This is to certify that the
dissertation entitled

STRUCTURAL AND FUNCTIONAL STUDIES OF THE ENZYMES INVOLVED IN A BACTERIAL GDP-D-RHAMNOSE BIOSYNTHETIC PATHWAY

presented by

NICOLE A. WEBB

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Biochemistry and Molecular Biology

Major Professor's Signature

12-12-05

Date

MSU is an Affirmative Action/Equal Opportunity Institution

STRUCTURAL AND FUNCTIONAL STUDIES OF THE ENZYMES INVOLVED IN A BACTERIAL GDP-D-RHAMNOSE BIOSYNTHETIC PATHWAY

By

Nicole A. Webb

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry and Molecular Biology

2005

ABSTRACT

STRUCTURAL AND FUNCTIONAL STUDIES OF THE ENZYMES INVOLVED IN A BACTERIAL GDP-D-RHAMNOSE BIOSYNTHETIC PATHWAY

By

Nicole A. Webb

D-rhamnose is a rare 6-deoxy sugar found primarily in the lipopolysaccharide molecules of pathogenic gram-negative bacteria, where it is involved in host-bacterium interactions and the establishment of infection. The biosynthetic pathway of D-rhamnose may be a potential therapeutic target, as it is expected that inhibitors of the enzymes would not affect human metabolic pathways. The biosynthesis of D-rhamnose proceeds through the conversion of GDP-D-mannose by GDP-D-mannose 4,6-dehydratase (GMD) to GDP-4-keto-6-deoxy-D-mannose, which is subsequently reduced to GDP-D-rhamnose by the reductase (RMD). This study focuses on the structural and enzymatic characterization of the two enzymes.

The X-ray crystal structure of GMD from *Pseudomonas aeruginosa* has been determined in complex with NADPH and GDP. GMD belongs to the NDP-sugar modifying subfamily of the short-chain dehydrogenase/reductase (SDR) enzymes, all of which exhibit bidomain structures and a conserved catalytic triad (Ser/Thr, Tyr-XXX-Lys). Although most members of this enzyme subfamily display homodimeric structures, this bacterial GMD forms a tetramer in the same fashion as the plant *Arabidopsis thaliana* GMD isoform, MUR1. Based on the conservation of subunit interactions and sequences in GMDs, evidence suggests that the tetrameric form of this enzyme is the preferred, and perhaps functionally relevant, oligomeric state for most bacterial and eukaryotic GMDs.

Initial X-ray analysis has been completed on RMD from *P. aeruginosa* as well as a homolog from the bacterium *Aneurinibacillus thermoaerophilus*. While crystals of RMD from *P. aeruginosa* diffracted to 3.7 Å, crystals of RMD from *A. thermoaerophilus* diffracted to 1.8 Å. Complete data sets were collected on crystals from both species. The *P. aeruginosa* RMD crystal belongs to the tetragonal space group P422 (or higher symmetry) (a = b = 182, c = 250 Å), with 12 molecules in the asymmetric unit. The *A. thermoaerophilus* RMD crystal belongs to the triclinic space group P1 (a = 47, b = 56, c = 79 Å, α = 72, β = 83, γ = 76°), with 2 molecules in the asymmetric unit. Amino acid sequence homology details indicate that RMD is also a member of the SDR family, whose structures are quite similar. Therefore, full structure determination by the molecular replacement method using the coordinates from the most closely related SDR protein, GMD from *P. aeruginosa*, is in progress. As no structure is currently available for RMDs, crystallographic analysis of RMD should provide detailed information on the active site of the enzyme and facilitate structure-based inhibitor design.

Finally, using a capillary-electrophoresis-based enzymatic assay, we have shown that recombinant GMD and RMD from *P. aeruginosa* are able to convert GDP-D-mannose to GDP-D-rhamnose, confirming the product structure by NMR analysis. Additionally, we present a method for the enzymatic synthesis of GDP-D-rhamnose, an important glycobiological building block not commercially available. Synthesis of GDP-D-rhamnose is a crucial prerequisite for (1) the identification and biochemical studies of corresponding glycosyltransferases, as well as (2) the determination of how this deoxysugar, as a component of cell wall glycoconjugates, contributes to the virulence of plant and human pathogens.

for my husband,
who has inspired me
to achieve more than I
ever dreamed possible

for my daughter
whose life
has provided new
meaning and purpose
to mine

The proper rewards are not simply tacked on to the activity for which they are given,

but are the activity itself in consummation.

-C.S. Lewis, The Weight of Glory

ACKNOWLEDGMENTS

One of the pleasures of finally finishing my graduate program is the opportunity to show my appreciation to those who were influential in my career as a graduate student. I am grateful to many people in the academic community at Michigan State University. I would first like to express my gratitude to my advisor and mentor, Dr. R. Michael Garavito, whose guidance and encouragement provided me with an opportunity to grow as a scientist. My past and current colleagues have imparted much needed support: Ms. Amy Sharmen, Ms. Karen Poster-Verrill and Mr. Micheal Dumond shared with me their extensive knowledge of lab techniques, Dr. Anne Mulichak helped walk me through solving my first crystal structure, Dr. Ling Qin was always a great source of assistance in collecting data, and Dr. Christine Harman and Dr. Rachel Powers were an invaluable source of support both academically as well as personally. I also wish to thank my committee members, Dr. Christoph Benning, Dr. James Geiger, Dr. Jack Priess and Dr. John Wang, who provided a great deal of insight into my project. I could also not due without the kindness and computer-savvy of Dr. Kaillathe "Pappan" Padmanabhan.

There are several members of the scientific community outside of Michigan State University I would also like to acknowledge. My collaborator on the functional studies was Dr. Karen Poon, a member of the lab of Dr. Joseph Lam, Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada. Furthermore, the staff at the Advanced Photon Source, Argonne National Laboratory assisted in collecting data and data processing. Specifically I would like to thank Dr. Stephan Ginell at the

Structural Biology Center Collaborative Access Team and Dr. Zdzislaw Wawrzak at the DuPont-Northwestern-Dow Collaborative Access Team.

I am blessed with a supportive family. I would like to thank those members who guided my path to graduate school: my father- and mother-in-law, Dr. and Mrs. Charles and Philippa Webb for planting the idea and my parents, Mr. and Mrs. Philip and Beverley Cowan for their encouragement to attend. My parents have instilled in me the importance of maintaining a high level of integrity and pressed me to use my God-given talents to the best of my abilities. My parents-in-law have shown confidence in me even when I have had none. My siblings and their families have been understanding of the commitment level of the path I have chosen as well as the distance it has taken me, which has often led to my absence in family activities.

Finally I am most grateful to my core family. When I first began at MSU, we were a family of two. Now we are three. My daughter Elizabeth, or as we call her "Bizzy", has brought a new and unexpected level of joy to my life. Her sense of exploration and experimentation warms a scientist's heart. Ultimately, I would like to express my appreciation for the friendship, encouragement and advice of my husband, David. I know very few people as intelligent, logical, decisive, and devoted. His unique outlook on life has not only influenced my graduate career but also my world-view, and I consider it a privilege that he accepted me as his wife.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADP	adenine diphosphate
AGME	ADP-L-glycero-D-mannoheptose 6`-epimerase
CDP	cytidine diphosphate
CE	capillary electrophoresis
CF	cystic fibrosis
dTDP	deoxythymidine diphosphate
E_1	CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydrase
E_3	E_1 reductase
E_{ep}	CDP-3,6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase
E_{OD}	CDP-D-glucose 4,6-dehydratase
E_P	glucose-1-phosphate cytidylyltransferase
E_{red}	CDP-3,6-dideoxy-D-glycero-L-glycero-4-hexulose-4-reductase
FPLC	fast protein liquid chromatography
GalE	UDP-galactose epimerase
GDP	guanosine diphosphate
GFS	GDP-fucose synthase
GMD	GDP-D-mannose 4,6-dehydratase
GMER	GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase/reductase
GTS	GDP-6-deoxy-D-talose synthetase
HEPES	4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid
HPLC	high-pressure/performance liquid chromatography
IPTG	isopropyl-beta-D-thyogalactopyranoside
LB	Luria broth
LPS	lipopolysaccharide
MES	4-morpholineethanesulfonic acid
MPD	2-methyl-2,4-pentanediol
MUR1	GMD isoform from *Arabidopsis thaliana*
NAD(H)	nicotinamide-adenine dinucleotide
NADP(H)	nicotinamide-adenine dinucleotide diphosphate
NDP	nucleoside disphosphate
Ni-NTA	nickel nitrilo-triacetic acid
NMR	nuclear magnetic resonance

OD	optical density
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
PEP	pentaerythritol propoxylate
PL	phospholipid
PLSCR1	phospholipid scramblase 1
RMD	GDP-4-keto-6-deoxy-D-mannose reductase
RmlA	glucose-1-phosphate thymidylyltransferase
RmlB/dTGDH	dTDP-D-glucose 4,6-dehydratase
RmlC	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase
RmlD	dTDP-4-keto-6-deoxy-L-mannose reductase
RMSD	root mean square deviation
SDR	short chain dehydrogenase/reductase
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SH3	Src homology 3
SQD1	UDP-sulfoquinovose synthase
T11	dTDP-4-keto-6-deoxy-L-mannose reductase
Tris	tris(hydroxymethyl)aminomethane
UDP	uridine diphosphate
UV	ultraviolet

INTRODUCTION

Deoxysugars and their biosynthesis

Prevalence of deoxyhexoses

Lipopolysaccharide (LPS) molecules are major components of the outer membrane of the cell walls of gram-negative bacteria and are intimately involved in host-bacterium interactions and the viability of the pathogen. The permeability restrictions imposed by the outer membrane contribute to the intrinsic resistance of gram-negative bacteria to antibiotics. The LPS molecules, in particular, play a role in outer membrane stability and impede the destruction of bacterial cells by serum components and phagocytic cells. In addition, the LPS molecules may be involved in colonization or antigenic shifts that determine the course and outcome of the infection.

LPS molecules share a common tripartite structure of a lipid A region, a core oligosaccharide region and an O-antigen region. The O-antigen, which is the most exposed region, consists of repeating units of varying monosaccharides. O-antigen can be a linear or branched homopolymer or heteropolymer, and can be strain-specific or vary within a strain. A tremendous amount of effort has been exerted to determine the exact chemical structures of O-antigen due to its contribution to bacterial virulence. In general, the structures consist of neutral sugars, amino sugars, sugar acids and many different deoxysugars. More recently, studies have expanded to include the enzymology involved in the biosynthesis and assembly of O-antigen. Specifically, the biosynthesis of 6-deoxyhexoses has received heightened interest after a few of them have shown to be substantial LPS O-antigen components of several human pathogens. Humans lack

metabolism for many deoxyhexoses found in bacterial cell walls and would most likely remain unaffected by inhibitors of bacterial cell wall metabolism. Since many of the currently available antibiotics target the same cellular process or even the same target enzyme, multiple-drug resistance of pathogens has become a serious problem, placing a high demand on new therapeutic targets and/or strategies. The biosynthetic pathways of 6-deoxyhexoses may be targets for novel antibacterial therapeutics.

6-deoxyhexoses are formed from a common monosaccharide by replacing the 6-hydroxyl group with a hydrogen atom, which can dramatically affect its biological function. The committed step in the 6-deoxyhexose biosynthetic pathway is the conversion of the precursor, the nucleotide diphosphate (NDP)-activated hexose, to an NDP-activated 4-keto-6-deoxyhexose by an NDP-sugar 4,6-dehydratase. The resulting NDP-4-keto-6-deoxyhexose intermediate serves as a branching point for D- and L-deoxysugar pathways. The intermediate may undergo subsequent steps such as epimerization, reduction or methylation to produce mono-, di-, tri- and tetradeoxysugars and branched-chain deoxysugars. More than one nucleotidyl sugar can be associated with a given deoxyhexose. For instance, both dTDP-D-glucose and UDP-D-glucose serve as precursors for L-rhamnose. However, it appears that only GDP-D-mannose serves as a precursor for D-rhamnose. In addition to being present as constituents of bacterial LPS molecules, deoxyhexoses can also be found in macrolide antibiotics. The general pathways to several 6- and 3,6-di-deoxyhexoses are shown in Figure 1. The structures of those deoxyhexoses are shown in Figure 2.

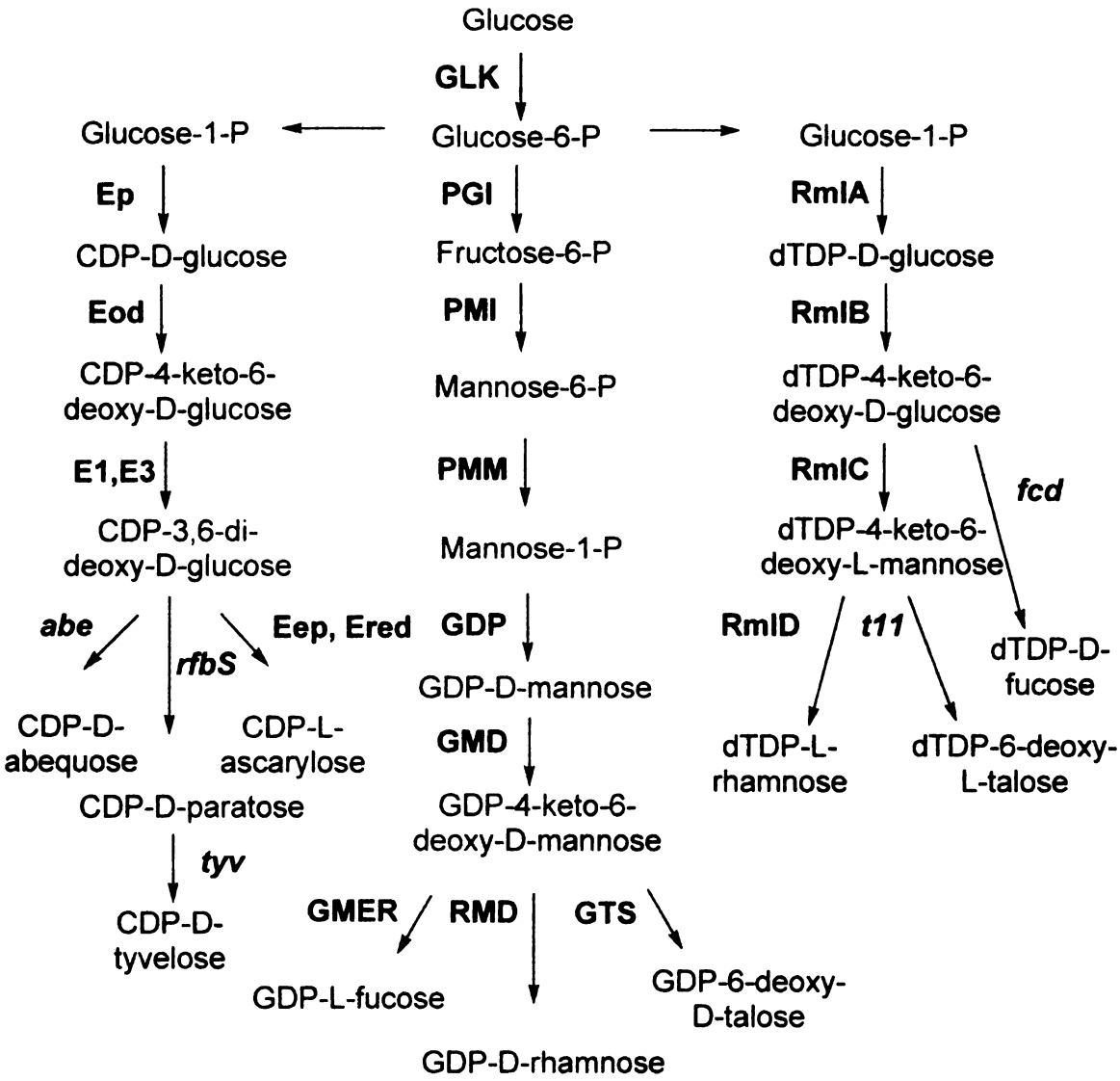
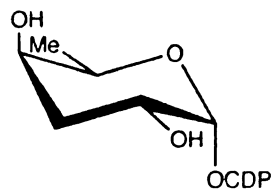
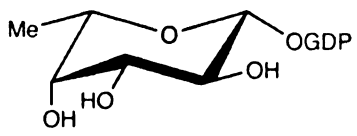


Figure 1. Biosynthetic pathways for 6- and 3,6-di-deoxyhexoses

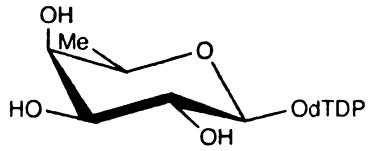
GLK, glucokinase, PGI, phosphoisomerase, PMI, phosphomannoisomerase, PMM, phosphomannomutase, GDP, manno-1-phosphate guanylyltransferase, GMD, GDP-D-mannose 4,6-dehydratase, GMER, GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase/reductase, RMD, GDP-4-keto-6-deoxy-D-mannose reductase, GTS, GDP-6-deoxy-D-talose synthetase, PGM, phosphoglycomutase, RmlA, glucose-1-phosphate thymidylyltransferase, RmlB, dTDP-D-glucose 4,6-dehydratase, RmlC, dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase, RmlD, dTDP-4-keto-6-deoxy-L-mannose reductase, T11, dTDP-4-keto-6-deoxy-L-mannose reductase, gene product of *fcd*, dTDP-4-keto-6-deoxy-D-glucose reductase, E_P, glucose-1-phosphate cytidylyltransferase, E_{OD}, CDP-D-glucose 4,6-dehydratase, E_1, CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydrase, E_3, E_1 reductase, E_{ep}, CDP-3,6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase, E_{red}, CDP-3,6-dideoxy-D-glycero-L-glycero-4-hexulose-4-reductase, gene product of *rfbS*, CDP-D-paratose synthase, GDP, guanosine diphosphate, dTDP, deoxythymidine diphosphate, CDP, cytidine diphosphate.



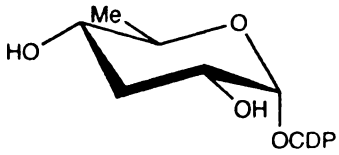
CDP-D-abequose



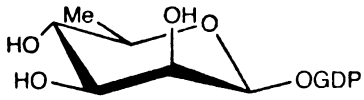
GDP-L-fucose



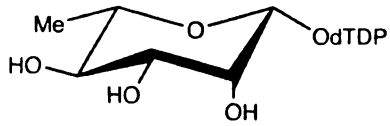
dTDP-D-fucose



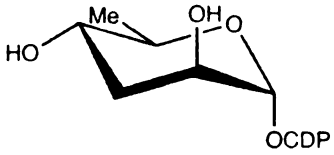
CDP-D-paratose



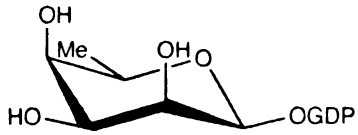
GDP-D-rhamnose



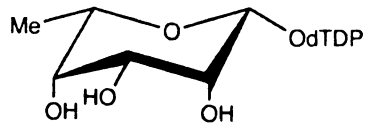
dTDP-L-rhamnose



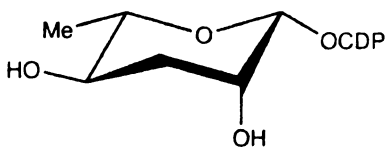
CDP-D-tylevose



GDP-6-deoxy-D-talose



dTDP-6-deoxy-L-talose



CDP-L-ascaryose

Figure 2. Structures of 6- and 3,6-di-deoxyhexoses

GDP-deoxyhexoses biosynthetic pathways

The 6-deoxyhexoses originating from the GDP-D-mannose precursor whose pathways have been studied include GDP-D-rhamnose, GDP-L-fucose and GDP-6-deoxy-D-talose (Figure 1, middle). The enzyme GDP-D-mannose 4,6-dehydratase (GMD) is responsible for catalyzing the conversion of GDP-D-mannose to GDP-4-keto-6-deoxy-D-mannose, which serves as the common intermediate to these 6-deoxyhexoses. Enzymes possessing GMD activity have been identified in bacteria [1-4], plants [5] and animals [6, 7]. For GDP-D-rhamnose synthesis, a 4-reductase (RMD) that targets the 4-keto group of the intermediate has been identified in *Pseudomonas aeruginosa* [8] and *Aneurinibacillus thermoaerophilus* [3, 8]. D-rhamnose is found mainly as a component of the LPS of gram-negative bacteria, such as the plant pathogen *Xanthomonas campestris* [9] and the human pathogens *Helicobacter pylori* [10] and *P. aeruginosa* [11]. The 3,5-epimerase/4-reductase (GMER) involved in the GDP-L-fucose pathway has been identified in bacteria [12], plants [13] and animals [14]. L-Fucose is a substantial cell wall component of bacteria, affects nodulation in rhizobial organisms [15, 16] and is important in stem development and strength in plants [5]. In humans, L-fucose is present in the human ABO blood group antigens and Lewis (Le) glycans, playing a role in inflammation and immune response [17]. The enzyme involved in the GDP-6-deoxy-D-talose pathway has been identified in bacteria, where 6-deoxy-D-talose is a component of LPS [18, 19]. The GDP-4-keto-6-deoxy-D-mannose reductase (GTS) involved in GDP-6-deoxy-D-talose biosynthesis has been characterized in *Actinobacillus actinomycetemcomitans* [18, 20]. In addition to 6-deoxyhexoses, GDP-4-keto-6-deoxy-D-mannose can also be converted to

the dideoxy amino sugar GDP-D-perosamine by GDP-D-perosamine synthetase (RfbE), which has been identified in *Vibrio cholerae* [19]

dTDP-deoxyhexoses biosynthetic pathways

The precursor dTDP-D-glucose leads to the following 6-deoxyhexoses: dTDP-L-rhamnose, dTDP-6-deoxy-L-talose and dTDP-D-fucose (Figure 1, right). Similar to the GDP-D-mannose pathways, these deoxyhexoses are synthesized from dTDP-D-glucose via dTDP-4-keto-6-deoxy-D-glucose with the assistance of dTDP-D-glucose 4,6-dehydratase (RmlB). The 4-keto-6-deoxyhexose again serves as a common intermediate in the pathways to these deoxyhexoses. The enzymes involved in the biosynthesis of dTDP-L-rhamnose have been studied extensively from *Salmonella enterica* [21-26] and *Escherichia coli* [27-32]. The catalytic action of RmlB is related to GMD, dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase (RmlC) performs a double epimerization at position C3 and C5 like GMER, and the direct analog to dTDP-4-keto-6-deoxy-L-mannose reductase (RmlD) is RMD. L-rhamnose is a known component of cell walls of several pathogens such as *P. aeruginosa*, *Mycobacterium tuberculosis*, *V. cholerae*, *Enterococcus feacalis* and *Streptococcus mutans* as reviewed in Maki *et al.* [33]. RmlB-C are also involved in the pathway leading to 6-deoxy-L-talose. The remaining enzyme in the pathway is a dTDP-4-keto-6-deoxy-L-mannose reductase like RmlD; however, it provides the stereoselectivity to reduce the intermediate to dTDP-6-deoxy-L-talose [34]. The gene for dTDP-6-deoxy-L-talose biosynthesis (product of the gene *tll*) and its corresponding protein has been identified in *A. actinomycetemcomitans*, one of the few species of bacteria where 6-deoxy-L-talose is a component of the cell wall [34]. Finally, in dTDP-D-fucose synthesis, the dTDP-4-keto-6-deoxy-D-glucose reductase (product of the gene

fcd) catalyzes the reduction of intermediate [35]. D-fucose is located in the cells and capsule structures of a limited number of gram-negative and gram-positive bacteria.

CDP-deoxyhexoses biosynthetic pathways

The deoxyhexoses originating from the precursor CDP-D-glucose shown in Figure 1 (left) are actually 3,6-dideoxyhexoses. Since 6-deoxyhexoses serve as the universal precursors for higher reduced di- and tri-deoxyhexoses, the initial enzymatic steps are the same. The enzymes involved in the pathway leading to L-ascarylose have been well studied from the bacteria *Yersinia pseudotuberculosis* [36]. Again the committed step is the conversion of CDP-D-glucose to the 4-keto-6-deoxy intermediate by the CDP-D-glucose 4,6-dehydratase (E_{OD}). The 4-keto-6-deoxyhexose is converted to the 3,6-dideoxyhexose by sequential activities of CDP-6-deoxy-L-threo-D-glycero-4-hexulose-3-dehydrase (E_1) and E_1 reductase (E_3). The 3,6-dideoxyhexose can undergo a C5 epimerization catalyzed by CDP-3,6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase (E_{ep}) and a C4 reduction catalyzed by CDP-3,6-dideoxy-D-glycero-L-glycero-4-hexulose-4-reductase (E_{red}) to yield CPD-L-ascarylose. The 3,6-dideoxyhexose can also undergo other subsequent epimerization and/or stereospecific reductions to yield paratose, abequose and tyvelose. Two reductases, CDP-D-paratose synthase (encoded by the *rfbS* gene) from *Salmonella typhi* [37] and *S. enterica* [38] and CDP-D-abequose synthase (encoded by the *abe* gene) from *S. enterica* [39] have been identified. Finally, tyvelose epimerase, which is involved in the last step of the conversion of CDP-D-paratose to CDP-D-tyvelose, has been isolated from *S. typhi* and its structure has been reported [40]. Each of these 3,6-dideoxyhexoses is known to be components of cell walls and capsules of gram-negative and gram-positive bacteria.

Structural biology of the deoxyhexose pathway enzymes

Of the enzymes involved in the deoxyhexose pathways mentioned above, several X-ray crystal structures are available (Table 1). Interestingly, despite functional diversity and low sequence identity, many of the enzymes exhibit strikingly similar three-dimensional (3D) structures. The monomers fold into two domains: the N-terminal NAD(H) or NADP(H) cofactor-binding domain and the C-terminal NDP-sugar substrate-binding domain. The N-terminal domain consists of an alternating α/β motif known as a Rossmann fold, which is commonly associated with dinucleotide binding. The cleft between the two domains is the site of dinucleotide cofactor binding, sugar nucleotide binding and where catalysis occurs. These structural characteristics place the enzymes in the NDP-sugar modifying subfamily of the short chain dehydrogenase/reductase (SDR) protein family. This is a large and diverse family of proteins, requiring NAD(H) or NADP(H) as their cofactor to carry out a variety of reactions including reductions, oxidations, dehydrations or epimerizations.

Although members of the NDP-sugar modifying subfamily of the SDR family share amino acid sequence identities of only 15-30%, distinct sequence motifs allow for tentative functional assignment of residues. For instance, a common N-terminal glycine-rich region is a part of the nucleotide-binding region; this allows for the close packing of the cofactor to the protein backbone. Also, the conserved catalytic triad of Thr/Ser and Tyr-XXX-Lys allows one to build a model of catalysis for the initiating step (Figure 3). Mutagenesis and kinetic studies on SDR proteins has revealed the importance of the catalytic triad member Tyr, which acts as an active site base deprotonating the O4 hydroxyl of the substrate as the C4 position is oxidized by NAD/NADP. The catalytic

Lys lowers the pKa of Tyr by stabilizing it in its negatively charged state [32, 41]. The role of the Ser/Thr catalytic triad member is less clearly defined. It has been proposed to help orient the substrate in the active site and/or facilitate proton transfer [42-45]. The oxidation at the C4 position prepares the sugar for further reactions by acidifying the hydrogen atoms at positions 3 and 5. The 4-keto intermediate acts as a springboard for other SDR reactions.

Table 1. X-ray crystal structures of NDP-sugar modifying enzymes involved in 6- and 3,6-di-deoxyhexose pathways

Enzyme	Substrate	Product	Protein Family	Number of Species
GDP-D-mannose 4,6-dehydratase (GMD)	GDP-D-mannose	GDP-4-keto-6-deoxy-D-mannose	SDR	4
GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER)	GDP-4-keto-6-deoxy-D-mannose	GDP-L-fucose	SDR	1
dTDP-D-glucose 4,6-dehydratase (RmlB)	dTDP-D-glucose	dTDP-4-keto-6-deoxy-D-glucose	SDR	4
dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase (RmlC)	dTDP-4-keto-6-deoxy-D-glucose	dTDP-4-keto-6-deoxy-L-mannose	new epimerase family	1
dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD)	dTDP-4-keto-6-deoxy-L-mannose	dTDP-L-rhamnose	SDR	1
CDP-D-glucose 4,6-dehydratase (E_{OD})	CDP-D-glucose	CDP-4-keto-6-deoxy-D-glucose	SDR	1
CDP-D-tyvelose 2-epimerase (*tyv*)	CDP-D-paratose	CDP-D-tyvelose	SDR	1

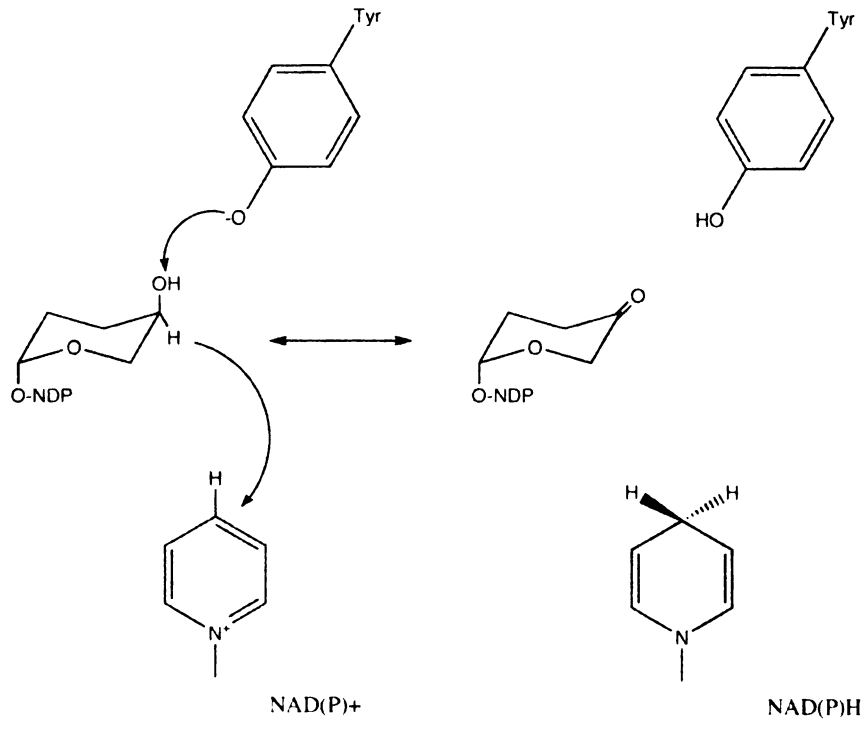


Figure 3. SDR model of catalysis

The SDR catalytic triad member Tyr deprotonates the O4 position of the sugar as a hydride is transferred from the C4 position of the sugar to NAD(P).

Nearly all the members of the SDR family exist as dimers where the monomers interact, forming a four-helix bundle at the interface involving 2 helices of each monomer. However, there are exceptions. RmlD exhibits a new Mg^{2+}-dependent dimerization mode and the GMDs seem to be following a new tetramerization pattern. So although we can make assumptions based on the highly conserved sequence motifs of some of these enzymes, we have already seen exceptions to the rule that makes the structure/function relationship of these enzymes an important topic to be studied. Furthermore, the fact that RmlC represents a new class of epimerases unrelated to the

SDR enzymes raises questions about other enzymes in the biosynthetic pathways of deoxyhexoses.

GDP-D-rhamnose biosynthesis

Rhamnose in P. aeruginosa LPS

The focus of this dissertation is on structure/function studies of the enzymes involved in the biosynthetic pathway of the NDP-activated 6-deoxyhexose, GDP-D-rhamnose, from the bacterium *P. aeruginosa.* This opportunistic pathogen causes infection in patients with impaired immune systems such as burn wound victims, cancer patients and especially those with cystic fibrosis (CF). Due to its intrinsic resistance to many antibiotics, infections by this bacterium are difficult to treat; therefore efforts have been made to understand the factors involved in host-bacterium interactions and the establishment of infection. The LPS, which plays a significant structural role in the outer membrane, contributes to the pathogenesis of *P. aeruginosa* [46]. It is involved in protecting the bacterium from phagocytosis [47] as well as serum-mediated killing [48].

As previously mentioned, the LPS of gram-negative bacteria is tripartite in nature: the hydrophobic lipid A region, the core oligosaccharide region and the O-antigen or O polysaccharide region (Figure 4). *P. aeruginosa* contains two variant forms of O-antigen known as A-band and B-band. The serotype-specific B-band is a heteropolymer composed of many different monosaccharides, while the common A-band LPS is a homopolymer of D-rhamnose arranged as trisaccharide repeat units linked $\alpha1\rightarrow2$, $\alpha1\rightarrow3$, $\alpha1\rightarrow3$ [8]. Interestingly, studies have shown that B-band O-antigen is absent or expressed in smaller amounts in chronic *P. aeruginosa* isolates from CF patients, while

the level of A-band O-antigen is maintained [8]. Furthermore, there is also a correlation between the presence of anti-A-band antibodies in CF patients with both lower pulmonary function and increased duration of *P. aeruginosa* infections [8]. Other organisms expressing the A-band O-antigen include two bacterial species known to be pathogenic in CF patients, *Burkholderia cepacia* and *Stenotrophomonas (Xanthomonas) maltophilia*, as well as *Pseudomonas syringae* pv. morsprunorum C28 and *P. syringae* pv. cerasi 435 [8]. As a result, the biosynthesis of A-band LPS has gained interest as a potential therapeutic target.

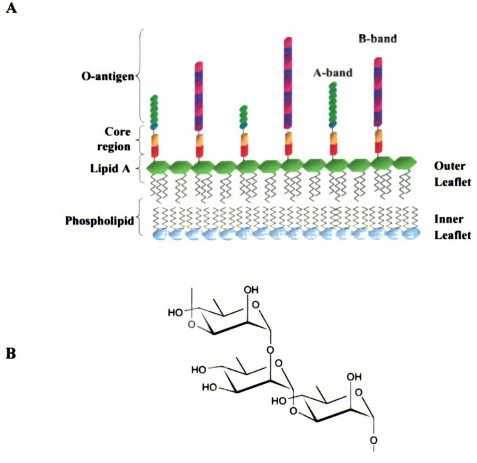


Figure 4. Lipopolysaccharide of *P. aeruginosa*[1]

(A) Schematic of the outer membrane and (B) chemical structure of A-band O-antigen of *P. aeruginosa*. Images in this dissertation are presented in color.

GDP-D-rhamnose biosynthetic pathway

The source of the D-rhamnose molecules found in *P. aeruginosa* A-band LPS is the nucleotide-activated GDP-D-rhamnose. Markovitz proposed the biosynthetic pathway to GDP-D-rhamnose in the 1960s [49]. The reaction proceeds through two steps: GDP-D-mannose 4,6-dehydratase (GMD) converts GDP-D-mannose to GDP-4-keto-6-deoxy-D-mannose, and the reductase (RMD) subsequently reduces the intermediate to GDP-D-rhamnose (Figure 5). The first step in the reaction is particularly important, as GMD is also part of the GDP-L-fucose pathway. Defects in GDP-L-fucose synthesis, particularly in GMD activity, have been linked to stem shoot development in plants [5]. In addition, deficiencies in the biosynthesis of GDP-L-fucose in humans have resulted in the rare immune disorder leukocytes adhesion deficiency type II (LADII) [50]. Characterization of GMD has been reported from bacterial sources *H. pylori* [4], *E. coli* [2], *Klebsiella pneumoniae* [1] and *A. thermoaerophilus* [3], plant source *Arabidopsis thaliana* (known as MUR1) [5], and mammalian sources porcine thyroid [6] and human [7, 51]. GMD amino acid sequences are quite similar and the molecular mass of the monomers generally lies between 40-55 kDa. Somoza *et al.* determined the 3D structure of GMD from *E. coli* and definitively confirmed the structural relationship of GMD to the NDP-sugar modifying subfamily of SDR enzymes [52]. However, GMDs from various species tend to differ in their multimerization. Like SDR members, dimeric structures have been reported for *E. coli* [52], *K. pneumonie* [1] and human GMDs [51]. Nevertheless, the more recently published structure of GDP-D-mannose 4,6-dehydratase MUR1 from *A. thaliana* [44] suggests it exists as a tetramer. Furthermore, there are reports that suggest

that *H. pylori* GMD is a tetramer [4] and that porcine thyroid GMD may be a hexamer [6].

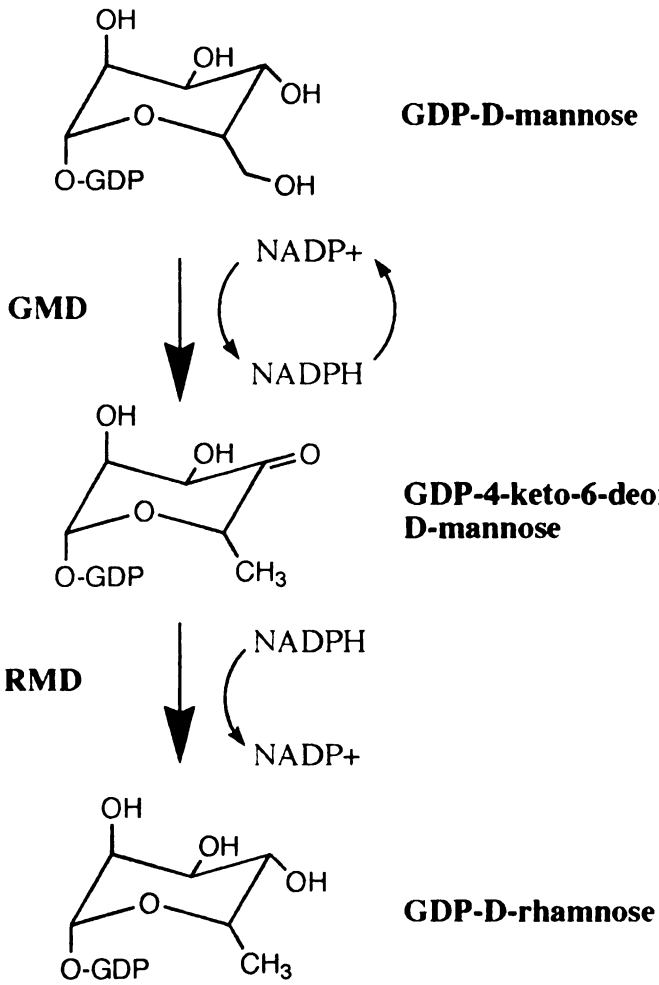


Figure 5. Biosynthetic pathway of GDP-D-rhamnose

GDP-D-mannose 4,6-dehydratase (GMD) catalyzes the conversion of GDP-D-mannose to the intermediate GDP-4-keto-6-deoxy-D-mannose. The 4-reductase (RMD) catalyzes the reduction of the intermediate to GDP-D-rhamnose.

A BLAST search with the amino acid sequence of *P. aeruginosa* GMD demonstrates that it most closely matches dTDP-glucose 4,6-dehydratases (like RmlB), RMDs and UDP-glucose 4-epimerases, which catalyze the interconversion of UDP-galactose and UDP-glucose. These enzymes are also members of the NDP-sugar modifying subfamily

of the SDRs and contain a catalytic triad that includes Ser/Thr and Tyr-XXX-Lys as well as a characteristic glycine-rich region at the N-terminus. Mutagenesis studies on *E. coli* GMD have confirmed the role of Ser/Thr-Tyr-Lys in catalysis [52]. Like the initial catalytic step of the SDR enzymes, Tyr deprotonates the O4 hydroxyl of the hexose ring while NADP oxidizes the C4 position forming the 4-keto intermediate. The ensuing dehydration step requires another active site base. Evidence suggests the second active site base is a conserved Glu, which effectively implies that a catalytic quartet is required for dehydratases [52]. To complete the action of oxidation and reduction, it has been demonstrated in a bacterial GMD that the mannose C4 hydride is transferred from NADPH back to the hexose C6 position to produce the 4-keto-6-deoxy intermediate product [53].

Much fewer studies have focused on the second step of the GDP-D-rhamnose biosynthetic pathway. In fact, RMD has only been characterized in the nonpathogenic, gram-positive bacteria *A. thermoaerophilus* [3] and in *P. aeruginosa* in a coupled reaction with *H. pylori* GMD [54]. The amino acid sequences of these RMDs indicate the molecular masses of the monomers are ~35 kDa. Currently there are no 3D structures available and the state of multimerization is unknown. A BLAST search with RMD demonstrates its relatedness to SDR members: GMDs, dTDP-glucose 4,6-dehydratases and UDP-glucose epimerases. The amino acid sequence of RMD most closely aligns with that of its biosynthetic pathway mate GMD (Figure 6). The conservation of the SDR residues is initially apparent: Gly-XX-Gly-XX-Gly and the catalytic triad of Ser and Tyr-XXX-Lys (corresponding to Gly9-XX-Gly12-XX-Gly15 and Thr126 and Tyr150-XXX-Lys154 of *P. aeruginosa* GMD). If RMD fits into the "functional-mold" of the

SDR enzymes, the catalytic triad member Tyr would be expected to activate the substrate by protonation facilitating the cofactor to reduce the 4-keto-6-deoxysugar to GDP-D-rhamnose. This reaction mechanism has been suggested for the related enzyme dTDP-6-deoxy-*lyxo*-4-hexulose reductase (RmlD), which is involved in dTDP-L-rhamnose biosynthesis [26]. It is interesting to note that one of the catalytic residues of GMD, Glu128 (GMD of *P. aeruginosa*), is not conserved across RMDs. In GMD catalysis, Glu128 is proposed to abstract a hydrogen atom from the hydroxyl group at C5 of the sugar. The C5 atom is not expected to be involved in the reduction as RMD functions by reducing the keto group at the C4 position.

[1] Figure courtesy of Dr. Joseph Lam, Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada.

Figure 6. The alignment of GMD and RMD amino acid sequences from various species

Conserved residues are highlighted in blue boxes; similar residues are highlighted in light blue.

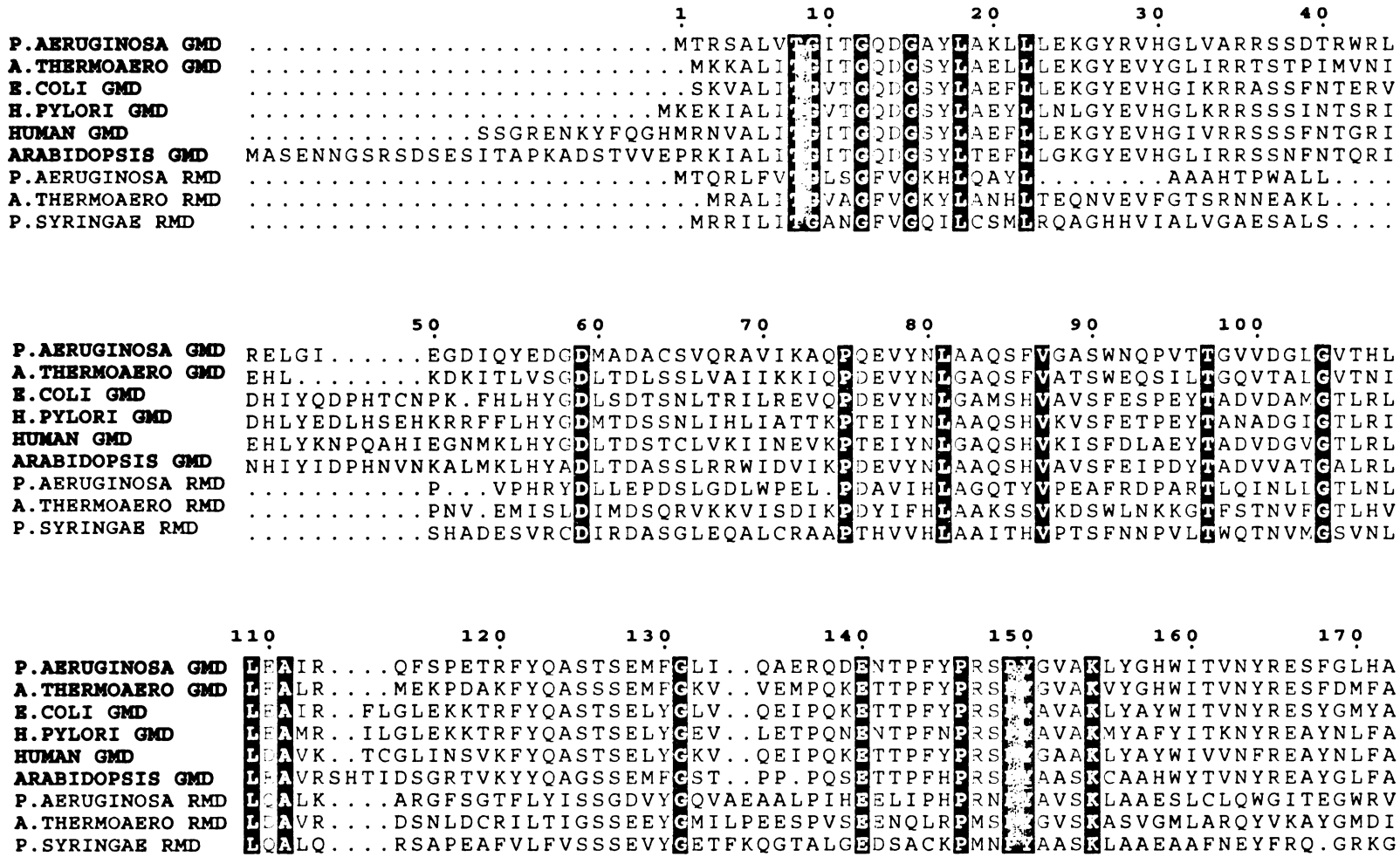
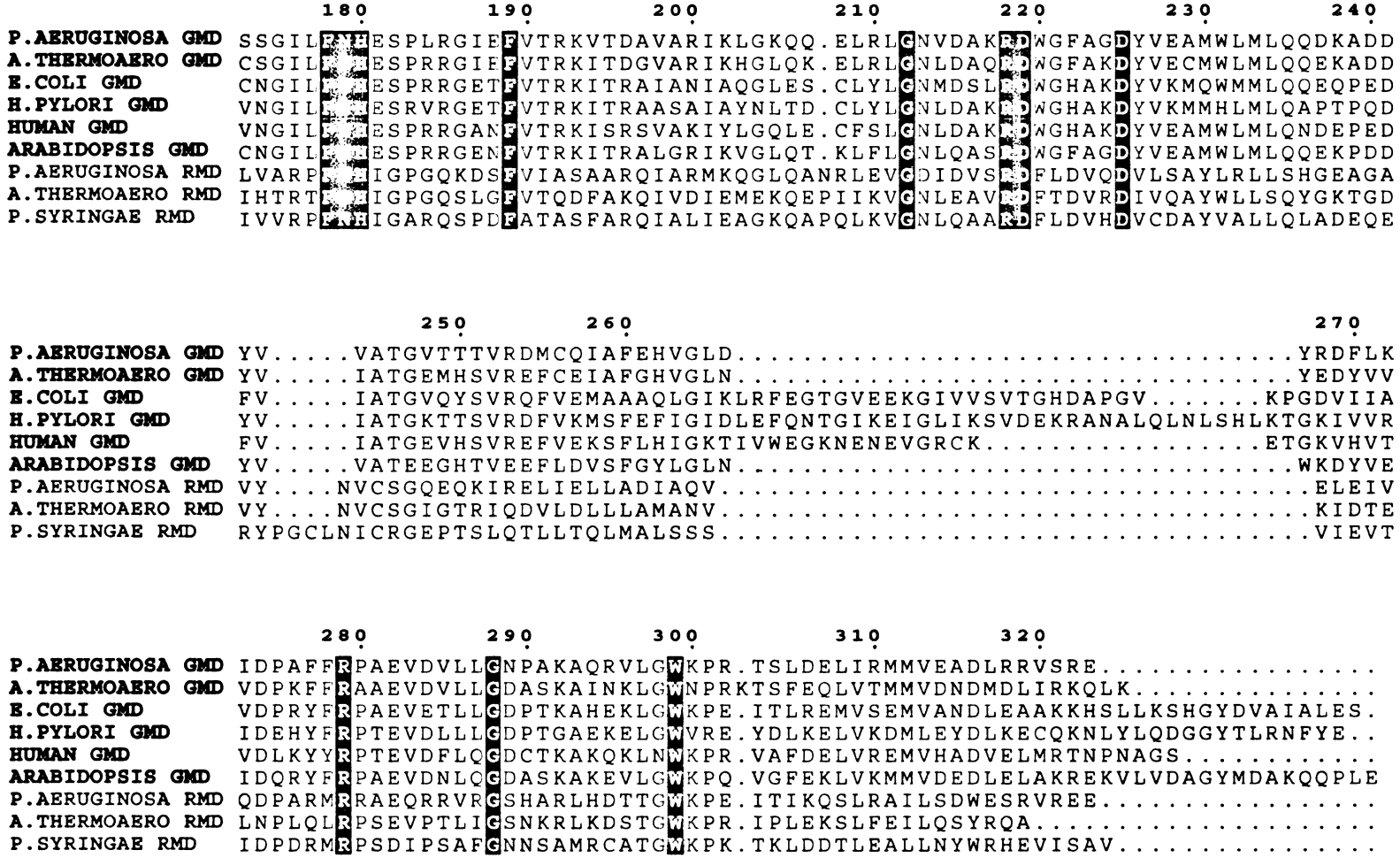

P.AERUGINOSA GMD
A.THERMOAERO GMD
E.COLI GMD
H.PYLORI GMD
HUMAN GMD
ARABIDOPSIS GMD
P.AERUGINOSA RMD
A.THERMOAERO RMD
P.SYRINGAE RMD

Figure 6. (cont'd)



CHAPTER 1: CRYSTAL STRUCTURE OF GMD[1]

1.1. Introduction

GDP-D-mannose 4,6-dehydratase (GMD) is a member of the NDP-sugar modifying subfamily of the SDR protein family. As a member of this subfamily, GMD binds its cofactor NADP(H) in the N-terminal portion of the molecule where a common glycine-rich region is present. Members share a handful of highly conserved residues that include a catalytic triad of Tyr-XXX-Lys and Ser/Thr that are all important for catalysis. Although the amino acid sequences of members of this subfamily are significantly different, the three dimensional structures are quite homologous. Closely related enzymes for which structures are known include UDP-glucose epimerase (GalE) from *E. coli* and human [45, 55-62], GDP-fucose synthetase (GFS) [63] and dTDP-D-glucose 4,6-dehydratase (dTGDH) [21, 42].

Somoza *et al.* [52] determined the three-dimensional structure of GMD from *E. coli* and confirmed the structural relationship of GMD to the NDP-sugar modifying subfamily of SDR enzymes. However their work suggested that active GMDs exist primarily as dimers. In contrast, Mulichak *et al.* [44] showed that the structure of the MUR1 isoform of GMD from *A. thaliana* was a tetramer where the NADP(H) binding site is intimately involved in creating the subunit interface. The question thus arises as to whether bacterial GMDs differ from eukaryotic GMDs in terms of their active oligomeric state and, perhaps, the regulation of enzyme activity. Reported here is the crystal structure of a GMD from the bacterium *P. aeruginosa* in the presence of the ligands NADPH and GDP. Our analysis again confirms the structural homology between the GMDs from all three species, but reveals that the GMD from the bacterium *P. aeruginosa* exists as a MUR1-

like tetramer. Based on the conservation of subunit interactions and sequence in GMDs, evidence suggests that the tetrameric form of this enzyme may be its active oligomeric state in prokaryotes and eukaryotes.

1.2. Experimental Procedures

1.2.1. Cloning and Expression of GMD. The *gmd* gene, originally isolated by Currie *et al.* [64], was amplified by PCR while incorporating a BamHI restriction site over the natural start codon and a HindIII restriction site over the natural stop codon. The new PCR fragment was ligated into the pQE30 vector (Qiagen), placing a 6x-His tag at the N-terminus. A glycerol stock of *E. coli* M15 cells harboring the recombinant plasmid was used to inoculate a starter culture of 50 ml Luria Bertani (LB) broth. Following cultivation at 37°C for 16 h, the starter culture was transferred to 1 L LB broth. Once an optical density at 600 nm (OD_{600}) of 0.5-0.6 was reached, the cells were induced with 0.25 mM IPTG and grown for an additional 16 h at room temperature before harvesting by centrifugation. Cells were re-suspended in Lysis Buffer (50 mM HEPES, 300 mM NaCl, 5 mM imidazole pH 8.0).

1.2.2. Fermentation of GMD. A low temperature glycerol stock of *E. coli* M15 cells harboring the recombinant plasmid was used to streak M9 plates containing 2.5% (w/v) glucose, 25 μg/ml kanamycin and 100 μg/ml ampicillin. After a growth period of about 16 h at 37°C, an individual colony was chosen to inoculate 5 ml of starter medium (M9 salts plus 2.5% (w/v) glucose, 2 mM $MgSO_4$, 25 μg/ml kanamycin, 100 μg/ml ampicillin). The starter culture was placed in an incubator for 12 h at 37°C at a shaker speed of 250 rpm. The culture was then transferred to a 250 ml Erlenmeyer flask containing 100 ml of growth medium (22 mM KH_2PO_4, 42 mM Na_2HPO_4, 8.3 mM NaCl,

18 mM NH_4Cl, 1 mM $MgSO_4$, 2.5% (w/v) glucose, 3 μM thiamine-HCl, 25 μg/ml kanamycin, 100 μg/ml ampicillin). The 100 ml culture was then placed back into the incubator at 37°C at a shaker speed of 250 rpm until an OD_{600} of between 2.0 and 3.0. The 100 ml culture was used to inoculate 1 L of fermentation medium (33 mM K_2HPO_4, 10 mM citric acid monohydrate, 21 mM H_2SO_4, 0.3 g ammonium iron III citrate, 25 μg/ml kanamycin, 100 μg/ml ampicillin, pH 7.0) in the fermentation vessel hooked to the BioFlow II Fermentor (New Brunswick Scientific). Using fed-batch fermentation, the conditions that produced the best protein yield consisted of growth for 12 h at 37°C in the fermentor followed by induction with 0.5 mM IPTG and a 12 h post induction growth period at 25°C. The dissolved oxygen was kept constant at 10%, pH at 7.0 and the max rpm at 750. Cells were harvested by centrifugation and re-suspended in Lysis Buffer.

1.2.3. Purification of GMD. Cells were lysed by sonication. The lysate was cleared by centrifugation at 10,000 x g for 20 min and loaded onto a column of 4 ml Ni-NTA resin (Qiagen). The column was washed with Wash Buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole pH 7.5) and protein was eluted with Elution Buffer (50 mM HEPES, 300 mM NaCl, 200 mM imidazole pH 7.5). Further purification was carried out by FPLC using a HiTrap™ Q anion exchange column (Amersham Pharmacia Biotech) equilibrated with Buffer A (20 mM Tris pH 8.5). Protein was eluted over a 40 ml salt gradient using Buffer B (20 mM Tris pH 8.5, 1 M NaCl). Fractions were analyzed by SDS-PAGE. Unless otherwise noted, protein samples for SDS-PAGE analysis were denatured using SDS and equal volumes were loaded onto a 4-12% acrylamide gradient gel. Those shown to be homogenous were pooled and concentrated to 10 mg/ml.

1.2.4. Crystallization. Purified GMD was used to set up crystallization screens (Crystal Screen 1 and 2 from Hampton Research) using the hanging drop vapor diffusion method. Prior to data collection, crystals were looped directly from the drop and flash-frozen in liquid propane.

1.2.5. Data Collection and Processing. X-ray diffraction data were collected to 2.2 Å on a MAR CCD detector at the Advanced Photon Source beamline 5-ID (DND), Argonne National Laboratory. During data collection the crystal was held at 100 K in a cryostream and radiation was used at a 1.0 Å wavelength. The data were processed using HKL version 1.6 software [65]. Data collection statistics are presented in Table 2.

1.2.6. Structure Determination and Refinement. The structure of GMD was determined by molecular replacement using the AMoRe programs [66] from the CCP4 suite [67], using a search model based on the known structure of MUR1 (PDB code 1N7H) [44]. Using a dimer as a search model, clear solutions for a tetramer in the asymmetric unit were obtained. CNS v1.1 [68] was used to perform several rounds of simulated annealing, followed by simple positional refinement and finally individual B-factor refinement. Manual building was carried out between each round of refinement using $2F_o$-F_c and F_o-F_c maps in CHAIN [69]. Throughout refinement non-crystallographic restraints were imposed on each of the four molecules in the asymmetric unit. The final model (R_{factor}=17.3, R_{free}=19.5) consists of a GMD tetramer, 693 water molecules and four molecules each of NADPH and GDP, despite the fact that no ligand or cofactor was added to the crystallization drop. There is clear electron density for residues 3-323 of 323 possible residues in each monomer. The final model has good

stereochemistry as determined by PROCHECK [70]. The refinement statistics are outlined in Table 2.

Table 2. GMD X-ray diffraction data and refinement statistics

Space group	$P3_221$
Resolution range (Å)	30.0-2.2
Cell parameters (Å, °)	a=b=125.7, c=220.0 α=ß=90, γ=120
Unique reflections	
Working set	110,070 (92.7%)
Test set	5,398 (4.9%)
Rmerge	3.9 (7.6)
Completeness (%)	98.3 (92.8)
Rfactor[a]	17.3
Rfree[b]	19.5
Residues	1282/1292
Water molecules	693
rmsd[c]	
Bonds (Å)	0.005
Angles (deg)	1.3
Dihedrals (deg)	23
Impropers (deg)	0.7
Ramachandran	
Most favored (%)	90.2
Allowed (%)	9.8

[a]R-factor = $\Sigma|F_O - F_C|/\Sigma F_O$, where F_O and F_C are observed and calculated structure factors, respectively. [b]R-free is the cross validation R-factor computed for the test set of reflections. [c]rmsd is the root-mean-square deviation.

1.3. Results and Discussion

1.3.1. Expression, Purification and Crystallization. A typical protein yield from the shake-flask method was 3-4 mg of purified protein per liter of cell culture. Fermentation increased the protein yield over the shake-flask method by 10-fold, producing about 30-40 mg of purified protein per liter of cell culture. The purification protocol described above produced protein of a quality that could be crystallized. SDS-PAGE gels of samples from a typical purification are shown in Figure 7. Protein after a Ni-NTA column step did not produce crystals, so the next step of purification was added. Bipyramidal crystals were grown in a drop containing equal parts protein and well solution (50% 2-methyl-2,4-pentanediol, 100 mM Tris pH 8.5, 200 mM $NH_4H_2PO_4$) and equilibrated against a reservoir of well solution (Figure 8). The resulting crystals (0.4 x 0.4 x 0.3 mm) belong to the trigonal space group $P3_221$ (a=b=125.7 Å, c=220.0 Å, α=ß=90°, γ=120°) with four molecules in the asymmetric unit.

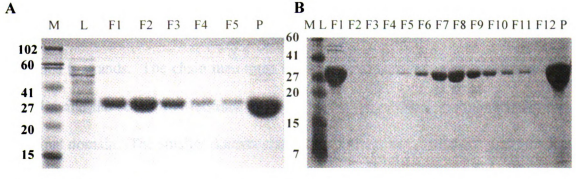


Figure 7. SDS-PAGE of samples from a typical GMD purification

(**A**) load (L), fractions (F) and pool (P) from the Ni-NTA column; (**B**) load (L), fractions (F) and pool (P) from the anion exchange column. Molecular weight marker (M) standards shown in kDa to the left of each gel.

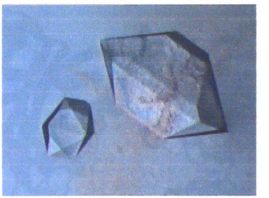


Figure 8. Crystals of GMD

Crystals were grown in 50% 2-methyl-2,4-pentanediol, 100 mM Tris pH 8.5, 200 mM $NH_4H_2PO_4$, 5 mM GDP-D-mannose, 2 mM NADP. The larger crystal is ~0.8 mm across at the largest point and the smaller crystal is ~0.35 mm across at the largest point.

1.3.2. Overall Structure. As is characteristic of members of the NDP-sugar modifying SDR subfamily, the GMD monomer folds into two domains: the N-terminal cofactor-binding domain and the C-terminal substrate-binding domain (Figure 9). The larger N-terminal domain consists of an alternating α/β motif of seven β-strands in the order 3-2-1-4-5-6-7 flanked by α-helices, yielding a modified Rossmann fold, an element generally associated with dinucleotide binding. Common in this subfamily is the transition into the C-terminal domain following β6, providing an extension of an α-helix and two β-strands. The chain then turns back to the N-terminal portion, adding another α-helix and the seventh N-terminal β-strand before returning to and completing the C-terminal domain. The smaller domain consists largely of three α-helices with two sets of mixed parallel and anti-parallel β-sheets. The cleft that exists between the two domains is the site of dinucleotide cofactor and nucleotide-sugar substrate binding and is where catalysis occurs.

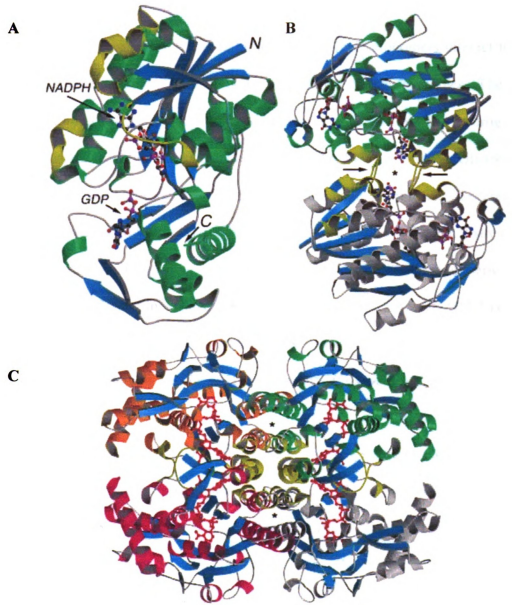


Figure 9. Overall Structure of GMD

Ribbon representation of the GMD monomer (A), the GMD dimer highlighting the tetramer interface (B), and the GMD tetramer (C). For each drawing β-strands are colored in blue, and the areas colored in yellow represent the regions involved at the tetramer interface. In the active site of the GMD monomer and dimer, the cofactor NADPH and substrate GDP are shown in backbone representation. Only the cofactor NADPH is included in the tetramer for clarity. Arrows in the GMD dimer indicate the RR loops, and the asterisk highlights the close approach of the cofactors from each monomer to one another. The asterisks in the tetramer highlight the four-helix bundle, the dimerization mode common among SDR enzymes.

GMD crystallizes with four molecules in the asymmetric unit, forming a homotetramer (Figure 9). Within the tetramer, two GMD monomers interact to form a four-helix bundle at the interface involving α4 and α5 of each monomer. The contacts between the long helices are mainly hydrophobic in nature, with a few hydrogen bonds between Asn163 and the main chain carbonyl of Arg147*, as well as Glu166 to both Tyr145* and Arg147* (*denotes residue from opposite monomer). Several closely related enzymes, as well as nearly all the members of the SDR family, share the four-helix bundle dimerization mode. GMD deviates from the typical homodimeric structure as two of these "SDR dimers" sandwich together to form a tetramer. GMD is only the third enzyme of the NDP-sugar modifying SDR subfamily to be observed as a tetramer next to its plant homolog MUR1 [44] and the recently published structure of tyvelose epimerase [40].

The tetramerization of GMD results in adjoining of the cofactor binding sites at the interface such that the adenosyl phosphate moieties of bound NADPH molecules fall within 7 Å and 7.5 Å respectively for GMD and MUR1. Both GMD and MUR1 share a flattened ellipsoidal shape (~ 95 x 75 x 60 Å and ~100 x 74 x 57 Å respectively). In contrast, tyvelose epimerase, which oligomerizes along the analogous surface, forms a less compact tetramer (~100 x 110 x 60 Å). Consequently, the tetramer interface of tyvelose epimerase is less extensive and the adenine rings of bound NAD molecules are separated by 11 Å.

1.3.3. Cofactor Binding Site. There is well-ordered electron density in the active site to unambiguously place an NADPH molecule in each monomer (Figure 10). The nonplanar electron density corresponding to the nicotinamide ring suggests that the

reduced form of the cofactor is present. NADPH is found in the cofactor-binding sites of the crystals when no cofactor was added to the crystallization conditions, suggesting tight binding of the cofactor throughout purification. The tight binding of reduced form of the cofactor is consistent with the characterization of both *E. coli* GMD and *Y. pseudotuberculosis* CDP-D-glucose 4,6-dehydratase, which reveals that they bind their reduced cofactor more tightly than they bind the oxidized form [52, 71]. This is clarified with the catalytic mechanism (discussed in detail in section 1.3.5.), which displays an intramolecular hydride transfer from the C4 carbon on the mannose ring to the C6 carbon, effectively regenerating NADP in the process and allowing it to remain bound [53]. However, NADPH is seen bound in the cofactor-binding site even with an abundance of NADP added to the crystallization conditions. One potential explanation for the presence of the reduced form of the cofactor is the inadvertent reduction of the dinucleotide by the buffer. Tris buffer has demonstrated reducing power in the case of the UDP-glucose epimerase/NAD complex in the presence of UMP [72]. A schematic diagram of the interactions of NADPH with the surrounding residues is shown in Figure 11.

NADPH binds in an extended conformation with the pyrophosphate positioned adjacent to the positive dipole of α1 in a manner consistent with dinucleotide binding in a Rossmann fold. The characteristic glycine-rich fingerprint sequence of Gly9-XX-Gly12-XX-Gly15 in this region allows for the close packing of the cofactor to the protein backbone. This leads to the formation of hydrogen bonds of the pyrophosphate to the main chain amide nitrogen atoms of Gln13 and Asp14. Further hydrogen bonding occurs from the pyrophosphate to Ser85.

Figure 10. Electron density at the GMD active site
NADPH (yellow) and GDP (orange) with residues surrounding the active site.

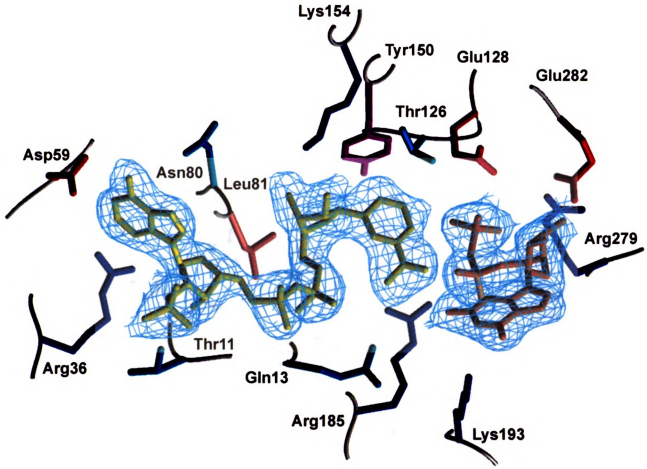
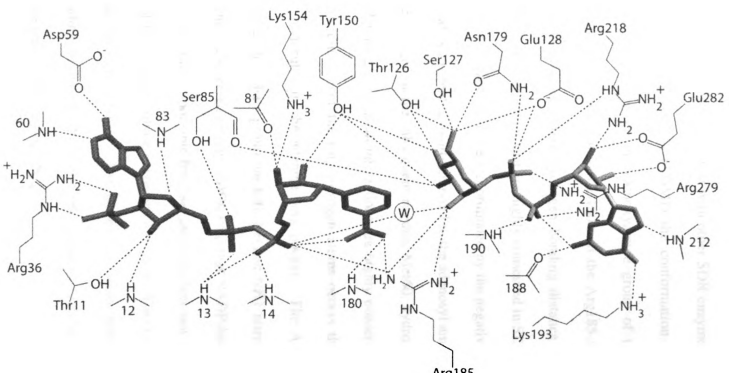


Figure 11. GMD active site interactions

Schematic diagram of the potential hydrogen bonds between the cofactor NADPH and substrate GDP-D-mannose with GMD. Mannose was modeled in using NADPH/GDP-D-rhamnose/MUR1 complex (1N7G) as a guide.



The nicotinamide moiety is bound in the *syn* conformation and may be stabilized in this orientation by hydrogen bonding between the carboxyamide nitrogen and the pyrophosphate. This conformation is consistent with other SDR enzymes, allowing a B-side hydride transfer during catalysis. In GMD the conformation may be further stabilized by hydrogen bonds between the carboxyamide group of the nicotinamide moiety to the main chain amide nitrogen of His180 and the Arg185 side chain. The nicotinamide ribose hydroxyls are within hydrogen bonding distance to the catalytic residues Tyr150 and Lys154, interactions that are highly conserved in SDR enzymes.

The adenine ring of the cofactor is largely coordinated by the negatively-charged side chain Asp59, which makes a potential hydrogen bond to the adenosyl amino group, while the main chain amide nitrogen of the subsequent residue, Met60, hydrogen bonds to the N1 nitrogen of the adenine group. Both interactions are highly conserved among SDR enzymes. Further coordination of the ring nitrogen atoms occurs through hydrogen bonding with water molecules in the adenine ring pocket. The Arg36 side chain hydrogen bonds to the 2' phosphate group via both NE and NH1 nitrogen atoms. The position of this arginine is conserved among GMDs and other NADP-binding SDRs such as GFS, and is suggested to be responsible for discriminating between NADP and NAD [63]. The hydroxyl of the adenosyl ribose is coordinated by the Thr11 side chain and by the main chain amide nitrogen of Gly12. The main chain amide nitrogen of Ala83 is within hydrogen bonding distance of the ribose ring oxygen. Similar interactions with the adenosyl ribose are seen in MUR1 and GFS [44, 63].

Most intriguing about the cofactor binding site is the involvement of the adenosyl end in the tetramer interface, a feature also seen in the plant homolog MUR1 [44]. The RR

loop, a segment of nine residues (Arg35-Arg43), stretches into the neighboring monomer making not only protein-protein interactions, but also contacts to the neighboring cofactor. Protein-protein interactions include Arg35 hydrogen bonding to Ser85* and Glu188*, Ser38 to Trp42*, and Arg43 to both Ser37* and Ser38*, as well as to the main chain carbonyl of Ser38*. Protein-cofactor interactions include hydrogen bonding of Ser37 to the neighboring adenosyl 2'-phosphate via the main chain nitrogen atom as well as the side chain hydroxyl group. Arg35 also hydrogen bonds to the pyrophosphate of the neighboring NADPH, further tethering the tetramer together. Furthermore, the Arg36 side chains from each monomer are involved in a parallel stacking arrangement between the two adenine rings. They also help to coordinate the 2'-phosphate of their own NADPH in a way similar to that observed in GFS and MUR1.

1.3.4. Substrate Binding Site. The electron density corresponding to the substrate reveals that GDP is present in the active site of each monomer (Figure 10). Attempts to replace GDP with GDP-D-mannose by adding excessive amounts of the natural substrate to the crystallization conditions have been unsuccessful. However, mannose could be modeled in using the NADPH/GDP-D-rhamnose/MUR1 (PDB code 1N7G) complex as a guide. Hydrogen bonding and van der Waals interactions were optimized while holding the GDP moiety constant. Binding interactions of the GDP-D-mannose with GMD are depicted in Figure 11. The orientation of the O6 hydroxyl of the GDP-D-mannose was chosen based on its ability to make potential hydrogen bonds to Thr126, Ser127 and Glu128. In the crystal structure, the presence of a water molecule close to the GDP-D-rhamnose C5 atom that makes hydrogen bonds to Ser127 and Glu128 further supports this proposed rotamer. Based on this model, both catalytic residues Thr126 and Tyr150

could hydrogen bond to the hexose O4 hydroxyl. Further coordination of the O2 and O3 hydroxyls and O5 of the hexose ring occurs through the side chains of Arg185, Tyr150, the main chain carbonyl oxygen of Ser85 and through water mediated hydrogen bonds.

The GDP moiety, which is completely buried in the small domain, is in the *syn* conformation. This is an unusual conformation for this nucleotide and may be related to substrate recognition. GDP is stabilized in this orientation by an intramolecular hydrogen bond between the guanine amine nitrogen atom and the phosphate. The pyrophosphate bridge of the GDP moiety abuts against the positive dipole of α6, in a manner similar to that of dinucleotide binding in a Rossmann fold. This allows for hydrogen bonding to the main chain amide nitrogen of Val190. Further contacts to the GDP moiety include Asn179, Lys193, Arg218, Arg279 and Glu282, all highly conserved residues among GMDs. The ability of Arg279 to hydrogen bond to both phosphates is a feature conserved among several SDRs, although the exact position of the arginine is not conserved. Furthermore, the hydrogen bonding of Glu282 to the ribose hydroxyls is also observed in UDP-binding SDRs [43, 57].

1.3.5. Catalytic Mechanism. The proposed catalytic mechanism for GMD occurs via a three-step process (Figure 12) [53]. The first step involves a hydride transfer from the mannose C4 to the C4 of the cofactor. An active site base deprotonates the O4 hydroxyl to yield the GDP-4-ketomannose intermediate (b). The second step involves the elimination of water from C6 to yield the GDP-4-keto-5,6-ene intermediate (c). In the final step, the C5-C6 double bond is reduced as the cofactor returns the hydride to the C6 position yielding GDP-4-keto-6-deoxy-D-mannose (d). The SDR enzymes share only a few conserved residues that include the catalytic triad Tyr-XXX-Lys and Ser/Thr that

are important in catalysis (Figure 13) [73]. Additionally, a highly conserved Glu among dehydratases has proven to be important in the dehydration step [42, 52]. Mutagenesis and kinetics studies on *E. coli* GMD [52] as well as other SDR enzymes have supported the roles of Tyr, Lys, Ser/Thr and Glu in the reaction mechanism. The GMD reaction resembles that of other nucleotide hexose dehydratases, such as CDP-D-glucose 4,6-dehydratase and dTDP-D-glucose 4,6-dehydratase that have been more extensively studied [27, 29-32, 71].

Figure 12. The mechanism proposed for GDP-D-mannose 4,6-dehydratase
The process can be divided into three steps: oxidation of GDP-D-mannose (A) to produce the 4-keto intermediate (B), dehydration into the 4-keto-5,6-ene intermediate (C), and, finally, reduction to yield GDP-4-keto-6-deoxy-D-mannose (D).

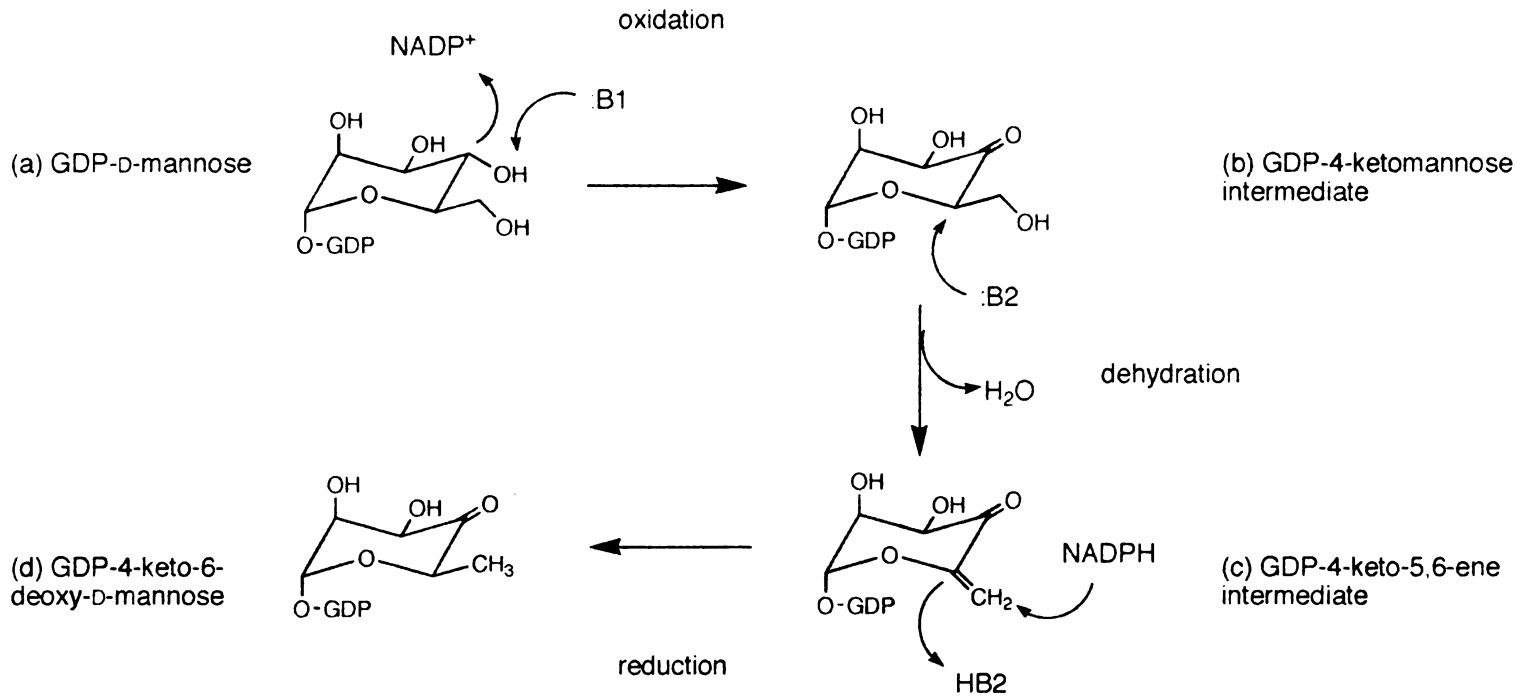


Figure 13. The alignment of GMD amino acid sequences from various species

Conserved residues are highlighted in blue; similar residues are highlighted in light blue. Secondary structural elements are marked according to *P. aeruginosa* GMD. Residues involved in the dimer interface are boxed in copper, residues involved in the tetramer interface are boxed in yellow. An asterisk marks catalytic residues Thr126, Tyr150, and Lys154.

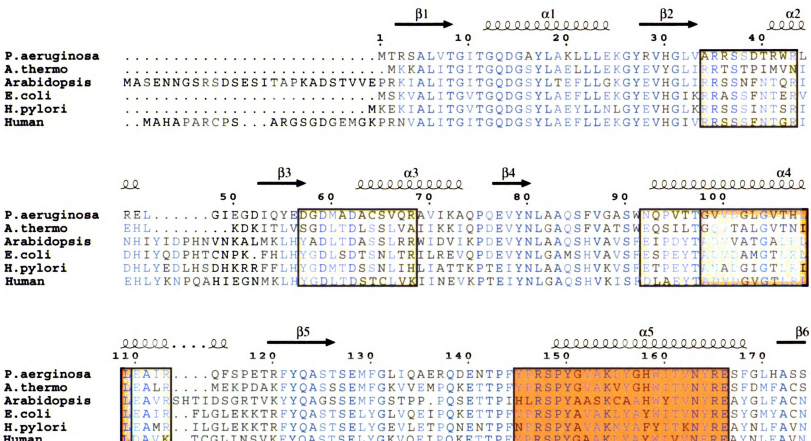
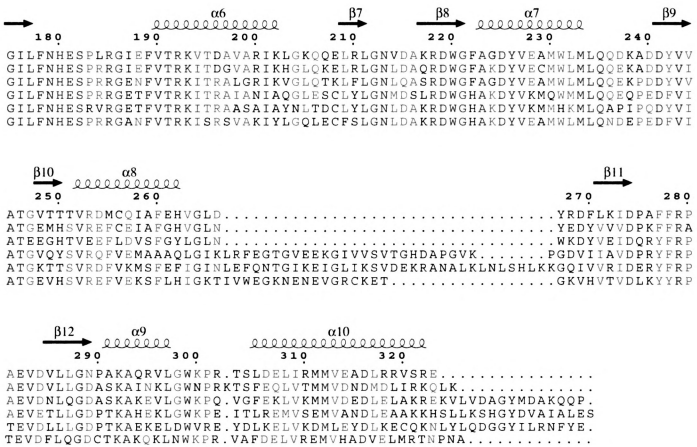

β1
α1
β2
α2
P.aeruginosa
A.thermo
Arabidopsis
E.coli
H.pylori
Human
β3
α3
β4
α4
β5
α5
β6

Figure 13. (cont'd)


α6
β7
β8
α7
β9
180
190
200
210
220
230
240
P.aeruginosa GILFNHESPLRGIEFVTRKVTDAVARIKLGKQQELRLGNVDAKRDWGFAGDYVEAMWLMLQQDKADDYVV
A.thermo GILFNHESPRRGIEFVTRKITDGVARIKHGLQKELRLGNLDAQRDWGFAKDYVECMWLMLQQEKADDYVI
Arabidopsis GILFNHESPRRGENFVTRKITRALGRIKVGLQTKLFLGNLQASRDWGFAGDYVEAMWLMLQQEKPDDYVV
E.coli GILFNHESPRRGETFVTRKITRAIANIAQGLESCLYLGNMDSLRDWGHAKDYVKMQWMMLQQEQPEDFVI
H.pylori GILFNHESRVRGETFVTRKITRAASAIAYNLTDCLYLGNLDAKRDWGHAKDYVKMMHKMLQAPIPQDYVI
Human GILFNHESPRRGANFVTRKISRSVAKIYLGQLECFSLGNLDAKRDWGHAKDYVEAMWLMLQNDEPEDFVI
β10
α8
β11
250
260
270
280
P.aeruginosa ATGVTTTVRDMCQIAFEHVGLD.....YRDFLKIDPAFFRP
A.thermo ATGEMHSVREFCEIAFGHVGLN.....YEDYVVVDPKFFRA
Arabidopsis ATEEGHTVEEFLDVSFGYLGLN.....WKDYVEIDQRYFRP
E.coli ATGVQYSVRQFVEMAAAQLGIKLRFEGTGVEEKGIVVSVTGHDAPGVK.....PGDVIIAVDPRYFRP
H.pylori ATGKTTSVRDFVKMSFEFIGINLEFQNTGIKEIGLIKSVDEKRANALKLNLSHLKKGQIVVRIDERYFRP
Human ATGEVHSVREFVEKSFLHIGKTIVWEGKNENEVGRCKET.....GKVHVTVDLKYYRP
β12
α9
α10
290
300
310
320
P.aeruginosa AEVDVLLGNPAKAQRVLGWKPR.TSLDELIRMMVEADLRRVSRE.....
A.thermo AEVDVLLGDASKAINKLGWNPRKTSFEQLVTMMVDNDMDLIRKQLK.....
Arabidopsis AEVDNLQGDASKAKEVLGWKPQ.VGFEKLVKMMVDEDLELAKREKVLVDAGYMDAKQQP.
E.coli AEVETLLGDPTKAHEKLGWKPE.ITLREMVSEMVANDLEAAKKHSLLKSHGYDVAIALES
H.pylori TEVDLLLGDPTKAEKELDWVRE.YDLKELVKDMLEYDLKECQKNLYLQDGGYILRNFYE.
Human TEVDFLQGDCTKAKQKLNWKPR.VAFDELVREMVHADVELMRTNPNA.....

Dimer Interactions

Tetramer Interactions

In the oxidation step, the NADP cofactor abstracts a hydride from the C4 hydroxyl of the mannose ring. The mannose modeled in the active site of GMD reveals a distance of 3.5 Å between the C4 of the nicotinamide ring and C4 of the hexose. The angle formed by the N1-C4 atoms of the nicotinamide ring and C4 of the hexose is 92°. Both values are consistent with the positioning of the cofactor and substrate in other related enzymes such as MUR1 [44], SQD1 [43] and human GalE [45], where the distances range from 3.4 Å to 3.7 Å and the angles fall in the range of 81° to 96°.

To complete the oxidation step, an active site Tyr removes a proton from the O4 hydroxyl forming the 4-keto intermediate. The catalytic Tyr150 in GMD is in proper position (distance of 2.7 Å) relative to the mannose model for it to directly attack the O4 hydroxyl. Early studies of *E. coli* GalE [55] and dTGDH [21] showed that the distance of Tyr to the O4 hydroxyl is too great for it to act directly as the base. The Ser/Thr catalytic triad member had been proposed to act as a proton shuttle to complete the oxidation step. However, more recent crystallographic studies of human GalE [45], dTGDH [42], SQD1 [43] and MUR1 [44], showed that Tyr is within proper hydrogen bonding distance to directly attack the O4 hydroxyl. The Ser/Thr instead may orient the substrate in the active site and/or facilitate proton transfer. The presence of hydrogen bonds between the O4 and O6 hydroxyl and the Thr126 hydroxyl suggests that both these roles may be accomplished. To further facilitate the oxidation step, the catalytic Lys may stabilize Tyr in its negatively charged state. Studies suggest that Lys lowers the pKa of Tyr [32], which is normally between 9-12. The measured pKa of Tyr in *E. coli* dTGDH [32] and *E. coli* GalE [41] is 6.4 and 6.1 respectively. The distance between the phenolic oxygen of Tyr150 and the amide nitrogen of Lys154 is 4.4 Å, a distance too far for hydrogen

bonding, but within the range of electrostatic interactions to effectively lower the pKa of Tyr150.

The 4-keto intermediate acts as a springboard to other SDR reactions. In the case of GMD, the ketone functionality serves to acidify the proton at C5 and permits the dehydration from C6 to form the GDP-4-keto-5,6-ene intermediate. The presence of the "ene" intermediate has recently been detected in the homologous dTGDH reaction [29]. For dehydration to occur, another active site base must be present to abstract a proton from the C5 position. Studies of *E. coli* GMD [52] and dTGDH [30] showed that a glutamic acid might fulfill this requirement. The corresponding Glu128 side chain in GMD is within 3.6 Å of the C5 carbon of the mannose model, a position that would enable it to deprotonate C5. To complete the dehydration reaction, the C6 hydroxyl must be protonated by an active site acid. An Asp residue has been proposed based on structural analysis of dTGDH [42] and supported by mutagenesis experiments [27, 30]. The corresponding GMD residue, Ser127, is within 2.8 Å of the modeled position of the hexose O6 hydroxyl, and may assume a similar role. Alternatively, Glu128 is within 2.6 Å, suggesting the possibility of this side chain playing a dual role in the dehydration step, acting as both a general base and a general acid, as has been suggested in dTGDH and MUR1 [30, 44]. Whereas the dehydration mechanism described here is the step-wise water elimination mechanism as seen in D135N and D135A mutants of dTGDH, dehydration may also occur through a concerted mechanism as seen in wild type dTGDH [31]. Further kinetic studies would need to be completed to determine which mechanism of dehydration GMD actually utilizes.

The final step of the GMD reaction involves a hydride transfer from NADPH back to the hexose C6 position. The distance between the nicotinamide C4 atom and the C4 and C6 atoms of the hexose moiety (3.5Å and 3.8Å respectively) suggests that only modest rotation of the hexose ring would be required to complete the hydride transfer. Interestingly, because NADP is regenerated, the cofactor may remain bound through each catalytic cycle. To finalize the reduction step, an active site acid is required for proton addition to the C5 position of the hexose. Proposed residues to fulfill this role based on structural analysis of dTGDH include the catalytic Tyr, Glu or Asp [42]. Of the corresponding residues in GMD, Tyr150 and Ser127 (aligning with Asp) are >4.4 Å to the C5 position of the hexose model. Glu128 is 3.6 Å away, but would move even further with the rotation of the hexose ring towards the nicotinamide ring. However, Thr126 of the catalytic triad and Asn179 are positioned such that they may be able to fulfill the role as the general acid to complete the reaction.

1.3.6. Structural Comparisons. The secondary structural elements between *P. aeruginosa* GMD and MUR1 superimpose well with a root mean square deviation (rmsd) of 1.2 Å over Cα atoms. The secondary structural elements between *P. aeruginosa* GMD and *E. coli* GMD do not superimpose as well (rmsd of 3 Å) because *E. coli* GMD has no substrate or cofactor bound in the active site. The main difference between the three is an area of disorder present in MUR1 residues 76-81 and *E. coli* GMD residues 35-55 which corresponds to a region that is highly variable in size and sequence among GMDs (Figure 13). This stretch in *P. aeruginosa* GMD is shorter and well ordered, forming a short helix (α2) between β2 and β3 of the Rossmann fold. Immediately preceding this region is a Gly-XX-ArgArg sequence that is conserved among all GMDs sequenced thus far

except for *P. aeruginosa* GMD. In contrast, *P. aeruginosa* GMD exhibits an arginine shift resulting in the sequence Gly-XXX-ArgArg. The two positively charged arginine residues mark the beginning of the RR loop that closes over the adenosyl phosphate end of the cofactor, are important in cofactor binding and are involved in the tetramer interface for *P. aeruginosa* GMD and MUR1. The RR shift causes an interesting rearrangement of interactions. In MUR1 the first arginine of the sequence, Arg60, adopts a parallel stacking arrangement with Arg60* of the neighboring monomer, simultaneously packing against the adenine ring and coordinating the 2'phosphate. In GMD the second arginine of the sequence Arg36 also adopts a parallel stacking arrangement with Arg36* of the neighboring monomer. However, due to the shift by two residues, the side chain is oriented almost perpendicular to the adenine ring (Figure 14). Despite the rearrangement in this region, Arg36 in *P. aeruginosa* GMD still maintains the electrostatic interactions to coordinate the 2'phosphate. Based on the role that the cofactor plays in the tetramer interface of *P. aeruginosa* GMD and MUR1, cofactor binding might also be expected to assist in ordering part of the RR loop in *E. coli* GMD and may be essential for tetramer formation.

Other differences between *P. aeruginosa* GMD and *E. coli* GMD include the positioning of the loop between the β4 strand and the α4 helix. This loop in the apo *E. coli* GMD occupies a portion of the cofactor-binding site and would have to move as much as 6 Å to make room for the cofactor. Secondly, the smaller substrate-binding domain in *E. coli* GMD adopts a more open conformation. The difference is apparent when comparing the rmsd of 2.4 Å with *P. aeruginosa* GMD for the C-terminal domain, as opposed to just 1.8 Å for the N-terminal domain. In addition to the more open

conformation of its C-terminal domain, *E. coli* GMD has an extended loop between the α8 helix and the β11 strand. The loop, which inserts an additional 27 residues compared to *P. aeruginosa* GMD, is variable in size among GMDs.

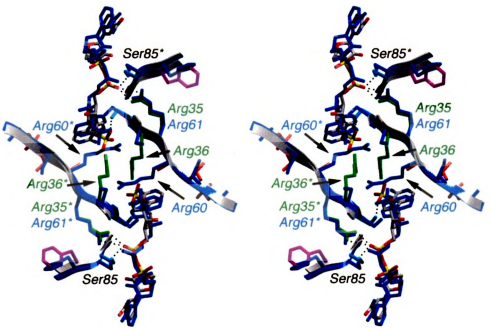


Figure 14. RR loop of GMD and MUR1

Stereoview of the superposition of the RR loops from GMD (highlighted in color) and MUR1 (blue). GMD side chains are labeled in green, MUR1 side chains are labled in blue, asterisks denote side chains from opposite monomer.

One of the intriguing features about *P. aeruginosa* GMD is its oligomeric state, as it deviates from the canonical homodimeric structures seen in most other related enzymes of the NDP-sugar modifying subfamily of the SDRs. The structures of all other members of this subfamily, including *E. coli* GMD, have been observed as dimers, with the exception of the previously mentioned tetrameric tyvelose epimerase [40] and ADP-L-glycero-D-mannoheptose-6-epimerase, which is a pentamer [74]. *P. aeruginosa* GMD and *A. thaliana* MUR1 can be seen as a dimer of canonical SDR dimers, which then

generates a new set of subunit interactions. An important subsequent question is whether a significant number of tetramer interactions between GMD and MUR1 are conserved. As previously mentioned, the RR loop of residues Arg35-Arg43 that is so intimately involved in the tetramer interface and the cofactor binding sites, makes protein-protein interactions as well as protein-cofactor interactions to the neighboring monomer. The Arg35 to Ser85* and Ser37 to Arg43* hydrogen bonds are conserved in GMD and MUR1. Furthermore, these residues are highly conserved among the GMD sequences. Also within the RR loop is hydrogen bonding between Ser38 and Arg43*. Although the residue corresponding to Ser38 is an Asn in MUR1, the interaction is conserved. The sequences of several GMDs reveal that in most cases, a Ser is present in this position. Away from the cofactor binding site overlap, Asp62 hydrogen bonds to the amide nitrogen atoms of Val95* and Thr96*. Asp62 is highly conserved among GMDs. One of the more interesting interactions involves residue Arg68, which is moderately conserved across several GMD sequences. This residue is involved in hydrogen bonding to the main chain carbonyl oxygen of Asn92*, an interaction also seen in MUR1. The same Arg in GMD and MUR1 extends towards the diagonally related monomer to hydrogen bond to Glu110 and Arg113, both highly conserved residues among GMD sequences.

1.3.7. Conclusions. In summary, the GDP-D-mannose 4,6-dehydratase MUR1 isoform from *Arabidopsis thaliana* was shown to be a tetramer, while the first GMD structure to be determined, *E. coli* apo-GMD, was observed as a dimer. This raised the question as to whether or not prokaryotic and eukaryotic GMDs differed in oligomeric state. We have determined the structure of *P. aeruginosa* GMD with NADPH and GDP bound in the active site and found it to exist as a tetramer. The tetramer arises from the

dimerization of the canonical dimer seen for most members of the SDR superfamily, but in a manner where the cofactor binding sites closely interact across the new interface. The residues involved in the tetramer interactions are well conserved between the prokaryotic GMD and the eukaryotic MUR1. Moreover, a high degree of sequence conservation among the residues within the tetramer interface is also observed across a broad range of GMDs. These observations suggest that the tetramer may be a more common oligomeric state for GMDs than previously thought.

[1]Portions of Chapter 1 were previously published, Webb *et al.*, (2004) *Protein Science.*

CHAPTER 2: CRYSTALLIZATION AND INITIAL X-RAY ANALYSIS OF RMD

2.1. Introduction

Genes coding for a particular bacterial polysaccharide are usually found in clusters. In the case of both *A. thermoaerophilus* and *P. aeruginosa*, the *rmd* gene is located adjacent to the *gmd* gene. The corresponding RMD enzymes have been characterized in both the gram-negative bacterium *P. aeruginosa* and gram-positive bacterium *A. thermoaerophilus* [3, 54]. A sequence alignment of the two RMDs shows 33% identity and 54% similarity between pairs. A BLAST search with RMD shows significant similarities to GDP-mannose dehydratases, dTDP-glucose dehydratases, UDP-galactose epimerases and GDP-fucose synthetases, all members of the NDP-sugar modifying subfamily of the SDRs. No RMD analogue has yet been found in the human genome using *P. aeruginosa* RMD as a probe, which is in agreement with the fact that humans lack rhamnosylation. Thus, the second step in GDP-D-rhamnose biosynthesis appears to be a potential target for enzyme inhibition to prevent the synthesis of D-rhamnose-containing LPS molecules in pathogenic bacteria. Structural information would aid in the design of inhibitors. Since no structure is currently available for RMDs, efforts were made to determine the X-ray structure of RMD from *P. aeruginosa*. As this proved to be a difficult task, an ancillary effort was made to determine the structure of RMD from *A. thermoaerophilus*, which has shown to be similar in function to the *P. aeruginosa* RMD.

Reported here are the cloning, expression, purification, crystallization and initial X-ray analysis of the RMD crystals from both bacterial species. The amino acid sequence of RMD most closely aligns with its biosynthetic pathway mate GMD. If the residues align at the tertiary level, then we would expect that the core structure would mimic that

of GMD, or other SDR enzymes, since the three-dimensional structures tend to be highly similar. Crystallographic analysis should allow us to see if RMD meets our expectations or breaks from the mold of the typical SDR protein. Preliminary data suggests that *P. aeruginosa* GMD may be an appropriate search model for the molecular placement method of phase determination using data from *A. thermoaerophilus* RMD crystals, indicating our hypothesis may be true.

2.2. Experimental Procedures

2.2.1. Cloning and Expression of RMD. The *rmd* gene from *P. aeruginosa*, originally isolated by Lightfoot and Lam [75], was amplified by PCR while incorporating a BamHI site over the natural start codon and a SalIII site over the natural stop codon. The new PCR fragment was ligated into the pQE30 vector (Qiagen), placing a 6x-His tag at the N-terminus. A glycerol stock of *E. coli* M15 cells harboring the recombinant plasmid was used to inoculate a starter culture of 50 ml Luria Bertani (LB) broth. Following cultivation at 37°C for 16 h, the starter culture was transferred to 1 L LB broth. Once an optical density at 600 nm of 0.5-0.6 was reached, the cells were induced with 0.25 mM IPTG and grown for an additional 16 h at room temperature before harvesting by centrifugation. Cells were re-suspended in Lysis Buffer (50 mM HEPES, 300 mM NaCl, 5 mM imidazole pH 8.0).

The *rmd* gene from *A. thermoaerophilus*, originally isolated by Kneidinger *et al.* [3], was also amplified by PCR while incorporating a BamHI site at the N-terminus and a KpnI site at the C-terminus. The PCR fragment was then ligated into the pQE80 vector (Qiagen), which placed a 6x-His tag at the N-terminus. The new plasmid was

transformed into *E. coli* BL21 DE3 cells. The same protocol as stated above was used to express the protein in shake flasks.

2.2.2. Purification of RMD. The same protocol as in section 1.2.3 was used to purify RMD resulting in the N-terminally His-tagged RMD from *P. aeruginosa* (PaRMD) or the N-terminally His-tagged RMD from *A. thermoaerophilus* (AtRMD).

2.2.3. Crystallization of RMD. Purified PaRMD or AtRMD at 10 mg/ml were used to set up crystallization screens using the microbatch method with the Impax I-5 robot (Douglas Instruments). A total of 198 conditions were screened by combining 1 μl of purified protein with 1 μl precipitating solution. Crystallization conditions were refined using the hanging drop vapor diffusion method. Prior to data collection, the PaRMD crystals were treated with 0.1% glutaraldehyde, briefly transferred to a cryoprotectant solution (10% PEG 10K, 100 mM Hepes pH 7.6, 7.5% glycerol, 2.75 M sodium formate, 5 mM GDP-D-mannose and 2 mM NADP) and flash-frozen in liquid nitrogen. The AtRMD crystals were flash-frozen in liquid nitrogen directly from the crystallization drop.

2.2.4. Cryoprotection Techniques. Through personal correspondence with Terese Bergfors [76], a technique for streak-seeding for cryoconditions was created. Purified PaRMD at 10 mg/ml was used to set up crystallization screens (Crystal Screen 1 and 2 from Hampton Research) using the hanging drop vapor diffusion method. A 1:1 mix of protein and precipitating solution in a drop size of 4 μl was placed on the cover slip and allowed to equilibrate over 500 μl of precipitating solution for one day. All drops were then streak-seeded with their respective protein crystals in an effort to determine possible cryoprotectants. This process was accomplished by touching a cat whisker to the side of

a previously obtained protein crystal in order to gather nuclei. Then the cat whisker was streaked through the drops on the cover slips whether there was precipitation or not. The drops were then monitored for their ability to grow crystals. Those conditions that supported any crystal growth, whose precipitating solution contained a component used for cryoprotecting, provided a starting point for optimizing the final cryosolution.

A technique was devised from an idea that stemmed from a method listed in the CCP4 bulletin board (www.ccp4.ac.uk) titled "Summary of oil and cryo-protectant combo". After crystal(s) were obtained in a hanging drop vapor diffusion screen, an o-ring greased with vacuum grease was placed on the cover slip around the drop containing the crystal(s). The resulting reservoir was then filled with paraffin oil, covering the drop. The crystal(s) were then directly looped out of the drop and flash-frozen, providing a layer of oil as the cryoprotectant.

2.2.5. Data Collection and Processing. X-ray diffraction data from both the PaRMD and AtRMD crystals were collected on a MAR CCD detector at the Advanced Photon Source beamline 5-ID (DND), Argonne National Laboratory. During data collection the crystals were held at 100 K in a cryostream and radiation was used at a 1.0 Å wavelength. The data were processed using XDS software [77]. The AtRMD crystal diffraction data were used to calculate a self-rotation function using the AMoRe programs [66] from the CCP4 suite [67]. See Table 3 for statistics.

2.2.6. Preparation of Phasing Models. The secondary structure of *P. aeruginosa* and *A. thermoaerophilus* RMD were predicted using the PSIPRED secondary structure prediction method [78] using the PSIPRED server [79]. The fold recognition server 3D-PSSM was used to predict tertiary structure by threading [80]. The tertiary structure was

also predicted using SWISS-MODEL [81], which uses comparative modeling. The coordinates were checked visually and compared to other structures of SDR members. Based on the secondary and tertiary structure prediction as well as sequence alignments, molecular replacement models were constructed using *P. aeruginosa* GMD (PDB code 1RPN), *E. coli* GMD (PDB code 1DB3) and *Arabidopsis* MUR1 (PDB code 1N7H) as starting models. All differing residues were changed to alanines and portions were removed in which 1) amino acid sequences differed significantly and/or 2) the predicted structure differed. The main structural feature of the SDR enzymes, the Rossmann fold, was maintained.

Table 3. RMD X-ray diffraction data

	PaRMD	AtRMD
Space group	P422 (potentially $P42_12$, $P4_122$, $P4_12_12$, $P4_222$, $P4_22_12$, $P4_322$, $P4_32_12$)	P1
Unit cell parameters (Å, °)	a=b=182, c=250 α=β=γ=90	a=46.88, b= 55.74, c=79.24, α=72.54, β=82.95, γ=75.61
Resolution range (Å)	30.0-3.7	30.0-1.8
No. observed reflections	209,657	215,748
No. unique reflections	44,486	64,129
Completeness (%)	93.9 (96.7)*	96.5 (95.2)*
Rmerge (%)	7.5 (21.8)*	10.6 (48.1)*
Average $I/\sigma(I)$	5.8 (1.2)*	13.4 (3.7)*
V_M (Å Da^{-1})	2.46	2.66
Molecules per asymmetric unit	12	2
Estimated solvent content (%)	50.1	53.8

*Indicates statistic from highest resolution shell.

2.3. Results and Discussion

2.3.1. Expression and Purification. RMD from both *P. aeruginosa* and *A. thermoaerophilus* were expressed in the pQE system as N-terminally His-tagged protein (PaRMD and AtRMD, respectively). Both constructs were readily purified to 90-95% by Ni-NTA column, and further purified to 95-98% by an anion exchange column. A typical protein yield from the PaRMD preparations was 9-12 mg of purified protein per liter of cell culture; the yield for AtRMD was less, only 2-3 mg of purified protein per liter of cell culture. SDS-PAGE gels of samples from a typical purification of PaRMD are shown in Figure 15 and AtRMD samples are shown in Figure 16. The gene products appear to run on the SDS-PAGE gels as a 35 kDa protein, which closely matches the molecular weight based on the amino acid sequence of 34.8 kDa for PaRMD and 35.9 kDa for AtRMD.

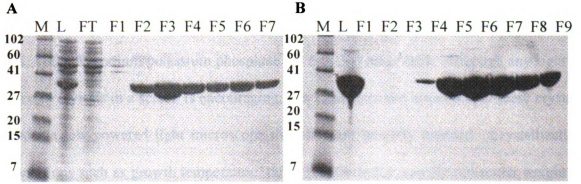


Figure 15. SDS-PAGE of samples from a typical PaRMD purification

(**A**) load (L), flow through (FT), wash (W) and fractions (F) from the Ni-NTA column; (**B**) load (L) and fractions (F) from the anion exchange column. Molecular weight marker (M) standards shown in kDa to the left of each gel. PaRMD runs as expected at ~35 kDa.

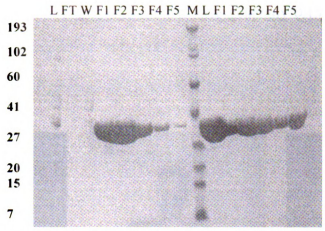


Figure 16. SDS-PAGE of samples from a typical AtRMD purification

The left-hand side of the gel shows the load (L), flow through (FT), wash (W) and fractions (F) from the Ni-NTA column; the right-hand side of the gel shows the load (L) and fractions (F) from the anion exchange column. Molecular weight marker (M) standards shown in kDa to the left of the gel. AtRMD runs as expected at ~35 kDa.

2.3.2. Crystallization. The initial screening of crystallization conditions at room temperature for PaRMD resulted in protein crystal clusters as shown in panels A, B and C of Figure 17. The crystals in panels A and B were obtained in the following conditions: 50:50 mix of protein sample and well solution where the protein sample contained 10 mg/ml protein, 5 mM GDP and 2 mM NADP and the well solution contained 10% PEG 8K, 100 mM sodium/potassium phosphate pH 6.0, 200 mM NaCl. Although any sight of protein crystals in a screen is encouraging, it is clear from the inspection of these crystals under a low-powered light microscope that they are severely twinned. Crystallization conditions such as growth temperature, the pH of the buffer and the molecular weight of the PEGs were modified to find the optimal crystal growing condition. Single crystals were obtained simply by replacing 5 mM GDP with the substrate analog 5 mM GDP-D-mannose to the protein sample. Examples of these crystals can be seen in panels D and E. The crystals in panel D are hexagonal in shape and about 0.2 x 0.2 x 0.1 mm in size.

Panel E shows thick rod-shaped crystals that were generally 0.15 mm in thickness and 0.4-0.5 mm in length.

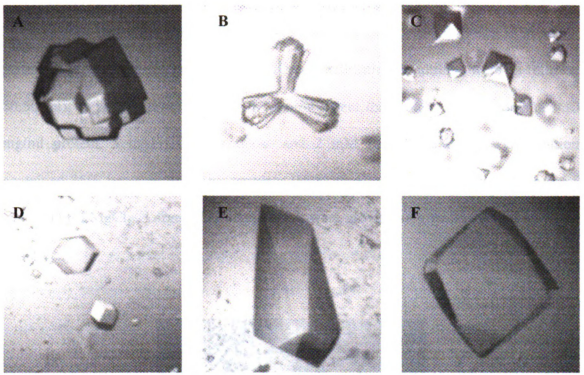


Figure 17. Crystals of PaRMD

Conditions giving crystals in panels A and B were optimized to yield single crystals shown in panels D and E by the addition of GDP-mannose instead of GDP. The conditions to grow the crystals in panel C were optimized by changing the buffer and the molecular weight of PEG to yield single crystals like that shown in panel F.

The PaRMD crystals in panel C of Figure 17 were obtained in a different screening condition: 50:50 mix of protein sample and well solution where the protein sample contained 10 mg/ml protein, 5 mM GDP-D-mannose and 2 mM NADP and the well solution contained 12% PEG 20K and 100 mM MES pH 6.5. By optimizing the PEG molecular weight and using a different buffer, single, non-twinned crystals were obtained in the following conditions: 50:50 mix of protein sample and well solution where the

protein sample contained 10 mg/ml protein, 5 mM GDP-D-mannose, 2 mM NADP and the well solution contained 19% PEG 10K and 100 mM HEPES pH 7.5. An example of these bipyramidal-like crystals, which were generally 0.4 x 0.4 x 0.4 mm in size, is shown in panel F.

A microbatch screen for crystallization conditions using AtRMD resulted in small, plate-like crystals shown in Figure 18. The crystallization conditions were as follows: 50:50 mix of protein sample and well solution where the protein sample contained 10 mg/ml protein, 5 mM GDP-D-mannose and 2 mM NADPH and the well solution contained 35% pentaerythritol propoxylate (PEP) (5/4 PO/OH), 100 mM MES pH 6.5 and 200 mM $MgCl_2$. Larger plate-like crystals were grown using the hanging drop vapor method.

Figure 18. Crystals of AtRMD
Small, plate-like crystals of AtRMD obtained in a microbatch screen.

The PaRMD crystals tended to grow within a week whereas the AtRMD crystals tended to take 3-4 weeks to grow. Reproducibility was a problem not only with protein originating from different purifications, but also with protein from the same batch. In addition, the crystals were prone to crumbling and dissolving, necessitating an arduous

search for a suitable cryoprotectant and/or cryoprotection technique that would not damage the crystals.

2.3.3. Cryoprotection. Generally before X-ray data collection, crystals are briefly transferred to cryoprotected mother liquor then flash-frozen to minimize radiation damage during the experiment. The cryoprotectant ideally prevents the formation of ice and maintains the crystallographic integrity of the crystal during the freezing process. The selection of the cryoprotectant(s) as well as the percent composition involves some trial and error. The PaRMD crystals grew in PEGs, which are known ice-preventing agents. The X-ray data of the crystals that were frozen directly from the drop revealed the formation of ice by the observed powder diffraction rings or "ice rings" on the X-ray diffraction screen. Data from crystals that had been transferred to a drop with a higher percentage of the respective PEG revealed poor-quality diffraction, potentially indicating poor cryoconditions. Thus an assay of several cryoprotectants, concentrations and a test of mixtures were required to determine the best cryocondition. To reduce the evaluation time it took to find the appropriate cryoprotectant, a screen of crystallization conditions was streak-seeded. First, a crystal screen is prepared and allowed to sit over night. Then a cat whisker is stroked over the surface of a previously obtained crystal to dislodge and collect the nuclei. Finally each drop of a screen is streaked with the cat whisker. The crystal trays are then monitored for any crystal growth, whether large or small, twinned or single, needles or chunks. If crystals appear, then it is assumed that the components of the precipitating solution would be more helpful than harmful to the crystal lattice. With any luck, one of the components is a cryoprotectant, giving a good starting point in the search for the appropriate cryosolution. Nevertheless streak-seeding PaRMD for

cryoconditions yielded no positive results, so several of the more common cryoprotectants were tested (i.e. sugars, MPD, ethylene glycol, glycerol and lower-molecular-weight PEGs). However, in each case, the crystals crumbled before flash-freezing.

Paraffin oil was also investigated as a suitable cryoprotectant, since it does not require the transfer of the crystal from one drop to another, which could disturb the crystal packing. A previous study has shown that crystals grown in PEG solutions are stripped of the thin film of mother liquor when passed through oil [83]. So a low-tech technique was created that would minimize manual manipulation of the PaRMD crystals. The cover slip on which the crystals were grown was turned drop side up. A greased o-ring was placed over the 4 μl drop and the reservoir was filled with paraffin oil. The crystal was then directly loop out of the drop, swiped through the oil and flash-frozen in liquid nitrogen. This method has the added advantage of providing extra handling-time as it protects the drop from drying out. Even though it seemed to be a less invasive procedure, the crystals were irreversibly damaged.

In the end, the best cryocondition for the PaRMD crystals was an initial treatment with glutaraldehyde, then the use of a combination of the same molecular weight PEG in which the crystals were grown, along with glycerol and sodium formate as well as GDP-D-mannose and NADP. Glutaraldehyde is known to cross-link proteins through the amino group of lysine, adding stability to fragile crystals for data collection. The final conditions served to improve the quality of the diffraction by reducing spot twinning and mosaicity.

The AtRMD crystals also proved to be very fragile. However, mesh litholoops (Protein Wave Corp.) provided support for the thin plates during freezing. Since the crystals grew in a high percentage of PEP, a known cryoprotectant, the crystals were frozen directly from the crystallization drop. Fortunately, this proved to be a suitable cryoprotectant and no further optimization was necessary.

2.3.4. Data Collection and Processing. A complete data set was collected from one of the bipyramidal PaRMD crystals, like the one shown in panel F of Figure 17. These crystals diffracted to 3.7 Å, while all other PaRMD crystals diffracted to 8-10 Å at best. Although the signal-to-noise ratio is 1.2 in the final resolution shell (≥2 is desired), the completeness (96.7%) and R_{merge} (21.8%) are respectable (see Table 3). These crystals belong to the P422 space group; it is unclear based on the systematic absences if the unit cell possesses symmetry elements that would place it in a higher symmetry space group. The number of molecules in the asymmetric unit was therefore estimated based on the current working space group (P422), unit cell parameters (a=b=182 Å, c=250 Å, $\alpha=\beta=\gamma=90°$) and molecular weight (34,825 Da) using the Matthews probability calculation [84]. The median Matthews coefficient (V_M) is 2.52 $Å^3$/Da with the vast majority of proteins falling into the 2-3.5 $Å^3$/Da range; the median solvent content is 47% with the vast majority of proteins falling in the 35-65% range [85]. For PaRMD a Matthews coefficient of 2.46 $Å^3$/Da yields 12 molecules in the asymmetric unit with a solvent content of 50.1%. However, certainly within reason is a V_M of 2.96 or 2.11 $Å^3$/Da, which would allow for 10 or 14 molecules respectively in the asymmetric unit with a solvent content of 58.4 or 41.8% respectively.

Though the resolution of the PaRMD data was high enough to determine phases, the data would not provide much detail beyond the backbone trace. While attempts were made to grow crystals in different conditions with the hope obtaining a different space group, these hopes were not realized. Since it is difficult to give up on crystals that have already been obtained, attempts were made to save the crystals and improve the resolution by post-crystallization treatments. Two separate annealing procedures were used: macromolecular crystal annealing (MCA) [86] and in situ annealing [86, 87]. MCA was able to improve the diffraction of the PaRMD crystals from 8-10 Å to 6-7 Å. In situ or flash-annealing was able to help decrease the ice-rings, but had little affect on the diffraction of the PaRMD. A "last resort" procedure, as described by the author of the paper in which the procedure was presented, was fast desiccation of protein crystals [88]. Although it was reported that poorly diffracting crystals showed "spectacular" improvement of diffraction with this procedure, the same results did not occur with any of the PaRMD crystals. This was the point at which efforts were directed at the RMD homolog from *A. thermoaerophilus.*

Up until the point of X-ray data collection, the RMD protein from *P. aeruginosa* and *A. thermoaerophilus* behaved in a similar fashion. Expression and purification were comparable, even the crystals had fragility in common. However, once a single frame of X-ray diffraction data was collected on the AtRMD crystals, the difference was readily apparent, as they diffracted to 1.8 Å. A complete data set was collected on the AtRMD crystals (see Table 3 for statistics). The crystal falls into the P1 space group and have a much smaller unit cell than that of the PaRMD crystals. As a result, based on a Matthews coefficient of 2.66 $Å^3$/Da, the number of molecules per asymmetric unit is 2, with a

solvent content of 53.8%. A self-rotation function indicates that the molecules in the asymmetric unit are related by a non-crystallographic twofold axis, suggesting AtRMD exists as a dimer.

2.3.6. Initial Phasing Results. Structure determination by the molecular replacement method using coordinates from various homologous structures was initially attempted with data from the PaRMD crystals. With the use of the programs AMoRe [66] and Phaser [89] from the CCP4 suite [67], no clear solution was evident in any of the possible P422 space groups. On the other hand, using a dimer as a search model based on the known structure of GMD from *P. aeruginosa* (PDB code 1RPN), a solution for a dimer in the asymmetric unit was obtained using the Phaser program from the CCP4 suite [67]. Currently efforts are underway to obtain experimental phases for comparison to those determined by molecular replacement.

2.3.7. Conclusions. To summarize, the process of crystal structure determination of RMD from *P. aeruginosa* and *A. thermoaerophilus* has been initiated. The cloning, expression, and purification were successful, yielding plenty of protein pure enough for crystallization trials. Single crystals were obtained of protein from both bacterial species and X-ray diffraction data has been collected. While PaRMD crystals diffracted weakly to only 3.7 Å, AtRMD crystals diffracted to 1.8 Å. Full structure determination by the molecular replacement method using the coordinates from GMD from *P. aeruginosa* is in progress. Our hope is that the AtRMD structure will serve as a successful search model for the PaRMD data, allowing a comparison of the two structures.

CHAPTER 3: FUNTIONAL CHARACTERIZATION OF GMD AND RMD[1]

3.1. Introduction

Rhamnose is a deoxyhexose found in glycoconjugates of bacteria and plants but not in humans. L-rhamnose is the more common isoform, found in cell walls and capsules of a variety of bacteria. On the other hand, D-rhamnose has mainly been described as a constituent of the LPS of gram-negative bacteria such as plant pathogens *X. campestris* [9], and human pathogens, for example *P. aeruginosa* [11], *H. pylori* [10] and *Campylobacter fetus* [90]. D-rhamnose molecules are arranged as repeating trisaccharide units in the A-band LPS O-antigen of *P. aeruginosa.* The LPS molecules of this opportunistic pathogen contribute to its virulence in immunocompromised patients including those with cystic fibrosis, cancer and burn-wounds. Interestingly, the same D-rhamnose polysaccharide structure seen in A-band LPS of *P. aeruginosa* has also been detected in other opportunistic pathogens *B. cepacia* and *S. maltophilia* that are problematic in cystic fibrosis patients [8].

The source of D-rhamnose comes from the nucleotide activated GDP-D-rhamnose, which is made in two steps: GDP-D-mannose 4,6-dehydratase (GMD) first converts GDP-D-mannose to GDP-4-keto-6-deoxy-D-mannose, which is subsequently reduced GDP-D-rhamnose by the reductase (RMD). The genetics of GDP-D-rhamnose biosynthesis in *P. aeruginosa* has been extensively studied. Lightfoot and Lam determined that the *P. aeruginosa gmd* gene coded for a protein that functions as a GDP-D-mannose dehydratase based on amino acid homology to other dehydratases. In addition, paper chromatography experiments showed the conversion of ^{14}C-GDP-D-mannose to GDP-D-rhamnose from supernatants of an A$^-$ strain carrying the *gmd* gene on

a plasmid [91]. Rocchetta *et al.* confirmed the role of *P. aeruginosa* RMD in the pathway from work that revealed that *rmd* knockout mutants lacked A-band O-antigen [92]. Amino acid homology of RMD shows that it is similar to enzymes that modify sugars at the C4 and C6 positions. Maki *et al.* verified the role of *P. aeruginosa* RMD in the conversion of the 4-keto-6-deoxy intermediate to GDP-D-rhamnose by co-expressing the *Helicobacter pylori gmd* and *P. aeruginosa rmd* genes in *Saccharomyces cerevisiae* [54]. However, GMD from *P. aeruginosa* has not yet been functionally characterized. Furthermore, the entire pathway has only been functionally characterized in *A. thermoaerophilus*, where D-rhamnose is a constituent of the surface layer glycoprotein of this bacterium [3].

In this study we developed a capillary-electrophoresis (CE) -based enzymatic assay to confirm the involvement of the enzymes GMD and RMD from *P. aeruginosa* in the biosynthesis of GDP-D-rhamnose. To supplement the X-ray analysis experiments (Chapter 2), assays were also completed on RMD from *A. thermoaerophilus.* We demonstrate dehydratase activity in GMD and reductase activity in both GMD and RMD. Furthermore, we present a method for the enzymatic synthesis of GDP-D-rhamnose, an important glycobiological building block.

3.2. Experimental Procedures

3.2.1. Cloning, Expression and Purification of GMD and RMD. Procedures outlined in Section 1.2.1. and 1.2.3. were used to produce purified GMD. Likewise, procedures stated in Sections 2.2.1. and 2.2.2. were used to produce purified RMD from both *P. aeruginosa* and *A. thermoaerophilus.* (Unless otherwise noted, discussion will center on RMD from the bacterial source *P. aeruginosa.*)

3.2.2. Determination of the Oligomeric State by Gel Filtration. A HiPrep 16/60 Sephacryl S-200 High Resolution gel filtration column (Amersham Biosciences), calibrated with high and low molecular weight calibration kits (Amersham Biosciences), was used to determine the oligomerization of both GMD and RMD. A 1 ml sample was prepared using protein in 20 mM Tris pH 8.5 and 300 mM NaCl with or without 10 mM GDP-D-mannose or 5 mM NADP. Samples were applied to the column and protein elution was detected at 280 nm. Molecular weights of the samples were estimated based on the calibration curve of K_{av} [(Ve-Vo)/(Vt-Vo)] vs. log molecular weight (where Ve is the elution volume for the protein, Vo is the void volume and Vt is the total bed volume).

3.2.3. Assay of GMD and RMD activity. The standard reaction buffer was 40 mM Tris-HCl (pH 7.5) and 10 mM $MgCl_2$ with a total reaction volume of 35 to 60 μL. Typically, the initial GMD reactions contained 1.0 μM GDP-D-mannose and 0.1 mM or 0.01 mM NADP. The reaction was initiated by the addition of 1 to 10 μg of purified protein and incubated at 37°C. The completion of the GMD reaction was judged by CE analysis (~1-2 h). For the coupled assay, 1 to 10 μg of RMD was added directly to the reaction tube along with 0.1 mM or 0.01 mM NADPH. For the sequential assay, GMD was removed by filtration using a microcon YM-10 filter cartridge; the filtrate containing the labile 4-keto intermediate was used as a substrate for the RMD/NADPH reaction. For the determination of the pH optimum for GMD, the following reaction buffers were used: MES (pH 5, 5.5, 6.5), Tris (pH 7.0, 7.5, 8.0) and Bis-Tris propane (pH 9.0, 10.0).

3.2.4. Analysis of Reaction Products by Capillary Electrophoresis. CE analysis was performed with a P/ACE MDQ Glycoprotein System (Beckman Coulter) with UV detection. A 75 μm x 50 cm bare silica capillary was used with the UV detector mounted

at 50 cm of the capillary. The CE analysis was run in 25 mM sodium tetraborate (pH 9.5) at 25°C, and sample was introduced by pressure injection for 8 sec. The separation was performed at 22 kV, and the sugar-nucleotide substrate and product were detected at 254 nm. Peak integration was performed using the Beckman 32 Karat Software.

3.2.5. Purification of Sugar Nucleotide Product via HPLC. A preparative scale enzymatic reaction containing 30 μmol of GDP-D-mannose and 1.5 mg of purified GMD was incubated for 2 h at 37°C in 40 mM Tris pH 7.5 and 10 mM Mg_2Cl. Protein was subsequently removed from the completed reaction by unltrafiltration through a Centriplus YM-3 cartridge (Millipore). The filtrate was subjected to an Econo-Pac High Q anion exchange column over a 0-500 mM triethylammonium bicarbonate buffer (pH 8.0) gradient. Fractions that contained GDP-D-rhamnose were pooled, acidified to pH 4.8 with AG 50W-X4 resin (Bio-Rad) and lyophilized in preparation for NMR analysis.

3.2.6. NMR Analysis of Reactions Product. The lyophilized sample containing GDP-D-rhamnose was resuspended in 160 μl of 99% D_2O and analyzed by NMR spectroscopy. All spectra were acquired using a Varian Inova 500 MHz spectrophotometer equipped with a Z-gradient triple resonance (^1H, ^{13}C, ^{31}P) probe. The experiments were performed at 15°C with suppression of the water resonance. The methyl resonance of acetone was used as an internal reference at δ_H 2.225 ppm and δ_C 31.07 ppm. The COSY, TOCSY, HMQC and ^{31}P HSQC experiments were used to assign the resonances [93].

3.3. Results

3.3.1 Protein Expression, Purification and Determination of Oligomeric Status. The results for GMD and RMD cloning, expression and purification are summarized in

Section 1.3.1. and 2.3.1. respectively. Gel filtration analysis of recombinant GMD suggested that it exists as a tetramer in similar buffer conditions used for crystallization (Chapter 1) as well as functional studies (Figure 19a). In contrast, recombinant RMD appears to exist as a dimer (Figure 19b). See Table 4 for elution volumes.

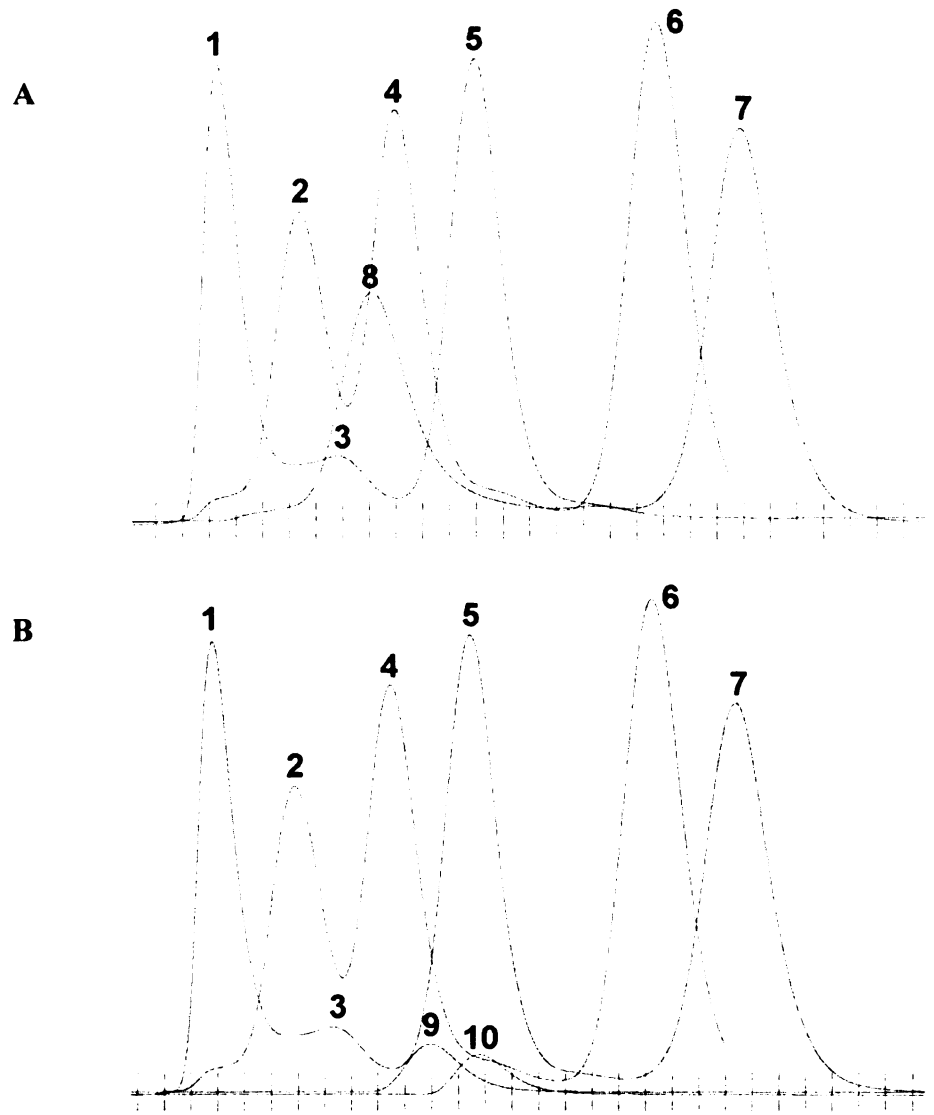


Figure 19. Gel filtration analysis of GMD and RMD

(**A**) Gel filtrations analysis of GMD and standards, and (**B**) RMD and standards. Peaks: (1) ferritin, (2) catalase, (3) aldolase, (4) albumin, (5) ovalbumin, (6) chymotrypsinogen A, (7) ribonuclease, (8) GMD, (9) RMD, (10) RMD/GDP-mannose/NADPH.

Table 4. Gel filtration results of GMD and RMD

Peak	Sample	Molecular Weight (Da)	Elution Volume, Ve (ml)
1	Ferritin	440,000	36.5
2	Catalase	232,000	42.7
3	Aldolase	158,000	45.5
4	Albumin	67,000	49.8
5	Ovalbumin	43,000	55.8
6	Chymotrypsinogen A	25,000	69.4
7	Ribonuclease	13,700	75.7
8	GMD		48.1
9	RMD		52.9
10	RMD/NADPH/GDP-man		56.5

3.3.2. GMD catalyzes the dehydration of GDP-D-mannose. CE analysis of the GMD-catalyzed reaction containing GDP-D-mannose in the presence or absence of NAD or NADP resulted in the appearance of a new peak (Figure 20). This new peak putatively corresponds to the formation of the intermediate product of GMD, GDP-4-keto-6-deoxy-D-mannose. No standard is available for this intermediate product, which is known to be labile making it difficult to analyze. The apparent dehydratase reaction of GMD proceeded quantitatively and irreversibly and required no exogenous cofactor for activity. To test the pH optimum for GMD activity, the reaction was completed under pH range of 5 to 10; GMD showed optimum activity between pH 7 and 8 (Figure 20).

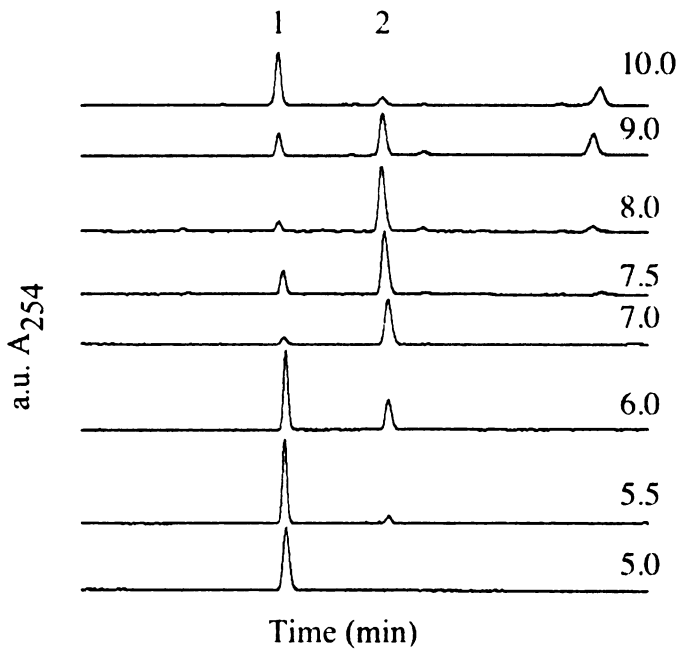


Figure 20. CE analysis of GMD dehydratase activity from pH 5 to 10
(1) GDP-D-mannose, (2) putative GDP-4-keto-6-deoxy-D-mannose.

3.3.3. RMD catalyzes the reduction of GDP-4-keto-6-deoxy-D-mannose. Due to the instability of the intermediate product of GMD, the 4-keto-6-deoxy-D-mannose intermediate was prepared *in situ*. First the GMD-catalyzed reaction was incubated for 2 h, and the enzyme was removed by filtration. The filtrate containing the 4-keto-intermediate was followed by CE analysis after incubation with NADPH (Figure 21f), NADP (Figure 21e), NADPH/RMD from *P. aeruginosa* (Figure 21c) and NADPH/RMD from *A. thermoaerophilus* (Figure 21b). No new peak was observed in the control reactions with just NADPH or NADP, ruling out the possibility that reduction of the labile intermediate was either spontaneous or by any reducing power of the buffer. However, a new peak, the putative GDP-D-rhamnose, was observed upon the addition of NADPH/RMD from either bacterial species. The novel peak was verified by spiking the

final reaction with GDP-D-mannose (Figure 21a). Similar results were observed in the coupled reactions (data not shown). The apparent reductase activity of RMD proceeded quantitatively and required exogenous NADPH. No reverse reaction was observed.

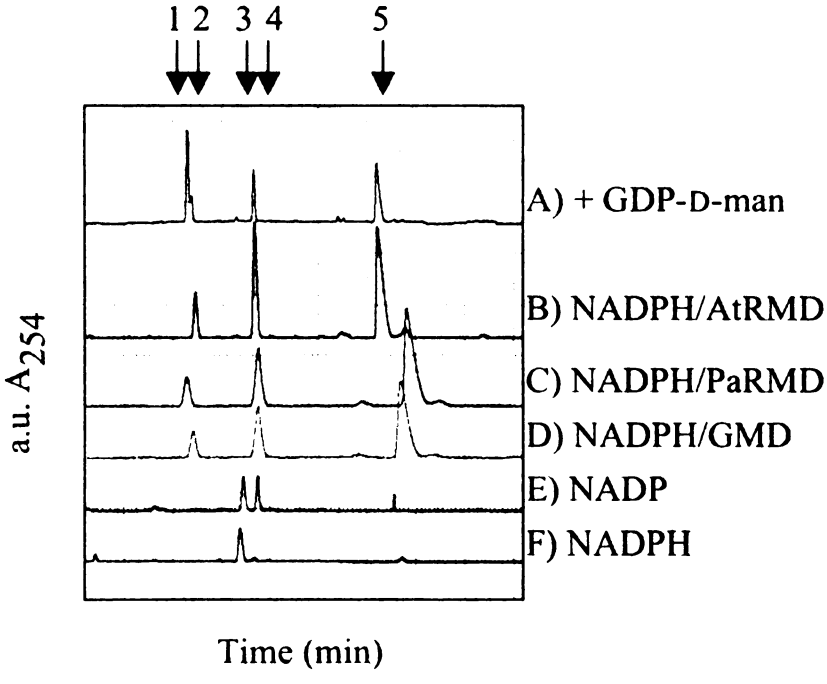


Figure 21. CE analysis of GMD/RMD reactions

The GDP-4-keto-6-deoxy intermediate was treated with the components noted; the final reaction was spiked with GDP-D-mannose (A). Arrows: (1) GDP-D-mannose, (2) GDP-D-rhamnose, (3) GDP-4-keto-6-deoxy-D-mannose, (4) NADP, (5) NADPH.

3.3.3. Is GMD bifunctional? Interestingly, CE analysis revealed that the same peak corresponding to GDP-D-rhamnose was observed when GMD/NADPH was added to the filtrate (containing just the 4-keto-intermediate) (Figure 21d) as when RMD/NADPH was added, indicating GMD displays bifunctionality *in vitro*. To test this observation further, the conversion of the 4-keto-intermediate by GMD was followed in a time-dependent fashion at 0 h (Figure 22d), 1 h (Figure 22c) and 2.5 h (Figure 22b), revealing the

quantitative conversion of the 4-keto-intermediate to GDP-D-rhamnose. The novel peak was again verified by spiking the reaction with GDP-D-mannose (Figure 22a). The reductase activity of GMD was dependent upon the addition of exogenous NADPH to the reaction and no reverse reaction was detected.

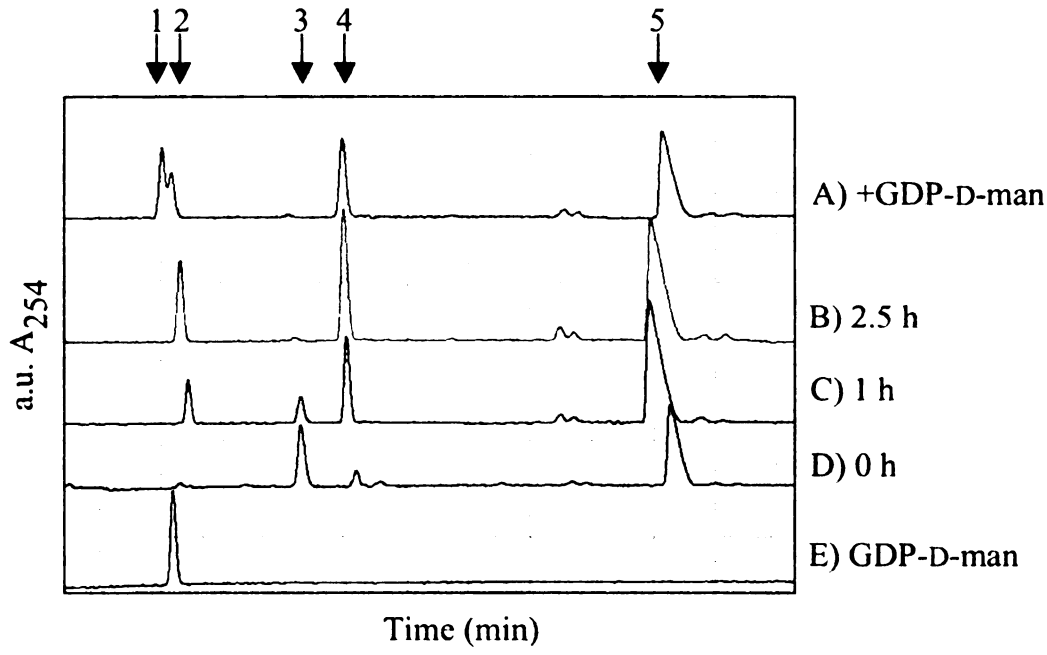


Figure 22. CE analysis of the time-dependent GMD conversion of the keto-intermediate in the presence of NADPH

GMD reductase activity at (D) 0 h, (C) 1 h, (B) 2.5 and (A) final reaction spiked with GDP-D-mannose; (E) GDP-D-man. Arrows: (1) GDP-D-mannose, (2) GDP-D-rhamnose, (3) GDP-4-keto-6-deoxy-D-mannose, (4) NADP, (5) NADPH.

3.3.4. Isolation and identification of the reaction product. No commercial standard for GDP-D-rhamnose was available, so the product had to be further characterized. The nucleotide sugar was purified by HPLC using anion exchange chromatography (Figure 23). Two major peaks were observed in the elution profile. Peak 1 was subjected to CE analysis to reveal that a small amount of NADP co-eluted with the product. However, the reaction product was unequivocally identified as GDP-D-rhamnose by NMR spectroscopy. The NMR spectrum of the nucleotide sample are presented in Figure 24. From a ^{31}P-HSQC experiment a correlation was observed to the anomeric resonance of the sugar resonance at 5.43 ppm. Proton assignments for the sugar were made using the COSY experiment. Assignments for ^{13}C were made from an HMQC experiment. Proton coupling constants were typical of a *manno*-pyranose configuration. The ^{1}H, ^{13}C, and ^{31}P chemical shifts for the sugar nucleotide were similar to the ones reported before for GDP-D-rhamnose [3]. The chemical shifts for the carbon, proton and coupling constants for the rhamnose are shown in Table 5. The values are in congruence with those previously reported for GDP-D-rhamnose by Maki *et al.* [18].

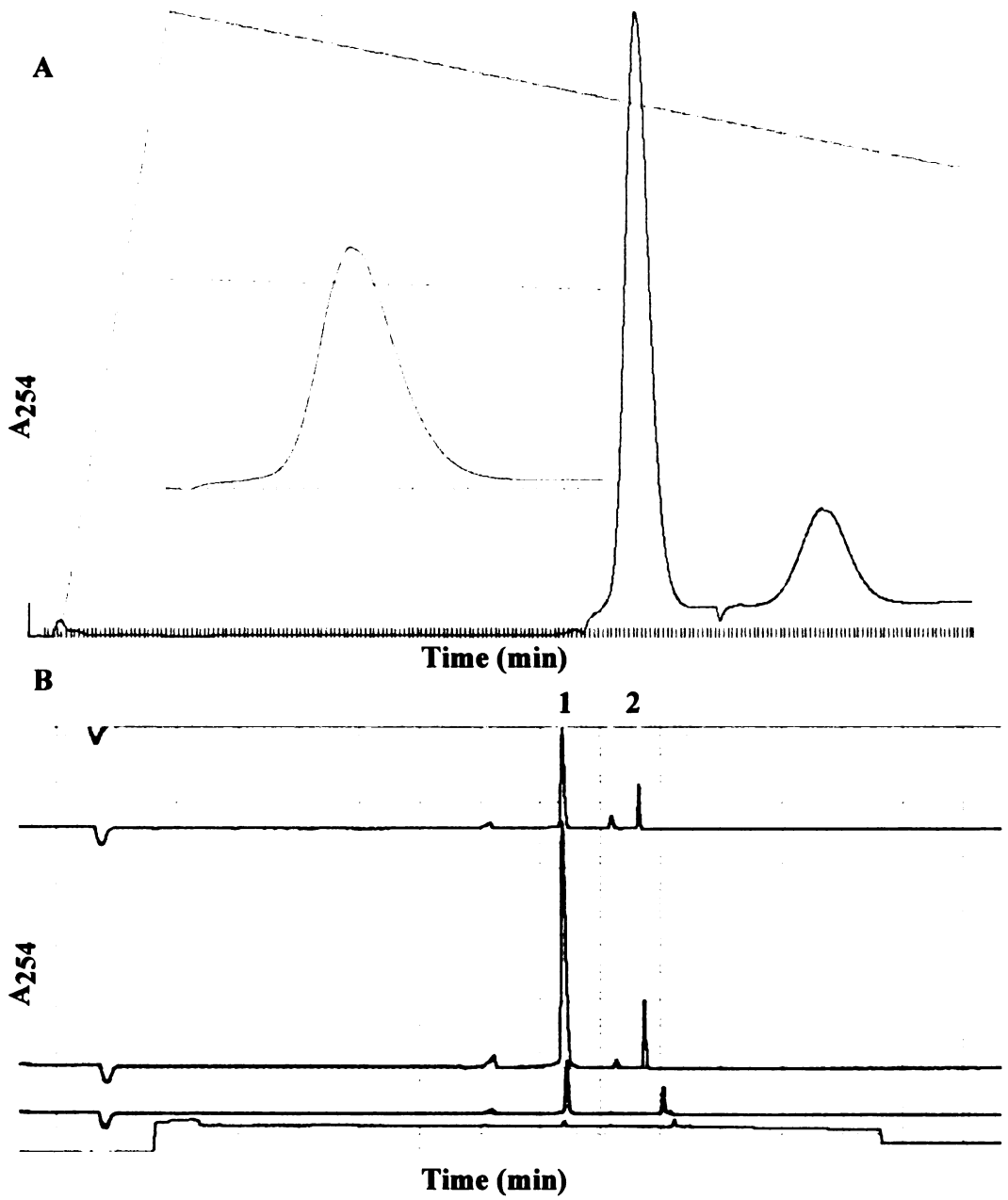


Figure 23. HPLC purification of the GMD/RMD reaction product followed by CE analysis

(**A**) Anion exchange chromatograph of reaction product (inset is enlarged view of the first peak), (**B**) CE analysis of fractions in the first peak. CE peaks: (1) GDP-D-rhamnose, (2) NADP.

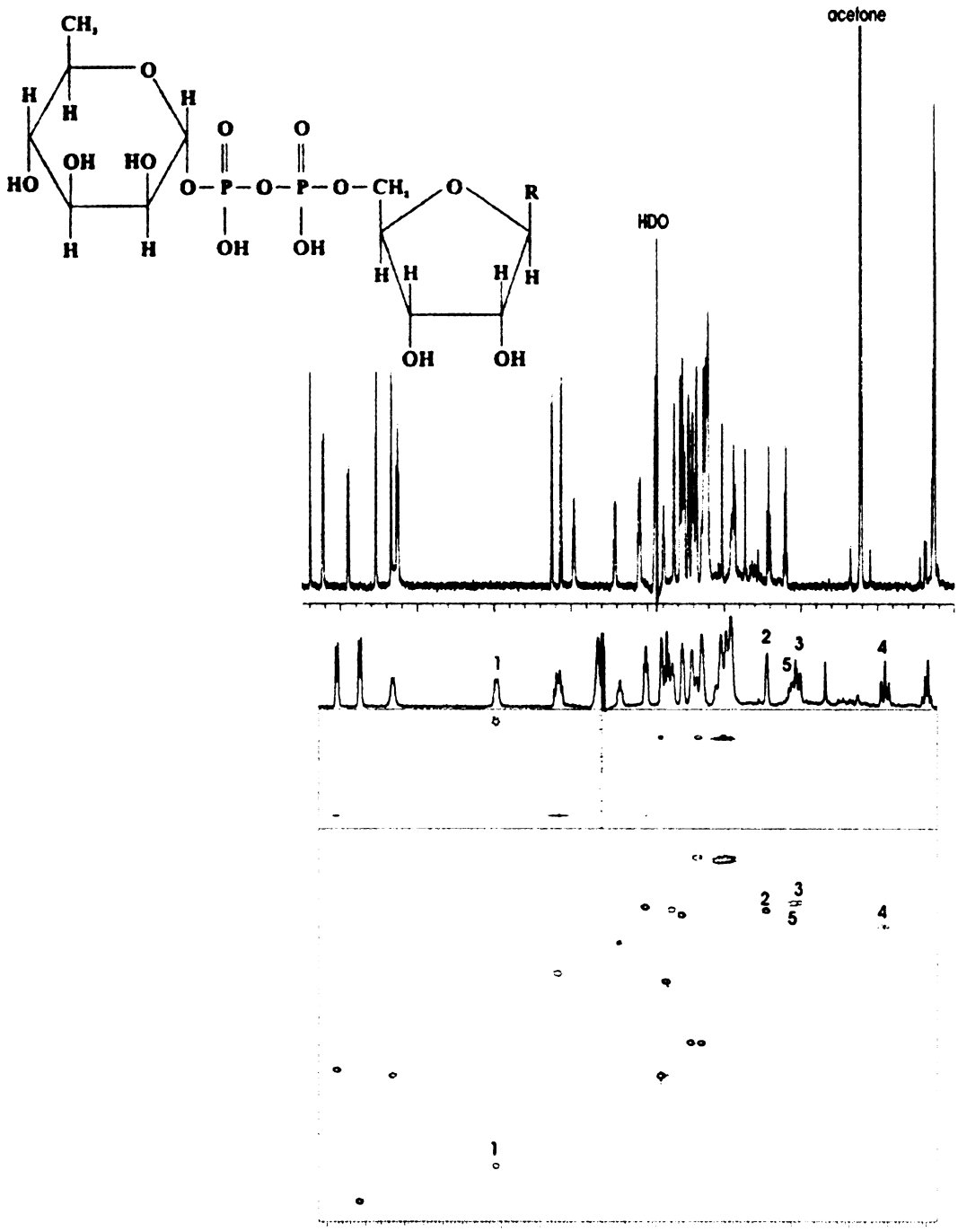


Figure 24. NMR spectra for GDP-D-rhamnose

(**A**) Partial proton spectra, (**B**) one-dimensional-TOCSY for selective excitation of the H-1 sugar resonance with a mixing time of 150 ms, (**C**) ^{31}P HMQC spectrum, (**D**) ^{13}C HSQC spectrum.

Table 5. NMR analysis of GDP-D-rhamnose

^{1}H and ^{13}C chemical shifts, δ(ppm) and proton coupling constants, $J_{H,H}$(Hz)							
Compound		H-1	H-2	H-3	H-4	H-5	H-6
		C-1	C-2	C-3	C-4	C-5	C-6
		$J_{(1,2)}$	$J_{(2,3)}$	$J_{(3,4)}$	$J_{(4,5)}$	$J_{(5,6)}$	
GDP-D-rhamnose	δ_H	5.43	4.03	3.87	3.42	3.90	1.26
	δ_C	97.3	71.2	70.4	72.9	60.6	17.6
	$J_{H,H}$	1.2	3.5	9.7	9.8	6.1	

3.4. Discussion

In this study, we developed a CE-based assay for the functional characterization of the recombinant enzymes GMD and RMD from bacterial source *P. aeruginosa*. Enzymes were expressed separately in *E. coli* as 6His-tagged protein and purified to homogeneity. Using CE, HPLC and NMR analyses, we were able to demonstrate and confirm their involvement in converting GDP-D-mannose to GDP-D-rhamnose.

Our results have shown that using recombinant GMD that was fresh or frozen for over 1 year resulted in the conversion of GDP-D-mannose to the intermediate product GDP-4-keto-6-deoxy-D-mannose. The long-term stability of GMD is in contrast to previous studies, which show the inherent instability of GMDs, except in the case of the plant GDP-D-mannose 4,6-dehydratase MUR1 [5]. The pH optimum for GMD activity is between 7 and 8, which is comparable to the pH optimums reported for GMD from *K. pneumoniae* (pH 7-7.5) [1], porcine thyroid (pH 6.5-8.0) [6] and recombinant GMD from *E. coli* (pH 8.0) [2] and human (pH 7.5) [51]. In most cases, the GMDs use NADP(H), which is in contrast to the dTDP- and CDP-glucose 4,6-dehydratases that show preference for NAD(H). Some GMDs (i.e. *K. pneumoniae*, *A. thermoaerophilus*, and *H. pylori*) are activated or require exogenous cofactor for activity, indicating that the

cofactor may be loosely bound [1, 3, 4]. However, no exogenous cofactor was required for *P. aeruginosa* GMD dehydratase activity, which is consistent with the purification data, the known mechanism and the reported structure. *P. aeruginosa* GMD crystallizes with cofactor bound in the active when none was added to the crystallization conditions, indicating that it was tightly bound throughout purification. This is consistent with reports that the cofactor was difficult to remove from MUR1, *E. coli* GMD and *P. aeruginosa* GMD [44, 52, 94], and the cofactor was described as tightly bound in porcine thyroid and human GMD as well. Considering the mechanism, the cofactor transfers a hydride from the C4 to the C6 position of the sugar resulting in its regeneration, suggesting that the cofactor may remain bound throughout the catalytic cycle. Many enzymes of the NDP-sugar modifying subfamily of SDRs that internally recycle their cofactor show that the cofactor is tightly bound in the active site [31, 60]. Furthermore, structural evidence of *P. aeruginosa* GMD suggests that the cofactor is involved in creating the tetramer interface. To support the crystallography data, gel filtration experiments have shown that *P. aeruginosa* GMD is functional as a tetramer. The cofactor seems to be important for tetramerization, and perhaps, the regulation of enzyme activity.

Regarding the RMD-catalyzed reaction, the apparent instability of the substrate, the 4-keto-intermediate product of GMD, was addressed by preparing the intermediate *in situ*. In both the coupled and sequential reactions, we demonstrated the reduction of the intermediate to GDP-D-rhamnose upon the addition of recombinant RMD from either *P. aeruginosa* or *A. thermoaerophilus*. Data suggest that it is an NADPH-dependent reaction, which is consistent with previous reports on RMD from *A. thermoaerophilus*

[3]. The requirement of exogenous NADPH suggests that the cofactor must be readily accessible to solvent to be released, allowing a new cofactor to bind for restoration of enzymatic activity, as it is suspected that there is not a second substrate in the reaction that regenerates the cofactor NADP. Finally, gel filtration experiments have revealed that RMD from *P. aeruginosa* is functional as a dimer and that the dimer is its assumed biological molecule. Interestingly, the data suggest that RMD adopts a more compact structure upon binding cofactor and substrate, perhaps indicating domain movement.

P. aeruginosa GMD not only exhibited dehydratase activity, but also NADPH-dependent reductase activity *in vitro*, creating the final product GDP-D-rhamnose. GMDs possessing bifunctionality have also been demonstrated in *A. thaliana* and *A. thermoaerophilus in vitro* [3, 5]. Whereas the reductase activity of *A. thermoaerophilus* GMD is quite low compared to its RMD, the reductase activity of *P. aeruginosa* GMD seems to be comparable to that of its RMD. We expect no GDP-D-rhamnose contamination from the *E. coli* expression system by activity of an enzyme encoded by the bacteria, as no RMD homolog exists in the *E. coli* genome. Based on the CE analysis of the enzyme-substrate reactions, the pathway for GDP-D-rhamnose biosynthesis in *P. aeruginosa* could be revised to show that GMD can be a bifunctional enzyme *in vitro*, converting GDP-D-mannose to the 4-keto-intermediate, then to the product GDP-D-rhamnose.

To conclude, GDP-D-rhamnose is an important component of LPS molecules, whose biosynthetic pathways may be targets for antimicrobial agents. The enzymatic synthesis of GDP-D-rhamnose would allow further studies on the assemblies of LPS molecules. We have shown that recombinant GMD and RMD are able to convert GDP-D-mannose to

GDP-D-rhamnose, confirming the product structure by NMR analysis. With the enzymatic activities of GMD and RMD characterized, these proteins are ideal candidates for scaling up the production of GDP-D-rhamnose for use in studying rhamnosyltransferases, WbpX, WbpY and WbpZ, involved in poly-D-rhamnan biosynthesis.

[1]Portions of this Chapter are from a manuscript in preparation under the following title: Poon, K.K.H.[1], Webb, N.A.[2], McNally, D.[3], Brisson, J-R.[3], Garavito, R.M.[2] and Lam, J.S.[1], "Functional characterization of GMD and RMD involved in the biosynthesis of GDP-D-rhamnose from *Pseudomonas aeruginosa*".

[1]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

[2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824-1319, USA

[3]Institute for Biological Sciences, National Research Council, Ottawa, Ontario K1A 0R6, Canada

CHAPTER 4: PROTEIN ENGINEERING WITH THE GOAL OF IMPROVING PROTEIN EXPRESSION, PURIFICATION AND/OR CRYSTALLIZATION

4.1. Introduction

The road to obtaining a protein crystal structure can have several obstacles along the way. The protein must be expressed to several milligram quantities and it must be protected from proteolysis and must fold properly, sometimes requiring chaperones. A purification scheme must also be devised that leads to a sample suitable for crystallization. Conditions must then be determined in which reproducible crystals can be grown that lead to X-ray diffraction data of a quality acceptable for structure determination. A protein whose characteristics present problems in any step of the way is typically referred to in the structural genomics world as a "high-hanging fruit". This term has been given primarily to membrane proteins; however, it can also refer to other proteins with low solubility and/or stability.

The use of affinity tags for purification of recombinant proteins has become a widespread tool since its inception in 1983 with Protein A. Some large affinity tags like glutathione-S-transferase (GST) and maltose-binding protein (MBP) have the added advantage as they may provide increased expression, greater solubility, protection from cleavage by proteases, and assistance in proper folding of the protein to which it is fused. For crystallization purposes, large affinity tags have generally been cleaved from the protein of interest. However, a handful of crystal structures of chimeric proteins with large affinity tags have been reported; human T cell leukemia virus type I gp21 ectodomain fragment (gp21) [95], *Staphylococcus* accessory regulator R (SarR) [96] and the MATa1 protein [97], which were resistant to crystallizing by themselves. It has been

suggested that this strategy may have a greater application in structural genomics on challenging proteins [98].

We have created a strategy that makes use of a large affinity tag along with an additional stabilizer protein for increasing the expression, enhancing the purification and facilitating the successful crystallization of problematic proteins. The pMal protein fusion and purification system (New England Biolabs), which makes use of the affinity tag MBP, was chosen as a starting point. In addition to the aforementioned advantages, MBP allows for affinity purification of the fusion protein over amylose resin. However, low affinity of the fusion protein to the resin may be caused by interactions between MBP and the protein of interest that block or distort the maltose-binding site. For this reason, the length of the linker between MBP and the protein of interest was maintained to reduce the likelihood of any inhibitory interactions. Yet, the presence of the linker adds flexibility and may prevent a stable conformation of the partners that is needed to facilitate crystallization. To address this issue, the linker in the pMal-C2 vector was replaced with an Src homology 3 (SH3) domain-binding site. The stabilizer peptide, the SH3 domain of Abelson [99] tyrosine kinase, which is known to tightly bind the SH3 domain-binding site, was added to the purified fusion protein before crystallization to serve as a clamp, providing rigidity to the new complex. This approach for problematic proteins was successful in the crystallization of *P. aeruginosa* RMD. Although reproducibility was a problem, the results presented in this chapter lead to promising suggestions for further optimization of this system, which may have a place in expressing, purifying and crystallizing "high-hanging fruit" in the structural genomics world.

4.2. Experimental Procedures

4.2.1. Construction of pMal-C2L3 Vector. The following oligos were used to replace 50 nucleotides between the SacI and XmnI restriction sites in the linker region of pMal-C2: 5'-CGGCAGCACCGACTTACAGCCCACCACCGCCGCCGGGAAGG-3' and 5'-CCTTCCCGGCGGCGGTGGTGGGCTGTAAGTCGGTGCTGCCGAGCTC-3'. The vector pMal-C2 was first digested with SacI and XmnI, isolated by electrophoresis on 0.8% agarose gel and purified using the QIAquick gel extraction kit (Qiagen). The following reaction was allowed to incubate at room temperature overnight: 1 μg each of the oligos, 100 ng of the cut vector, 1 μl of 10x ligation buffer (New England Biolabs), 1 μl T4 DNA ligase (New England Biolabs) and ddH_2O to 10 μl. The reaction was transformed into DH5α *E. coli* cells and plated on an LB agar plate with 100 μg/ml ampicillin. Overnight cultures were grown from the resulting colonies; DNA was isolated and sequenced for confirmation of the insertion.

4.2.2. Cloning of RMD into pMal-C2 and pMal-C2L3. The RMD-pQE30 (section 2.2.1.) construct and was digested with BamHI and SalI, isolated by electrophoresis on 0.8% agarose gel and purified using the QIAquick gel extraction kit (Qiagen). The fragment was cloned into either pMal-C2 or pMal-C2L3 digested with BamHI and SalI. The new constructs were confirmed via sequencing.

4.2.3. Expression and Purification of RMD-pMalC2 and RMD-pMal-C2L3. A glycerol stock of *E. coli* BL21 (DE3) cells transformed with RMD-pMalC2 or RMD-pMal-C2L3 was used to inoculate 50 ml LB broth with 2 mg/ml glucose and 100 μg/ml ampicillin. Following cultivation at 37°C for 16 h, the starter culture was transferred to 1 L LB broth with 2 mg/ml glucose and 100 μg/ml ampicillin. Once an optical density at

600 nm of 0.5-0.6 was reached, the cells were induced with 0.3 mM IPTG and grown for an additional 4 h at 37° before harvesting by centrifugation. Cells were re-suspended in Column Buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, 1 mM DTT pH 7.4). After a freeze/thaw cycle and sonication, the lysate was cleared by centrifugation at 10,000 x g for 20 min and loaded onto a column of 10 ml amylose resin. The column was washed with 20 ml Column Buffer; then the protein was eluted with Elution Buffer (Column Buffer plus 10 mM maltose). Those fractions shown to be homogenous were pooled and concentrated to 10 mg/ml. The protocol recommended in the pMal Protein Fusion and Purification System manual for cleavage of the fusion protein with Factor Xa was used to cleave MBP from MBP-RMD.

4.2.4. Expression and Purification of Abl-SH3. The SH3 domain of human Abl tyrosine kinase had been cloned into pET15b (Novagen) over the 6-His tag via the NcoI/BamHI sites; the construct was a gift from Dr. Rhea Hudson, Department of Structural Biology and Biochemistry, Hospital for Sick Children, Toronto, Ontario, Canada. A glycerol stock of *E. coli* BL21 (DE3) cells harboring the Abl-SH3-pET15b construct was used to inoculate 50 ml LB broth with 100 μg/ml ampicillin. After growth at 37° for 16 h, this starter culture was used to inoculate 1 L of LB broth with 100 μg/ml ampicillin. Once an optical density at 600 nm of 0.5-0.6 was reached, the cells were induced with 0.3 mM IPTG and grown for an additional 6 h at room temperature before harvesting by centrifugation. The cells were re-suspended in Buffer A (20 mM Tris pH 8.5) and exposed to a freeze/thaw cycle. After sonication, the lysate was cleared by centrifugation at 10,000 x g for 20 min. The supernatant was filtered over a 30 kDa centricon; then the flow-through was concentrated over a 5 kDa centricon. The sample

was purified by FPLC using a Resource Q anion exchange column (Amersham Pharmacia Biotech) equilibrated with Buffer A. Protein was eluted over a 60 ml salt gradient using Buffer B (20 mM Tris pH 8.5, 1 M NaCl). Fractions were analyzed by SDS-PAGE. Those shown to be homogenous were pooled and concentrated to 10 mg/ml.

4.2.5. *In vitro* binding of MBP-SH3-RMD and the Abl-SH3 domain. The Abl-SH3 domain was added to MBP-SH3-RMD at an approximately 1:1 molar ratio and allowed to sit on ice for 1 h. Unbound Abl-SH3 domain was separated from the MBP-SH3-RMD:Abl-SH3 domain complex by FPLC on a Sephacryl S200 gel filtration column (Amersham Biosciences) equilibrated with S200 Column Buffer (20 mM Tris pH 8.5, 300 mM NaCl, 10 mM maltose, 0.3% sarcosyl). Fractions were analyzed by SDS-PAGE.

4.2.6. Crystallization of MBP-RMD, MBP-SH3-RMD and MBP-SH3-RMD:Abl-SH3 domain complex. Purified MBP-RMD, MBP-SH3-RMD or MBP-SH3-RMD:Abl-SH3 domain complex at 10 mg/ml was used to set up crystallization screens using the microbatch method with the Impax I-5 robot (Douglas Instruments). A total of 198 commercially available conditions were screened by combining 1 µl of purified protein with 1 µl precipitating solution. Crystallization conditions were refined using the handing drop vapor diffusion method. To aid in crystallization, 0.5 mM maltose, 5 mM GDP and/or 2 mM NADP(H) were added to the protein sample.

4.3. Results and Discussion

4.3.1. Crystallization trials on the MBP-RMD chimera. Previous attempts at obtaining high quality crystals of *P. aeruginosa* RMD have been unsuccessful (see

Chapter 2). Therefore, RMD became the trial protein on which to experiment new strategies for expression, purification and crystallization. Though solubility is not a problem with RMD, the large affinity tag MBP was chosen to create a fusion protein for added stability. MBP is known to increase expression, assist in folding, protect proteins from proteolysis and facilitate affinity purification over amylose resin. The MBP tag should also change the crystallization environment as a fusion with RMD, hopefully facilitating the growth of crystals of higher quality. In addition, the three dimensional structure of MBP may allow for phase determination of the fusion protein by molecular replacement or provide known locations of heavy atom-binding sites for phase determination via multiple isomorphous replacement. The structures of *E.coli* MBP in apo- [100] and maltose-bound [101] forms have been reported.

The gene encoding *P. aeruginosa* RMD was initially cloned into pMal-C2 to check for expression of the fusion protein. Purification over amylose resin yielded 8 mg of purified protein per liter of cell culture that was 90-95% pure. Fractions analyzed by SDS-PAGE (Figure 25a) show the molecular weight of the fusion protein to be roughly 80 kDa, which closely matches 76 kDa, the molecular weight predicted from the amino acid sequence. Attempts were made to cleave MBP from the fusion protein using Factor Xa. The cleavage was carried out at a ratio of 1 mg Factor Xa per 100 mg protein at room temperature and was followed by SDS-PAGE (Figure 25b). The results indicated that partial cleavage began at 0.5 h and was nearly 75% complete at 18 h. However, the cost of the protease makes this process somewhat prohibitive for the amount of protein needed for crystallization purposes. The expense, along with the tedious task of optimizing the cleavage conditions as well as precipitation of the target protein are

common problems encountered during the cleavage step. To avoid the cleavage process, re-purification and the loss of protein, the tag may be left on for crystallization trials. This approach was attempted for the full-length MBP-RMD fusion; however, no crystals were observed in any of the conditions. This is consistent with the results from the crystallization trials of the gp21 ectodomain fragment that had also been cloned into pMal-C2 [102], and is most likely due to the flexibility caused by the linker region.

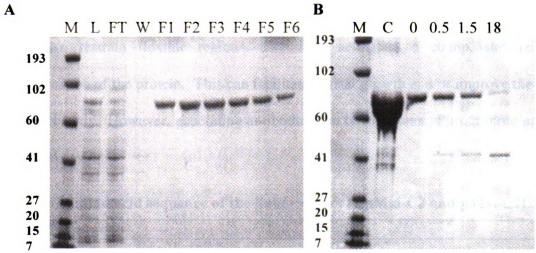


Figure 25. SDS-PAGE of samples from a MBP-RMD purification and Factor Xa cleavage

(**A**) MBP-RMD amylose column purification: load (L), flow through (FT), wash, (W) and fractions (F); (**B**) Factor Xa cleavage of MBP-RMD: control (C) and samples at 0, 0.5, 1.5 and 18 h. Molecular weight marker (M) standards shown in kDa to the left of each gel (MBP, 41 kDa, MBP-RMD, 76 kDa).

4.3.2. Reconstruction of the linker region. The amino acid sequence of the linker region in the original pMal-C2 vector is shown in Table 6. This represents the region between MBP and the protein of interest, which is 22 amino acids in length and contains a Factor Xa cleavage site at the C-terminus. The length of the linker is important for some fusion proteins to bind tightly to the amylose resin. Since the binding affinity is already only in the micromolar range, the separation of MBP to the protein of interest is

crucial to avoid any interactions that might block MBP from binding to the amylose. However, the length of the linker introduces conformational heterogeneity. To address this issue, the concept of a clamp or stabilizer was intriguing when considering how to restrain the flexibility of the MBP-fusion protein. Antibody fragments have been used in co-crystallization experiments for this purpose, specifically with membrane proteins with some success. While antibodies can cover a portion of the hydrophobic surface of the membrane protein converting it to a region more likely to make crystal contacts, they may also restrain flexible regions, essentially acting as a clamp, stabilizing the conformation of the protein. This can facilitate crystal growth and/or improve the quality of the crystals. However, generating antibodies can be laborious. Plus, a more universal clamp is desirable.

Table 6. Amino acid sequence of the linker region in pMal-C2 and pMal-C2L3

pMal-C2

				SacI					XmnI		
malE	ACT	AAT	TCG	AGC	TCG	AAC	AAC	AAC	AAC	AAT	AAC
MBP	Thr	Asn	Ser	Ser	Ser	Asn	Asn	Asn	Asn	Asn	Asn
AAT	AAC	AAC	AAC	CTC	GGG	ATC	GAG	GGA	AGG	ATT	TCA
Asn	Asn	Asn	Asn	Leu	Gly	Ile	Glu	Gly	Arg	Ile	Ser
								Factor Xa ↑			

pMal-C2L3

				SacI							
malE	ACT	AAT	TCG	AGC	TCG	GCA	GCA	CCG	ACT	TAC	AGC
MBP	Thr	Asn	Ser	Ser	Ser	Ala	Ala	Pro	Thr	Tyr	Ser
							XmnI				
CCA	CCA	CCG	CCG	CCG	GGA	AGG	ATT	TCA			
Pro	Pro	Pro	Pro	Pro	Gly	Arg	Ile	Ser			
					Factor Xa ↑						

While working on a side project, the expression, purification and crystallization of human phospholipid scramblase 1 (HuPLSCR1) (see Appendix A), an idea presented itself. HuPLSCR1 contains multiple proline-rich motifs resembling SH3 domain-binding sites. SH3 domains are non-catalytic protein modules present in a wide variety of unrelated proteins that can be independently expressed, purified and even crystallized. HuPLSCR1 is known to have marked specificity for the Abl-SH3 domain; in fact, studies suggest that HuPLSCR1 is the normal substrate of Abl tyrosine kinase [103]. The Abl-SH3 domain was available in milligram quantities for the studies with HuPLSCR1. Work with HuPLSCR1 and the Abl-SH3 domain was the inspiration for the following idea: replace the linker in the pMal-C2 vector with an SH3 domain-binding site and the Abl-SH3 domain could be used as the clamp or stabilizer for crystallization trials of the fusion protein. An added advantage is that the crystal structure of the proline-rich peptide from the SH3-binding protein 3BP-1 bound to the Abl-SH3 domain is available [104].

In general, SH3 domains bind proline-rich peptides with affinities of 0.2-50 μM and are quite promiscuous. Pisabarro and Serrano [105] rationally designed a peptide ligand for the Abl-SH3 domain with high-affinity and specificity by mutating the 3BP-1 peptide (APTMPPPLPP). Mutations were chosen based on the tendency of the peptide to adopt a polyproline helix II (PPII) conformation and the ability of the residues to form favorable interactions to the Abl-SH3 domain. In doing so they created a peptide APTYSPPPPP that binds the Abl-SH3 domain with an affinity 100 times higher than the original peptide. This sequence was chosen for the linker region.

The structures of the reported MBP fusion proteins along with the structure of MBP, which are shown in Figure 26, were consulted before deciding on the final linker sequence. In all three cases, the linker region had been shortened. The linker region was substituted with AAA for gp21, AAAEF for SarR and AAAAA for MATa1. The rationale behind the short AAA linker in gp21 was to connect the C-terminal helix of MBP to the predicted N-terminal helix of gp21. However, instead of forming a continuous helix, the short linker provides a 90° turn between the two. None of the three structures showed that a continuous helix was necessary for forming a rigid connection, so we were confident with the sequence chosen for the linker, which was not necessarily optimized for such a linkage. The SacI site at the N-terminus of the pMal-C2 linker region along with the XmnI site at the C-terminus was used to incorporate the SH3 domain-binding site linker-coding region (shown in Table 6) creating the new vector pMal-C2L3.

The expected tertiary structure of MBP plus the SH3 domain-binding site in complex with the Abl-SH3 domain is shown in Figure 27. The structure of MBP is taken from the coordinates of the Protein Data Bank entry 1anf. The break in the final helix of MBP is predicted based on the structures of MBP, SarR and Mata1; gp21 broke one residue proximal. The linker region is the ten-residue proline-rich peptide (APTMPPPLPP) from the SH3-binding protein 3BP-1, whose structure is known in complex with the Abl-SH3 domain [104]. Although the sequence of the peptide encoded by pMal-C2L3 is different by three residues (APTYSPPPPP), the Abl-SH3 domain should bind to the peptide in a similar fashion based on modeling work of Pisabarro and Serrano [105]. The very C-terminal portion of the linker marks the beginning of the target protein. One would

expect that MBP would assist in the formation of the tertiary structure of the target protein. However, there is certainly a concern for the potential affect of MBP and/or the Abl-SH3 domain on the quaternary structure of the target protein. The three MBP-fusion protein structures that have been solved are all in their biologically relevant quaternary state. Our hope is that our protein will assume its biologically relevant quaternary state before adding the Abl-SH3 domain, and the addition of the peptide will stabilize the overall structure to facilitate crystallization.

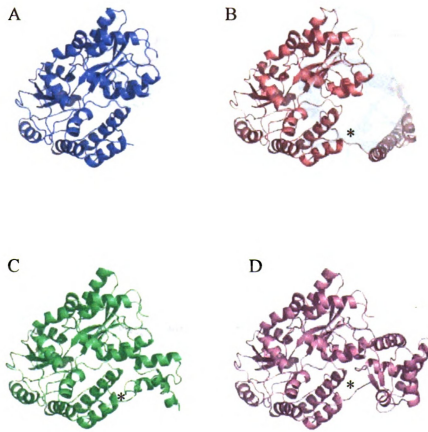


Figure 26. Crystal structures of MBP and MBP-fusion proteins

(A) MBP, (B) MBP-gp21, (C) MBP-Mata1 and (D) MBP-SarR. MBP is shown in the same orientation in each panel. The asterisks in panels B, C and D highlight the linker region.

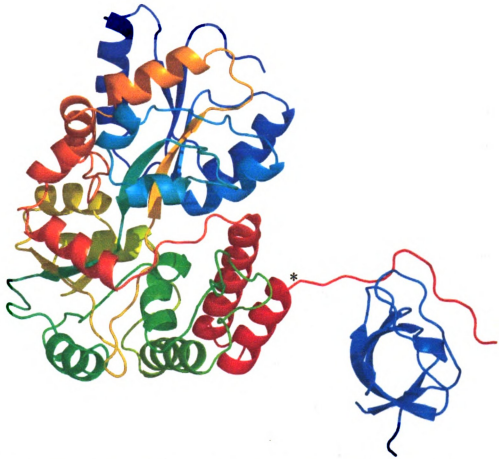


Figure 27. Model of MBP-SH3 binding domain site:Abl-SH3 domain complex
The MBP portion is rainbow-colored. An asterisk denotes the beginning of the linker region, or the SH3-binding domain site, colored in red. The Abl-SH3 domain is colored in blue.

4.3.3. Crystals of the suspected MBP-SH3-RMD:Abl-SH3 domain complex.

RMD was cloned into the newly constructed vector pMal-C2L3. Expression and purification trials were performed on both RMD-pMal-C2L3 as well as the control pMal-C2L3. SDS-PAGE analysis of samples from the amylose column are shown in Figure 28. MBP-SH3-RMD elution fractions run around 80 kDa, which matches closely to 78 kDa, the predicted molecular weight based on the amino acid sequence. MBP-SH3 elution fractions run around 45 kDa, which matches closely to its predicted molecular

weight of 43 kDa. While pMal-C2L3 yielded 20 mg of purified MBP-SH3 per liter of cell culture, RMD-pMAL-C2L3 yielded only 8 mg of purified MBP-SH3-RMD per liter of cell culture. This was most likely due in part to an unexpected result: the fusion protein tended to aggregate. This is contrary to the results of other proteins expressed as fusions with MBP, which tended to become more soluble upon the addition of MBP. Nonetheless, it was obvious that the buffer would need some tweaking, i.e. addition of detergent, to keep the complex soluble. Several detergents were tested; the best seemed to be sarcosyl at a concentration of 0.3%. Also, it was noted that increasing the pH of the Tris buffer from 7.4 to 8.5 helped keep the protein from aggregating.

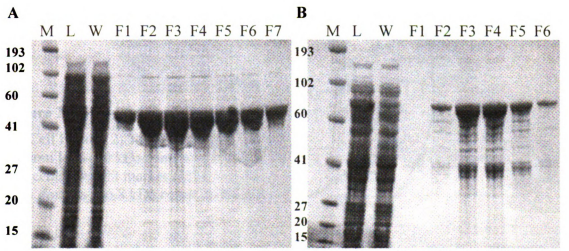


Figure 28. SDS-PAGE of samples from a typical MBP-SH3 and MBP-SH3-RMD purification

(**A**) MBP-SH3 amylose column purification: load (L), wash (W) and fractions (F); (**B**) MBP-SH3-RMD amylose column purification: load (L), wash (W) and fractions (F). Molecular weight marker (M) standards shown in kDa to the left of each gel.

The gene encoding the Abl-SH3 domain had been cloned into the pET15b vector over the code for the 6-His tag, so it was purified over an anion exchange column rather than a Ni-NTA column (Figure 29). Before loading the column, the sample was filtered over a

30 kDa centricon to decrease the amount of unwanted protein. Although this step resulted in the loss of a significant amount of desired protein along with the unwanted protein (see lane 3 of the SDS-PAGE gel in Figure 29), the sample in the end was more pure. The purification process for the Abl-SH3 domain resulted in about 1.5 mg of purified protein per liter of cell culture.

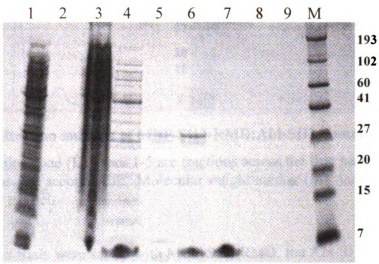


Figure 29. SDS-PAGE of samples from a typical Abl-SH3 domain purification
(1) Cell lysate, (2) 30 kDa centricon flow-through, (3) 30 kDa centricon retentate, (4) column load (30 kDa centricon flow-through concentrated to 2 ml) and (5-9) fractions. Molecular weight marker (M) standards shown in kDa to the right of the gel. The Abl-SH3 domain is 6.8 kDa based on the amino acid sequence.

It is known that the SH3 binding-domain peptide encoded in the linker region of pMal-C2L3 binds the Abl-SH3 domain with the affinity of K_d=0.4 μM [105]. The Abl-SH3 domain was added to the pool of the MBP-SH3-RMD fractions from the amylose resin at a 1:1 molar ratio and allowed to bind before running a gel filtration column. The results from the S200 gel filtration column are seen in Figure 30. A nice, single peak of the complex is seen in the column trace, along with a smaller peak corresponding to MBP. The Abl-SH3 domain does not show up on the SDS-PAGE gel shown most likely

because it is very low in concentration. Not having an antibody available made this difficult to prove definitively.

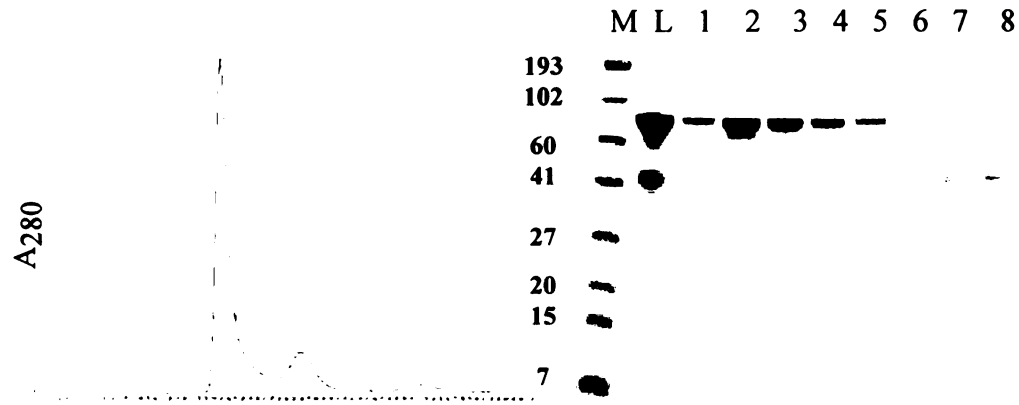


Figure 30. Gel filtration analysis of MBP-SH3-RMD:Abl-SH3 domain

SDS-PAGE samples: load (L), lanes 1-5 are fractions across the first peak and lanes 6-8 are fractions across the second peak. Molecular weight marker (M) standards shown in kDa to the left of the gel.

Crystallization trials were set up with MBP-SH3-RMD, the Abl-SH3 domain and the complex. While crystals were observed in a small number of conditions for the Abl-SH3 domain, none were observed for MBP-SH3-RMD alone. When crystallization trials were set up with the complex, needles and small, plate-like crystals were observed (Figure 31). The best results occurred upon the addition of 5 mM GDP and 2 mM NADP to the protein sample before the trays were set up. Since MBP adopts a different conformation upon binding maltose, 0.5 mM maltose was used in some of the trials as a variable in the protein sample. However, no obvious differences were observed with or without it in the sample. This is most likely because MBP had retained the maltose molecules during purification. This was observed in the structures of MBP-gp21 and MBP-SarR where MBP was seen in a closed conformation [95, 96]. In contrast, MBP was observed in an

open conformation in the MBP-Mata1 structure, most likely because maltose was stripped during purification over a strong cation exchange column [97]. The crystallization condition of the best-formed crystals was 30% PEG 400, 100 mM sodium acetate trihydrate pH 4.6 and 100 mM cadmium chloride dihydrate. This condition is consistent with the conditions of the other MBP-fusion proteins: PEGs seem to be the most successful precipitants, and the pH of the buffer in the solution is generally low. Although several rounds of optimization trials were completed, reproducibility of the crystals of the suspected complex was clearly a problem. In addition, the crystals obtained were too small for x-ray analysis.

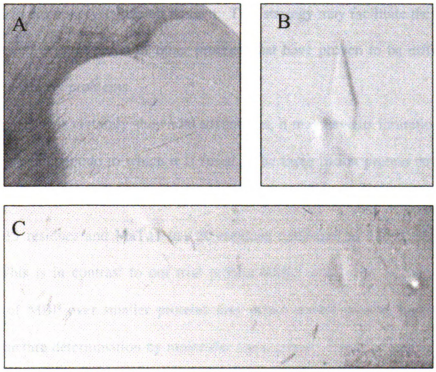


Figure 31. Crystals of MBP-SH3-RMD:Abl-SH3

Panel A: needle-like crystals in 28% PEG 400, 100 mM HEPES pH 7.5, 200 mM $CaCl_2$ dihydrate. Panel B: plate-like crystals in 18% PEG 8K, 100 mM sodium cacodylate pH 6.5. Panel C: rod-shaped crystals in 30% PEG 400, 100 mM sodium acetate trihydrate pH 4.6, 100 mM cadmium chloride dihydrate.

4.3.4 Conclusions. To conclude, we have presented here a strategy of employing (1) a large affinity tag, MBP, to assist in the expression and purification, along with (2) a stabilizer protein, the Abl-SH3 domain, to assist in the crystallization of problematic proteins. While many of the other strategies to improve expression, purification and/or crystallization require the modification of the protein itself (i.e. point mutations, truncations, etc.), this strategy was designed so that the protein of interest is left unhindered in its secondary, tertiary and most likely quaternary structure formation. Although reproducibility has been a problem, we were successful in obtaining crystals upon the addition of the stabilizer protein, the Abl-SH3 domain, to the MBP-SH3-RMD fusion protein. It appears the Abl-SH3 domain is providing rigidity to the MBP-fusion protein, which did not crystallize on its own. This strategy may facilitate the expression, purification and crystallization of other proteins that have proven to be difficult due to solubility or stability problems.

Though MBP has certainly shown its advantages, it may have its limitations based on the size of protein/peptide to which it is fused. The three fusion protein structures that have been solved are all small in comparison to the size of MBP; gp21 has 88 residues, SarR has 115 residues and MaTa1 has 50 residues compared to MBP, which has 368 residues. This is in contrast to our trial protein RMD, which has 313 residues. The dominance of MBP over smaller proteins may direct crystal packing and more easily facilitate structure determination by molecular replacement. Therefore, this method may be more conducive to small proteins. Since it seems impractical to keep up a 3:1 tag to protein size ratio (based on the largest fusion protein structure solved, SarR), perhaps a

different strategy would work for larger proteins. Work on new strategies in our lab is ongoing.

CHAPTER 5: FUTURE DIRECTIONS

- Is the cofactor necessary for GMD to form a tetramer?

It was shown that unlike the typical homodimeric SDR structures, GMD crystallizes as a tetramer where the cofactor is intimately involved in the tetramer interface. It would be interesting to see if removal of the cofactor causes the tetramer to fall apart. To test for this, one would need to remove the cofactor, which has proven to be difficult to do in the case of *P. aeruginosa* GMD and *A. thaliana* MUR1. A protocol has been reported for stripping CDP-D-glucose 4,6-dehydratase of its cofactor by dialyzing it against phosphate buffer containing potassium bromide for several days [71]. The same protocol could be used on GMD to produce the apoenzyme and the process could by followed by gel filtration experiments.

- Can the substrate specificity of GMD be determined and/or altered?

Efforts could be made to determine substrate specificity. It does not appear that GMDs are particular about the sugar moiety they accept. They are able to bind their natural substrate and intermediates, as well as GDP-D-rhamnose [5], GDP-D-fucose [2, 51, 52] and GDP-D-glucose [6]. However, GMDs are considerably more particular about nucleotide specificity. This is supported by work completed on porcine thyroid GMD where the following list of nucleotides were tested as potential inhibitors: GMP, GDP, GTP, AMP, ADP, ATP, CMP, CDP, CTP, UMP, UDP UTP and ITP. Only GDP and GTP showed any appreciable inhibitory action [6]. Initial look at the interactions show that several residues in contact with GDP are conserved across GMDs: Asn179, Lys193, Arg218, Arg279 and Glu282. Mutations on these residues could be generated to see which are necessary for GDP binding. The activity of the resulting mutants could be

tested via CE (protocol presented in Chapter 3) and the structures could be studied via X-ray crystallography. It would also be interesting to see whether or not the substrate-binding site could be altered so that GMD could accept other nucleotide diphosphates or NDP-sugars. If GMD could be modified to accept different substrates, it opens up the potential to utilize GMD in making alternate deoxyhexoses for the synthesis of glycoconjugates.

- Do the structural characteristics of RMD place it into the SDR protein family?

Specific stretches of conserved amino acid sequences (i.e. G-XX-G-XX-G, S/T, Y-XXX-L) suggest that RMD fits into the SDR protein family. It will be interesting to see if the monomeric structure reveals that it folds into two domains, the N-terminal cofactor binding domain and the C-terminal substrate binding domain, like the NDP-sugar modifying subfamily of SDRs. X-ray structure determination using the data set from AtRMD is currently in progress. A respectable molecular replacement solution has been obtained using coordinates from structure of GMD from *P. aeruginosa*, suggesting that the core structure of RMD will be similar to the SDR protein structures. Certainly the first order of business is to perform model building/refinement on the molecular replacement solution for AtRMD to check the solution. In addition, heavy-atom derivatives will be prepared for use in experimental phase determination by isomorphous replacement for comparison to those determined by the molecular replacement method.

- What residues are involved in the active site of RMD?

AtRMD crystals were obtained in the presence of NADPH and product analog GDP-D-mannose. Hopefully GDP-D-mannose or at least GDP (as seen in GMD) is bound in the active site. The 4-keto-6-deoxy intermediate could also be modeled in using MUR1

as a guide (as in *P. aeruginosa* GMD). Our hope is to get a picture of the GDP-sugar bound in the active site to get an idea of what residues may be involved in substrate binding and/or catalysis. Furthermore, we hope to see if the GDP moiety is bound in a *syn* conformation, as seen in GMD and MUR1, or the more commonly seen *anti* conformation. The ability of the nucleotide to adopt this unusual, strained conformation may be related to the substrate specificity for GMD or RMD enzymes.

- Will the structures of AtRMD and PaRMD be comparable?

As previously mentioned, a data set was also collected on one of the PaRMD crystals. The diffraction data is only 3.7 Å and phase determination by the molecular replacement method has been unsuccessful. However, AtRMD may serve as a better search model. Although the resolution of the data will not be able to reveal much beyond the backbone trace, perhaps we will be able to see evidence as to why the crystals diffract so poorly compared to AtRMD (i.e. flexible loops, surface areas interfering with crystal packing, etc.). This may lead to ideas for modifying the PaRMD in hopes to obtaining a version of the protein that would provide better diffracting crystals.

- What? GMD is bifunctional?

Recombinant GMD was able to convert GDP-D-mannose to GDP-D-rhamnose, exhibiting both dehydratase and reductase activities. However, this raises several questions. (1) Is there a minimal amount of reduced cofactor where GMD is still bifunctional? Experiments were performed in excess NADPH conditions, far in excess of natural conditions *in vivo*. The CE-based assays could be completed in various NADPH concentrations to determine the minimal concentration necessary. (2) Can it be assumed that the GMD tetramer falls apart to bind the reduced form of the cofactor? The

catalytic mechanism of GMD indicates that the NADP cofactor is regenerated with each round of catalysis and remains tightly bound. It would be interesting to follow the oligomeric state through these experiments via gel filtration, dynamic light scattering or ultracentrifugation, and also determine if GMD possesses reductase activity as a dimer or tetramer. (3) Last, but certainly not least, does GMD possess bifunctionality *in vivo*? Our collaborators at the University of Guelph have a *P. aeruginosa* A-band LPS knockout mutant available. The plasmid containing the gmd gene could be transformed into the bacterium and the LPS analyzed for the presence of D-rhamnose to see if GMD saves the less virulent phenotype.

APPENDIX

APPENDIX A: MBP-HuPLSCR1 AND THE Abl-SH3 DOMAIN

A.1. Introduction

Human phospholipid (PL) scramblase 1 (HuPLSCR1) is a plasma membrane protein that has been implicated in the transbilayer movement of plasma membrane PLs upon the increase of intracellular Ca^{2+} due to cell activation, cell injury or apoptosis [106]. The newly surface-exposed PLs play a central role in promoting blood coagulation and have been implicated in the clearance of injured or apoptotic cells. HuPLSCR1 is a 35 kDa type 2 membrane protein with a single transmembrane segment near the C-terminus. An EF-hand-like Ca^{2+}-binding segment immediately precedes the transmembrane domain. The cytoplasmic domain also includes multiple proline-rich motifs resembling Src homology 3 (SH3) domain-binding sites. HuPLSCR1 has marked specificity for binding to the SH3 domain of Abelson [99] tyrosine kinase, which phosphorylates Tyr residues of HuPLSCR1 [103]. In fact, studies suggest that HuPLSCR1 is the normal substrate of Abl tyrosine kinase [103].

The DNA encoding HuPLSCR1 and the Abl-SH3 domain were given to our lab for the purpose of crystallizing the proteins that they encode. The original goal of this project was to obtain the crystal structures of the MBP-HuPLSCR1 fusion protein in complex with the Abl-SH3 domain. This appendix contains a short summary of the work that was completed to express, purify, crystallize and test the binding of MBP-HuPLSCR1 and the Abl-SH3 domain. This work was the inspiration that led to the strategy for expression, purification and crystallization of problematic proteins presented in Chapter 4.

A.2. Experimental Procedures

A.2.1. Expression and Purification of HuPLSCR1. The cDNA encoding HuPLSCR1 had been cloned into pMAL-C2 (New England Biolabs) as previously reported [106] and provided to our lab from Dr. Peter Sims of Scripps Research Institute, La Jolla, California. The expression protocol in section 4.2.3. was used to express MBP-HuPLSCR1. The same protocol in section 4.2.3. for running the amylose column was used for the first step in the purification of MBP-HuPLSCR1. Further purification was carried out by FPLC using a HiTrap™ Q anion exchange column (Amersham Pharmacia Biotech) equilibrated with Buffer A (20 mM Tris pH 8.5). Protein was eluted over a 40 ml salt gradient using Buffer B (20 mM Tris pH 8.5, 1 M NaCl). Fractions were analyzed by SDS-PAGE.

A.2.2. Expression and Purification of the Abl-SH3 domain. The Abl-SH3 domain was expressed and purified using the protocol under section 4.2.4.

A.2.3. *In vitro* binding of MBP-HuPLSCR1 to the Abl-SH3 domain. For binding, 100 µl of MBP-HuPLSCR1 (10 mg/ml) was mixed with 10 µl of the Abl-SH3 domain (10 mg/ml) for an approximately 1:1 molar ratio. The mixture was allowed to sit on ice for 2 h, and run over a Sephadex G-50 spin column to clear any excess Abl-SH3 domain; then concentrated over a 30 kDa centricon. The samples were analyzed by an SDS-PAGE gel.

A.2.4. Crystallization of MBP-HuPLSCR1, the Abl-SH3 domain and MBP-HuPLSCR1:Abl-SH3 domain complex. Purified MBP-HuPLSCR1, the Abl-SH3 domain or MBP-HuPLSCR1:Abl-SH3 domain complex at 10 mg/ml was used to set up crystallization screens using the microbatch method with the Impax I-5 robot (Douglas

Instruments). A total of 198 commercially available conditions were screened by combining 1 µl of purified protein with 1 µl precipitating solution. Crystallization conditions were refined using the handing drop vapor diffusion method. To aid in crystallization, 0.5 mM maltose and/or 0.5 mM $CaCl_2$ were added to the protein sample.

A.3. Results

Active HuPLSCR1 had been previously cloned, expressed and purified as an MBP-fusion protein [106]. A similar expression and amylose column purification scheme was used. An additional step of purification over a HiTrap Q anion exchange column was added in preparation for crystallization trials. The purification was followed by SDS-PAGE (Figure 32). About 3-4 mg of purified MBP-HuPLSCR1 (78 kDa) was obtained from 1 L of cell culture. Results from the Abl-SH3 domain expression and purification can be found in section 4.3.3 and Figure 29.

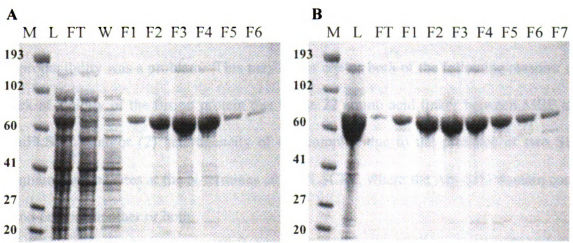


Figure 32. SDS-PAGE of samples from a typical MBP-HuPLSCR1 purification

(**A**) Samples from the amylose column purification: load (L), flow through (FT), wash (W) and elution fractions (F). (**B**) Samples from the anion exchange column purification: load (L), flow through (FT), elution fractions (F). Molecular weight marker (M) standards shown in kDa to the left of each gel.

Immunoprecipitation experiments have shown that HuPLSCR1 interacts with Abl tyrosine kinase, most likely through the interaction of the SH3 domain of Abl with one or more of the proline-rich domains of HuPLSCR1 [103]. Purified MBP-HuPLSCR1 was bound to the purified Abl-SH3 domain at a 1:1 molar ratio and run over a gel filtration column to clear any excess Abl-SH3 domain. The complex was then concentrated using a 30 kDa centricon. The process was followed by SDS-PAGE (Figure 33). MBP-HuPLSCR1 (78 kDa) and the Abl-SH3 domain (6 kDa) are seen in Lane 3 as the mixture before Sephadex G-50 clarification and concentration over a 30 kDa centricon. Lane 4 shows the Sephadex G-50 flow through. After concentration of the flow through over a 30 kDa centricon, the Abl-SH3 domain can clearly be seen (Lane 5). As a control, Lane 7 shows the Abl-SH3 domain at the same concentration as that in Lane 3. Once certain that MBP-HuPLSCR1 was binding the Abl-SH3 domain, crystallization trials were set up with the complex as well as MBP-HuPLSCR1 and the Abl-SH3 domain as controls. Small crystals were seen with all three samples; of note, thin rods were seen in differing conditions for MBP-HuPLSCR1 and the complex (data not shown). However, reproducibility was a problem. This may be for one or both of the following reasons: (1) lack of rigidity of the fusion protein due to the 22 amino acid linker between MBP and HuPLSCR1 and/or (2) heterogeneity of the complex due to the presence of two SH3 domain-binding sites at the N-terminus of HuPLSCR1, where the Abl-SH3 domain could bind one or the other or both.

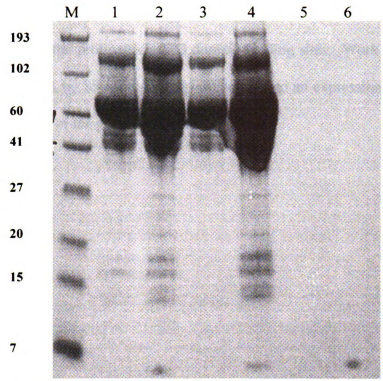


Figure 33. SDS-PAGE of MBP-HuPLSCR1/Abl-SH3 domain binding

(1) 1:1 molar ratio of MBP-scarmblase:Abl-SH3; (2) the sample in (1) concentrated by half; (3) the gel filtration flow through; (4) the 30 kDa retentate; (5) the 30 kDa flow through; (6) the Abl-SH3 domain at a comparable concentration to that of (2). MBP-HuPLSCR1 is 78 kDa and the Abl-SH3 domain is 6 kDa. Molecular weight marker (M) standards shown in kDa to the left of the gel.

To address the flexibility of the system, work was completed to cleave the linker and the MBP tag from HuPLSCR1 using Factor Xa. However, as mentioned in Chapter 4, the protease cleavage step is expensive and optimization is tedious; also in this case, it added heterogeneity to the system because cleavage was incomplete. Although we would have liked to pursue the structure of the HuPLSCR1:Abl-SH3 domain complex with or without MBP, it was apparent that it was problematic. Still interested in the structure of HuPLSCR1 alone, a different approach was considered. An alignment of HuPLSCR1 to HuPLSCR2 and two mouse homologs MuPLSCR1 and MuPLSCR2 indicates that the significant similarities between the amino acid sequences starts at residue number 85

(numbering based on HuPLSCR1). In fact, HuPLSCR2 lacks the entire N-terminal portion containing the proline-rich SH3 domain-binding sites. Work has been started to shorten HuPLSCR1 by 85 residues and re-clone it into an expression vector in hopes of obtaining a version of the protein that will crystallize.

LITERATURE CITED

LITERATURE CITED

1. Yamamoto, K., Katayama, I., Onoda, Y., Inami, M., Kumagai, H. & Tochikura, T. (1993) Evidence that the enzyme catalyzing the conversion of guanosine diphosphate D-mannose to a 4-keto sugar nucleotide intermediate requires nicotinamide adenine dinucleotide phosphate, *Arch. Biochem. Biophys. 300*, 694-698.

2. Sturla, L., Bisso, A., Zanardi, D., Benatti, U., De Flora, A. & Tonetti, M. (1997) Expression, purification and characterization of GDP-D-mannose 4,6-dehydratase from *Escherichia coli*, *FEBS Lett. 412*, 126-130.

3. Kneidinger, B., Graninger, M., Adams, G., Puchberger, M., Kosma, P., Zayni, S. & Messner, P. (2001) Identification of two GDP-6-deoxy-D-*lyxo*-4-hexulose reductases synthesizing GDP-D-rhamnose in *Aneurinibacillus thermoaerophilus* L420-91T, *J. Biol. Chem. 276*, 5577-5583.

4. Wu, B., Zhang, Y. & Wang, P. G. (2001) Identification and characterization of GDP-D-mannose 4,6-dehydratase and GDP-L-fucose synthetase in a GDP-L-fucose biosynthetic gene cluster from *Helicobacter pylori*, *Biochem. Biophys. Res. Commun. 285*, 364-371.

5. Bonin, C. P., Potter, I., Vanzin, G. F. & Reiter, W.-D. (1997) The MUR1 gene of *Arabidopsis thaliana* encodes an isoform of GDP-D-mannose-4,6-dehydratase, catalyzing the first step in the *de novo* synthesis of GDP-L-fucose, *Proc. Natl. Acad. Sci. USA. 94*, 2085-2090.

6. Broschat, K. O., Chang, S. & Serif, G. (1985) Purification and characterization of GDP-D-mannose 4,6-dehydratase from porcine thyroid, *Eur. J. Biochem. 153*, 397-401.

7. Sullivan, F. X., Kumar, R., Kriz, R., Stahl, M. L., Xu, G.-Y., Rouse, J., Chang, X.-j., Boodhoo, A., Potvin, B. & Cumming, D. A. (1998) Molecular cloning of human GDP-mannose 4,6-dehydratase and reconstitution of GDP-fucose biosynthesis *in vitro*, *J. Biol. Chem. 273*, 8193-8202.

8. Rocchetta, H. L., Burrows, L. L. & Lam, J. S. (1999) Genetics of O-antigen biosynthesis in *Pseudomonas aeruginosa*, *Microbiol. Mol. Biol. Rev. 63*, 523-553.

9. Molicaro, A., Silip, A., Lanzetta, R., Newman, M. A., Dow, J. M. & Parrilli, M. (2003) Structural elucidation of the O-chain of the lipopolysaccharide from *Xanthomonas campestris* strain 8004, *Carbohydr. Res. 338*, 277-281.

10. Kocharova, N. A., Knirel, Y. A., Widmalm, G., Jansson, P. E. & Moran, A. P. (2000) Structure of an atypical O-antigen polysaccharide of *Helicobacter pylori* containing a novel monosaccharide 3-C-methyl-D-mannose, *Biochemistry. 39*, 4755-4760.

11. Yokata, S., Kaya, S., Sawada, S., Kawamura, T., Araki, Y. & Ito, E. (1987) Characterization of a polysaccharide component of lipopolysaccharide from *Pseudomonas aeruginosa* IID 1008 (ATCC 27584) as D-rhamnan, *Eur. J. Biochem. 167*, 203-209.

12. Ginsburg, V. (1960) Formation of guanosine diphosphate L-fucose from guanosine diphosphate D-mannose, *J. Biol. Chem. 253*, 2196-2201.

13. Liao, T.-H. & Barber, G. A. (1972) Purification of guanosine 5'-diphosphate D-mannose oxidoreductase from *Phaseolus vulgaris*, *Biochim. Biophys. Acta. 70*, 794-798.

14. Overton, K. & Serif, G. (1981) Synthesis of L-fucose in thyroid tissue, *Biochim. Biophys. Acta. 675*, 281-284.

15. Lopez-Lara, I. M., Blok-Tip, L., Quinto, C., Garcia, M. L., Bloemberg, G. V., Lamers, G. E., Lugtenberg, B. J., Thomas-Oates, J. & Spaink, H. P. (1996) NodZ of Bradyrhizobium extends the nodulation host range of Rhizobium by adding a fucosyl residue to nodulation signals, *Mol. Microbiol. 21*, 397-408.

16. Mergaert, P., Van Mantago, M. & Holsters, M. (1997) The modulation gene *nolK* of *Azorizobium caulinodans* is involved in the formation of GDP-fucose from GDP-mannose, *FEBS Lett. 409*, 312-316.

17. Varki, A. (1994) Selectin ligands, *Proc. Natl. Acad. Sci. USA. 91*, 7390-7497.

18. Maki, M., Jarvinen, N., Rabina, J., Maaheimo, H., Mattila, P. & Renkonen, R. (2003) Cloning and functional expression of a novel GDP-6-deoxy-D-talose synthetase from *Actinobacillus actinomycetemcomitans*, *Glycobiology. 4*, 295-303.

19. Albermann, C. & Piepersberg, W. (2001) Expression and identification of the RfbE protein from *Vibrio cholerae* O1 and its use for the enzymes synthesis of GDP-D-perosamine, *Glycobiology. 11*, 655-661.

20. Suzuki, N., Nakano, Y., Yoshido, Y., Nezu, T., Terada, Y., Yamashita, Y. & Koga, T. (2002) Guanosine diphosphate-4-keto-6-deoxy-D-mannose reductase in the pathway for the synthesis of GDP-6-deoxy-D-talose in *Actinobacillus actinomycetemcomitans*, *Eur. J. Biochem. 269*, 5963-5971.

21. Allard, S. T. M., Giraud, M.-F., Whitfield, C., Graninger, M., Messner, P. & Naismith, J. H. (2001) The crystal structure of dTDP-D-glucose 4,6-dehydratase (RmlB) from *Salmonell enterica* serovar *typhimurium*, the second enzyme in the dTDP-L-rhamnose pathway, *J. Mol. Biol. 307*, 283-295.

22. Allard, S. T. M., Giraud, M.-F., Whitfield, C., Messner, P. & Naismith, J. H. (2000) The purification, crystallization and structural elucidation of dTDP-D-glucose 4,6-

dehydratase (RmlB), the second enzyme of the dTDP-L-rhamnose synthesis pathway from *Salmonella enterica* serovar *typhimurium*, *Acta Crystallogr. D. 56*, 222-225.

23. Giraud, M.-F., Leonard, G. A., Field, R. A., Berlind, C. & Naismith, J. H. (2000) RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase, *Nat. Struct. Biol. 7*, 398-402.

24. Giraud, M.-F., McMiken, H. J., Leonard, G. A., Messner, P., Whitfield, C. & Naismith, J. H. (1999) Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from *Salmonella enterica* serovar *typhimurium*., *Acta Crystallogr. D. 55*, 2043-2046.

25. Graninger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C. & Messner, P. (1999) Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* Serovar *Typhimurium* LT2, *J. Biol. Chem. 274*, 25069-25077.

26. Blankenfeldt, W., Kerr, I. D., Giraud, M.-F., McMiken, H. J., Leonard, G. A., Whitfield, C., Messner, P., Graninger, M. & Naismith, J. H. (2002) Variation on a theme of SDR: dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase (RmlD) shows a new Mg^{2+}-dependent dimerization mode, *Structure. 10*, 773-786.

27. Gross, J. W., Hegeman, A. D., Gerratana, B. & Frey, P. A. (2001) Dehydration is catalyzed by glutamate-136 and aspartic acid-135 active site residues in *Escherichia coli* dTDP-glucose 4,6-dehydratase., *Biochemistry. 40*, 12497-12504.

28. Gross, J. W., Hegeman, A. D., Vestling, M. M. & Frey, P. A. (2000) Pre-steady-state kinetics of deoxythymidine diphosphate glucose 4,6-dehydratase: detection and quantification of reaction intermediates by mass spectroscopy, *Biochemistry. 39*, 1559.

29. Gross, J. W., Hegeman, A. D., Vestling, M. M. & Frey, P. A. (2000) Characterization of enzymatic processes by rapid mix-quench mass spectrometry: the case of dTDP-glucose 4,6-dehydratase, *Biochemistry. 39*, 13633-13640.

30. Hegeman, A. D., Gross, J. W. & Frey, P. A. (2001) Probing catalysis by *Escherichia coli* dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of functional amino acid residues at the active site, *Biochemistry. 40*, 6598-6610.

31. Hegeman, A. D., Gross, J. W. & Frey, P. A. (2002) Concerted and stepwise dehydration mechanisms observed in wild-type and mutated *Escherichia coli* dTDP-glucose 4,6-dehydratase, *Biochemistry. 41*, 2797-2804.

32. Gerratana, B., Cleland, W. W. & Frey, P. A. (2001) Mechanistic roles of Thr134, Tyr160, and Lys164 in the reaction catalyzed by dTDP-glucose 4,6-dehydratase, *Biochemistry. 40*, 9187-9195.

33. Maki, M. & Renkonen, R. (2004) Biosynthesis of 6-deoxyhexose glycans in bacteria, *Glycobiology. 14*, 1R-15R.

34. Nakano, Y., Suzuki, N., Yoshida, Y., Nezu, T., Yamashita, Y. & Koga, T. (2000) Thymidine diphosphate-6-deoxy-L-*lyxo*-4-hexulose reductase synthesizing dTDP-6-deoxy-L-talose from *Actinobacillus actinomycetemcomitans*, *J. Biol. Chem. 275*, 6806-6812.

35. Yoshida, Y., Nakano, Y., Nezu, T., Yamashita, Y. & Koga, T. (1999) A novel NDP-6-deoxyhexosyl-4-ulose reductase in the pathway for the synthesis of thymidine diphosphate-D-fucose, *J. Biol. Chem. 274*, 16933-16939.

36. Liu, H.-w. & Thorson, J. S. (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria, *Annu. Rev. Microbiol. 48*, 223-256.

37. Hallis, T. M., Lei, Y., Que, N. L. S. & Liu, H.-w. (1998) Mechanistic studies of the biosynthesis of paratose: purification and characterization of CDP-paratose synthase, *Biochemistry. 37*, 4935-4945.

38. Verma, N. & Reeves, P. R. (1989) Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae, *J. Bacteriol. 171*, 5694-701.

39. Wyk, P. & Reeves, P. R. (1989) Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B salmonellae: homology with galactose epimerase, *J. Bacteriol. 171*, 5687-93.

40. Koropatkin, N. M., Liu, H.-w. & Holden, H. M. (2003) High resolution X-ray structure of tyvelose epimerase from *Salmonella typhi*, *J. Biol. Chem. 278*, 20874-20881.

41. Liu, Y., Thoden, J. B., Kim, J., Berger, E., Gulick, A. M., Ruzicka, F. J., Holden, H. M. & Frey, P. A. (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry. 36*, 10675-10684.

42. Allard, S. T. M., Beis, K., Giraud, M.-F., Hegeman, A. D., Gross, J. W., Wilmouth, R. C., Whitfield, C., Graninger, M., Messner, P., Allen, A. G., Maskell, D. J. & Naismith, J. H. (2002) Toward a structural understanding of the dehydratase mechanism, *Structure. 10*, 81-92.

43. Mulichak, A. M., Theisen, M. J., Essigmann, B., Benning, C. & Garavito, R. M. (1999) Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose, *Proc. Natl. Acad. Sci. USA. 96*, 13097-13102.

44. Mulichak, A. M., Bonin, C. P., Reiter, W.-D. & Garavito, R. M. (2002) Structure of MUR1 GDP-mannose 4,6-dehydratase from *Arabidopsis thaliana*: implications for ligand binding specificity, *Biochemistry. 41*, 15578-15589.

45. Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. & Holden, H. M. (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase, *Biochemistry. 39*, 5691-5701.

46. Cryz, S. J., Pitt, T. L., Furer, E. & Germanier, R. (1984) Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*, *Infect. Immun. 44*, 508-513.

47. Engels, W., Endert, J., Kamps, M. A. & van Boven, C. P. (1985) Role of lipopolysaccharide in opsonization and phagocytosis of *Pseudomonas aeruginosa*, *Infect. Immun. 49*, 182-189.

48. Dasgupta, R., de Kievit, T. R., Masoud, H., Altman, E., Richards, J. C., Sadovskaya, I., Speert, D. P. & Lam, J. S. (1994) Characterization of lipopolysaccharide-deficient mutants of *Pseudomonas aeruginosa* derived from serotypes O3, O5, and O6, *Infect. Immun. 62*, 809-817.

49. Markovitz, A. (1964) Biosynthesis of guanosine diphosphate D-rhamnose and guanosine diphosphate D-talomethylose from guanosine diphosphate a-D-Mannose, *J. Biol. Chem. 239*, 2091-2098.

50. Etzioni, A., Frydman, M., Pollack, S., Avidor, I., Phillips, M. L., Paulson, J. C. & Gershoni-Baruch, R. (1992) Brief report: recurrent severe infections caused by a novel leukocyte adhesion deficiency, *N. Engl. J. Med. 327*, 1789-1792.

51. Bisso, A., Sturla, L., Zanardi, D., De Flora, A. & Tonetti, M. (1999) Structural and enzymatic characterization of human recombinant GDP-D-mannose-4,6-dehydratase, *FEBS Lett. 456*, 370-374.

52. Somoza, J. R., Menon, S., Schmidt, H., Joseph-McCarthy, D., Dessen, A., Stahl, M. L., Somers, W. S. & Sullivan, F. X. (2000) Structural and kinetic analysis of *Escherichia coli* GDP-mannose 4,6 dehydratase provides insights into the enzyme's catalytic mechanism and regulation by GDP-fucose, *Structure. 8*, 123-135.

53. Oths, P. J., Mayer, R. M. & Floss, H. G. (1990) Stereochemistry and mechanism of the GDP-mannose dehydratase reaction, *Carbohydr. Res. 198*, 91-100.

54. Maki, M., Jarvinen, N., Rabina, J., Roos, C., Maaheimo, H., Mattila, P. & Renkonen, R. (2002) Functional expression of *Pseudomonas aeruginosa* GDP-4-keto-6-deoxy-D-mannose reductase which synthesizes GDP-rhamnose, *FEBS Lett. 269*, 593-601.

55. Thoden, J. B., Frey, P. A. & Holden, H. M. (1996) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol, *Protein Sci. 5*, 2149-2161.

56. Thoden, J. B., Frey, P. A. & Holden, H. M. (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from *Escherichia coli*: implications for the catalytic mechanism, *Biochemistry. 35*, 5137-5144.

57. Thoden, J. B., Frey, P. A. & Holden, H. M. (1996) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from *Escherichia coli*, *Biochemistry. 35*, 2557-2566.

58. Thoden, J. B., Gulick, A. M. & Holden, H. M. (1997) Molecular structures of the S124A, S124T, and S124V site-directed mutants of UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry. 36*, 10685-10695.

59. Thoden, J. B., Hegeman, A. D., Wesenberg, G., Chapeau, M. C., Frey, P. A. & Holden, H. M. (1997) Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry. 36*, 6294-6304.

60. Thoden, J. B. & Holden, H. M. (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*, *Biochemistry. 37*, 11469-11477.

61. Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. & Holden, H. M. (2001) Molecular basis for severe epimerase-deficiency galactosemia: X-ray structure of the human V94M-substituted UDP-galactose 4-epimerase, *J. Biol. Chem. 23*, 20617-20623.

62. Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. & Holden, H. M. (2001) Human UDP-galactose 4-epimerase. Accommodation of UDP-n-acetylglucosamine within the active site, *J. Biol. Chem. 276*, 15131-15136.

63. Somers, W. S., Stahl, M. L. & Sullivan, F. X. (1998) GDP-fucose synthetase from *Escherichia coli*: structure of a unique member of the short-chain dehydrogenase/reductase family that catalyzes two distinct reactions at the same active site, *Structure. 6*, 1601-1612.

64. Currie, H. L., Lightfoot, J. & Lam, J. S. (1995) Prevalence of gca, a gene involved in synthesis of A-band common antigen polysaccharide in *Pseudomonas aeruginosa*, *Clin. Diagn. Lab. Immunol. 2*, 554-562.

65. Otwinowski, Z. & Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307-326.

66. Navaza, J. (1994) AMoRe: and automated package for molecular replacement, *Acta Crystallogr. A. 50*, 157-163.

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

68. Brunger, A. T. (1998) Crystallography and NMR system (CNS): a new software system for macromolecular structure determination, *Acta Crystallogr. D. 54*, 905-921.

69. CHAIN. (1995) *CHAIN: Crystallographic Modeling Program Version 7.0*, Baylor College of Medicine, Waco, TX.

70. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283-291.

71. He, X., Thorson, J. S. & Liu, H.-w. (1996) Probing the coenzyme and substrate binding events of CDP-D-glucose 4,6-dehydratase: mechanistic implications, *Biochemistry. 35*, 4721-4731.

72. Konopka, J. M., Halkides, C. J., Vanhooke, J. L., Gorenstein, D. G. & Frey, P. A. (1989) UDP-galactose 4-epimerase. Phosphorus-31 nuclear magnetic resonance analysis of NAD and NADH bound at the active site, *Biochemistry. 28*, 9071-9080.

73. Persson, B., Krook, M. & Jornvall, H. (1991) Characteristics of short-chain alcohol dehyrdogenases and related enzymes, *Eur. J. Biochem. 200*, 537-543.

74. Deacon, A. M., Ni, Y. S., Coleman Jr., W. G. & Ealick, S. E. (2000) The crystal structure of ADP-L-glycero-D-mannoheptose 6-epimerase: catalysis with a twist, *Structure. 8*, 453-462.

75. Lightfoot, J. & Lam, J. S. (1991) Molecular cloning of genes involved with expression of A-band lipopolysaccharide, an antigenically conserved form in *Pseudomonas aeruginosa*, *J. Bacteriol. 173*, 5624-5630.

76. Bergfors, T. (2003) Seeds to crystals, *J. Struct. Biol. 142*, 66-76.

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

78. Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices, *J. Mol. Biol. 292*, 195-202.

79. McGuffin, L. J., Bryson, K. & Jones, D. T. (2000) The PSIPRED protein structure prediction server, *Bioinformatics. 16*, 404-405.

80. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM, *J. Mol. Biol. 299*, 499-520.

81. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003) SWISS-MODEL: an automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381-3385.

82. Luft, J. R. & DeTitta, G. T. (1995) Chaperone salts, polyethylene glycol and rates of equilibration in vapor-diffusion crystallization, *Acta Crystallogr. D. 51*, 780-785.

83. Riboldi-Tunnicliffe, A. & Hidgenfeld, R. (1999) Crystallography with oil - an old idea revived, *J. Appl. Cryst. 32*, 1003-1005.

84. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491-197.

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

86. Harp, J. M., Timm, D. E. & Bunick, G. J. (1998) Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography, *Acta Crystallogr. D. 54*, 622-628.

87. Yeh, J. L. & Hol, W. G. (1998) A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase, *Acta Crystallogr. D. 54*, 479-480.

88. Abergel, C. (2004) Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals, *Acta Crystallogr. D. 60*, 1413-1416.

89. Storoni, L. C., McCoy, A. J. & Read, R. J. (2004) Likelihood-enhanced fast rotation functions, *Acta Crystallogr. D. 60*, 432-438.

90. Senchenkova, S. N., Shashkov, A. S., Knirel, Y. A., McGovern, J. J. & Moran, A. P. (1996) The O-specific polysaccharide chain of *Campylobacter fetus* serotype B lipopolysaccharide is a D-rhamnan terminated with 3-O-methyl-D-rhamnose (D-acofriose), *Eur. J. Biochem. 239*, 434-438.

91. Lightfoot, J. & Lam, J. S. (1993) Chromosomal mapping, expression and synthesis of lipopolysaccaride in *Pseudomonas aeruginosa*: a role for guanosine diphosphate (GDP)-D-mannose, *Mol. Microbiol. 8*, 771-782.

92. Rocchetta, H. L., Pacan, J. C. & Lam, J. S. (1998) Synthesis of the A-band polysaccharide sugar D-rhamnose required RMD and WbpW: indentification of multiple AlgA homologues, WbpW and ORF488, in *Pseudomonas aeruginosa*, *Mol. Microbiol. 29*, 1419-1434.

93. Brisson, J. R., Sue, S. C., Wu, W. G., McManus, G., Nghia, P. T. & Uhrin, D. (2002) in *NMR spectroscopy of glycoconjugates* (Jimenez-Barbero, J. & Peters, T., eds) pp. 59-93, Wiley-VCH, Weinhem.

94. Webb, N. A., Mulichak, A. M., Lam, J. S., Rocchetta, H. L. & Garavito, R. M. (2004) Crystal structure of a tetrameric GDP-D-mannose 4,6-dehydratase from a bacterial GDP-D-rhamnose biosynthetic pathway, *Protein Sci. 13*, 529-539.

95. Kobe, B., Center, R. J., Kemp, B. E. & Poumbourios, P. (1999) Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins, *Proc. Natl. Acad. Sci. USA. 96*, 4319-4324.

96. Liu, Y., Manna, A., Li, R., Martin, W. E., Murphy, R. C., Cheung, A. L. & Zhang, G. (2001) Crystal structure of the SarR protein from *Staphylococcus aureus*, *PNAS. 98*, 6877-6882.

97. Ke, A. & Wolberger, C. (2003) Insights into binding cooperativity of MATa1/MATa2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera, *Protein Sci. 12*, 306-312.

98. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags, *Protein Sci. 12*, 1313-1322.

99. Zablackis, E., York, W. S., Pauly, M., Hantus, S., Reiter, W.-D., Chapple, C. C. S., Albersheim, P. & Darvill, A. G. (1996) Substitution of L-fucose by L-galactose in cell walls of *Arabidopsis* mur1, *Science. 272*, 1808-1810.

100. Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. (1992) Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis, *Biochemistry. 31*, 10657-10663.

101. Spurlino, J. C., Lu, G.-Y. & Quiocho, F. A. (1991) The 2.3 Å resolution structure of the maltose- and maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis, *J. Biol. Chem. 266*, 5202-5219.

102. Center, R. J., Kobe, B., Wilson, K. A., Teh, T., Howlett, G. J., Kemp, B. E. & Poumbourios, P. (1998) Crystallization of a trimeric human T cell leukemia virus type I gp21 ectodomain fragment as a chimera with maltose-binding protein, *Protein Sci. 7*, 1612-1619.

103. Sun, J., Zhao, J., Schwartz, M. A., Wang, J. Y. J., Wiedmer, T. & Sims, P. J. (2001) c-Abl tyrosine kinase binds and phosphorylates phospholipid scramblase I, *J. Biol. Chem. 276*, 28984-28990.

104. Musacchio, A., Saraste, M. & Wilmanns, M. (1994) High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides, *Struct. Biol. 1*, 546-551.

105. Pisabarro, M. T. & Serrano, L. (1996) Rational design of specific high-affinity peptide ligands for the Abl-SH3 domain, *Biochemistry. 35*, 10634-10640.

106. Zhou, Q., Zhao, J., Stout, J. G., Luhm, R. A., Wiedmer, T. & Sims, P. J. (1997) Molecular cloning of human plasma membrane phospholipid scramblase, *J. Biol. Chem. 272*, 18240-18244.