells were harvested by centrifugation and disrupted by sonication on ice. The lysate was cleared by centrifugation, and the 3,4-ketoisomerase domain was purified utilizing Ni-NTA resin (Qiagen) according to the manufacturer's instructions. For removal of the N-terminal His_6-tag, a solution containing a 30:1 molar ratio (domain:TEV protease) was allowed to digest at 4° C for 48 hours. Uncleaved protein

and the TEV protease were removed by passage over Ni-NTA resin. The protein was dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl and concentrated to 20.0 mg/mL based on an extinction coefficient of 0.87 $(mg/mL)^{-1}cm^{-1}$.

2.3.5 Kinetic Analyses

The kinetic parameters for the *A. thermoaerophilus* FdtA, the *T. thermosaccharolyticum* QdtA, the 3,4-isomerase domain of the *S. denitrificans* FdtD, and the eight site-directed mutant proteins were evaluated according to Scheme 2.4 and using a Beckman Coulter DU-640 spectrophotometer as previously described (11). Specifically, enzymatic activities were measured by monitoring the decrease in absorbance at 340 nm as NADPH is oxidized to $NADP^+$ due to the action of KijD10. Previous studies from the laboratory have verified that KijD10 from *Actinomadura kijaniata* functions as an NADPH-dependent C-3' ketoreductase (12). The reaction mixtures contained 50 mM HEPES (pH 7.5), 0.2 mM NADPH, 1 mg/mL KijD10, and dTDP-4-keto-6-deoxyglucose varying from 0.005 to 5.0 mM, depending on the enzyme/variant being evaluated.



Scheme 2.4 Schematic of enzyme turnover assay. KijD10 is a C-3' ketoreductase which is able to act upon either of the keto sugars presented. The loss of absorbance at 340 nm depicts the relative turnover of the ketoisomerase.

The reactions were initiated by the addition of the 3,4-ketoisomerases to the reaction mixtures in the following concentrations: 0.33 μM (FdtA wild-type), 0.19 μM (FdtA Y35F), 2.2 μM (FdtA R15Y), 0.11 μM (FdtA H95R) 6.0 μM (FdtA R15Y/H95R), 0.26 μM (QdtA wild-type), 21 μM (QdtA Y37F), 1.2 μM (QdtA Y17R), 0.79 μM (QdtA R97H), 8.8 μM (QdtA Y17R/R97H), and 1.2 μM (3,4-ketoisomerase domain of FdtD).

Plots of concentrations versus initial rates were analyzed using PRISM (GraphPad Software, Inc.) and were fitted to the equation $v_o=(V_{max}[S])/(K_M+[S])$. Kinetic parameters are listed in Table 2.1. The required *A. thermoaerophilus* FdtA and *T. thermosaccharolyticum* QdtA enzymes were purified as previously described (6, 7).

Table 2.1 Kinetic Constants for Isomerases and Mutant Enzymes

Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	% original sugar epimer[1]
FdtA wild-type	0.50 ± 0.03	10.5 ± 0.3	2.1×10^4	100
FdtA Y35F	0.15 ± 0.01	3.3 ± 0.2	2.2×10^4	80
FdtA R15Y	0.41 ± 0.05	0.19 ± 0.02	4.6×10^2	100
FdtA H95R	0.045 ± 0.005	1.23 ± 0.08	2.7×10^4	40
FdtA R15Y/H95R	0.37 ± 0.04	0.18 ± 0.02	4.9×10^2	0
QdtA wild-type	0.14 ± 0.01	3.85 ± 0.25	2.8×10^4	100
QdtA Y37F	0.034 ± 0.007	0.023 ± 0.004	6.9×10^2	20
QdtA Y17R	0.37 ± 0.06	0.32 ± 0.04	8.6×10^2	90
QdtA R97H	0.061 ± 0.009	0.40 ± 0.06	6.5×10^3	65
QdtA Y17R/R97H	0.021 ± 0.005	0.026 ± 0.005	1.2×10^3	5
Isomerase domain from FdtD	0.79 ± 0.12	4.85 ± 0.56	6.1×10^3	100

[1]Determined by NMR analysis of products from separate reactions. For wild-type FdtA, the sugar product is in the "galacto" configuration. For the FdtA Y35F variant, 80% of the product was still in the "galacto" configuration whereas 20% was in the "gluco" configuration. For the FdtA R15Y/H95R variant, all sugar product was in the "gluco" configuration. For wild-type QdtA, the sugar product is in the "gluco" configuration. In the case of the QdtA Y17R/R97H variant, only 5% of the resulting product was in the "gluco" configuration whereas 95% of it was in the "galacto" configuration.

2.3.6 Sugar Product Analyses

^{1}H NMR spectra were recorded using a Varian Inova 500 MHz spectrometer in D_2O solutions at 25°C with acetone standard (2.225 ppm for ^{1}H). The relative content of the *gluco* and *galacto* isomers in the reaction products was estimated by the integration of α-anomeric ^{1}H NMR 500 MHz signals of free Qui3N (5.81 ppm) and Fuc3N (5.86 ppm), which were present as decomposition products of dTDP derivatives. Integration of the dTDP-sugar signals was problematic because of the overlap: H-1 were at 5.59 ppm (dTDP-Qui3N) and 5.61 ppm (dTDP-Fuc3N) (half overlapped), and worse with H-6, which mixed dTDP-sugar signals with that of free sugars. The integral of glycosyl phosphate H-1 relative to free α-anomeric signals showed that the samples contained about 25% of free sugars relative to their nucleotide derivatives. Visual estimate showed that the ratio of free sugars reflected their ratio as dTDP derivatives, thus both products decomposed to the same extent.

2.3.7 Crystallization of Enzymes

Crystallization conditions for the FdtD 3,4-ketoisomerase domain were surveyed by the hanging drop method of vapor diffusion using a laboratory-based sparse matrix screen. The enzyme was initially tested either in the presence of 5 mM dTDP or in the absence of any ligands. The TEV-cleaved version of the protein yielded the best crystals from the initial screens. X-ray diffraction quality crystals of the protein in the absence of

ligands were subsequently grown from precipitant solutions composed of 16 - 18% poly(ethylene glycol) 8000, 200 mM NaCl, and 100 mM MOPS (pH 7). The crystals belonged to the monoclinic space group $P2_1$ with unit cell dimensions of a = 85.4 Å, b = 64.6 Å, c = 113.0 Å and β= 111.4°. The asymmetric unit contained three dimers. The crystals were prepared for X-ray data collection by serially transferring them from a synthetic mother liquor composed of 20% poly(ethylene glycol) 8000, 200 mM NaCl, 3 mM dTDP-4,6-dideoxy-4-formamido-glucose, and 100 mM MOPS (pH 7) to a cryoprotectant solution containing 25% poly(ethylene glycol) 8000, 300 mM NaCl, 3 mM dTDP-4,6-dideoxy-4-formamido-glucose, 18% ethylene glycol, and 100 mM MOPS (pH 7).

Crystallization conditions for the QdtA Y17R/R97H variant were surveyed similarly either in the presence of 5 mM dTDP or in the absence of any ligands. The C-terminally tagged histidine version of the protein yielded the best crystals from the initial screens. X-ray diffraction quality crystals of the protein in the absence of ligands were subsequently grown from precipitant solutions composed of 20 - 24% pentaerythritol propoxylate (5/4 PO/OH) and 100 mM MES (pH 6). The crystals belonged to the monoclinic space group *C2* with unit cell dimensions of a = 136.6 Å, b = 96.2 Å, c = 111.8 Å and β= 114.5°. The asymmetric unit contained four dimers. The crystals were prepared for X-ray data collection by serially transferring them to a cryoprotectant solution containing 35% pentaerythritol propoxylate (5/4 PO/OH), 200

mM NaCl, 5% ethylene glycol, and 100 mM MES (pH 6). The X-ray diffraction pattern was somewhat anisotropic in nature.

X-ray diffraction quality crystals of the C-terminally tagged histidine QdtA apoenzyme were grown from precipitant solutions composed of 18 - 22% pentaerythritol propoxylate (5/4 PO/OH) and 100 mM Homo-pipes (pH 5). The crystals belonged to the tetragonal space group $P42_12$ with unit cell dimensions of $a = b = 93.7$ Å, and $c = 95.6$ Å. The asymmetric unit contained one dimer. The crystals were prepared for X-ray data collection by serially transferring them to a cryoprotectant solution containing 35% pentaerythritol propoxylate (5/4 PO/OH), 200 mM NaCl, 5% ethylene glycol, and 100 mM Homo-pipes (pH 5).

2.3.8 X-ray Data Collection and Processing

All X-ray data sets were collected in house using a Bruker AXS Platinum 135 CCD detector controlled with the APEX software suite (Bruker AXS Inc.) The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. The data sets were processed with SAINT and scaled with SADABS (Bruker AXS Inc.). Relevant X-ray data collection statistics are listed in Table 2.2.

Table 2.2: X-ray Data Collection Statistics.

	QdtA Y17R/R97H	QdtA Apoenzyme	Isomerase domain from FdtD
resolution limits (Å)	30.0-2.3 (2.4-2.3)[b]	50-1.80 (1.86-1.80)[b]	30-1.70 (1.76-1.70)[b]
number of independent reflections	52869 (5554)	37504 (3408)	115507 (1254)
completeness (%)	90.3 (79.3)	93.5 (87.1)	91.8 (82.4)
redundancy	2.9 (1.5)	12.4 (5.5)	3.3 (2.1)
avg I/avg σ(I)	10.1 (1.8)	44.5 (1.7)	46.4 (6.2)
R_{sym} (%)[a]	6.5 (24.4)	4.8 (35.3)	5.2 (13.2)

[a] $R_{sym} = (\sum|\Sigma I - \overline{I}\ |/ \sum I) \times 100$.

[b] Statistics for the highest resolution bin

The structure of the FdtD 3,4-ketoisomerase domain was solved by molecular replacement with PHASER (13). The search probe was a fragment of the full length FdtD model corresponding to residues Ser 160 to Asn 304 (11). Iterative cycles of model building with COOT (14, 15) and refinement with REFMAC (16) reduced the R_{work} and R_{free} to 17.1% and 20.7%, respectively, from 30 - 1.7 Å resolution.

The QdtA Y17R/R97H protein structure was solved by molecular replacement with PHASER using as a starting model PDB entry 4O9G (7). Iterative cycles of model building with COOT and refinement with REFMAC reduced the R_{work} and R_{free} to 19.2% and 25.6%, respectively, from 30 – 2.3 Å resolution.

The QdtA apoenzyme structure was solved by molecular replacement with PHASER using as a starting model PDB entry 4O9G (7). Iterative cycles of model building with COOT and refinement with REFMAC reduced the R_{work} and R_{free} to 20.5% and 23.7%, respectively, from 50 – 1.8 Å resolution. Model refinement statistics for all structures are listed in Table 2.3.

Table 2.3: Refinement Statistics.

	QdtA Y17R/R97H	QdtA Apoenzyme	Isomerase domain from FdtD
resolution limits (Å)	30-2.3	50.0-1.8	30-1.7
[a]*R*-factor (overall)%/no. reflections	19.6/52868	20.6/37504	17.3/115507
R-factor (working)%/no. reflections	19.2/50208	20.5/35613	17.1/109726
R-factor (free)%/no. reflections	25.6/2660	23.7/1889	20.7/5781
number of protein atoms	9148	2285	7090
number of heteroatoms	404	140	934
average B values			
protein atoms ($Å^2$)	41.4	40.3	27.7
ligand ($Å^2$)	57.3 (dTDP)	33.8 (thymidine)	27.1 (dTDP-sugar)
solvent ($Å^2$)	34.5	43.4	35.1
weighted RMS deviations from ideality			
bond lengths (Å)	0.013	0.012	0.013
bond angles (º)	1.78	1.93	1.94
planar groups (Å)	0.007	0.012	0.011
Ramachandran regions (%)[b]			
most favored	86.8	88.6	91.5
additionally allowed	12.3	11.0	8.1
generously allowed	0.9	0.4	0.4

R-factor = $(\Sigma|F_O - F_C| / \Sigma|F_O|) \times 100$ where F_O is the observed structure-factor amplitude and F_C is the calculated structure-factor amplitude.

[b]Distribution of Ramachandran angles according to PROCHECK (17).

2.4 RESULTS AND DISCUSSION

2.4.1 Biochemical Analyses

Both FdtA and QdtA catalyze 3,4-isomerization reactions by employing a conserved histidine as the catalytic base required to remove the proton from the C-3' carbon (Schemes 2.2 and 2.3). The stereochemistry about C-4' of their resulting products is different, however. To address the manner in which these two enzymes yield different products starting from the same substrate we first determined their kinetic parameters. Plots of initial velocities versus substrate concentrations for wild-type FdtA and QdtA are shown in Figure 3. Relevant kinetic parameters are listed in Table 2.1. Both FdtA and QdtA show similar catalytic efficiencies of 2.1 x 10^4 s^{-1} M^{-1} and 2.8 x 10^4 s^{-1} M^{-1}, respectively.

FdtA is a "galacto" enzyme, and thus it was reasoned that replacement of Tyr 35 with a phenylalanine would not affect the overall stereochemical outcome of the product [Figure 1b]. Analysis of the sugar product demonstrated that the FdtA Y35F variant produced a mixture that contained 80% of the expected sugar in the "galactose" configuration, but interestingly 20% of it was in the "glucose" configuration indicating that proton transfer had occurred on the opposite side of the pyranosyl ring (Table 2.1). The source of the proton is not known but clearly the change from a tyrosine to a phenylalanine affected the overall electrostatic distribution of the active site pocket.

Although both the K_m and k_{cat} for the Y35F mutant protein decreased, the catalytic efficiency remained the same as that for the wild-type enzyme. Most likely the Y35F mutation resulted in non-productive substrate binding. This would lead to a decrease in K_m due to the availability of more substrate binding modes, but also to a decrease in k_{cat} because a certain percentage of the substrate is not bound correctly for productive catalysis. As a control, sugar analysis of the product from wild-type FdtA showed 100% of it in the galactose configuration.

In QdtA, which is a "gluco" enzyme, Tyr 37 is thought to function as the active site acid [Scheme 2.3, Figure 2.2a]. By replacing tyrosine with a phenylalanine, it was reasoned that QdtA could be changed to a "galacto" enzyme. Sugar analysis of the resulting product of the QdtA Y37F variant demonstrated that while 20% of the product was in the "glucose" configuration, more importantly 80% was now in the "galactose" configuration. As a control, sugar analysis of the resulting product from wild-type QdtA showed 100% of it in the glucose configuration as expected. Thus by simply changing the identity of the residue at position 35/37 in FdtA and QdtA, it was possible to affect the ratio of the product outcome for these two enzymes. In the case of the QdtA Y37F mutant protein, the catalytic efficiency dropped by a factor of ~40.

As highlighted in Figure 2.2b, there are two positions that differ significantly between FdtA and QdtA (Arg 15/Tyr 17 and His 95/Arg 97, respectively). To explore the roles of these residues in product outcome, four site-directed mutant variants were

subsequently constructed and tested: FdtA R15Y, FdtA H95R, QdtA Y17R, and QdtA R97H. As can be seen from Table 2.1, the FdtA R15Y and QdtA Y17R variants produced either 100% or 90% of the sugars expected for the wild-type enzymes. These residues are positioned somewhat away from the active site cleft so the results were not entirely unexpected. The FdtA H95R and QdtA R97H mutations had a more drastic effect, however. For the FdtA H95R variant, 60% of the sugar product was now in the "glucose" configuration, whereas for the QdtA R97H protein, 35% of the sugar product was in the "galactose" configuration. Again, these mutations changed the charge distribution in the active sites and thus could have perturbed the pK_a of Tyr 35 (FdtA) or Tyr 37 (QdtA).

Strikingly, as presented in Table 2.1, the double mutant variants FdtA R15Y/H95R and QdtA Y17R/R97H showed nearly complete conversions, namely the FdtA R15Y/H95R protein now behaved as a "gluco" enzyme and the QdtA Y17R/R97H protein functioned as a "galacto" enzyme. Plots of initial velocities versus substrate concentrations for these double mutant variants are shown in Figure 2.3. The catalytic efficiencies of the FdtA R15Y/H95R and the QdtA Y17R/R97H mutant proteins were reduced by factors of 43 and 23, respectively.

To ensure that the double mutations did not affect the overall structures of the proteins, crystallization trials were initiated on both enzymes. Whereas it was not possible to obtain X-ray quality crystals of the FdtA R15Y/H95R double mutant protein,

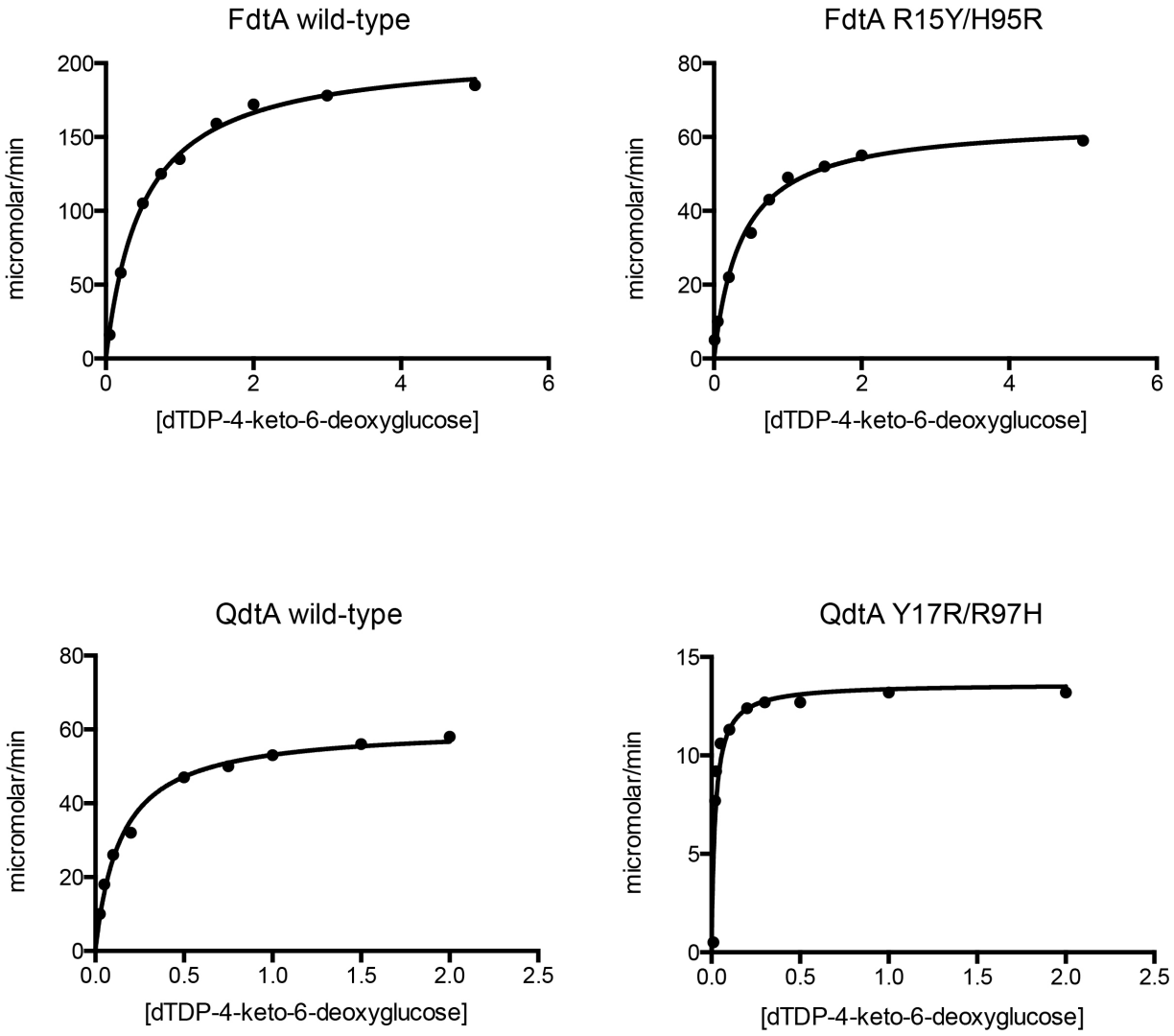


Figure 2.3 Plots of initial velocities versus substrate concentrations. The substrate concentrations (dTDP-4-keto-6-deoxyglucose) are in mM. The rate of product formation is given as micomolar/min.

single crystals were obtained for the QdtA Y17R/R97H enzyme that diffracted to 2.3-Å resolution. These crystals contained four dimers in the asymmetric unit with the α-carbons of subunit 1 superimposing on those of the other subunits with root-mean-square deviations of between 0.18 Å and 0.27 Å. Although the protein was crystallized in the absence of dTDP, it was clear that each subunit in the asymmetric unit contained ligands with varying degrees of occupancies. The electron density corresponding to the ligand in subunit 5 is presented in Figure 2.4. Clearly the mutant protein acquired dTDP during protein expression and purification. The α-carbons for the wild-type and double mutant protein superimpose with a root-mean-square deviation of 0.25 Å, indicating that little structural perturbation occurred upon the substitution of Tyr 17 with an arginine and Arg 97 with a histidine residue.

Curious as to whether the wild-type QdtA enzyme also fortuitously binds nucleotides during protein expression and purification in *E. coli*, it was also crystallized, and its structure solved and refined to 1.8-Å resolution and an *R*-factor of 20.6%. In electron density maps calculated with F_o-F_c coefficients, there was some residual density in both subunits of the dimer that was suggestive of thymidine moieties bound at low occupancies, but there was no clear density for dTDP ligands.

The QdtA Y17R/R97H structure emphasizes the need for not only measuring the kinetic parameters of site-directed mutant proteins but also for determining their three-

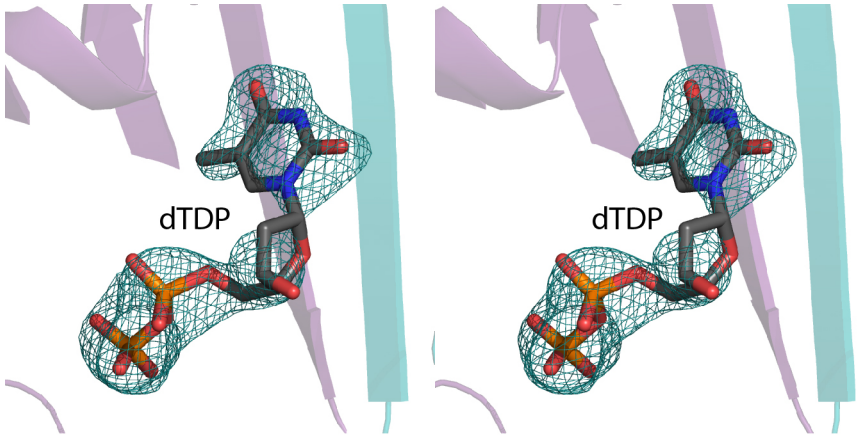


Figure 2.4 Observed electron density for the dTDP moiety bound to the QdtA Y17R/R97H mutant enzyme. The electron density shown corresponds to that observed in the fifth subunit of the asymmetric unit. The map, contoured at 2σ was calculated with coefficients of the form Fo - Fc, where Fo was the native structure factor amplitude and Fc was the calculated structure factor amplitude.

dimensional architectures. The manner in which the accompanying dTDP molecule affects the overall kinetic parameters of the mutant protein is unclear.

2.4.2 Structure of the FdtD 3,4-Ketoisomerase Domain/dTDP-Sugar Complex

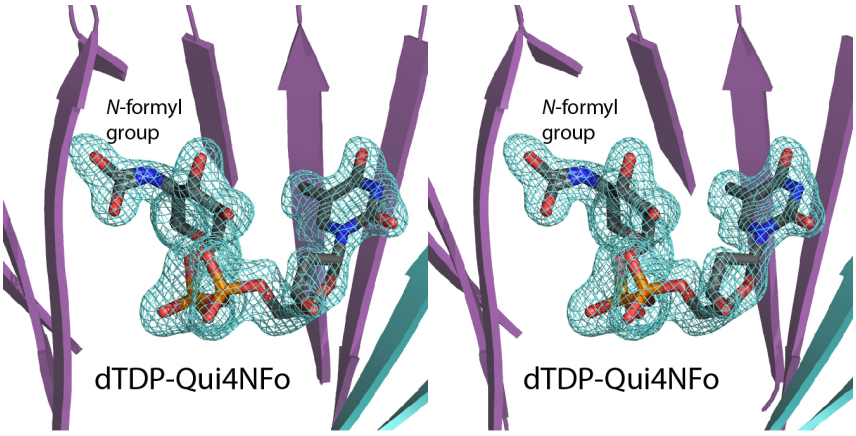
For nearly eight years the laboratory has attempted to solve the structure of the *A. thermoaerophilus* FdtA in the presence of a dTDP-sugar substrate or analog with no success. In addition, more than five homologs of FdtA were cloned but again none of the resulting proteins yielded X-ray quality crystals. However, the three-dimensional structure of a bifunctional ketoisomerase/*N*-acetyltransferase, referred to as FdtD, was known from our recent research (11). FdtD, like the *A. thermoaerophilus* FdtA, is a "galacto" enzyme (11). Given the overall structure of the bifunctional FdtD, it was possible to clone and express only its 3,4-ketoisomerase domain, which extends from Ser 160 to Asn 304. To ensure that the isolated domain retained catalytic activity, its kinetic parameters were determined and are listed in Table 2.1. The catalytic efficiency of the isolated 3,4-ketoisomerase domain is 6.1×10^3 s^{-1} M^{-1}. When it is part of the bifunctional enzyme, the catalytic efficiency is 9.8×10^4 s^{-1} M^{-1} (11).

Crystals of the isolated domain were then soaked in various dTDP-linked sugars in an effort to obtain a binary complex. These ligands included dTDP-4-keto-6-deoxyglucose, dTDP-4-amino-4,6-dideoxyglucose, dTDP-3-amino-3,6-dideoxyglucose, dTDP-3-amino-3,6-dideoxygalactose, and dTDP-4,6-dideoxy-4-formamido-glucose.

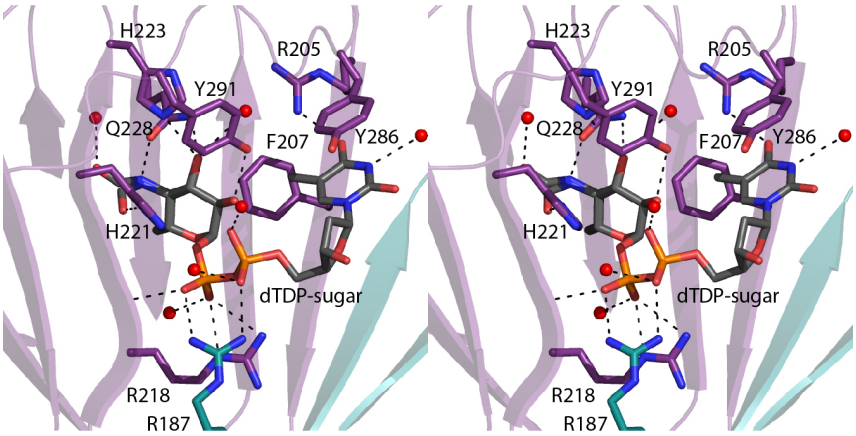
Only the latter sugar bound in the crystalline lattice without destroying its diffraction properties.

The crystals of the isomerase domain used in this investigation belonged to the space group $P2_1$ with three dimers in the asymmetric unit. Electron density corresponding to the dTDP-sugar in subunit 1 is shown in Figure 2.5a. As can be seen, the *N*-formyl group adopts alternate conformations in the active site. Whereas the electron density for the dTDP-sugar was unambiguous in subunit 1, only partial electron densities were observed for the ligands in subunits 2 and 3. No ligands could be modeled into subunits 4, 5, and 6. As such the following discussion refers only to the first dimer in the X-ray coordinate file and specifically to subunit 1.

A close-up view of the 3,4-ketoisomerase domain is displayed in Figure 2.5b. As typically observed for these enzymes, the thymine ring of the dTDP-sugar is sandwiched between two aromatic residues, Phe 207 and Tyr 286. The side chain of Arg 205 and a water molecule lie within hydrogen bonding to the thymine ring as well. The pyrophosphoryl moiety of the dTDP-sugar participates in electrostatic interactions with Arg 218 from subunit 1 and Arg 187 from subunit 2. Three additional water molecules lie within 3.2 Å of the phosphoryl oxygens. Finally, the pyranosyl and *N*-formyl moieties are surrounded by His 221, His 223, Gln 228, Tyr 291, and two water molecules.



(a)



(b)

Figure 2.5 Structure of the 3,4-ketoisomerase domain of FdtD. The structure of the enzyme was solved in the presence of dTDP-4,6-dideoxy-4-formamido-D-glucose. Shown in (a) is the observed electron density corresponding to the dTDP-sugar in subunit 1. The map, contoured at 3σ was calculated with coefficients of the form Fo - Fc, where Fo was the native structure factor amplitude and Fc was the calculated structure factor amplitude. A close-up view of the active site in subunit 1 is presented in (b). Possible hydrogen bonding interactions are indicated by the dashed lines. Water molecules are depicted as red spheres. Arg 187, highlighted in teal, belongs to subunit 2 of the dimer.

As described in the Introduction, the first 3,4-ketoisomerase structure solved in the presence of a dTDP-sugar ligand was QdtA from *T. thermosaccharolyticum*. For this investigation, the H51N variant was utilized for crystallization, and the structure was solved in the presence of the true substrate, dTDP-4-keto-6-deoxyglucose. Shown in Figure 2.6 is a superposition of the active sites of QdtA and the ketoisomerase domain of FdtD with bound ligands. The α-carbons for these two enzymes superimpose with a root-mean-square deviation of 1.2 Å. As can be seen, the two nucleotide-linked sugar ligands adopt similar conformations when bound in the active sites of QdtA and FdtD, and thus the dTDP-4,6-dideoxy-4-formamido-glucose used in the soaking experiments represents an appropriate substrate analog. A close-view of the region surrounding the C-3' and C-4' carbons of the substrate analog when bound to the isolated isomerase domain is depicted in Figure 2.7. His 221 is in the correct position to function as the active site base to remove the C-3' proton and shuttle it to the C-4' carbon whereas His 223 is ideally situated to move a proton from the C-3' hydroxyl to the C-4' keto oxygen.

In conclusion, the sugar 3,4-ketoisomerases represent a fascinating family of enzymes that catalyze reactions on the same substrate but result in products with either the “glucose” or “galactose” configuration about the C-4' carbons. The active sites for the two classes of isomerases have evolved to tightly control the stereochemistry of the resulting products. Simple changes to the active sites of these enzymes via site-directed mutagenesis results in loss of this fine-tuning.

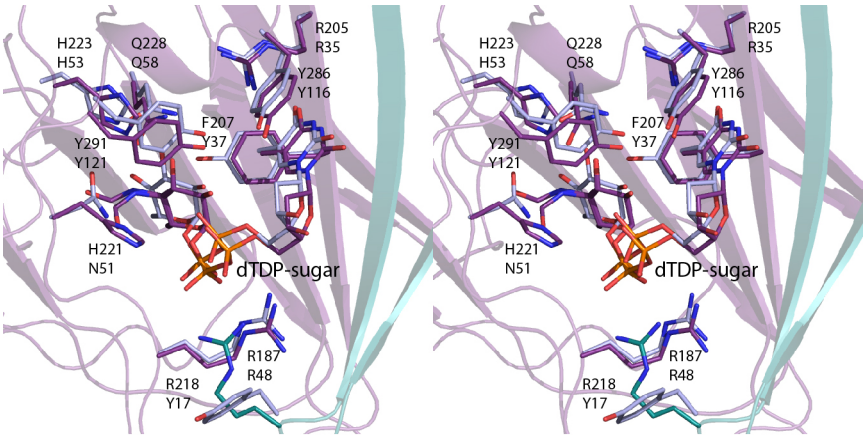


Figure 2.6 Comparison of dTDP-sugar binding in the isomerase domain of FdtD and QdtA. The side chains and dTDP-sugar ligands for the FdtD isomerase domain and QdtA are displayed in violet and blue, respectively. The top and bottom residue labels correspond to the FdtD isomerase domain and QdtA. Note that in the wild-type form of QdtA, Asn 51 corresponds to a histidine. In order to trap a dTDP-sugar in the active site, the H51N variant was employed for crystallizations (7).

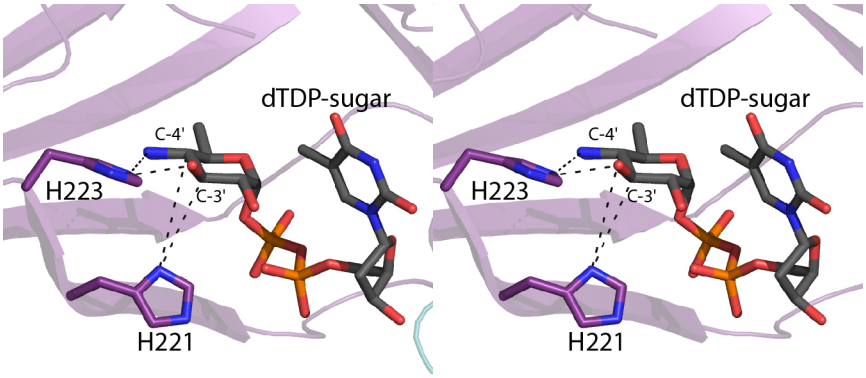


Figure 2.7 The positions of the conserved histidines in the FdtD isomerase domain. The conserved histidines are properly positioned to function in proton transfer as indicated by the dashed lines. For the sake of clarity, the formyl group of the ligand was removed.

By the judicious choice of site-directed mutations, it has, indeed, been possible to convert a “gluco” enzyme into a “galacto” one, and vice versa. In addition, the investigation described herein has finally led to the structure of a “galacto” enzyme with a bound dTDP-linked substrate analog, which further supports the original catalytic proposal put forth in 2007 (6).

2.5 ACKNOWLEDGMENTS

We thank Professor Grover Waldrop for helpful comments.

2.6 REFERENCES

1. Raynaud C, Meibom KL, Lety MA, Dubail I, Candela T, Frapy E, & Charbit A (2007) Role of the *wbt* locus of *Francisella tularensis* in lipopolysaccharide O-antigen biogenesis and pathogenicity. *Infect Immun* 75(1):536-541.

2. Bengoechea JA, Najdenski H, & Skurnik M (2004) Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. *Mol Microbiol* 52(2):451-469.

3. Plainvert C, Bidet P, Peigne C, Barbe V, Médigue C, Denamur E, Bingen E, & Bonacorsi S (2007) A new O-antigen gene cluster has a key role in the virulence of the *Escherichia coli* meningitis clone O45:K1:H7. *J Bacteriol* 189(23):8528-8536.

4. Pfoestl A, Hofinger A, Kosma P, & Messner P (2003) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-galactose in *Aneurinibacillus thermoaerophilus* L420-91T. *J Biol Chem* 278(29):26410-26417.

5. Pföstl A, Zayni S, Hofinger A, Kosma P, Schäffer C, & Messner P (2008) Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose. *Biochem J* 410(1):187-194.

6. Davis ML, Thoden JB, & Holden HM (2007) The X-ray structure of dTDP-4-keto-6-deoxy-D-glucose-3,4-ketoisomerase. *J Biol Chem* 282(26):19227-19236.

7. Thoden JB & Holden HM (2014) The molecular architecture of QdtA, a sugar 3,4-ketoisomerase from *Thermoanaerobacterium thermosaccharolyticum*. *Protein Sci* 23(6):683-692.

8. Tello M, Rejzek M, Wilkinson B, Lawson DM, & Field RA (2008) Tyl1a, a TDP-6-deoxy-D-xylo-4-hexulose 3,4-isomerase from *Streptomyces fradiae*: structure prediction, mutagenesis and solvent isotope incorporation experiments to investigate reaction mechanism. *Chembiochem* 9(8):1295-1302.

9. Zimmer AL, Thoden JB, & Holden HM (2014) Three-dimensional structure of a sugar *N*-formyltransferase from *Francisella tularensis*. *Protein Sci* 23:273-283.

10. Thoden JB & Holden HM (2005) The molecular architecture of human *N*-acetylgalactosamine kinase. *J Biol Chem* 280(38):32784-32791.

11. Chantigian DP, Thoden JB, & Holden HM (2013) Structural and biochemical characterization of a bifunctional ketoisomerase/*N*-acetyltransferase from *Shewanella denitrificans*. *Biochemistry* 52(46):8374-8385.

12. Kubiak RL & Holden HM (2011) Combined structural and functional investigation of a C-3''-ketoreductase involved in the biosynthesis of dTDP-L-digitoxose. *Biochemistry* 50(26):5905-5917.

13. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, & Read RJ (2007) Phaser crystallographic software. *J. Appl. Cryst.* 40:658-674.

14. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.

15. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486-501.

16. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.

17. Laskowski RA, Moss DS, & Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231(4):1049-1067.

18. Galperin MY & Koonin EV (2012) Divergence and convergence in enzyme evolution. *J Biol Chem* 287(1):21-28.

19. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. *The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA*.

CHAPTER 3:

BIOCHEMICAL STUDIES ON WBCA, A SUGAR EPIMERASE FROM *YERSINIA ENTEROCOLITICA*

This work was published as: Salinger, A.J., Brown, H.A., Thoden, J.B., and Holden, H.M. (2015) Biochemical Studies On Wbca, A Sugar Epimerase from *Yersinia enterocolitica*, *Protein Sci. 10, 1633-1639.*

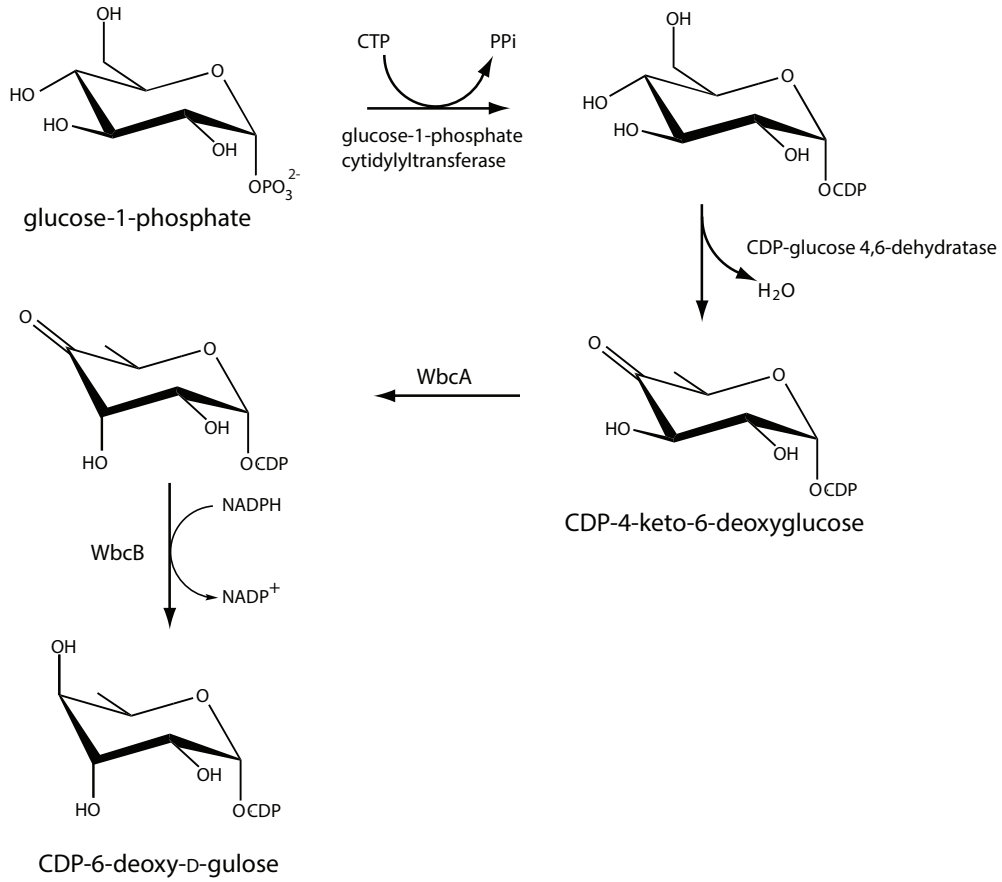
3.1 ABSTRACT

Yersinia enterocolitica is a Gram-negative bacterium that causes yersiniosis, a zoonotic disease affecting the gastrointestinal tract of humans, cattle, and pigs, among others. The lipopolysaccharide of *Y. enterocolitica* O:8 contains an unusual sugar, 6-deoxy-D-gulose, which requires four enzymes for its biosynthesis. Here, we describe a combined structural and functional investigation of WbcA, which catalyzes the third step in the pathway, namely an epimerization about the C-3' carbon of a CDP-linked sugar. The structure of WbcA was determined to 1.75-Å resolution, and the model was refined to an overall R-factor of 19.5%. The fold of WbcA places it into the well-defined cupin superfamily of sugar epimerases. Typically, these enzymes contain both a conserved histidine and a tyrosine residue that play key roles in catalysis. On the basis of amino acid sequence alignments, it was anticipated that the "conserved" tyrosine had been replaced with a cysteine residue in WbcA (Cys 133), and indeed this was the case. However, what was not anticipated was the fact that another tyrosine residue (Tyr 50) situated on a neighboring β-strand moved into the active site. Site-directed mutant proteins were subsequently constructed and their kinetic properties analyzed to address the roles of Cys 133 and Tyr 50 in WbcA catalysis. This study emphasizes the continuing need to experimentally verify assumptions that are based solely on bioinformatics approaches.

3.2 INTRODUCTION

Yersinia is a genus consisting of enteric rod-shaped Gram-negative bacteria. Of the various species identified thus far, only three are known human pathogens: *Yersinia pestis,* the causative agent of bubonic plague, and *Yersinia pseudotuberculosis* and *Yersinia enterocolitica,* both of which can lead to severe gastrointestinal distress (3). Like most Gram-negative bacteria, members of this genus contain on their outer membranes complex glycoconjugates referred to as lipopolysaccharides. These remarkable molecules consist of a lipid A component, a core oligosaccharide, and an O-antigen, which often contains highly unusual deoxysugars. Whereas the core oligosaccharides are typically conserved within a specific genus, the O-antigens are quite variable and clearly contribute to the wide species variation observed in nature (4). Much research has focused on the lipid A component of the lipopolysaccharide and its role in bacterial virulence (5-7). Recent investigations are beginning to unravel the biological role of the O-antigen in bacterial toxicity as well. Indeed, studies suggest that the O-antigen of *Y. enterocolitica* O:8 is essential for virulence (8, 9).

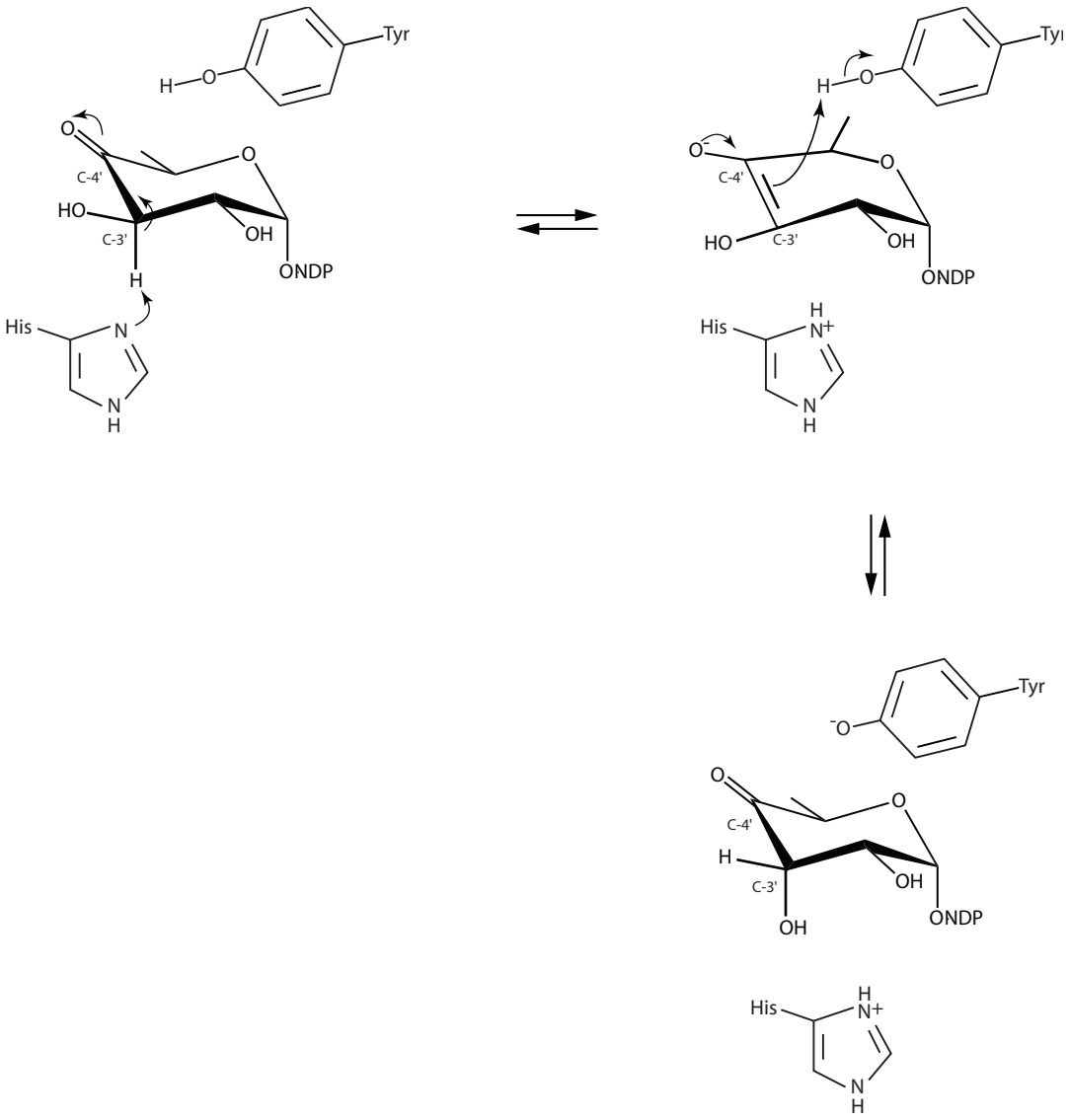
In 1997, the chemical composition of the O-antigen from *Y. enterocolitica* O:8 was reported to consist of repeat units containing N-acetyl-D-galactosamine, D-galactose, D-mannose, L-fucose, and the unusual sugar, 6-deoxy-D-gulose (10). Subsequently, on the basis of genome analysis, the biosynthetic pathway for the production of 6-deoxy-D-gulose was suggested as presented in Scheme 3.1 The first step involves the attachment



Scheme 3.1 Predicted biosynthetic pathway for the production of CDP-6-deoxy-D-gulose.

of a nucleoside monophosphate to glucose-1-phosphate, a reaction catalyzed by glucose-1-phosphate cytidylyltransferase. In the next step, the C-6' hydroxyl is removed and the C-4' hydroxyl is converted to a keto moiety by the action of CDP-glucose 4,6-dehydratase. Both the cytidylyltransferase and the dehydratase have been well characterized structurally and biochemically from either *Salmonella typhi* or *Y. pseudotuberculosis* (11-14). The third step in the pathway, namely a C-3' epimerization, is reportedly catalyzed by WbcA, the focus of this investigation. The pathway is thought to be completed by the action of WbcB, a C-4' ketoreductase.

On the basis of amino acid sequence alignments, WbcA was thought to belong to the "cupin" family of sugar epimerases that catalyze reactions about the C-3' or C-5' carbons. Some function as monoepimerases whereas others catalyze epimerization reactions about both carbons. In all the enzymes that have been structurally characterized to date, two conserved residues, a histidine and a tyrosine, are thought to function as catalytic bases and acids, respectively (15-20). The reaction mechanism for a C-3' monoepimerase is outlined in Scheme 3.2. The conserved histidine abstracts the proton from the C-3' carbon resulting in an enolate anion intermediate. The conserved tyrosine provides a proton to the opposite face of the pyranosyl ring resulting in the epimerization about the C-3' carbon. It is not known whether the reaction mechanism is stepwise or concerted.



Scheme 3.2 Possible catalytic mechanism for a C-3' monoepimerase.

According to amino acid sequence alignments [Figure 3.1], the conserved tyrosine in WbcA is apparently replaced with a cysteine residue (Cys 133). In order to explore the active site architecture of a sugar epimerase lacking the conserved tyrosine, we initiated a combined structural and functional investigation of WbcA as reported herein. For this analysis, the structure of WbcA was determined to 1.75 Å resolution. Strikingly, the structure revealed that whereas the conserved histidine (His 63) was present in WbcA, the cysteine residue was not in the appropriate position to function as an active site acid. Rather, a tyrosine residue (Tyr 50), located on a neighboring β-strand, moved into the proper position. Site-directed mutant proteins were subsequently constructed and their kinetic parameters measured to confirm the catalytic roles of His 63, Tyr 50, and Cys 133. Details concerning the structure and kinetic properties of WbcA are presented.

```
ChmJ            ----MHPLSIEGAWSQEPVIHSDHRGRSHEWFRGESFRQAFG-HDFPVAQVNVAVSHRGA 55
NovW            --MRLRPLGIEGVWEITPEQRADPRGVFLDWYHVDRFAEAIG-RPLRLAQANLSVSVRGV 57
EvaD            --MQARKLAVDGAIEFTPRVFADDRGLLILPYQEEAFVEAHGGPLFRVAQTIHSMSKRGV 58
RmlC            -MMIVIKTAIPDVLILEPKVFGDERGFFFESYNQQTFEELIG-RKVTFVQDNHSKSKKNV 58
WbcA            MIINITELNISGCYLIESPIFSDERGEFVKTHHQEIFKNFGL--EIPSAEEYYSRSKNNV 58

                    #
ChmJ            LRGIHYTEIPPGQAKYSVCVRGAGLDVVVDVRIGSPTFGRWEIVPMDAERNTAVYLTAGL 115
NovW            VRGIHFVDVPPGQAKYVTCVRGAVFDVVVDLRVGSPTYGCWEGTRLDDVSRRAVYLSEGI 117
EvaD            VRGIHYTVTPPGTAKYVYCARGKAMDIVIDIRVGSPTFGQWDSVLMDQQDPRAVYLPVGV 118
RmlC            LRGLHFQRGENAQGKLVRCAVGEVFDVAVDIRKESPTFGQWVGVNLSAENKRQLWIPEGF 118
WbcA            IRGMHFQQYPDDHNKLVFCPEGEVLDVFLDIRKDSNTYGQFMSFILNPHNRRSIFLAKGI 118

                             *
ChmJ            GRAFLSLTDDATLVYLCSSGYAPAREHSVNPLDPDLGIAWPDDIEPLLSDRDENAPTLAT 175
NovW            GHGFCAISDEATLCYLSSGTYDPATEHGVHPLDPELAIDWPTG-TPLLSPRDQDALLLAE 176
EvaD            GHAFVALEDDTVMSYMLSRSYVTQDELALSALDPALGLPIDIGVEPIVSDRDRVAITLAE 178
RmlC            AHGFVTLSEYAEFLYKATNYYSPSSEGSILWNDEAIGIEWPFSQLPELSAKDAAAPLLDQ 178
WbcA            AHGFLSMKDNTLIVCKTSTVHSPSRDSGIHWN--SFGFKWPVE-NPIISDKDRNLDCFF- 174

ChmJ            AERLGLLPTYQAWQEQQQAQR---------- 196
NovW            ARDAGLLPTYATCQAVTVPSPAPGSVGDPGP 207
EvaD            AQRQGLLPDYTTSQEIERRLTA--------- 200
RmlC            ALLTE------------------------- 183
WbcA            ------------------------------
```

Figure 3.1 Sequence alignment of WbcA and other 3'-monoepimerases. Enzymes from *Streptomyces bikiniensis* (ChmJ), *Streptomyces spheroids* (NovW), A*mycolatopsis orientalis* (EvaD), and *Salmonella enterica* (RmlC). The conserved, active-site histidine and conserved, active-site tyrosine, are highlighted in yellow. WbcA is aligns a cysteine (C133) in place of the tyrosine.

3.3 MATERIALS AND METHODS

3.3.1 Cloning wbcA from *Yersinia enterocolitica* (type O:8)

The wbcA gene from *Y. enterocolitica* ATCC 9610 served as the starting template for PCR using platinum Pfx DNA polymerase (Invitrogen). Primers were designed that incorporated NheI and XhoI restriction sites. The PCR product was digested with NheI and XhoI and ligated into pET28T, a laboratory pET28b(+) vector that had been previously modified to incorporate a TEV protease cleavage recognition site after the N-terminal polyhistidine tag (21).

3.3.2 Protein Expression and Purification

The pET28T-wbcA plasmid was utilized to transform Rosetta2(DE3) *Escherichia coli* cells (Novagen). The cultures were grown in lysogeny broth supplemented with kanamycin (35 μg/mL) and chloramphenicol (25 μg/mL) at 37° C with shaking until an optical density of 0.8 was reached at 600 nm. The flasks were cooled in an ice bath, and protein expression was initiated with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside.

The cells were allowed to express protein at 16° C for 20 h and then harvested by centrifugation, flash frozen, and subsequently disrupted by sonication on ice. The lysate was cleared by centrifugation, and WbcA was purified with Ni-NTA resin (Qiagen) according to the manufacturer's instructions. The histidine tag was removed by digestion

with TEV protease. The protein was dialyzed against 10 mM Tris–HCl (pH 8.0) and concentrated to 18 mg/mL based on an extinction coefficient of 1.1 $(mg/mL)^{-1}cm^{-1}$. The 4-ketoreductase (WbcB) required for the kinetic assays and the CDP-sugar synthesis was cloned and purified in a similar manner.

CDP-6-deoxy-D-gulose was enzymatically prepared using 30 mL reactions containing 1.2 M CDP glucose, 0.5 mg/mL CDP-D-glucose 4,6 dehydratase (12), 1 mg/mL WbcA, 0.5 mg/mL WbcB (Scheme 3.1), 5 mM NADPH, and 50 mM HEPPS (pH 8). The reaction was incubated for 6 h at 21° C. Enzymes were removed by passage through an Amicon Ultra 10 kDa cutoff membrane. The reaction mixture was diluted and loaded on to a HiLoad 26/10 Q-sepharose column and purified by HPLC at a buffered pH of 4.0 with an ammonium acetate gradient of 0–1.2M over 12 column volumes. The product sugar, which eluted at approximately 600 mM ammonium acetate, was pooled and lyophilized.

3.3.3 Crystallization and Structural Analysis

Crystallization conditions were surveyed by the hanging drop method of vapor diffusion using a laboratory based sparse matrix screen at both room temperature and 4° C. X-ray diffraction quality crystals of the enzyme were subsequently grown from precipitant solutions containing 14-18% O-methylpoly (ethylene glycol) 5000, 5 mM CDP-6-deoxy-D-gulose, 2% glycerol, and 100 mM MOPS (pH 7.0) at 4° C. The crystal

morphology was hexagonal with the largest crystals showing deep indentations. Moderately sized crystals (~0.4 mm in length) were used for X-ray data collection. The crystals belonged to the space group *C*2 with unit cell dimensions of a = 79.2 Å, b = 42.7 Å, c = 150.5 Å, and β = 90.48°. The asymmetric unit contained three subunits.

For X-ray data collection, the crystals were transferred gradually to a cryoprotectant solution containing 25% O-methyl-poly(ethylene glycol) 5000, 100 mM NaCl, 16% glycerol, 5 mM CDP-6-deoxy-D-gulose, and 100 mM MOPS (pH 7.0). X-ray data were collected at 100 K with a Bruker AXS Platinum 135 CCD detector controlled by the Proteum software suite (Bruker AXS). The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. These X-ray data were processed with SAINT version 7.06A (Bruker AXS) and internally scaled with SADABS version 2005/1 (Bruker AXS). Relevant X-ray data collection statistics are listed in Table 3.1. The structure was determined via molecular replacement with the software package PHASER and using as a search model the X-ray coordinates 4HN0 from the Protein Data Bank (18). Iterative cycles of model building with COOT and refinement with REFMAC reduced the R_{work} and R_{free} to 19.5 and 24.8%, respectively, from 30 to 1.75 Å resolution (22-24).

Table 3.1 X-ray Data Collection Statistics and Model Refinement Statistics

	WbcA
Resolution limits (Å)	30.0-1.75 (1.80-1.75)[b]
Number of independent reflections	50180 (3422)
Completeness (%)	98.1 (95.0)
Redundancy	2.5 (1.5)
avg I/avg σ(I)	6.9 (2.0)
R_{sym} (%)[a]	8.2 (34.1)
[c]*R*-factor (overall)%/no. reflections	19.5/50180
R-factor (working)%/no. reflections	19.2/47639
R-factor (free)%/no. reflections	24.8/2541
number of protein atoms	4285
number of heteroatoms	445
average B values	
protein atoms (Å^2)	20.5
solvent (Å^2)	26.9
weighted RMS deviations from ideality	
bond lengths (Å)	0.011
bond angles (°)	2.0
planar groups (Å)	0.01
Ramachandran regions (%)[d]	
most favored	92.9
additionally allowed	7.1
generously allowed	0.0

[a] R_{sym} 5(R|I 2I |/R I) 3 100.

[b] Statistics for the highest resolution bin.

[c] R-factor 5(R|F_o 2F_c|/R|F_o|) 3 100 where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.

[d] Distribution of Ramachandran angles according to PROCHECK (2)

3.3.4 Site-Directed Mutagenesis

All site-directed mutant proteins were generated via the QuikChange method of Stratagene and purified in a similar manner to that of the wild-type enzyme including the removal of the N-terminal histidine tags.

3.3.5 Kinetic Analyses

Steady-state kinetic parameters for WbcA, and the site-directed mutant variants were determined via a coupled spectrophotometric assay, which followed the conversion of NADPH to $NADP^+$ by the action of WbcB (Scheme 3.1). The starting substrate was CDP-4-keto-6-deoxyglucose. All reaction mixtures contained 50 mM HEPES (pH 8), 0.2 mM NADPH, and 1 mg/mL WbcB.

The CDP-4-keto-6-deoxyglucose and WbcA wild-type and mutant protein concentrations varied between reactions as follows. For the wild-type enzyme, the WbcA concentration was 0.0025 mg/mL in the reaction mixture, with CDP-4-keto-6-deoxyglucose concentrations ranging between 0.001 and 8 mM. The C133S variant required a concentration of 0.018 mg/ mL, and the ketosugar concentration was varied between 0.05 and 8 mM. Both the H63N and Y50F mutant proteins were assayed at 5 mg/mL, with CDP- 4-keto-6-deoxyglucose concentrations ranging from 0.1 and 20 mM. However, no activity was seen above background.

All reactions were initiated by the addition of WbcA at 25° C and followed using a Beckman Coulter DU-640 spectrophotometer for 5 min. Reduction of the CDP-4-keto-6-deoxygulose sugar and concurrent oxidation of NADPH to $NADP^+$ was monitored by a decrease in absorbance at 340 nm. Plots of concentrations versus initial rates were analyzed using PRISM (GraphPad Software) and were fitted to the equation $v_\circ = (V_{max}[S])/(K_M + [S])$. Kinetic parameters are listed in Table 3.2.

Table 3.2 Kinetic Parameters

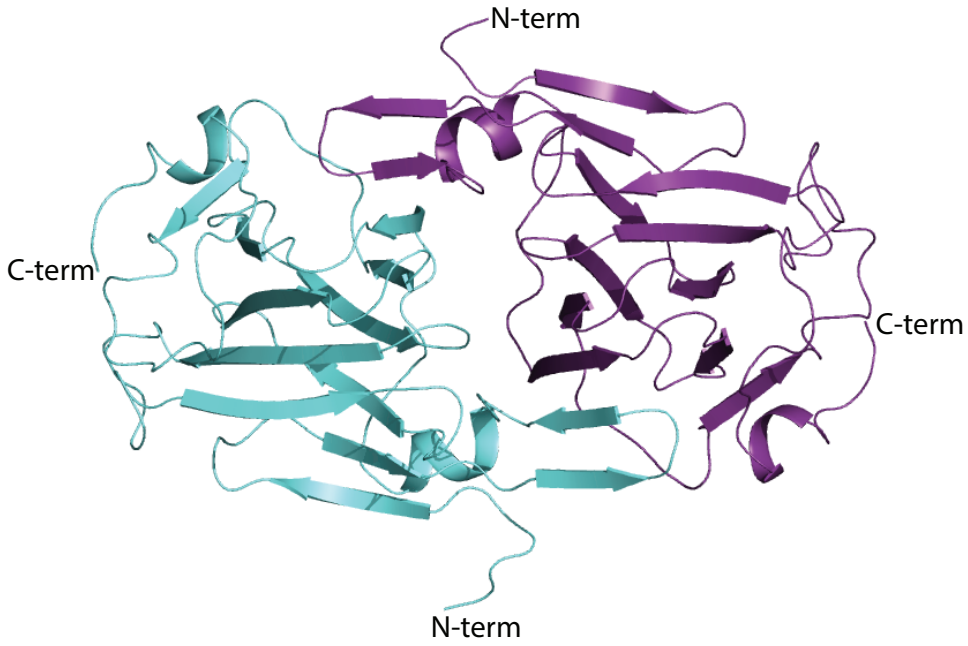
Enzyme	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
Wild-type	0.69 ± 0.08	7.8 ± 0.3	1.1×10^4
H63N	nd[a]	nd	nd
C133S	0.61 ± 0.12	2.0 ± 0.1	3.3×10^3
Y50F	nd	nd	nd

[a] Could not be determined under the assay conditions employed

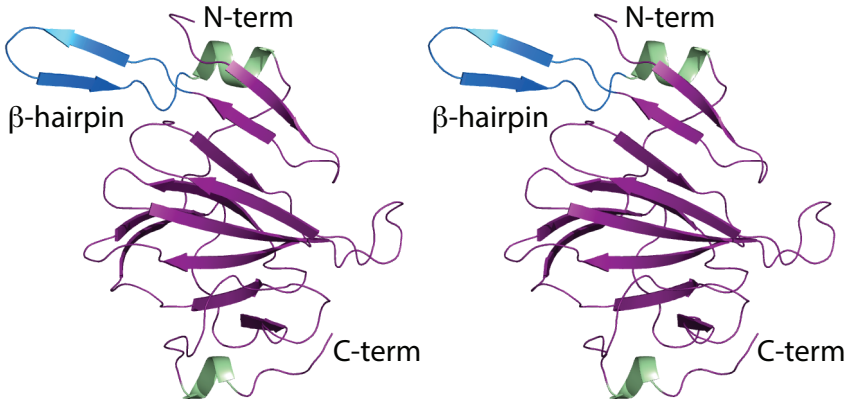
3.4 RESULTS AND DISCUSSION

The WbcA model reported here was solved to 1.75-Å resolution and refined to an overall R-factor of 19.5%. Relevant X-ray data collection and model refinement statistics are given in Table 3.1. Like other members of the cupin superfamily, WbcA functions as a dimer. The crystals utilized for this analysis belonged to the space group C2 with three monomers in the asymmetric unit. Two of the monomers forming the dimer were related by a local twofold rotational axis whereas the third monomer packed along a crystallographic dyad. All three subunits adopted similar molecular architectures such that their α-carbons superimpose with root-mean-square deviations of between 0.20 and 0.25 Å. For the sake of clarity, the following discussion refers to the dimer that packed along the local twofold rotational axis.

Shown in Figure 3.2(a) is a ribbon representation of the WbcA dimer, which has overall dimensions of ~50 × ~55 × ~60 Å. The buried surface area is 3800 Å^2. As observed for members of the cupin fold, domain swapping occurs whereby a β-hairpin motif from one subunit projects towards the active site of another subunit. A close-up stereo view of one subunit is displayed in Figure 3.2(b). The polypeptide chain initiates with an extended coil and two antiparallel β-strands. The β-hairpin motif (Ile 20–Val 29), responsible for domain swapping, is followed by the first of two a-helices. An additional nine β-strands wrap around to form the cupin barrel. The barrel can be envisioned as two



(a)



(b)

Figure 3.2 The structure of WbcA. A ribbon representation of the WbcA dimer is shown in (a) with the two subunits displayed in teal and purple, respectively. A stereo view of one subunit is presented in (b). The two α-helices and the β-hairpin motif are highlighted in green and blue, respectively. This figure and Figure 3.3 were prepared with PyMOL (1)(24).

layers of antiparallel β-sheet. There is a final small α-helix formed by Asp 165 to Arg 168. Both Met 62 and Pro 68 adopt the cis conformation.

The crystallization conditions for WbcA included 5 mM CDP-6-deoxy-D-gulose in an attempt to trap a stable substrate analog in the active site. No clear electron density was observed for the CDP-sugar ligand. However, the structure of another C-3' epimerase from *Streptomyces bikiniensis* was determined in the laboratory in 2012 as a complex with dTDP-6-deoxy-D-glucose (18). This enzyme, referred to as ChmJ, has an amino acid sequence identity and similarity to WbcA of 30 and 50%, respectively. The α-carbons for these two enzymes correspond with a root-mean-square deviation of 1.2 Å. In ChmJ, His 60 serves as the catalytic base whereas Tyr 130 functions as the active site acid (18). Shown in Figure 3.3 is a superposition of His 60 and Tyr 130 of ChmJ onto the structural equivalent residues in WbcA. The histidine residues in WbcA (His 63) and ChmJ (His 60) align well, and in both enzymes the residue leading up to the histidine adopts the cis conformation (Met 62 in WbcA or Ile 59 in ChmJ). As noted in the Introduction, it was assumed that Cys 133 in WbcA would align with Tyr 130 in ChmJ, and indeed it does as can be seen in Figure 3.3. What was not expected, however, was that in WbcA Tyr 50, which is an asparagine residue in ChmJ, would project into the active site in the appropriate position to serve as the active site acid. This residue is situated on β-strand 5 whereas in ChmJ the active site tyrosine is located on β-strand 12.

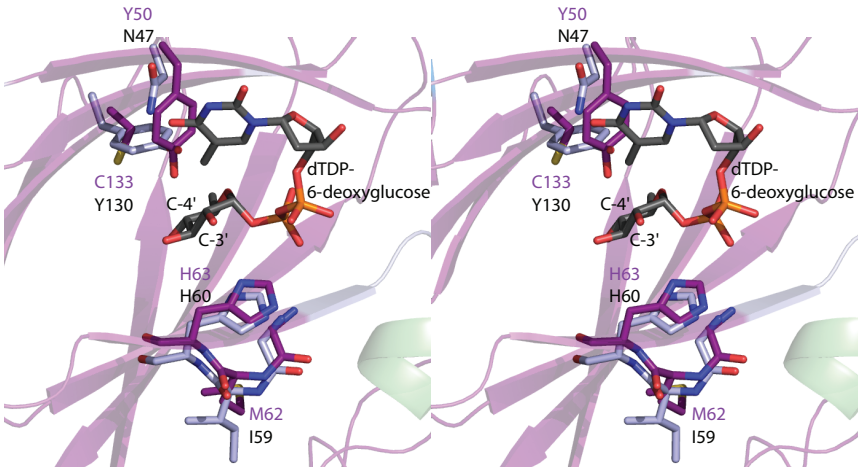
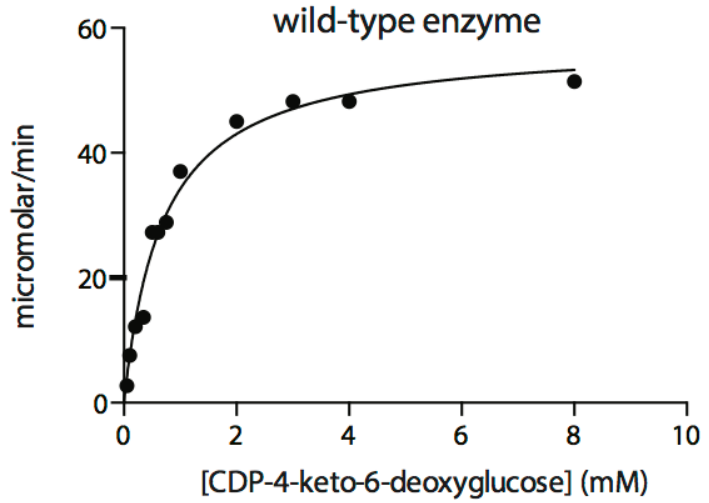


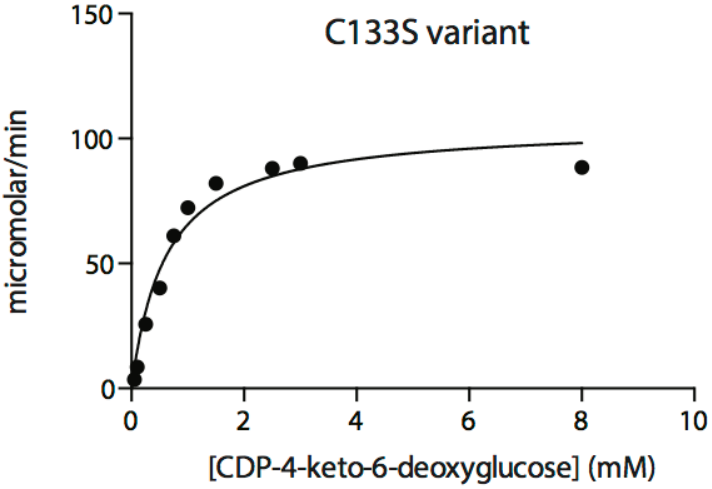
Figure 3.3 Superposition of the active sites for WbcA and ChmJ. Those residues belonging to WbcA are highlighted in purple whereas those from ChmJ are displayed in light blue. The top and bottom labels refer to amino acids in WbcA and ChmJ, respectively. The dTDP-6-deoxyglucose ligand is from the ChmJ model (PDB code 4HMZ).

To further test the roles of His 63, Tyr 50, and Cys 133 in the catalytic activity of WbcA, the following site-directed mutant variants were constructed: H63N, Y50F, and C133S. Their kinetic parameters were measured and are presented in Table 3.2. Plots of initial velocities versus substrate concentrations are presented in Figure 3.4. As expected, substituting His 63 with an asparagine residue resulted in a catalytically inert enzyme under the assay conditions employed. Likewise, replacement of Tyr 50 with a phenylalanine residue resulted in the complete abrogation of enzymatic activity. In the case of the C133S mutation, however, the K_m for the substrate was identical to that of the wild-type enzyme within experimental error. The overall k_{cat} was slightly reduced. As a consequence, the C133S mutant protein demonstrated an approximate threefold reduction in catalytic efficiency. This small reduction in catalytic efficiency is not unexpected given the change in overall hydrophobicity of the active site upon substitution of a cysteine to a serine residue.

In summary, the overall molecular architecture of WbcA has been defined. Importantly, this structural/biochemical analysis is yet another example of the perils in relying only on amino acid sequence homologies. Indeed, it was expected from such analyses that Cys 133 was the active site acid, and in fact it is a tyrosine residue from a neighboring β-strand.



(a)



(b)

Figure 3.4 Plots of initial velocities versus substrate concentration.

3.5 ACKNOWLEDGEMENTS

We gratefully acknowledge Dr. Evgeny Vinogradov at Human Health Therapeutics, National Research Council, Canada for confirming by NMR the identity of the CDP-6-deoxy-D-gulose ligand.

3.6 REFERENCES

1. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. *The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA*.

2. Laskowski RA, Macarthur MW, Moss DS, & Thornton JM (1993) Procheck - A Program To Check The Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* 26:283-291.

3. Brubaker RR (1991) Factors promoting acute and chronic diseases caused by yersiniae. *Clinical Microbiology Reviews* 4(3):309-324.

4. Jansson P-E, Lindberg B, Lindberg AA, & Wollin R (1981) Structural Studies on the Hexose Region of the Core in Lipopolysaccharides from *Enterobacteriaceae*. *European Journal of Biochemistry* 115(3):571-577.

5. Galanos C, Lüderitz O, Rietschel ET, Westphal O, Brade H, Brade L, Freudenberg M, Schade U, Imoto M, Yoshimura H, Kusumoto S, & Shiba T (1985) Synthetic and natural Escherichia coli free lipid A express identical endotoxic activities. *European Journal of Biochemistry* 148(1):1-5.

6. Raetz CRH, Reynolds CM, Trent MS, & Bishop RE (2007) Lipid Modification Systems in Gram-Negative Bacteria. *Annual review of biochemistry.* 76:295-329.

7. Raetz CRH & Whitfield C (2002) Lipopolysaccharide Endotoxins. *Annual review of biochemistry* 71:635-700.

8. Bengoechea JA, Najdenski H, & Skurnik M (2004) Lipopolysaccharide O antigen status of *Yersinia enterocolitica* O:8 is essential for virulence and absence of O antigen affects the expression of other *Yersinia* virulence factors. *Molecular Microbiology* 52(2):451-469.

9. Zhang L, Radziejewska-Lebrecht J, Krajewska-Pietrasik D, Toivanen P, & Skurnik M (1997) Molecular and chemical characterization of the lipopolysaccharide O-

antigen and its role in the virulence of *Yersinia enterocolitica* serotype O:8. *Molecular Microbiology* 23(1):63-76.

10. Tomshich SV, Gorshkova RP, & Ovodov YS (1987) Structural Studies on the Lipopolysaccharide from *Y. enterocolitica* serovar O:8. *Khimia Prirodnykh Soedinenii*:657-664.

11. Koropatkin NM, Cleland WW, & Holden HM (2005) Kinetic and structural analysis of alpha-D-glucose-1-phosphate cytidylyltransferase from Salmonella typhi. *J. Biol. Chem.* 280(11):10774-10780.

12. Koropatkin NM & Holden HM (2004) Molecular structure of alpha-D-glucose-1-phosphate cytidylyltransferase from Salmonella typhi. *J. Biol. Chem.* 279(42):44023-44029.

13. Koropatkin NM & Holden HM (2005) Structure of CDP-D-glucose 4,6-dehydratase from Salmonella typhi complexed with CDP-D-xylose. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 61:365-373.

14. Vogan EM, Bellamacina C, He X, Liu HW, Ringe D, & Petsko GA (2004) Crystal structure at 1.8 A resolution of CDP-D-glucose 4,6-dehydratase from Yersinia pseudotuberculosis. *Biochemistry* 43(11):3057-3067.

15. Dong C, *et al.* (2007) RmlC, a C3′ and C5′ Carbohydrate Epimerase, Appears to Operate via an Intermediate with an Unusual Twist Boat Conformation. *J Mol. Biol.* 365(1):146-159.

16. Jakimowicz P, Tello M, Meyers CL, Walsh CT, Buttner MJ, Field RA, & Lawson DM (2006) The 1.6-A resolution crystal structure of NovW: A 4-keto-6-deoxy sugar epimerase from the novobiocin biosynthetic gene cluster of Streptomyces spheroides. *Proteins* 63(1):261-265.

17. Kubiak RL, Phillips RK, Zmudka MW, Ahn MR, Maka EM, Pyeatt GL, Roggensack SJ, Holden HM (2012) Structural and Functional Studies on a 3′-Epimerase Involved in the Biosynthesis of dTDP-6-deoxy-D-allose. *Biochemistry* 51(46):9375-9383.

18. Merkel AB, Major LL, Errey JC, Burkart MD, Field RA, Walsh CT, & Naismith JH (2004) The position of a key tyrosine in dTDP-4-Keto-6-deoxy-D-glucose-5-epimerase (EvaD) alters the substrate profile for this RmlC-like enzyme. *J. Biol. Chem.* 279(31):32684-32691.

19. Naismith JH (2004) Chemical insights from structural studies of enzymes. *Biochem. Soc. Trans.* 32:647-654.

20. Tello M, Jakimowicz P, Errey JC, Freel Meyers CL, Walsh CT, Buttner MJ, Lawson DM, & Field RA (2006) Characterisation of Streptomyces spheroides NovW and revision of its functional assignment to a dTDP-6-deoxy-D-xylo-4-hexulose 3-epimerase. *Chem. Commun.* (10):1079-1081.

21. Thoden JB & Holden HM (2005) The molecular architecture of human *N*-acetylgalactosamine kinase. *J. Biol. Chem.* 280(38):32784-32791.

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

23. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 66:486-501.

24. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* 53:240-255.

CHAPTER 4:

STRUCTURAL AND FUNCTIONAL INVESTIGATION OF FDHC FROM *ACINETOBACTER NOSOCOMIALIS*: A SUGAR *N*-ACYLTRANSFERASE BELONGING TO THE GNAT SUPERFAMILY

This work was published as: Salinger, A.J., Thoden, J.B., and Holden, H.M. (2016) Structural and Functional Investigation of FdhC from *Acinetobacter Nosocomialis*: A Sugar *N*-Acyltransferase Belonging to The Gnat Superfamily *Biochemistry. 10, 1633-1639.*

4.1 ABSTRACT

Enzymes belonging to the GNAT superfamily are widely distributed in nature where they play key roles in the transfer of acyl groups from acyl-CoAs to primary amine acceptors. The amine acceptors range the gamut from histones to aminoglycoside antibiotics to small molecules such as serotonin. Whereas those family members that function on histones have been extensively studied, the GNAT enzymes that employ nucleotide-linked sugars as their substrates have not been well characterized. Indeed, though the structures of two of these "amino sugar" GNAT enzymes have been determined within the last ten years, details concerning their active site architectures have been limited due to a lack of bound nucleotide-linked sugar substrates. Here we describe a combined structural and biochemical analysis of FdhC from *Acinetobacter nosocomialis* O2. On the basis of bioinformatics, it was postulated that FdhC catalyzes the transfer of a 3-hydroxybutanoyl group from 3-hydroxylbutanoyl-CoA to dTDP-3-amino-3,6-dideoxy-D-galactose, to yield an unusual sugar that is ultimately incorporated into the surface polysaccharides of the bacterium. We present data confirming this activity. In addition, two ternary complexes of FdhC, in the presence of CoA and either 3-hydroxybutanoylamino-3,6-dideoxy-D-galactose or 3-hydroxybutanoylamino-3,6-dideoxy-D-glucose, were solved by X-ray crystallographic analyses to high resolution. Kinetic parameters were determined, and activity assays demonstrated that FdhC can also utilize acetyl-CoA, 3-methylcrotonyl-CoA, or hexanoyl-CoA as acyl donors, albeit at

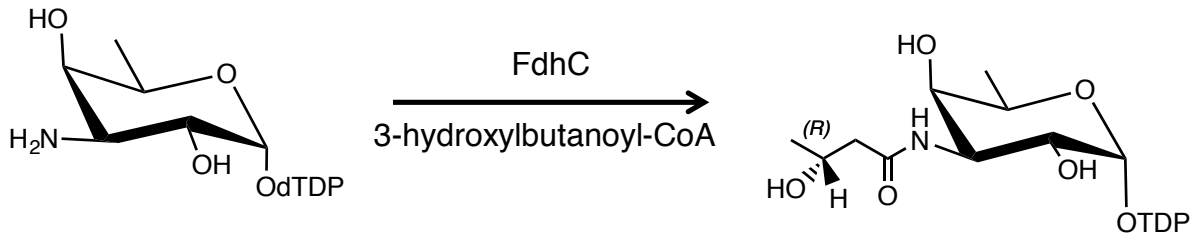
reduced rates. Site-directed mutagenesis experiments were conducted to probe the catalytic mechanism of FdhC. Taken together, the data presented herein provide significantly new molecular insight into those GNAT superfamily members that function on nucleotide-linked amino sugars.

N-acyltransferases are widespread in nature where they catalyze the transfer of acyl groups from acyl-CoAs to primary amine acceptors. Two structural superfamilies of *N*-acyltransferases have been identified thus far. One family contains the GCN5-related *N*-acetyltransferases, or GNAT enzymes, whereas the other is composed of the left-handed β-helix, or LβH enzymes. In recent years, our understanding of the *N*-acyltransferases belonging to the LβH superfamily, specifically those involved in the biosynthesis of unusual sugars, has dramatically increased as various structures have been reported (1-7). All of these enzymes contain a repeated isoleucine-rich, hexapeptide motif and a remarkable β-helix motif with exceedingly rare left-handed crossover connections.

Less is known, however, regarding those GNAT superfamily members that employ nucleotide-linked sugars as substrates. dTDP-fucosamine acetyltransferase (WecD) from *Escherichia coli*, which catalyzes the acetylation of dTDP-4-amino-4,6-dideoxy-D-galactose, was the first "amino sugar" GNAT superfamily member to have its structure determined (8). This initial report was followed by a hiatus for nine years until the structures of PseH from *Campylobacter jejuni* and *Helicobacter pylori* were solved (9, 10). Whereas the WecD and PseH structures provided considerable molecular insight into the "nucleotide-linked sugar" GNAT superfamily, none of the available structures contained a substrate or substrate analog.

Given our long standing interest in enzymes involved in the biosynthesis of unique sugars found on the lipopolysaccharides of Gram negative bacteria, and curious as to the manner in which a GNAT superfamily member can accommodate a nucleotide-linked sugar, we undertook an X-ray crystallographic analysis of the hypothetical protein FdhC from *Acinetobacter nosocomialis* O2 (11). Infections caused by such organisms as *A. nosocomialis* are becoming increasingly prevalent, and it is of particular concern that some isolates display resistance to all clinically available antibiotics (12). Although still a topic of debate, it is thought that *Acinetobacter* spp. produce lipooligosaccharides rather than lipopolysaccharides. Interestingly, the gene clusters required for the biosynthesis of the lipooligosaccharide sugars are extremely diverse in *Acinetobacter* spp (13). Some of the sugars in these organisms are quite unique, including 3-[(R)-3-hydroxybutanoylamino]-3,6-dideoxy-D-galactose and 3-[(R)-3-hydroxybutanoylamino]-3,6-dideoxy-D-glucose, hereafter referred to as Fuc3N(R3Hb) and Qui3N(R3Hb), respectively (11).

On the basis of its amino acid sequence, *A. nosocomialis* FdhC was predicted to catalyze the transfer of a 3-hydroxybutanoyl moiety from 3-hydroxylbutanoyl-CoA to dTDP-3-amino-3,6-dideoxy-D-galactose (dTDP-Fuc3N) to yield dTDP-Fuc3N(R3Hb) as indicated in Scheme 4.1 (11). Here we describe a combined structural and functional investigation of FdhC. For this analysis, three X-ray structures were determined to high resolution, and kinetic parameters were measured. The kinetic data demonstrate that FdhC can function on both dTDP-Fuc3N and dTDP-3-amino-3,6-dideoxy-D-glucose (dTDP-Qui3N), albeit



dTDP-Fuc3N

dTDP-Fuc3N(R3Hb)

Scheme 4.1 Reaction catalyzed by the enzyme FdhC. The enzyme of interest is predicted to be a C-3′ *N*-hydroxybutanoyl transferase.

with significantly different catalytic efficiencies. In addition, two site-directed mutant variants, E131L and E131Q, were constructed to test the role of Glu 131 in catalysis. Importantly, two of the structures solved were that of the enzyme in complex with CoA and either dTDP-Fuc3N(R3Hb) or dTDP-Qui3N(R3Hb). These models revealed for the first time the manner in which any GNAT superfamily member accommodates a nucleotide-linked sugar.

4.3 MATERIALS AND METHODS

4.3.1 Cloning, Expression, and Purification

The gene encoding FdhC from *A. nosocomialis* serogroup O2 was synthesized by DNA2.0 for optimized *E. coli* codon usage and was provided in a pJ201 plasmid with 5'-NdeI and 3'-XhoI restriction sites. The *fdhC* gene was subsequently digested and ligated into a pET31b expression vector (Novagen) whereupon the STOP codon was removed to yield a protein with a C-terminal hexahistidine tag. For a tagless version of FdhC, the *fdhC* gene was ligated into pET28JT, a laboratory pET28b(+) vector that had been previously modified to incorporate a TEV protease cleavage recognition site after the N-terminal hexahistidine tag (14).

Both vectors were utilized to transform Rosetta(DE3) *E. coli* cells (Novagen). The cells harboring the pET31-*fdhC* plasmid were cultured in lysogeny broth supplemented with ampicillin (100 mg/L) and chloramphenicol (25 mg/L). The cells transformed with the pET28JT-*fdhC* vector were cultured in lysogeny broth supplemented with kanamycin (35 mg/L) and chloramphenicol (25 mg/L). Both cell lines were grown with shaking at 37° C. When the optical density reached 0.8 at 600 nm, the flasks were cooled in an ice water bath, and the cultures were subsequently induced with 1 mM isopropyl β-D-1-thiogalactopyranoside and transferred to a refrigerated shaker at 16° C. The cells were allowed to express protein at 16° C for 18 hours after induction.

FdhC with the C-terminal tag was purified by standard procedures using Ni-nitrilotriacetic acid resin. Following purification, the protein was dialyzed against 10 mM Tris-HCl and 200 mM NaCl at pH 8.0, and concentrated to 18 mg/mL based on the calculated extinction coefficient of 0.62 $(mg/mL)^{-1}$ cm^{-1}. N-terminally tagged FdhC was purified in a similar manner. For removal of the tag, a solution containing a 30:1 molar ratio (enzyme:TEV protease) was allowed to digest at 4° C for 48 hr. Uncleaved protein and the TEV protease were removed by passage over Ni-nitrilotriacetic acid resin. Wild-type FdhC was dialyzed against 10 mM Tris-HCl (pH 8.0) and 200 mM NaCl and concentrated to 11 mg/mL based on the calculated extinction coefficient of 0.62 $(mg/mL)^{-1}$ cm^{-1}.

To express selenomethionine-labeled FdhC, the *E. coli* Rosetta(DE3) cells containing the pET31-*fdhC* plasmid were grown in M9 minimal media at 37° C. Aliquots of 25 mL of this culture were then used to inoculate 12 2-liter baffled flasks containing 1 L M9 minimal medium supplemented with ampicillin (100 mg/L) and chloramphenicol (50 mg/L). The cultures were grown at 37° C to an optical density of 0.9 at 600 nm. Subsequently, the flasks were cooled on ice for 5 min. 50 mg/L each of L-lysine, L-threonine, and L-phenylalanine, and 25 mg/L each of L-leucine, L-isoleucine, L-valine, and L-selenomethionine were added to each flask. After 20 min of additional growth, the cells were induced with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside, and

the protein was purified as described above. Mass spectrometry showed full incorporation of the selenium atoms.

4.3.2 Site-Directed Mutagenesis

All site-directed mutant variants of FdhC were generated via the QuikChange method of Stratagene. The protein variants were expressed and purified as described for the wild-type enzyme.

4.3.3 Protein Crystallization and X-ray Structural Analysis

Crystallization conditions were initially surveyed by the hanging drop method of vapor diffusion using a sparse matrix screen developed in the laboratory. The first experiments were conducted with either the apoprotein or a selenium-labeled form at room temperature. X-ray diffraction quality crystals of both forms appeared as thin rods after three weeks and were grown by mixing 1:1 the protein sample at 18 mg/mL with a precipitant solution composed of 2.4 - 2.8 M ammonium sulfate, 2% dimethyl sulfoxide, and 100 mM HEPPS (pH 8.5). The crystals grew to maximal dimensions of 1.0 x 0.3 x 0.3 mm and belonged to the space group $P6_322$ with unit cell dimensions of $a = b = 69.5$ Å and $c = 137.9$ Å. The asymmetric unit contained one subunit. Prior to X-ray data collection, the crystals of either the apoprotein or selenium-labeled enzyme were transferred to

solutions composed of saturated ammonium sulfate, 2% dimethyl sulfoxide, 200 mM NaCl, 100 mM HEPPS (pH 8.5), and 10% sucrose.

X-ray diffraction quality crystals of FdhC with bound products were grown at 4° C via batch experiments. Specifically, protein solutions containing 5 mM 3-hydroxylbutanoyl-CoA and either 5 mM of the dTDP-Fuc3N or dTDP-Qui3N substrates were mixed 1:1 with precipitant solutions composed of 26 - 30% monomethylether poly(ethylene glycol) 5000, 200 mM LiCl, and 100 mM Homo-PIPES (pH 5.0). The enzyme turned over the substrates during the crystallization experiments, and the products were trapped in the active site. Both complexes crystallized in the space group $P4_12_12$ with unit cell dimensions of $a = b =$ ~76 Å, and $c = 75$ Å and a monomer in the asymmetric unit.

Crystals of the FdhC ternary complexes were flash cooled for X-ray data collection via serial transfer into solutions composed of 25% monomethylether poly(ethylene glycol) 5000, 300 mM NaCl, 100 mM LiCl, 100 mM Homo-PIPES, and 16% ethylene glycol (pH 5.0). The solutions also included 5 mM of each respective ligand.

X-ray data from the selenium-labeled protein crystals were collected at the Structural Biology Center beamline 19-BM at a wavelength of 0.9794 Å (Advanced Photon Source). The X-ray data were processed and scaled with HKL3000. The structure was solved via multiple wavelength anomalous dispersion phasing. The program CRANK from the CCP4 software suite was utilized to determine and refine the position of the single selenium atom (15). The preliminary electron density map allowed for a complete

tracing of the polypeptide chain with the software package COOT (16). The model was partially refined to 2.1-Å resolution with the software package REFMAC (17). This structure served as the search model for all subsequent X-ray analyses.

X-ray data sets from crystals of the apoprotein and the ternary complexes were collected in-house at 100 K using a Bruker AXS Platinum 135 CCD detector controlled with the Proteum software suite (Bruker AXS Inc.). The X-ray source was Cu Kα radiation from a Rigaku RU200 X-ray generator equipped with Montel optics and operated at 50 kV and 90 mA. These X-ray data were processed with SAINT version 7.06A (Bruker AXS Inc.) and internally scaled with SADABS version 2005/1 (Bruker AXS Inc.). The structures of the apoprotein and the ternary complexes were solved by molecular replacement using PHASER (18). The resulting models were subjected to alternating cycles of manual model building with COOT and refinement with REFMAC. Relevant X-ray data collection and model refinement statistics are provided in Tables 4.1 and 4.2, respectively.

4.3.4 Size Exclusion Chromatography

A Superdex™ 200 10/300 GL (GE Healthcare) column was equilibrated at room temperature with a buffer solution of 10 mM Tris-HCL (pH 8.0) and 200 mM NaCl on an ÄKTA HPLC system. Purified C-terminally tagged protein (diluted to 10 mg/mL) was applied to the column. The retention was monitored via ultraviolet absorption at 280 nm. The results were compared to standards of known molecular mass.

Table 4.1. X-ray Data Collection Statistics.

	Apoenzyme	FdhC complex with dTDP-Fuc3N(R3Hb) and CoA	FdhC in complex with dTDP-Qui3N(R3Hb) and CoA
resolution limits	50.0 - 1.89 (1.99 - 1.89)[b]	50.0 - 1.80 (1.90 - 1.80)	50.0 - 1.90 (2.0 - 1.90)
number of independent reflections	16130 (2061)	20344 (2775)	17505 (2444)
completeness (%)	98.0 (93.9)	98.2 (91.9)	99.7 (98.4)
redundancy	6.9 (5.1)	9.5 (3.8)	29.5 (15.4)
avg I/avg σ(I)	17.1 (2.9)	17.6 (3.8)	21.1 (5.2)
R_{sym} (%)[a]	8.0 (41.1)	7.4 (27.5)	10.8 (42.5)

[a] $R_{sym} = (\sum|I - \overline{I}\ |/ \sum I) \times 100$.

[b] Statistics for the highest resolution bin

Table 4.2. Refinement Statistics.

	Apoenzyme	FdhC in complex with dTDP-Fuc3N(R3Hb) and CoA	FdhC in complex with dTDP-Qui3N(R3Hb) and CoA
space group	$P6_322$	$P4_12_12$	$P4_12_12$
unit cell dimensions (Å)	$a = b = 69.5$, $c = 137.9$	$a = b = 76.3$, $c = 73.5$	$a = b = 75.8$, $c = 73.5$
resolution limits (Å)	50 - 1.9	50.0 - 1.8	50.0 - 1.9
[a]*R*-factor (overall)%/no. reflections	19.8/16107	19.7/20303	20.4/17457
R-factor (working)%/no. reflections	19.5/15290	19.4/19260	20.0/16577
R-factor (free)%/no. reflections	25.8/817	24.7/1043	26.9/880
number of protein atoms	1486	1583	1583
number of heteroatoms	111	242	196
average B values			
protein atoms (Å^2)	24.1	20.6	23.2
ligands (Å^2)	n/a	19.5	21.5
solvent (Å^2)	28.0	23.2	22.6
weighted RMS deviations from ideality			
bond lengths (Å)	0.018	0.019	0.017
bond angles (º)	1.9	2.0	1.9
general planes (º)	0.009	0.009	0.008
Ramachandran regions (%)[b]			
most favored	89.1	90.2	91.3
additionally allowed	10.9	9.2	8.1
generously allowed	-	0.6	0.6

[a]*R*-factor = $(\Sigma|F_o - F_c| / \Sigma|F_o|) \times 100$ where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude.

[b]Distribution of the Ramachandran angles according to PROCHECK.(26)

4.3.5 Activity of FdhC

To confirm the *N*-acyltransferase activity of FdhC, simple activity assays were first conducted. Specifically, 1 mL reactions were prepared using various CoA substrates (3-hydroxylbutanoyl-CoA, acetyl-CoA, 3-methylcrotonyl-CoA, or hexanoyl-CoA) and either dTDP-Fuc3N or dTDP-Qui3N. Each reaction condition contained FdhC at 1 mg/mL, 50 mM HEPPS (pH 8.0) and 0.5 mM of the CoA derivatives and the dTDP-sugars. The reaction products were separated from the protein by filtration through a 10 kDa cutoff ultrafiltration membrane, diluted with two volumes of water, and loaded onto an ÄKTA HPLC system equipped with a 1 mL Resource-Q column. Elution with a 20 mL gradient at pH 8.5 from 0 to 0.6 M ammonium bicarbonate showed the loss of the dTDP-sugar and CoA substrates (retention 13.5 mL and 18 mL, respectively) and the generation of a new peak with a retention volume of approximately 9 mL. The identity of this peak as a modified dTDP-sugar was verified by electrospray ionization mass spectrometry (Mass Spectrometry/Proteomic Facility at the University of Wisconsin). The dTDP-Fuc3N and dTDP-Qui3N ligands required for the assays were prepared as previously described (19).

4.3.6 Determination of the Kinetic Constants for FdhC

The *N*-acyltransferase activity of FdhC was monitored spectrophotometrically by following the increase in absorbance at 412 nm due to the reaction of the free sulfhydryl

group of the CoASH product with 5,5′-dithiobis(2-nitrobenzoic acid) (DTNB). This reaction results in a disulfide interchange that leads to the formation of 5-thio-2-nitrobenzoic acid, which has a characteristic absorbance at 412 nm and an extinction coefficient of 14150 M^{-1} cm^{-1}. The use of this compound for quantification of free CoASH was first reported by Albers *et al.* (20), and our assay method was similar to that described by Magalhaes *et al.,* 2005 (21). Reactions were continuously monitored with a Beckman DU 640B spectrophotometer. All reaction mixtures contained 50 mM HEPES (pH 7.5) and 5 mM DTNB in addition to enzyme and substrates.

For determination of the K_m of the enzyme for dTDP-Fuc3N, the concentration of 3-hydroxybutyryl-CoA was held constant at 1 mM while the dTDP-Fuc3N concentration varied from 0.02 to 5 mM. To measure the K_m of the enzyme for 3-hydroxylbutanoyl-CoA, the dTDP-Fuc3N concentration was held constant at 5 mM while the 3-hydroxylbutanoyl-CoA concentration varied from 0.02 to 3 mM. All reactions were initiated by the addition of FdhC (or mutant variants) to a final concentration of 0.5 μg/mL.

To measure the K_m of the enzyme for dTDP-Qui3N, the concentration of 3-hydroxylbutanoyl-CoA was held constant at 1 mM while the dTDP-Qui3N concentration varied from 0.02 to 10 mM. The reactions were initiated by the addition of FdhC to a final enzyme concentration of 0.5 μg/mL.

Finally, for the determination of the K_m of the enzyme for acetyl-CoA, the dTDP-Fuc3N concentration was held constant at 5 mM while the acetyl-CoA concentration varied from 0.002 to 1 mM. The final enzyme concentration was 0.5 mg/mL.

All data were fitted by initial velocity Michaelis–Menten kinetics to the following equation:

$$v = (V_{max}[S])/(K_M + [S]).$$

The k_{cat} values were calculated according to the equation $k_{cat} = V_{max}/[ET]$. The reactions of FdhC with 3-methylcrotonyl-CoA and hexanoyl-CoA as substrates were too slow to be measured using the spectrophotometric assay. However, turnover was observed in the activity assays, and the identities of the products were confirmed via electrospray ionization mass spectrometry.

4.4 RESULTS AND DISCUSSION

4.4.1 Kinetic Analysis of Wild-Type FdhC

On the basis of bioinformatics, FdhC was proposed to function as an *N*-acyltransferase, although this activity had never been biochemically verified (11). Thus the first step in our enzymological analysis of FdhC was to conduct simple activity assays using 3-hydroxylbutanoyl-CoA, acetyl-CoA, 3-methylcrotonyl-CoA, or hexanoyl-CoA as acyl donors and either dTDP-Qui3N or dTDP-Fuc3N as acyl acceptors. On the basis of electrospray ionization mass spectrometry of the FdhC products it was shown that the enzyme could function on all of the above listed CoA and dTDP-sugar substrates [Figure 4.1].

The next step was to determine the kinetic constants of the enzyme for these acyl donors and acceptors [Table 4.3]. In keeping with the predicted function of FdhC, the catalytic efficiency of the enzyme for dTDP-Fuc3N versus dTDP-Qui3N was approximately 17-fold higher. The catalytic efficiency of FdhC for acetyl-CoA, using dTDP-Fuc3N as the acceptor, was low at 8.5×10^{1} M^{-1} sec^{-1}. Reactions with hexanoyl-CoA and 3-methylcrotonyl-CoA as the acyl donors were too slow to be kinetically analyzed under the assay conditions employed. Note that the protein utilized for the kinetic analyses did not contain a histidine tag. As described below, it was the C-terminal histidine-tagged version of the protein that crystallized. To ensure that the C-terminal tag did not affect the activity of the enzyme, the kinetic parameters for the tagged version

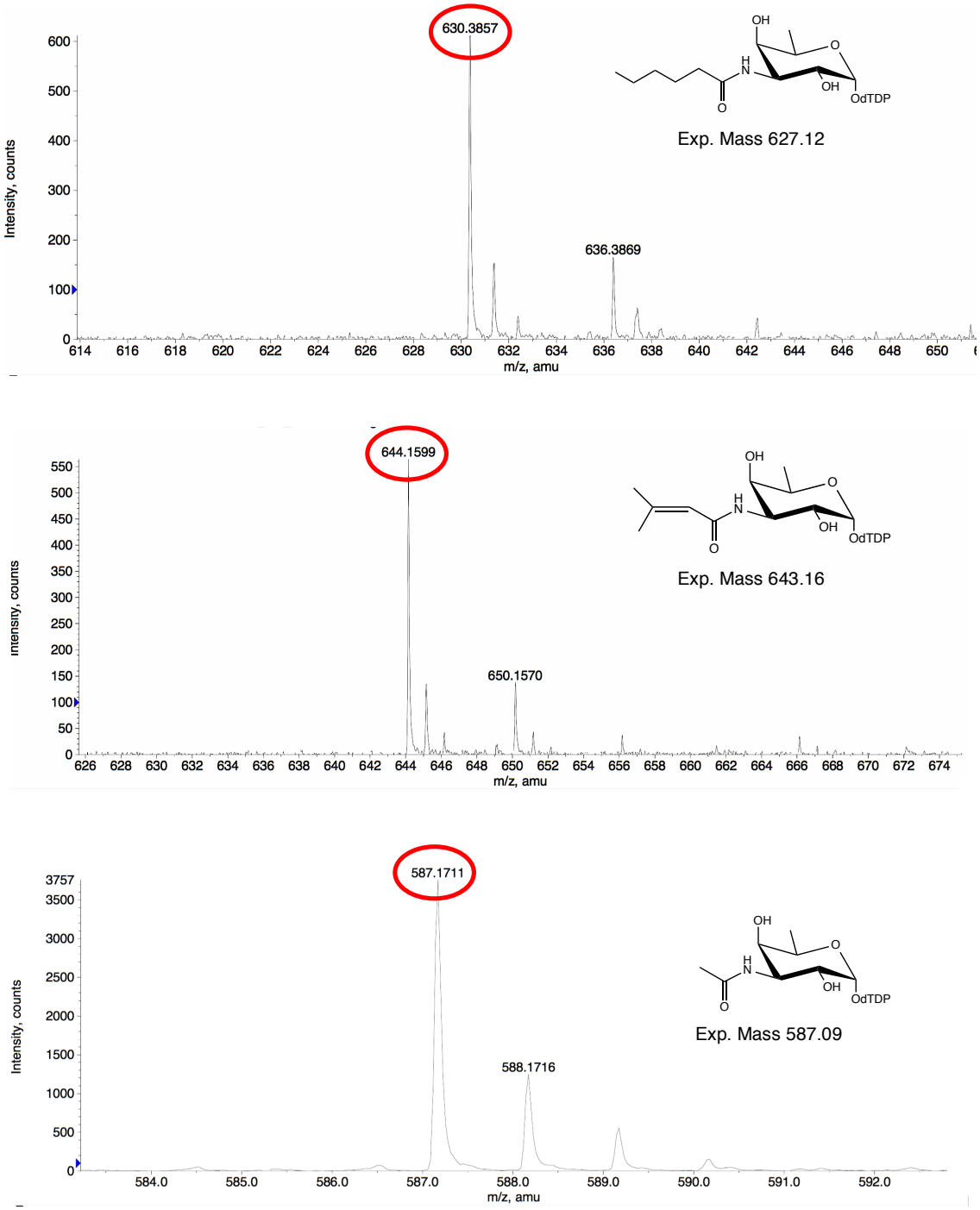


Figure 4.1 FdhC can transfer various acyl groups. While FdhC can, indeed, transfer 3-hydroxybutanoyl groups, it also has the capacity to transfer β-methylcrotonyl groups, hexanoyl groups, and simple acetyl groups.

Table 4.3. Kinetic Parameters for Wild-type FdhC.

Substrate	K_m (mM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($M^{-1}sec^{-1}$)
dTDP-Fuc3N	0.65 ± 0.07	52.7 ± 2.1	8.1×10^4
3-hydroxylbutanoyl-CoA	0.69 ± 0.11	63.2 ± 3.6	9.2×10^4
dTDP-Qui3N	5.3 ± 0.8	24.7 ± 1.7	4.7×10^3
acetyl-CoA	0.04 ± 0.01	0.0034 ± 0.0002	8.5×10^1

were also determined as listed in Table 4.4. The kinetic constants were similar for the two versions of FdhC.

4.4.2 Overall Structure of FdhC

The first structure to be determined in this investigation was that of the unliganded form of FdhC (with a C-terminal hexahistidine tag). The protein crystallized in the space group $P6_322$ with one molecule in the asymmetric unit. The structure was determined to 1.9-Å resolution, and the model was refined to an overall *R*-factor of 19.8%. Although GNAT superfamily members typically function as dimers, there are clearly cases where they assume monomeric quaternary structures. On the basis of both size exclusion gel filtration chromatography [Figure 4.2] as well as crystalline packing considerations, FdhC is a monomer in solution.

Shown in Figure 4.3 is a ribbon representation of the FdhC monomer. With overall dimensions of ~47 Å x 49 Å x 43 Å, the FdhC architecture is quite compact. The polypeptide chain initiates with a β-hairpin motif, which is followed by two α-helices that are oriented nearly perpendicular to one another. The polypeptide chain then folds into a six-stranded mixed β-sheet flanked on one side by three α-helices and on the other side by two α-helices. As is typical of GNAT superfamily members, β-strands three and four of the mixed β-sheet splay apart. Electron density for the polypeptide chain backbone was continuous from Val 2 to Gly 179. Proline 89 adopts the *cis* conformation

Table 4.4 Kinetic Parameters for the Mutant Variants of FdhC.

	Substrate	K_m (mM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($M^{-1}sec^{-1}$)
FdhC-Cterm tag	dTDP-Fuc3N	1.5 ± 0.2	38.3 ± 1.9	2.6×10^4
	3-hydroxylbutanoyl-CoA	0.20 ± 0.04	23.5 ± 1.2	1.2×10^5
E131Q variant	dTDP-Fuc3N	0.24 ± 0.02	21.7 ± 0.4	9.0×10^4
	3-hydroxylbutanoyl-CoA	0.64 ± 0.08	79.9 ± 3.7	1.2×10^5
E131L variant	dTDP-Fuc3N	1.6 ± 0.2	23.3 ± 1.2	1.5×10^4
	3-hydroxylbutanoyl-CoA	0.65 ± 0.11	22.7 ± 1.5	3.5×10^4

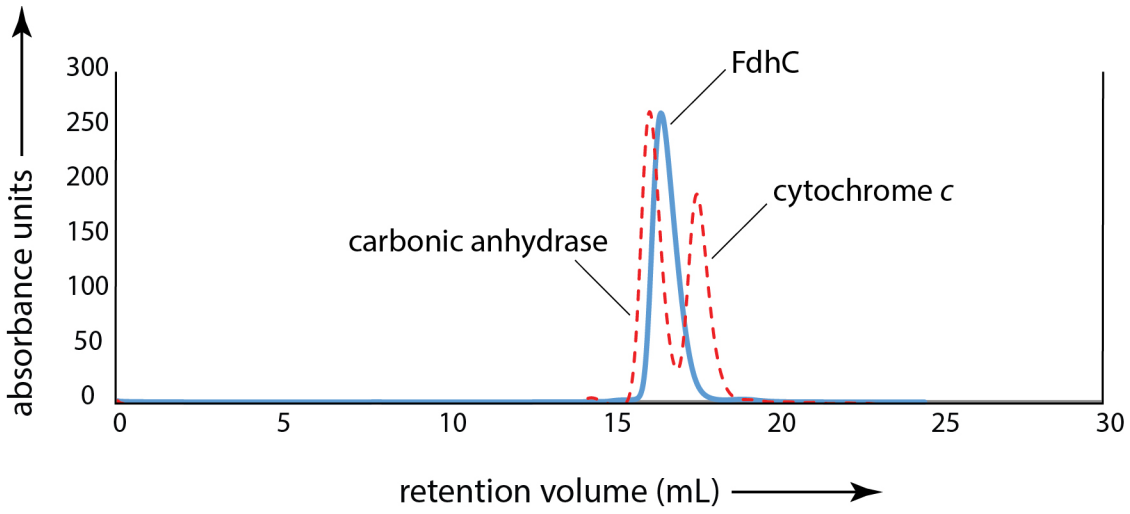


Figure 4.2 Gel filtration chromatography of FdhC. Shown in the blue trace is the retention volume of the FdhC sample versus absorbance. The red dashed lines show the retention times for horse heart cytochrome *c* (molecular weight of ~12,400) and bovine erythrocyte carbonic anhydrase (molecular weight of ~29,000).

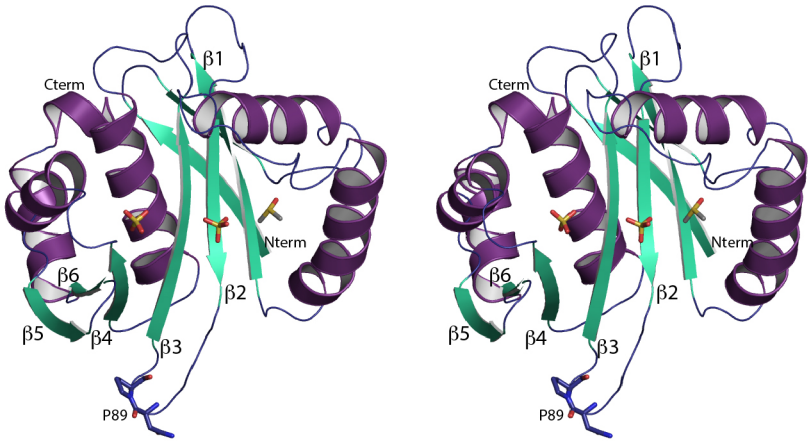


Figure 4.3 Stereo ribbon representation of FdhC. The α-helices and β-strands are depicted in purple and green, respectively. The sulfate anions and the DMSO molecule observed binding in the interior of the protein are displayed as sticks. This figure and all others were prepared with PyMOL (27).

and lies in the loop connecting β-strands 2 and 3 [Figure 4.3]. Two ordered sulfates and a dimethyl sulfoxide molecule were observed binding in the interior of the protein. The electron density for Cys 25 suggested that it had been oxidized to sulfenic acid.

4.4.3 Structure of FdhC in Complex with CoA and either dTDP-Fuc3N(R3Hb) or dTDP-Qui3N(R3Hb)

Crystals of the FdhC/CoA/dTDP-Fuc3N(R3Hb) complex belonged to the space group $P4_12_12$, again with one molecule in the asymmetric unit. The crystals diffracted to a nominal resolution of 1.8 Å, and the overall *R*-factor for the final model was 19.7%. Shown in Figure 4.4a is the observed electron density for the CoA and the dTDP-Fuc3N(R3Hb) ligands. The ordered sulfates and the dimethyl sulfoxide molecule observed in the apoenzyme structure are displaced by the dTDP-sugar. As can be seen, the pyranosyl moiety of the substrate adopts the $^4C_1$ conformation whereas the ribose assumes the C2'-*endo* pucker. A close-up view of the active site surrounding the dTDP-sugar ligand is depicted in Figure 3b. The thymine ring is anchored into the active site by the side chains of Gln 47 and Tyr 65. In addition, it is sandwiched between the indole ring of Trp 50 and the guanidinium group of Arg 28, which forms a cation-π stacking interaction. Two arginine residues, Arg 28 and Arg 81, interact with the β- and α-phosphoryl groups of the dTDP-sugar, respectively.

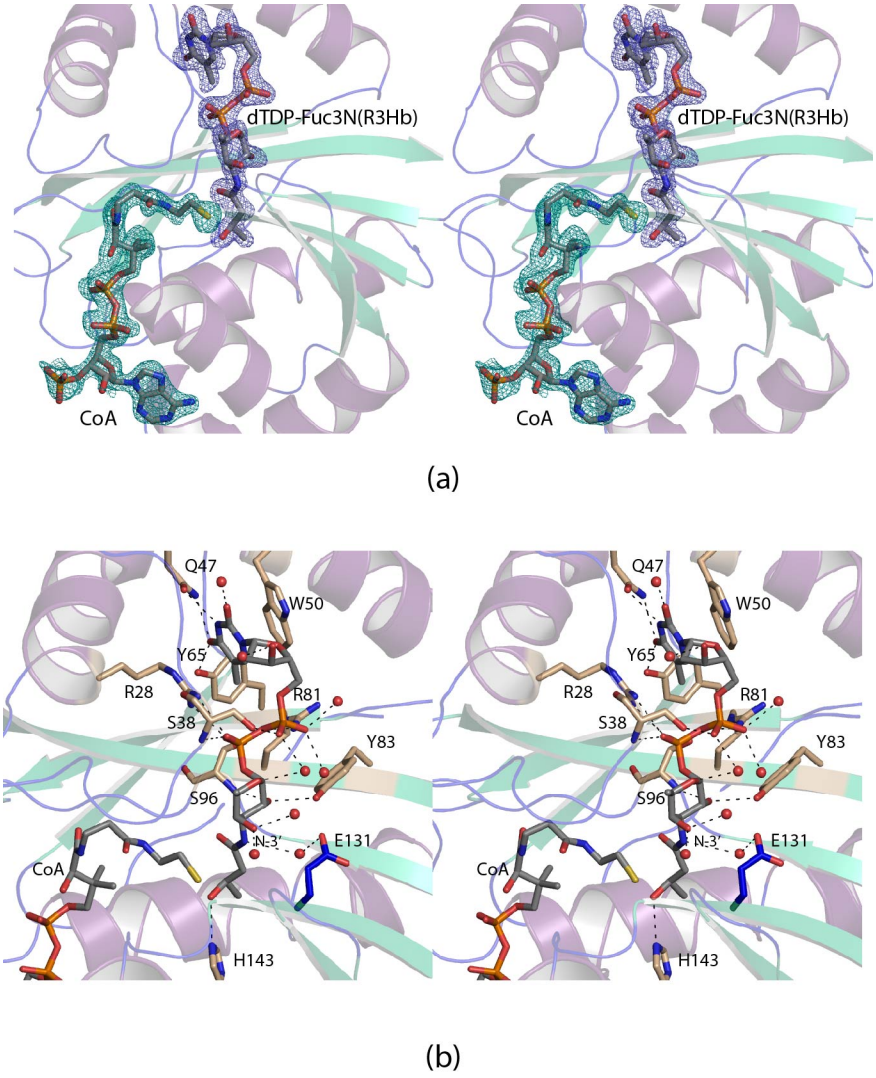
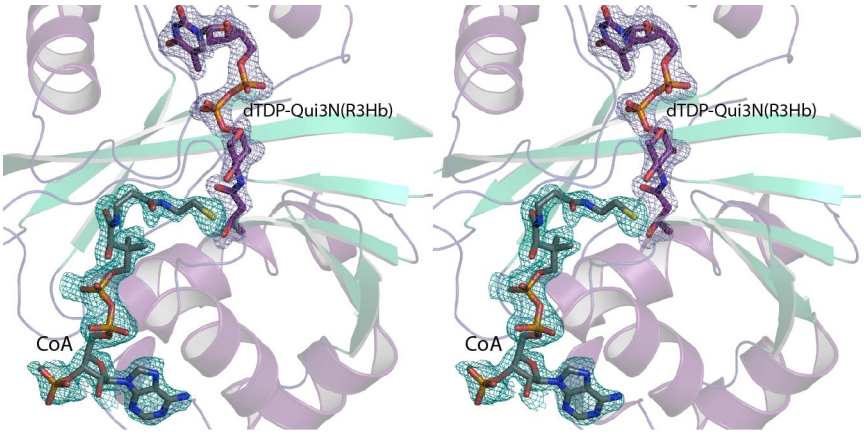


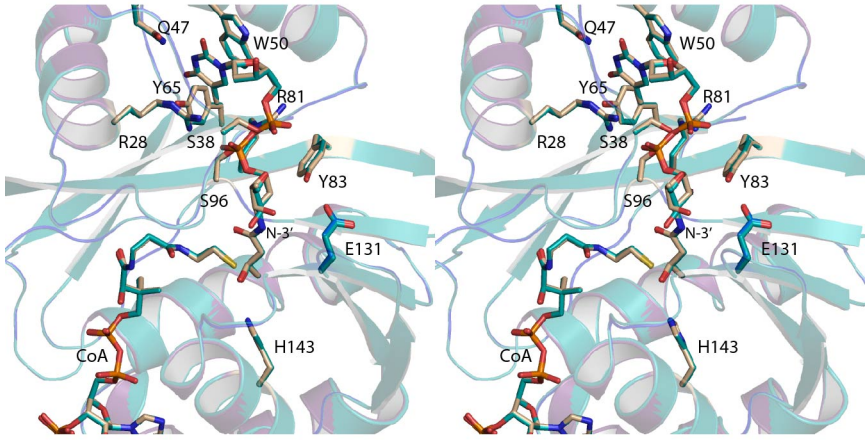
Figure 4.4 Active site of FdhC. Observed electron densities corresponding to the CoA and dTDP-Fuc3N(R3Hb) ligands are shown in (a) in stereo. The omit map, contoured at ~3σ, was calculated with coefficients of the form (F_o - F_c), where F_o was the native structure factor amplitude and F_c was the calculated structure factor amplitude. Note that the map was calculated before the ligands had been included in the X-ray coordinate file, and thus there is no model bias. A close-up view of the region surrounding the dTDP-Fuc3N(R3Hb) ligand is presented in stereo in (b). Possible hydrogen bonds are indicated by the dashed lines. The amino acid side chains that interact with the dTDP-sugar ligand are colored in wheat. Glu 131, which interacts with the amino nitrogen of the product indirectly via an ordered water molecule, is displayed in blue.

The side chain and backbone amide group of Ser 38 lie within 3.2 Å of an α-phosphoryl and a β-phosphoryl oxygen, respectively. The only side chain that interacts with the pyranosyl moiety of the dTDP-sugar is Tyr 83. Note that the C-4' hydroxyl group lies within hydrogen bonding distance to only water molecules. The side chain of His 143 is situated within 3.2 Å of the hydroxyl of the 3-hydroxybutanoyl group. There are no potential catalytic bases lying within 3.2 Å of the N-3' nitrogen of the dTDP-sugar. Glu 131 is the closest residue that could possibly function as a catalytic base via an ordered water molecule. This residue was targeted for further investigation as described below. The sulfur of the CoA sits at 3.6 Å from the N-3' nitrogen.

Given the lack of protein interactions between the C-4' hydroxyl group of the dTDP-Fuc3N(R3Hb) product, and the fact that the catalytic efficiency of FdhC with dTDP-Qui3N as a substrate is substantially reduced, the next structure determined in this investigation was that of FdhC complexed with CoA and dTDP-Qui3N(R3Hb). The model was refined at 1.9 Å resolution to an overall *R*-factor of 20.4%. Shown in Figure 4.5a is the electron density observed for the bound ligands. Within experimental error, the two complex models are virtually identical such that their α-carbons superimpose with a root-mean-square deviation of 0.1 Å. A superposition of the active sites for these two models is displayed in Figure 4.5b. As can be seen, there are no major changes in the positions of either the ligands or the protein side chains. Indeed, the positions of the C-4'



(a)



(b)

Figure 4.5 Active site of FdhC with an alternative substrate. Observed electron densities corresponding to the CoA and dTDP-Qui3N(R3Hb) ligands are shown in (a) in stereo. The omit map was calculated as described in the legend to Figure 3. A superposition of the FdhC active sites with bound CoA and either dTDP-Fuc3N(R3Hb) or dTDP-Qui3N(R3Hb), colored in wheat and cyan, respectively is displayed in (b) in stereo.

carbons of the sugars differ by only 0.5 Å and those of the C-4' hydroxyl oxygens by 1.8 Å. The reason for the variations in catalytic efficiencies of FdhC for the two substrates is not obvious. Most likely there are subtle differences that cannot be explained by time averaged crystal structures.

Shown in Figure 4.6 is a superposition of the ribbon drawings for the apoenzyme and the FdhC/CoA/dTDP-Fuc3N(R3Hb) complex. Excluding the loop defined by Asn 30 to Asp 43, the α-carbons for these two models superimpose with a root-mean-square deviation of 0.5 Å. The Asn 30 to Asp 43 loop, which connects the first two α-helices of the polypeptide chain, begins to diverge between the two models at Asn 29. Indeed, there are significant changes in the dihedral angles for Asn 29 (apoenzyme: $\phi = -66^{o}$, $\psi = -56^{o}$, ternary complex: $\phi = -100^{o}$, $\psi = 7.8^{o}$). As a consequence, some of the α–carbons moved by as much as ~13 Å when ligands bind to the FdhC active site. The driving force for this loop closure is apparently Ser 38, which forms hydrogen bonding interactions with both the α–and β-phosphoryl groups of the dTDP-sugar ligand through its side chain and its backbone amide group, respectively [Figure 3b]. Although caution must be used in the interpretation of B-values, it is noteworthy that the average B-factors for the loops in the apoenzyme and the ternary complex models are 48 $Å^2$ and 24 $Å^2$, respectively. Whereas it cannot be ruled out that this loop movement is simply an artifact of crystal packing, it does suggest protein flexibility in this region. This type of movement is also not without precedence in the *N*-acetyltransferases. For example, in the apo and liganded structures

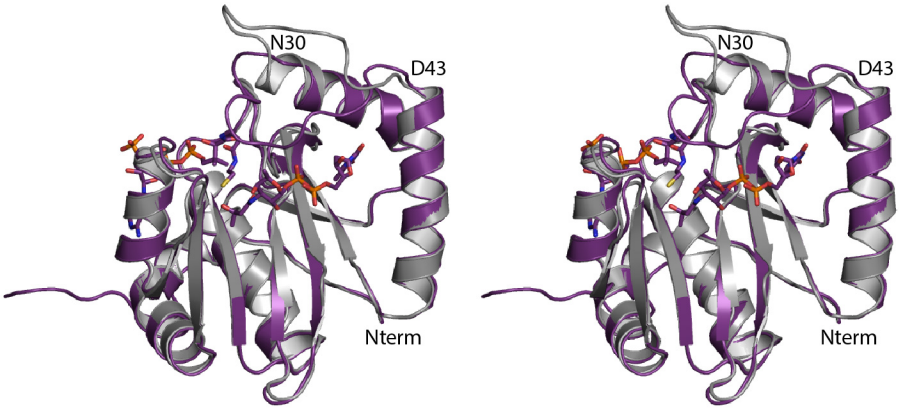


Figure 4.6 Comparison of the FdhC apoenzyme and ternary complex structures. The apoenzyme is depicted in white whereas the ternary complex of FdhC/CoA/dTDP-Fuc3N(R3Hb) is drawn in violet. A dramatic difference between the two forms of the enzyme occurs in the loop conformation defined by Asn 30 to Asp 43. The bound ligands are drawn in stick representations. The figure is presented in stereo.

of serotonin *N*-acetyltransferase, there is a large rearrangement of the loop connecting the first two α–helices of the polypeptide chain as well (22).

FdhC has a similar overall molecular architecture to that observed for PseH from *C. jejuni* and *H. pylori* and WecD from *E. coli*. The following comparison will focus on PseH from *H. pylori* given that it is of similar size to FdhC and that the X-ray model for the protein had no breaks in the polypeptide chain. PseH catalyzes the *N*-acetylation of UDP-4-amino-4,6-dideoxy-*N*-acetyl-L-altrosamine, and thus it functions on a C-4' rather than C-3' amino group. Overall the amino acid sequence identity and similarity between FdhC and PseH is quite low at 21% and 37%, respectively. A superposition of the ribbon representations for these two enzymes is presented in Figure 4.7. As can be seen, the overall folds of the two enzymes are strikingly similar. Indeed, the α–carbons for these two proteins superimpose with a root-mean-square deviation of 1.6 Å for 150 target pairs. The structure of PseH was solved in the presence of acetyl-CoA. The 4′-phosphopantetheine units align well between the two proteins. The adenosine groups, however, are oriented at nearly 180° from one another. Although the structure of PseH was solved in the absence of a UDP-sugar ligand, based on the superposition shown in Figure 4.7 it can be postulated that, similar to FdhC, the side chains of Arg 30 and Phe 52 abut either side of the uracil ring. The equivalent residues in FdhC are Arg 28 and Trp 50. Likewise, in FdhC the side chains of Gln 47 and Tyr 65 form hydrogen bonding

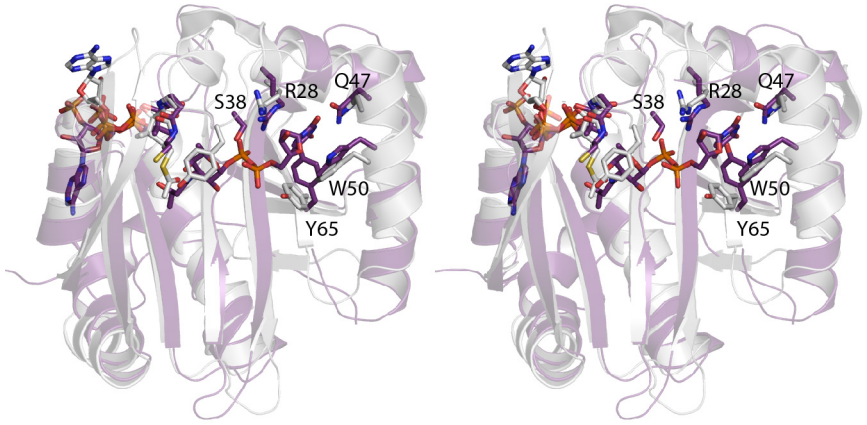


Figure 4.7 Comparison of the FdhC and PseH structures. Coordinates for the PseH model were obtained from the Protein Data Bank (accession code: 4RI1) (10). The ribbon representations for FdhC and PseH are displayed in violet and white, respectively. Those ligands and side chains belonging to FdhC are highlighted in purple where those belonging to PseH are displayed in white. The figure is presented in stereo.

interactions with the thymine ring of the substrate. In PseH the structurally equivalent residues are His 49 and Tyr 64, which might also serve a role in nucleobase positioning. As noted above, Ser 38 in FdhC serves to anchor the phosphoryl groups of the dTDP-sugar substrate into the active site. It is possible that Tyr 40 functions in a similar manner in PseH. A more detailed comparison of the active sites of these two enzymes will require an X-ray analysis of PseH as a ternary complex.

4.4.4 Catalytic Mechanism of FdhC

There has been considerable debate in the literature regarding the catalytic mechanisms of GNAT superfamily members and the need for a catalytic base to remove a proton from the amino group that is ultimately acylated (23-25). Inspection of the FdhC active site revealed no side chains located within ~3.2 Å that could function as a catalytic base. There was an ordered water molecule, however, located near the C-3' nitrogens of the dTDP-sugar ligands in both ternary complex models. As can be seen in Figure 4.4b, this solvent is positioned between the C-3' nitrogen of the dTDP-sugar substrate and the carboxylate of Glu 131. To test whether a water-mediated proton relay was playing a role in FdhC catalysis, two site-directed variants were constructed and kinetically studied. Both the E131Q and E131L were able to turn over the substrates, and in fact, their catalytic efficiencies were not significantly altered [Table 4.4]. Clearly from the site-directed mutagenesis experiments, Glu 131 does not function as a catalytic base.

Although it can be argued that at physiological pH the C-3' amino group of the FdhC substrate would be positively charged, the hydrophobic nature of the active site surrounding the pyranosyl group suggests that the neutral amino sugar is bound. A catalytic mechanism can thus be envisioned whereby the neutral amine functions as the nucleophile that attacks the carbonyl carbon of acyl-CoA. This attack leads to a cationic intermediate that spontaneously deprotonates. In addition, the thiolate of CoA is an excellent leaving group so there may be no need for a catalytic acid. The reaction mechanism of FdhC, as well as perhaps many other GNAT superfamily members, most likely proceeds via catalysis by approximation.

In summary, the research described here demonstrates that FdhC is, indeed, an *N*-acyltransferase as originally suggested on the basis of bioinformatics. Importantly, for the first time, the structure of a GNAT superfamily member that functions on a nucleotide-linked sugar has been obtained as a ternary complex. As a consequence, key interactions between the dTDP-sugar and the protein have been identified. This study thus provides new molecular insight into the GNAT superfamily in general.

4.5 Acknowledgments

A portion of the research described in this paper was performed at Argonne National Laboratory, Structural Biology Center at the Advanced Source (U. S. Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357). We gratefully acknowledge Dr. Norma E. C. Duke and Dr. Stephan L. Ginell for assistance during the X-ray data collection at Argonne.

4.6 References

1. Rangarajan ES, Ruane KM, Sulea T, Watson DC, Proteau A, Leclerc S, Cygler M, Matte A, & Young, NM (2008) Structure and active site residues of PglD, an *N*-acetyltransferase from the bacillosamine synthetic pathway required for *N*-glycan synthesis in *Campylobacter jejuni*. *Biochemistry* 47(7):1827-1836.

2. Olivier NB & Imperiali B (2008) Crystal structure and catalytic mechanism of PglD from *Campylobacter jejuni*. *J Biol Chem* 283(41):27937-27946.

3. Thoden JB, Cook PD, Schaffer C, Messner P, & Holden HM (2009) Structural and functional studies of QdtC: an *N*-acetyltransferase required for the biosynthesis of dTDP-3-acetamido-3,6-dideoxy-alpha-D-glucose. *Biochemistry* 48(12):2699-2709.

4. Chantigian DP, Thoden JB, & Holden HM (2013) Structural and biochemical characterization of a bifunctional ketoisomerase/*N*-acetyltransferase from *Shewanella denitrificans*. *Biochemistry* 52(46):8374-8385.

5. Thoden JB, Reinhardt LA, Cook PD, Menden P, Cleland WW, & Holden HM (2012) Catalytic mechanism of perosamine *N*-acetyltransferase revealed by high-resolution X-ray crystallographic studies and kinetic analyses. *Biochemistry* 51(16):3433-3444.

6. Kubiak RL & Holden HM (2012) Structural studies of AntD: an *N*-Acyltransferase involved in the biosynthesis of D-anthrose. *Biochemistry* 51(4):867-878.

7. Thoden JB & Holden HM (2010) Molecular structure of WlbB, a bacterial *N*-acetyltransferase involved in the biosynthesis of 2,3-diacetamido-2,3-dideoxy-D-mannuronic acid. *Biochemistry* 49(22):4644-4653.

8. Hung MN, Rangarajan E, Munger C, Nadeau G, Sulea T, & Matte A (2006) Crystal structure of TDP-fucosamine acetyltransferase (WecD) from *Escherichia coli*, an enzyme required for enterobacterial common antigen synthesis. *J Bacteriol* 188(15):5606-5617.

9. Song WS, Nam MS, Namgung B, & Yoon SI (2015) Structural analysis of PseH, the *Campylobacter jejuni* *N*-acetyltransferase involved in bacterial *O*-linked glycosylation. *Biochem Biophys Res Commun* 458(4):843-848.

10. Ud-Din AI, Liu YC, & Roujeinikova A (2015) Crystal structure of *Helicobacter pylori* pseudaminic acid biosynthesis *N*-acetyltransferase PseH: implications for substrate specificity and catalysis. *PloS one* 10(3):e0115634.

11. Hu D, Liu B, Dijkshoorn L, Wang L, & Reeves PR (2013) Diversity in the major polysaccharide antigen of *Acinetobacter baumannii* assessed by DNA sequencing, and development of a molecular serotyping scheme. *PloS one* 8(7):e70329.

12. Weber BS, Harding CM, & Feldman MF (2015) Pathogenic *Acinetobacter*: from the Cell Surface to Infinity and Beyond. *J Bacteriol* 198(6):880-887.

13. Kenyon JJ, Nigro SJ, & Hall RM (2014) Variation in the OC locus of *Acinetobacter baumannii* genomes predicts extensive structural diversity in the lipooligosaccharide. *PloS one* 9(9):e107833.

14. Thoden JB & Holden HM (2005) The molecular architecture of human *N*-acetylgalactosamine kinase. *J Biol Chem* 280(38):32784-32791.

15. Ness SR, de Graaff RA, Abrahams JP, & Pannu NS (2004) CRANK: new methods for automated macromolecular crystal structure solution. *Structure* 12(10):1753-1761.

16. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.

17. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.

18. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, & Read RJ (2007) Phaser crystallographic software. *J. Appl. Cryst.* 40:658-674.

19. Woodford CR, Thoden JB, & Holden HM (2015) New role for the ankyrin repeat revealed by a study of the *N*-formyltransferase from *Providencia alcalifaciens*. *Biochemistry* 54(3):631-638.

20. Alpers DH, Appel SH, & Tomkins GM (1965) A spectrophotometric assay for thiogalactoside transacetylase. *J Biol Chem* 240:10-13.

21. Magalhaes ML & Blanchard JS (2005) The kinetic mechanism of AAC3-IV aminoglycoside acetyltransferase from *Escherichia coli*. *Biochemistry* 44(49):16275-16283.

22. Hickman AB, Namboodiri MA, Klein DC, & Dyda F (1999) The structural basis of ordered substrate binding by serotonin *N*-acetyltransferase: enzyme complex at 1.8 A resolution with a bisubstrate analog. *Cell* 97(3):361-369.

23. Dyda F, Klein DC, & Hickman AB (2000) GCN5-related *N*-acetyltransferases: a structural overview. *Annu Rev Biophys Biomol Struct* 29:81-103.

24. Vetting MW1, S de Carvalho LP, Yu M, Hegde SS, Magnet S, Roderick SL, & Blanchard JS (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* 433(1):212-226.

25. Favrot L, Blanchard JS, & Vergnolle O (2016) Bacterial GCN5-Related *N*-acetyltransferases: from resistance to regulation. *Biochemistry* 55(7):989-1002.

26. Laskowski RA, Moss DS, & Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231(4):1049-1067.

27. DeLano WL (2002) The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. *The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA*.

CHAPTER 5

PROJECT CRYSTAL (COLLEAGUES RESEARCHING WITH YOUNG SCIENTISTS TEACHING AND LEARNING)

5.1 GOALS OF THE OUTREACH

Project CRYSTAL (Colleagues Researching with Young Scientists Teaching And Learning) was first established because there was a need for local middle school students to develop a love for science which seemed lacking previously. While this project originated as a result of a local issue, it appears that this lack of appreciation for science among young students spans the nation to the point of concern. For the past several decades, American middle school students have ranked below 20th in the world when assessing their understanding of basic science. As middle school is a time for immense maturation (academically, physically, and emotionally), this time in a student's development can impact them for the rest of their lives. In fact, if students do not discover an appreciation for science by the age of 13, studies show that the likelihood of sparking scientific curiosity later on in life is minimal (1-3).

The primary objective of Project CRYSTAL is to engage middle school students in applicable scientific discovery at a high level. By exposing these students to graduate level research accompanied by a comprehensive lecturing series, this outreach program aims to have students start to understand how science plays a role in their everyday lives. Success of this is analyzed at the end of both the Fall and the Spring semesters. At the end of the Fall semester, the students play a Jeopardy-style game designed to assess their retention of basic materials. At the end of the Spring semester, students present their research project as a group at the Madison Middle School Science Symposium.

This experience shows the cumulative comprehension of their research experience and its impact.

Goals for Project CRYSTAL reach beyond the realm of educating middle school students. Another important objective of this program is to train graduate students to teach complicated science to a novice audience. Graduate students are most often required to be a Teacher's Assistant (TA) for one or more semesters during their duration at an institution. While this experience sometimes allows students to lecture to an audience of college students, the TA experience does not give graduate students an opportunity to teach at the most basic level; college students are, more-or-less, standardized in their science instruction by the time they reach higher education.

Bringing graduate students into a classroom where they need to develop a curriculum and present science to young students is a challenging learning experience. The mentor needs to be able to bring complicated subject matter, like protein chemistry and valence electron reactivity, down to an understandable level for 13-year-olds. It is a strong lesson in teaching; explaining science is not a script that can be recited independent of the composition of the audience. Different groups of students need to be taught via different methods.

As Project CRYSTAL has become more and more successful at the University of Wisconsin-Madison, laboratories from other universities have begun to take notice. Therefore, another goal of the program has been to construct a curriculum which can

easily be shared with other research groups interested in starting their own version of the outreach. While organizing an original project every year for the students garnered published manuscripts, it also sometimes resulted in complications incurring from normal research stumbling blocks such as difficulties with cloning, insoluble proteins, impure protein purifications, or a lack of protein crystals. This often forced mentors to fumble week-to-week trying to get experiments to work for the students. Furthermore, as graduate students turned over, and other laboratories became interested in adopting Project CRYSTAL, consistency became a priority.

During its most recent year, the Project CRYSTAL team worked to establish a curriculum that could be repeated year after year and could easily be shared with other institutions. This curriculum revolved around cloning, expression, and purification of fluorescent protein variants eGFP, mCitrine, mCherry, and mPlum. Furthermore, lectures were created with all new original artwork and photography in order to convey information in an original way that any laboratory could adopt. Since the implementation of this goal in 2014, two more groups (one from Indiana University-Indianapolis and another from Marquette University) have begun their own form of Project CRYSTAL.

5.2 MENTORING EXPERIENCE

While Project CRYSTAL meets once a week with its students, there is a great deal of preparation for the mentors week-to-week. When Project CRYSTAL was in its infancy, the mentor's work began early in the summer when a new project was chosen. Once a target had been selected, much of the summer was designated to running through the expansive experimentation that the Project CRYSTAL students would replicate later on in the year. On top of this, all lecture media were re-oriented to emphasize the new project.

In its most recent year, because of the replicable project that was chosen, these summer experiments were not as necessary. However, what still needed to be accomplished was to organize the lectures and prepare the laboratory for the young scientists. The lecture order was edited for the year's specific schedule given holidays, and a syllabus dictating who would be presenting certain material was composed. For the laboratory, specific bench space, pipettes, and reagents were placed aside for use exclusively by the middle schoolers. Because of their inexperience, this specification of space and materials is necessary to prevent contamination and breakage of the group's common equipment.

The first meeting of the mentors and the Project CRYSTAL students is at an initial gathering where the students are accompanied by their parents. This is a chance to give a first impression to the students while being reassuring and professional in

front of the parents. During this meeting, each mentor takes a moment to introduce themselves, their background, and what their role is in the laboratory. Finally, information is gathered on the new students to create a profile for each of them on the Project CRYSTAL website.

Each Friday during the school year, the mentors and Dr. Holden convene to discuss a plan for the upcoming week. The meeting starts by dictating a detailed timeline for the Project CRYSTAL students: from the time they arrive for class to the time their parents pick them up. As a result of this discussion, the mentors are able to focus their goals for what should be accomplished in the given class period. Once the timeline has been compiled, those who are not lecturing in the coming week list the materials and reagents that need to be prepared for the laboratory. They then decide who will be responsible for this preparation. This part of the meeting also contains a discussion of group dynamics, and how the young scientists should be split during their time in the wet lab.

Finally, during these meetings, the lecturer goes through their slides in front of the other mentors and Dr. Holden. The slides are critiqued for clarity, originality, and age-adaptiveness of the material. Often, the lecturer will meet in private with Dr. Holden once more before the class in order to go through the changes that were made. Furthermore, the lecturer is expected to practice and be prepared to present their slides without stumbling.

On the day of the class, mentors spend the majority of the morning preparing for the student's arrivals. Everything that had been discussed in the Friday meeting is accomplished while normal laboratory duties continue simultaneously. All of the mentors meet the students as they enter the science complex, and the students are always accompanied to the lecture room. Class is 90 minutes long and includes both a lecture and research protocol experience in the laboratory. Each day is designed such that the students are not in a didactic lecture setting for longer than 20 minutes at a time.

After the students are escorted by the mentors to where their parents pick them up, Dr. Holden and the graduate students reconvene to discuss the highs and lows of the day's class. Group dynamics are once again discussed and responsibilities for the remainder of the week are divided amongst the mentors. One graduate student is to update the Project CRYSTAL website with a picture or two from the week's class. Another mentor is to update the Project CRYSTAL social media pages to reflect the day's accomplishments.

At the end of the semester the mentors assess the progress of the young scientists. In the winter, a trivia game is made to test the concepts learned in the first half of the year. Emphasis is given to the ideas that the students had a more difficult time grasping. In the Spring, mentors guide the Project CRYSTAL students in creating a poster to present at the Madison Middle School Science Symposium. Mentors ensure

that their presentation is well organized and that information is conveyed in an appropriate way such that their audience remains engaged, however impressed.

Finally, in late Spring, the mentors meet with Mr. Ropa to receive applications for the following year's students. Applications are quite basic; they merely ask students what other extracurricular activities they are involved in, and why they want to be in Project CRYSTAL. Students are picked based on their responses to these questions and a confirmation of their good behavior in a classroom setting by Mr. Ropa.

5.3 References

1. Anonymous (2003) NSTA Position Statement.

2. Anonymous (2014) Chapter 1. Elementary and Secondary Mathematics and Science Education. in *Science and Engineering Indicators* (National Science Fund).

3. Desilver D (2015) U.S. students improving – slowly – in math and science, but still lagging internationally. (Pew Research Center).