Characterization of a *Shigella boydii* B14 Galactosyltransferase

Involved in O-antigen Synthesis

By

CHANGCHANG XU

A thesis submitted to the Department of Biochemistry in conformity with the requirements for the degree of Master of Science

Queen’s University

Kingston, Ontario, Canada

September, 2010

Copyright © Changchang Xu, 2010
ABSTRACT

The O-antigenic glycan chains (O-antigen) of lipopolysaccharides (LPS) in Gram-negative bacteria are synthesized by glycosyltransferases (GTs), most of which are yet to be characterized. The LPS are important for cell surface functions in bacteria and for the host immune system, through influencing the ability of bacteria to invade tissues and to evade the immune system of the host. *Shigella* is a Gram-negative pathogen that causes diarrhea or dysentery in humans. The O-antigenic chain of *Shigella boydii* (*S. boydii*) serotype B14 consists of repeating units with the structure \([\rightarrow6-D-Galp\alpha1\rightarrow4-D-Glc\alpha1\rightarrow6-D-Galp\beta1\rightarrow4-D-Galp\beta1\rightarrow4-D-Glc\alpha1\rightarrow]_n\). None of the genes in the *S. boydii* serotype B14 O-antigen gene cluster, which includes putative GT genes, have been functionally characterized. We have developed the technology to biochemically identify these novel GTs. The *wfeD* gene in the B14 O-antigen gene cluster was proposed to encode a galactosyltransferase (GalT) involved in O-antigen synthesis. We confirmed here that the *wfeD* gene product is a β4GalT that synthesizes the Galβ1-4GlcNAcα-R linkage. WfeD was expressed in *E.coli* and the activity characterized using UDP-[³H]Gal as the donor substrate and the synthetic acceptor substrate GlcNAcα-pyrophosphate-phenyl-undecyl (GlcNAc-PP-PhU). A His-tagged version of the enzyme was purified via Ni²⁺-affinity chromatography, and its disaccharide product was analyzed by liquid chromatography-mass spectrometry (LC-MS), high pressure liquid chromatography (HPLC), 1D and 2D nuclear magnetic resonance (NMR), and galactosidase digestion. The enzyme was shown to be specific for the UDP-Gal donor substrate and required pyrophosphate in the acceptor substrate. Divalent metal ions such as Mn²⁺, Ni²⁺, and surprisingly also Pb²⁺, were able to enhance enzyme activity. Mutational analysis showed
that the Glu101 residue within a DxD motif is essential for activity, possibly by forming the catalytic nucleophile. The Lys211 residue within a cluster of positively charged amino acids was also found to be required for activity. This latter residue may be involved in binding the negatively charged acceptor substrate. Our study revealed that the β4GalT WfeD is a previously unknown enzyme that has extremely low sequence similarities to any other GalT and has an unusual preference for metal ion cofactors.
ACKNOWLEDGEMENTS

First, I would like to thank Dr. Inka Brockhausen for her advice and help through the whole project. I also would like to thank Dr. John Allingham and Dr. Peter Davies for their help and advice on protein purification, Dr. John Schutzbach and Mike Nesheim for their help on enzymology, Darlene Davis from Allingham’s lab for the molecular biology work, Sherry Gauthier from from Davies’ lab for the help on the western blot experiments, Dr. Xiaojing Yang for the help on the bacterial culture and the enzyme assay, and Dr. Walter Szarek, Dr. Pedro Montoya, Dr. John Riley, Dr. Ole Hindsgaul (Carlsberg Inst. Denmark) for the compounds they provided. I am grateful to Dr. Francoise Sauriol in the Chemistry department for the help with NMR experiments, and David McLeod from PFD for performing the MALDI-TOF MS for my protein. I am also very grateful for the enzyme plasmid and sequence analysis provided by Bin Liu, Bo Hu, Lu Feng and Lei Wang from Nankai University, Tianjin, China.

This work was supported by GIDRU (Gastro-Intestinal Diseases Research Unit) and NSERC (Natural Sciences and Engineering Research Council).
TABLE OF CONTENTS

ABSTRACT.........................................................................................................................ii
ACKNOWLEDGEMENTS........................................................................................................iv
TABLE OF CONTENTS........................................................................................................v
LIST OF ABBREVIATIONS....................................................................................................viii
LIST OF TABLES..................................................................................................................xii
LIST OF FIGURES.................................................................................................................xiii
CHAPTER 1 INTRODUCTION................................................................................................. - 1 -
  1.1 Lipopolysaccharides (LPS) and O-antigen in the cell environment......................... - 1 -
    1.1.1 LPS and O-antigen structures............................................................................. - 1 -
    1.1.2 LPS as an endotoxin that induces intestinal disease ....................................... - 5 -
  1.2 *Shigella* are involved in gastrointestinal disease .................................................. - 7 -
    1.2.1 Pathogenicity of Shigellosis .............................................................................. - 7 -
    1.2.2 Role of O-antigen in Shigellosis and O-antigen vaccine development ........... - 9 -
  1.3 Biosynthesis of O-LPS in *Shigella* ........................................................................ - 10 -
  1.4 Identification of glycosyltransferases by sequence similarity .................................. - 12 -
    1.4.1 Large, uncharacterized glycosyltransferase pool.............................................. - 12 -
    1.4.2 Identification of *Shigella* B14 putative glycosyltransferases ......................... - 15 -
    1.4.3 Comparison of mammalian and bacterial galactosyltransferases for O-
      polysaccharide synthesis...................................................................................... - 15 -
  1.5 The DxD motif is involved in the catalytic mechanism of glycosyltransferases - 17 -
  1.6 Project overview......................................................................................................... - 22 -
    1.6.1 Project rationale .............................................................................................. - 22 -
    1.6.2 Project hypothesis .......................................................................................... - 22 -
    1.6.3 Objectives ....................................................................................................... - 23 -
CHAPTER 2 MATERIALS AND METHODS...................................................................... - 24 -
  2.1 DNA plasmid construction and expression of the wfeD gene ......................... - 24 -
2.1.1 Construction of the mutants .................................................................- 24 -
2.1.2 Transformation of plasmid constructs into the expression cell lines.........- 26 -
2.1.3 Expression of wild type and mutant wfeD in E.coli BL21 cells ..............- 27 -
2.2 Galactosyltransferase activity and inhibition assays .............................- 27 -
2.3 Structural analysis of the reaction product ..............................................- 29 -
  2.3.1 Product identification using mass spectrometry and HPLC .................- 29 -
  2.3.2 Product linkage confirmation by galactosidases .............................- 29 -
  2.3.3 Product analysis by NMR ..............................................................- 31 -
2.4 Enzyme purification and analysis of the protein by SDS-PAGE and mass spectrometry.................................................................- 31 -
2.5 Detergent test on enzyme activity and separation of protein and membrane aggregate pellet ..............................................................- 32 -
2.6 Western blot analysis of purified WfeD (wild type and mutants)..............- 33 -
2.7 Testing enzyme activity with amino acid modifying reagents..................- 34 -

CHAPTER 3 RESULTS ....................................................................................- 35 -

3.1 Enzyme characterization of crude and purified WfeD protein .................- 35 -
  3.1.1 Sequence hydrophobicity analysis for the potentially membrane-associated region .................................................................- 35 -
  3.1.2 Detergent test on enzyme activity and the separation of protein and membrane-aggregate pellet ......................................................- 35 -
  3.1.3 Enzyme purification by Ni²⁺-NTA chromatography .............................- 40 -
  3.1.4 Product confirmation by mass spectrometry and western blot ..........- 41 -
  3.1.5 WfeD kinetics and storage stability of crude and purified enzyme forms...- 46 -
3.2 Confirmation of Gal transfer and newly synthesized linkage of Galβ1-4GlcNAc.................................................................- 50 -
  3.2.1 Molecular weight confirmation of the reaction product ......................- 50 -
  3.2.2 Confirmation of the β-linkage in the reaction product .......................- 52 -
  3.2.3 Linkage identification of the product by NMR .................................- 52 -
3.3 Substrate specificity, enzyme inhibition and cofactor requirement of WfeD.....- 58 -
3.3.1 Donor substrate specificity .......................................................... 58 -
3.3.2 Role of pyrophosphate in acceptor substrate specificity .................. 60 -
3.3.3 Screening of compounds as potential inhibitors ............................. 60 -
3.3.4 Divalent metal ions enhance WfeD enzyme activity ........................ 66 -
3.4 Role of amino acid residues in enzyme activity .................................. 69 -
3.4.1 Effect of amino acid modifying reagents on WfeD activity ............... 69 -
3.4.2 Sequence comparison of WfeD and putative glycosyltransferases for the selection of crucial amino acids .......................................................... 72 -
3.4.3 Expression confirmation of WfeD mutant proteins ......................... 76 -
3.4.4 GalT activity of the WfeD mutants .................................................. 81 -
CHAPTER 4   GENERAL DISCUSSION AND FUTURE DIRECTIONS ............... 85 -
4.1 General discussion .................................................................................. 85 -
4.1.1 Membrane requirement of WfeD enzyme activity .......................... 85 -
4.1.2 Enzyme purification efficacy and enzyme stability ........................ 86 -
4.1.3 Enzyme kinetics and stability ........................................................... 87 -
4.1.4 GlcNAcα-PP-PhU, acts as a good natural substrate analog in vitro ...... 88 -
4.1.5 Insights into inhibitor development .................................................. 88 -
4.1.6 Activity stimulation by divalent metal ions ...................................... 89 -
4.1.7 Involvement of the DxD motif in WfeD activity ............................... 90 -
4.1.8 Involvement of RKR in WfeD activity ............................................. 91 -
4.1.9 Role of Cys residue(s) in enzyme activity or stability ...................... 91 -
4.2 Future directions ..................................................................................... 92 -
4.2.1 Further examination of Glu101 and Lys211 involvement in WfeD activity - 92 -
4.2.2 Further improvements of protein solubility and purity ....................... 93 -
4.2.3 Further investigations on the role of the two Cys residues in WfeD activity ............................................................... 94 -
4.2.4 WfeD cellular localization determination ........................................ 95 -
4.2.5 Significance for intestinal disease .................................................... 95 -
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>AG1</td>
<td>Affinity gel</td>
</tr>
<tr>
<td>Bn</td>
<td>Benzyl</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>CMP-SA</td>
<td>Cytidine monophosphate-neuraminic acid</td>
</tr>
<tr>
<td>COSY</td>
<td>Correlation spectroscopy</td>
</tr>
<tr>
<td>cpm</td>
<td>Counts per minute</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>Den</td>
<td>Density</td>
</tr>
<tr>
<td>DTNB</td>
<td>5,5’-dithiobis-2-nitrobenzoic acid</td>
</tr>
<tr>
<td>DxD</td>
<td>Aspartate-x-aspartate, x can be any amino acid</td>
</tr>
<tr>
<td>DxE</td>
<td>Aspartate-x-glutamate, x can be any amino acid</td>
</tr>
<tr>
<td>DYEIE</td>
<td>Aspartate-Tyrosine-Glutamate-Isoleucine-Glutamate</td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene diamine tetraacetate, C_{10}H_{14}N_{2}O_{8}Na_{2} •H_{2}O</td>
</tr>
<tr>
<td>EtOH</td>
<td>Ethanol</td>
</tr>
<tr>
<td>FAE</td>
<td>Follicle- associated epithelium</td>
</tr>
<tr>
<td>Fuc</td>
<td>Fucose</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>GalA</td>
<td>Galacturonic acid</td>
</tr>
<tr>
<td>Galf</td>
<td>Galactofuranose</td>
</tr>
<tr>
<td>Galp</td>
<td>Galactopyranose</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-acetyl galactosamine</td>
</tr>
<tr>
<td>β4 GalT</td>
<td>β4 galactosyltransferase</td>
</tr>
<tr>
<td>GI</td>
<td>Gastro-intestinal</td>
</tr>
<tr>
<td>Glc</td>
<td>Glucose</td>
</tr>
<tr>
<td>GlcA</td>
<td>Glucuronic acid</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl glucosamine</td>
</tr>
<tr>
<td>GlcNAc-P</td>
<td>GlcNAc-phosphate</td>
</tr>
<tr>
<td>GlcNAcα-PP-Und</td>
<td>GlcNAcα-PO_3-PO_3-undecaprenol</td>
</tr>
<tr>
<td>GlcNAcα-PP-PhU</td>
<td>GlcNAcα-PO_3-PO_3-(CH_2)_11-OPhenyl</td>
</tr>
<tr>
<td>GlcNBu</td>
<td>N-butyryl glucosamine</td>
</tr>
<tr>
<td>GT</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>HPG</td>
<td>p-Hydroxyphenylglyoxal</td>
</tr>
<tr>
<td>HPLC</td>
<td>High pressure liquid chromatography</td>
</tr>
<tr>
<td>HPR</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>HSQC</td>
<td>Heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>IAA</td>
<td>Iodoacetic acid</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IpaB</td>
<td>Invasion plasmid antigen B</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropylthiogalactoside</td>
</tr>
<tr>
<td>kDa</td>
<td>Kilo Dalton</td>
</tr>
<tr>
<td>KDO</td>
<td>Keto-deoxyoctulosonate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LBP</td>
<td>LPS binding protein</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharides</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix-assisted laser desorption/ionization-time of flight</td>
</tr>
<tr>
<td>Man</td>
<td>Mannose</td>
</tr>
<tr>
<td>MBP</td>
<td>Maltose binding protein</td>
</tr>
<tr>
<td>MDO</td>
<td>Membrane-derived oligosaccharide</td>
</tr>
<tr>
<td>MeOH</td>
<td>Methanol</td>
</tr>
<tr>
<td>MES</td>
<td>2-[N-Morpholine]ethanesulfonic acid, $C_6H_{13}NO_4S\cdot H_2O$</td>
</tr>
<tr>
<td>MHz</td>
<td>Megahertz</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td><em>Neisseria meningitidis</em></td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NTA</td>
<td>Nitrilotriacetic acid</td>
</tr>
<tr>
<td>NTP</td>
<td>Nucleoside triphosphate</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polycrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PPEtn</td>
<td>Pyrophosphorylethanolamine</td>
</tr>
<tr>
<td>Pyr</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>Rha</td>
<td>Rhamnose</td>
</tr>
<tr>
<td>RKR</td>
<td>Arginine-Lysine-Arginine</td>
</tr>
<tr>
<td>S. boydii</td>
<td><em>Shigella boydii</em></td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumor necrosis factor-alpha</td>
</tr>
<tr>
<td>TLR4</td>
<td>Toll-like receptor 4</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UndP</td>
<td>Undecaprenol phosphate</td>
</tr>
<tr>
<td>Und-PP</td>
<td>Undecaprenol pyrophosphate</td>
</tr>
<tr>
<td>UDP</td>
<td>Uridine diphosphate</td>
</tr>
<tr>
<td>vs</td>
<td>Versus</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Examples of functionally characterized mammalian and bacterial GalT enzymes from CAZy database.................................................................- 18 -

Table 2. List of forward and reverse primers used in single amino acid mutagenesis..- 25 -

Table 3. The pixel values of the bands on the western blot film (Figure 16C), and calculated relative band density. .......................................................................................- 45 -

Table 4. $^1$H and $^{13}$C NMR parameters established from 1D $^1$H NMR, 2D COSY and $^1$H-$^{13}$C HSQC.......................................................................................................- 57 -

Table 5. WfeD GalT acceptor substrate specificity.............................................- 61 -

Table 6. Screening of potential inhibitors of WfeD GalT. .....................................- 65 -

Table 7. Western blot (Figure 31) band density calculated by ImageJ and the estimated relative amount of His-tagged protein in Ni$^{2+}$-NTA purified WfeD mutants...........- 80 -
# List of Figures

Figure 1. A model of inner and outer cell membranes of *E.coli* K-12. .................................. 2 -

Figure 2. An example of a cell surface LPS, from *E.coli* O111:B4. ................................. 3 -

Figure 3. Three O-antigen repeating unit structures that are identical in different serotypes of *Shigellae* and *E.coli*. .................................................................................. 4 -

Figure 4. LPS stimulates the Toll-like receptor involved in signalling pathways in a macrophage cell. ........................................................................................................ 6 -

Figure 5. Scheme of *Shigella* pathogenicity: rupture, invasion, and inflammatory destruction of the intestinal barrier. ........................................................................... 8 -

Figure 6. One synthesis cycle of a single O-unit of *S. boydii* serotype B14. .............. 11 -

Figure 7. The proposed Wzx/Wzy dependent O-antigen synthetic pathway in *Shigella* B14. .................................................................................................................. 13 -

Figure 8. The O-antigen gene cluster of *S. boydii* serotype B14. ................................ 16 -

Figure 9. The overall structure of the LgtC monomer (1G9R) with bound substrate analogs. ................................................................................................................. 19 -

Figure 10. The DxD motif in the WfeD sequence and the mechanism involving the DxD motif. .................................................................................................................. 20 -

Figure 11. HPLC separation of the purified reaction product. ........................................ 30 -

Figure 12. Kyte & Doolittle hydropathy profile of the WfeD sequence (as shown in Figure 10A). .............................................................................................................. 36 -

Figure 13. Detergent effect on WfeD enzyme activity. .................................................... 38 -

Figure 14. SDS-PAGE using Coomassie Blue staining: detergent effect on the separation of protein and membrane aggregate pellet. .............................................. 39 -

Figure 15. Analysis of the Ni$^{2+}$-NTA purified WfeD with linear gradient of imidazole. ................................................................. 42 -

Figure 16. Analysis of small scale Ni$^{2+}$-NTA purified WfeD........................................ 44 -

Figure 17. Kinetics of crude WfeD enzyme. ..................................................................... 47 -

Figure 18. Kinetics of purified WfeD enzyme. ................................................................. 48 -
Figure 19. Stability of GalT activity of WfeD. .............................................................. - 49 -

Figure 20. Electrospray MS (negative ion mode) of the isolate reaction product Gal-GlcNAc-PP-PhU (MW. 789), containing substrate GlcNAc-PP-PhU (MW. 627)........ - 51 -

Figure 21. Galactose cleavage of the galactosidase-treated reaction product.............. - 53 -

Figure 22. 600 MHz proton NMR spectrum of the WfeD product............................... - 56 -

Figure 23. Donor specificity of the crude enzyme, using seven nucleotide sugars at the indicated concentrations................................................................. - 59 -

Figure 24. Cofactor (divalent metal ion) requirement of WfeD. ................................ - 67 -

Figure 25. Effect of metal ion cofactor on WfeD GalT activity. ................................... - 68 -

Figure 26. Further test of Pb$^{2+}$ effect on enzyme activity. ...................................... - 70 -

Figure 27. Effect of specific amino acid-modifying reagents on the activity of purified WfeD GalT. .......................................................................................... - 71 -

Figure 28. Effect of disulfide reducing agents on WfeD activity. ............................... - 73 -

Figure 29. Alignment of the complete WfeD sequence with other putative GTs that have high sequence similarity to WfeD. ......................................................... - 74 -75 -

Figure 30. SDS-PAGE of Ni$^{2+}$-NTA (small scale) purified mutants using Coomassie Blue staining................................................................................................. - 77 -

Figure 31. Western blot of Ni$^{2+}$-NTA (small scale) purified wild type and six mutant proteins of WfeD. ..................................................................................... - 79 -

Figure 32. Activity assay of WfeD mutant enzymes..................................................... - 83 -

Figure 33. Activity assay of two WfeD mutants after Pb$^{2+}$ and Mn$^{2+}$ stimulation. ...... - 84 -

xiv
CHAPTER 1 INTRODUCTION

1.1 Lipopolysaccharides (LPS) and O-antigen in the cell environment

1.1.1 LPS and O-antigen structures

LPS are large molecules consisting of a lipid and a polysaccharide joined by a covalent bond, and are found in the outer membrane of Gram-negative bacteria (Figure 1). LPS molecules in the cell membrane not only contribute to the structural integrity of the membrane, but also act as endotoxins which elicit strong immune responses in animals\(^1\). The complete LPS consists of a lipid A that embeds in the outer cell membrane, a core oligosaccharide and an O-polysaccharide (or O-antigen) side chain that are both exposed on cell surface\(^1,2\) (Figure 2).

The structures of the lipid A and core oligosaccharide portions have little diversity: the former is composed of phosphorylated \(\beta1,6\)-linked glucosamine diasaccharides, and the latter is commonly composed of heptose and keto-deoxyoctulosonate (KDO). However, the O-antigen portion has a fairly high structural diversity, which defines different serotypes of bacteria. For example, the O-antigen structures within three serotypes of *Shigella* can have different sugar compositions (Figure 3). *Shigella* is a subgroup of *Escherichia coli (E.coli)* as a result of evolution, thus some O-antigens in these two bacteria have been shown to possess the same structure\(^3\). The positioning of these oligosaccharide molecules permits the interaction of LPS with the external environment. The structure of the outer portion of LPS is involved in virulence of the bacterium by controlling interactions between the bacterium and the host\(^4,5\).
Figure 1. A model of inner and outer cell membranes of *E.coli* K-12.

The outer membrane is populated with lipopolysaccharides containing the O-antigen, and also contains membrane proteins such as porins. The inner membrane is occupied by incomplete LPS molecules as well as some membrane proteins. PPEtn—pyrophosphorylethanolamine, MDO—membrane derived oligosaccharides. (This figure is adapted from Raetz et al\(^1\))
Figure 2. An example of a cell surface LPS, from *E.coli* O111:B4.

LPS consists of lipid A (in cell membrane), core oligosaccharide and O-antigen (both on cell exterior). O-antigen: polysaccharide O-side chains. Core oligosaccharide: heptose & Keto-deoxyoctulosonate. Lipid A: a phosphorylated disaccharide of glucosamines anchored to the membrane by multiple hydrophobic fatty acid tails. (This figure is a redrawn and modified version of a figure by Raetz *et al*.)
Figure 3. Three O-antigen repeating unit structures that are identical in different serotypes of *Shigellae* and *E.coli*.

Pyr—pyruvate, GlcpNAc—N-acetylglucosamine pyranoside, Rhap—rhamnose pyranoside, Ala—alanine, GalpA—galacturonic acid pyranoside, Galf—galactose furanosi, GalpNAc-N-acetylglucosamine pyranoside, GlcpA—glucuronic acid pyranoside. (This figure is adapted from Liu *et al*).
1.1.2 LPS as an endotoxin that induces intestinal disease

The human intestine is occupied by a complex and dynamic microbial ecosystem, with communications occurring between microorganisms and the gastrointestinal (GI) epithelium. An endotoxin is a component of Gram-negative bacteria that triggers an innate immune response through Toll-like receptor 4 (TLR4) signalling pathways in myeloid cells, and a potent stimulus of pro-inflammatory response. The biological activity of an endotoxin in Gram-negative bacteria originates from LPS. Toxicity is associated with the lipid component (Lipid A) and immunogenicity is associated with the polysaccharide components. The O-antigen in Gram-negative bacteria is specifically responsible for recognition by the immune system. When bacterial pathogens invade the host intestine, the LPS molecules on cell surface are released upon the rupture of the bacterial cell wall. The released LPS are recognized and bound by LPS binding protein (LBP), which is then transferred to the CD14 receptor or TLR4 on macrophages (Figure 4), and induces secretion of cytokines, such as tumor necrosis factor α (TNFα), interleukin-1 (IL-1), IL-6, IL-8, and IL-10. These cytokines are markers of inflammatory responses against bacterial infections. However, if excessive LPS stimulation of the immune system occurs, the elevated level of activated immune cells can result in septic shock and even death. Therefore, the LPS mediated host-bacterial interactions play crucial roles in maintaining internal homeostasis of the body ecosystem and in regulating immune system activity.
Figure 4. LPS stimulates the Toll-like receptor involved in signalling pathways in a macrophage cell.

TLR4 of the innate immune system recognizes LPS and triggers an immune response. The TLRs are single membrane spanning, non-catalytic receptors that play a key role in the innate immunity by recognizing the structurally conserved molecules derived from microbes. The full ligand sensitivity of TLR4 for the recognition of LPS also depends on other co-receptors such as LBP, CD14 and MD-2. PG- proteoglycan; PGRP- proteoglycan receptor protein; LBP- LPS binding protein. (This figure is a redrawn and modified version of a figure by Allen et al\textsuperscript{4}.)
1.2 Shigellae are involved in gastrointestinal disease

1.2.1 Pathogenicity of Shigellosis

Shigella is a genus of highly adapted bacterial pathogens that causes bacillary dysentery in humans and an average of 1 million associated deaths every year\(^8\text{-}^{10}\). The first genus identified was Shigella dysenteriae at the end of 19\(^{th}\) century by a Japanese microbiologist named Kiyoshi Shiga\(^8\). Recently, there has been a resurgence of interest in Shigella as a human pathogen, partly driven by the availability of more precise data on the disease burden and the emerging antibiotic resistance\(^8\text{-}^{10}\). Upon invasion of Shigellae into the host tissue, they are first confronted with the innate immune response, the cells of which only recognize and respond to pathogens in a generic way. By recruiting immune cells, such as macrophages and DCs, to the site of infection and producing chemical factors (such as cytokines), the innate immune system aims to establish a physical barrier against the spread of infection\(^1\text{-}^{9}\text{-}^{10}\). Shigellae cross the intestinal epithelium in selected areas corresponding to M cells (micro-folding cells) of follicle-associated epithelium (FAE) that covers the mucosa-associated lymphoid follicles. Subsequently, the invasive Shigellae enter the dome area of the lymphoid follicle, which is populated with macrophages and DCs. Shigellae express the invasive phenotype by apoptotic killing of macrophages. The apoptosis of macrophages and DCs is caused by secreted IpaB (invasion plasmid antigen). The cell death of macrophages and DCs not only permits bacterial survival following the crossing of the FAE but is also essential to trigger the early stage of inflammation\(^4\text{-}^{9}\) (Figure 5).
Figure 5. Scheme of *Shigella* pathogenicity: rupture, invasion, and inflammatory destruction of the intestinal barrier.

After crossing M cells of the FAE, the invasive *Shigella* find themselves in the macrophages and dendritic cells (DCs) populated lymphoid follicle. They secrete invasion proteins (i.e. IpaA, IpaB, and IpaC) which lead to the activation of Cysteine protease caspase 1 and apoptosis of macrophages. As a consequence, the inflammatory cytokines IL-1 and IL-18 are released. The release of IL-1β leads to the rupture of epithelial barrier and destabilization of tissue homogeneity, while IL-18 induces the secretion of IFN-γ which allows the innate immune system to establish proper conditions to eradicate the *Shigella*. This figure is adopted from Sansonetti.10
In some virulent strains, such as *Shigella flexneri* and *Shigella dysenteriae*, their LPS are maximally stimulatory and induce intestinal leukocyte infiltration, which then leads to further disruption of epithelial membranes and further bacterial invasion\(^{10,11}\).

### 1.2.2 Role of O-antigen in Shigellosis and O-antigen vaccine development

Although the pathogenicity of *Shigella* has been described, the involvement of LPS in eliciting these consequences was not known until 1995\(^{12}\). In a study on *Shigella dysenteria*, a defect in O-antigen synthesis was made by deleting the *rfe* gene which encodes the GlcNAc-phosphate transferase that transfers the first O-chain sugar. This defect led to the loss of inter- and intra-cellular motility of *Shigellae*. As a result, the mutant *Shigellae* exhibited an impaired ability to invade the host cells as compared to the wild type *Shigellae*, and eventually failed to provoke an inflammatory reaction in the host. Another study by West *et al* showed that glucosylation of *Shigella* O-antigen (the cell surface portion of the LPS molecules) shortened the overall distance that LPS molecules extended from the cell surface, but increased the ability of bacterial invasion and evasion of innate immunity without compromising the protective properties of LPS\(^{13}\). Therefore, the investigation of O-antigen synthesis is of crucial importance to understanding the mechanism of bacterial pathogenicity\(^{14,15}\). Over several decades, clinical studies have shown that, following an initial wild type *Shigella* infection, strong responses of mucosal secretory anti-O-antigen IgA (sIgA) antibodies and serum anti-O-antigen IgG were observed, accompanied by a number of the gut-derived anti-O-antigen IgA antibody-secreting cells (ACSs)\(^{16,17}\). The importance of the mucosal antibodies in preventing shigellosis was demonstrated by Tacket *et al*\(^{18}\). Along with the antibody recognition, two vaccine strategies were developed: one used *Shigella* strains as live oral vaccines, and the
other used conjugates of *Shigella* O-polysaccharide covalently linked to a carrier protein. Understanding the biosynthesis of O-antigens is necessary to develop technologies of chemo-enzymatic synthesis of polysaccharide vaccines\textsuperscript{17,18}.

### 1.3 Biosynthesis of O-LPS in Shigella

The O-LPS or O-antigen consists of several repeats of an oligosaccharide unit (generally two to six sugar residues), also known as the O-repeating unit or O-unit. Among different organisms, the O-antigen structures are highly variable, due to the nature, order, and linkages of the different sugars within the O-unit. The genes encoding enzymes involved in O-antigen synthesis are usually found in a cluster\textsuperscript{1,19-24}. Within these clusters, the genes fall into three categories: a group involved in nucleotide sugar (or donor substrate) synthesis; a group involved in transferring single sugars to form the O-unit, the GTs; and a group involved in O-antigen processing. Among those three groups of enzymes, the GTs play the major role in determining the variations of the O-antigen structures\textsuperscript{19-24}. The synthesis of an O-unit is initiated by the transfer of a sugar phosphate to undecaprenol phosphate (UndP) via sugar-phosphate transferase (such as WecA). Only a limited number of sugar phosphates are known to function in this step; they are derived from nucleotide sugars such as UDP-Glucose (UDP-Glc) and UDP-N-Acetylglucosamine (UDP-GlcNAc)\textsuperscript{20-22}. The *Shigella* B14 O-antigen consists of the following O-unit (GlcA = glucuronic acid): [Gal\(\alpha\)1-4GlcA\(\beta\)1-6Gal\(\beta\)1-4Gal\(\beta\)1-4GlcNAc-]. The synthesis starts with the addition of GlcNAc-phosphate (GlcNAc-P) to UndP by WecA, and the remaining sugars are expected to be added sequentially by specific GTs (Figure 6).
Figure 6. One synthesis cycle of a single O-unit S. boydii serotype B14.

This repeating unit consists of five sugars\textsuperscript{3,20,24}. The synthesis starts from the lipid carrier Und-PP, and takes place at the cytosolic face of the inner cell membrane. After flipping, polymerization and ligation of the O-antigen to the inner oligosaccharide-lipid A to form LPS, Und-PP is recycled as UndP.
In the proposed O-antigen biosynthetic pathway in *E.coli* (most related to *Shigella* in genetic content), the O-antigen synthesis follows a Wzx/Wzy dependent pathway\(^{20,21}\). A similar synthesis pathway appears to be functional in *Shigella* B14 which has \(wzx\) and \(wzy\) genes in the O-antigen gene cluster\(^3\). The biosynthesis of an O-antigen is initiated by the addition of sugars to UndP at the cytoplasmic face of the inner cell membrane. The completed single O-unit will be flipped by Wzx (flippase) to the periplasmic face of the inner cell membrane, where it is polymerized to become a complete O-antigen-UndP by Wzy (polymerase). Wzx acts as a chain processing enzyme that controls the O-antigen length. After the complete O-chain is synthesized, it is transferred by a ligase to the core oligosaccharide and lipid A located at the periplasmic surface. The complete LPS molecule is finally translocated to the cell surface by a poorly understood complex process (Figure 7).

### 1.4 Identification of glycosyltransferases by sequence similarity

#### 1.4.1 Large, uncharacterized glycosyltransferase pool

Of many putative GTs identified only a small number of enzymes have been described biochemically, such as the sugar-phosphate transferase or the galactosyltransferase (GalT) WfeD\(^{1,2,20}\). The Carbohydrate-Active enZYmes database (CAZy) is the most complete internet database of GT sequences. Most of the GTs in the CAZy database (CAZy.org) have been classified according to their sequences and their predicted folds. According to their sequence similarity, the CAZy database has the GTs categorized into 92 GT families (GT1 to GT92) as well as one non-classified family, containing thousands of different GTs.
Figure 7. The proposed Wzx/Wzy dependent O-antigen synthetic pathway in *Shigella* B14.

For the synthesis of *Shigella* B14 O-antigen in particular, WfeD is proposed to transfer a Gal residue to GlcNac-P (transferred by WecA) on the lipid carrier\(^{20,24}\). Other GTs encoded by the O-antigen gene cluster finish the synthesis of one complete O-unit. The Wzx and Wzy work to flip the O-chain to the periplasmic space and start the polymerization of the O-unit. The Wzx protein regulates the O-chain length.
Based on the predicted structural folding, they are also classified into larger families, mainly GT-A (containing GT2, GT6, GT7, GT31, GT62, GT64) and GT-B superfamilies (containing GT9, GT26, GT41). The unique folding feature of GT-A superfamily proteins involves a single domain consisting of parallel β-strands flanked on either side by α-helices, such as the SpsA from *Bacillus subtilis* and the cellulose synthase\(^{24-25}\). Most GT-A enzymes also have a DxD motif in the catalytic site\(^{25-27}\). This DxD motif usually contains three amino acids in the order of Asp/Glu-X(any residue)-Asp/Glu, and either one of the first or third residue can be involved in the catalytic activity. The GT-B superfamily proteins (such as MurG, a GlcNAc bacterial GlcNAc-transferase) share three glycine-rich G-loops located at turns between the carboxyl ends of β-strands and the N-termini of the following α-helices in Rossman fold (β–α–β–α–β) domains\(^{28}\). A deep cleft in between two Rossman domains forms a binding site for the substrate\(^{28}\).

However, the functions of most of these putative GTs still remain to be defined by biochemical means. One of the limiting factors in the characterization of these GTs is the lack of biochemical assays. A defined synthetic substrate has only recently been described, and it is very difficult to synthesize highly specific substrates\(^{20-32}\). Separating the product from the unreacted substrates can also be challenging\(^{29-33}\). Devising ways to address some of these obstacles is a goal of my research. Furthermore, through the work of my project and other projects in our lab on enzyme characterization, we aim to develop specific inhibitors to block O-antigen synthesis or synthesize O-antigen conjugated vaccines as new approaches to therapy in shigellosis and related diseases.
1.4.2 Identification of Shigella B14 putative glycosyltransferases

Structural analysis of the O-LPS from *S. boydii* serotype B14 showed a repeating pentasaccharide structure, →6-D-Galpα1→4-D-GlcpAβ1→6-D-Galpβ1→4-D-Galpβ1→4-D-GlcpNAcβ1→. The O-antigen gene cluster of *S. boydii* B14 was sequenced by our collaborator Lei Wang in China. The gene cluster contains the wzy O-unit polymerase gene and the wzx O-unit flipase gene, as well as four putative GT genes, *wfeE, wfeD, wfeB, wfeA*. *WfeD* shares 38% identity and 57% similarity to the GT *Orf9* in the O-antigen gene cluster of *E. coli* O136 O-antigen (with unpublished nucleotide sequence). Since the O-antigens of *S. boydii* B14 and *E. coli* O136 only share one common linkage: D-Galpβ1→4-D-GlcpNAc, *wfeD* was proposed to encode a GalT that transfers Gal to GlcNAcα-PP-Und in β1-4 linkage, which is the second step in the biosynthetic pathway of the *S. boydii* B14 O-unit (Figure 8). The other putative GT genes of the B14 gene cluster may be responsible for transferring the remaining sugars to the O-unit. The enzyme WfeD from *S. boydii* B14 has some sequence similarity (by BLAST search) with other bacterial enzymes from the GT26 family with putative GT functions. WfeD seems to be the first enzyme within this group subjected to functional determination. This makes it hard to model the WfeD structure for predicting the catalytic site and amino acids involved in catalysis. Our work on WfeD is the first biochemical study of a GT involved in the biosynthesis of an O-antigen in *Shigellae*.

1.4.3 Comparison of mammalian and bacterial galactosyltransferases for O-polysaccharide synthesis
Figure 8. The O-antigen gene cluster of *S. boydii* serotype B14.

The genes labeled by yellow arrows are putative GTs. The functions of all the other genes in the cluster have only been suggested based on sequence comparison with characterized sequences of the O-antigen gene cluster library. The solid arrows indicate the biochemically confirmed (WfeD, in this thesis work) and proposed GT genes (WecA). Dotted arrows indicate unknown, putative GT genes. Genes that are homologous (but uncharacterized) to known genes include *wzx* (O-unit flippase), *wzy* (O-unit polymerase), *gnd* (involved in nucleotide sugar synthesis) and *galF* (phosphorylase involved in nucleotide sugar synthesis).
The mammalian O-linked oligosaccharides are usually found in soluble, secreted and membrane bound glycoprotein or proteoglycan forms. For example, the mucins are the main class of O-glycosylated glycoproteins, with heavily O-glycosylated Ser/Thr/Pro rich tandem repeat regions. The structures of O-glycans of mucins are diverse and the diversity is mainly controlled by the GTs that are encoded by cell-specific genes. In contrast to their counterparts in bacteria, the mammalian O-chain GTs have been more extensively studied in enzymatic functions and structures\textsuperscript{30,34-37}. As seen in Table 1, the structures of several mammalian GalTs have been solved by X-ray crystallography (Figure 9). So far, a \textit{Neisseria meningitidis} (\textit{N. meningitides}) \(\beta 4\text{GalT} \) (LgtC) is the only structure solved among all bacterial GalTs\textsuperscript{27}. Although the bacterial and mammalian GalTs have very different sequences, they may have a common fold and all contain the DxD motif. However, due to the low sequence similarity between any mammalian and bacterial GalT, no structural inference from a known enzyme can be easily obtained for studying a novel enzyme in this case.

1.5 The DxD motif is involved in the catalytic mechanism of glycosyltransferases

Similar to other \(\beta\text{GalTs} \), WfeD also contains a DxD resembling motif within the DYEIE sequence (Figure 10A), which is thought to be involved in binding divalent metal ions for positioning sugar nucleotides in the active site. A DxD motif is usually found flanked by apolar residues\textsuperscript{27,36} in totally different GT proteins with poor sequence alignment. The X-ray structure of a rabbit GlcNAc-transferase I illustrated the importance of the DxD motif in the interaction between the donor UDP-GlcNAc and the Mn\(^{2+}\) ion, which is shown by the schematic drawing in Figure 10B\textsuperscript{36}. 

- 17 -
Table 1. Examples of functionally characterized mammalian and bacterial GalT enzymes from CAZy database\textsuperscript{27,29,32-37}.

<table>
<thead>
<tr>
<th>Category</th>
<th>GalT Name</th>
<th>Origin</th>
<th>PDB Name of the Structures</th>
<th>CAZy Family</th>
<th>Fold</th>
<th>Catalytic Motif</th>
<th>Acceptor Substrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mammalian GalTs</td>
<td>β3GalT</td>
<td>Core 1, T-antigen</td>
<td>No</td>
<td>GT31</td>
<td>GT-A</td>
<td>DxD</td>
<td>GalNAcα-R</td>
</tr>
<tr>
<td></td>
<td>β4GalT</td>
<td>Bovine</td>
<td>1TVY</td>
<td>GT7</td>
<td>GT-A</td>
<td>DxD</td>
<td>GlcNAcβ-R</td>
</tr>
<tr>
<td></td>
<td>α3GalT</td>
<td>Blood group B</td>
<td>2RJ9</td>
<td>GT6</td>
<td>GT-A</td>
<td>DxD</td>
<td>(Fucα1-2)Galβ-R</td>
</tr>
<tr>
<td></td>
<td>α3GalT</td>
<td>Bovine</td>
<td>1VZT</td>
<td>GT6</td>
<td>GT-A</td>
<td>DxD</td>
<td>Galβ-R</td>
</tr>
<tr>
<td>Bacterial GalTs</td>
<td>β3GalT</td>
<td>\textit{E.coli}</td>
<td>No</td>
<td>GT2</td>
<td>GT-A</td>
<td>DxD</td>
<td>GlcNAcα-PP-R</td>
</tr>
<tr>
<td></td>
<td>β4GalT</td>
<td>\textit{N. meningitidis}</td>
<td>1G9R</td>
<td>GT8</td>
<td>GT-A</td>
<td>DxD</td>
<td>Lactose</td>
</tr>
</tbody>
</table>
Figure 9. The overall structure of the LgtC monomer (1G9R) with bound substrate analogs.

The substrates are depicted in CPK representation with the acceptor coloured dark gray, the donor light gray and the manganese pink. The nine β-strands and fourteen α-helices are labelled. (This figure is adapted from the figure by Persson et al.)
A)

MGSSHHHHHHPSGLVPRGSHM
  1 VIDNLKRTIP EIRLLENKR VTGVTVFNP YSYYKIEYN KISQLDYIVI DGILLKLFN
  61 FVNGTKIKRH SFYYSIQNKMM KIGLIGSKEYIE E@AVKNIR KKHGPIDISY
  121 FHSGYFESSLE EKSSVIDSVI KKSIIIICGL GTPAEEELAL DIKKSNEHL IFTCGGFTQ
  181 TASRADFYYP WIKRYNLMLW Qrivlykhvr KRFFIDYPKF IVRFISENLM KIFTRSN

B)

Figure 10. The DxD motif in the WfeD sequence and the mechanism involving the DxD motif.

A) Amino acid sequence of *Shigella* B14 WfeD (237aa), with His-tag fused at the N-terminus, and the DxD (DYEIE) motif underlined. B) Schematic representation of rabbit GlcNAc transferase I reaction mechanism. The acceptor substrate has a terminal mannose residue, and the donor substrate is UDP-GlcNAc. Asp291 from the DxD motif is involved in deprotonating a specific OH group on the acceptor, and Mn$^{2+}$ is the cofactor that stabilizes the leaving UDP group. (This figure is adapted from Unligil et al$^{26}$)
The D291 from the DxD motif acts as a general base to deprotonate a specific hydroxyl group on the acceptor sugar (Figure 10B). The deprotonated mannose in the acceptor becomes a nucleophile and attacks the C1 of the sugar in the nucleotide donor substrate. The role of the divalent metal ion is to neutralize the negative charge developed on the pyrophosphate of the UDP leaving group. An oxocarbenium-ion-like transition state is created, which leads to the transfer of GlcNAc from UDP-GlcNAc to a specific position of the acceptor substrate, by inverting the $\alpha$– to $\beta$– configuration. If the same mechanism applies, the WfeD enzyme should also utilize its Asp or Glu residue in the DYEIE sequence for deprotonating the 4-OH of the acceptor GlcNAc on GlcNAc-PP-Und, and a divalent cation would serve to neutralize the negative charge on UDP-Gal.

Previous work in our lab identified the substrate specificity and some inhibitors of bovine milk $\beta_{1,4}$-GalT1, using a series of synthetic GlcNAc-R analogs. Our research on substrate specificity, cofactor requirements, and inhibitor identification has formed the basis for re-engineering specific glycan structures. Although the bovine $\beta_{4}$GalT enzyme also contains the DxD motif and recognizes GlcNAc-R as its substrate, the sequence and substrate specificity of this mammalian enzyme is highly divergent from Shigella WfeD (similarity <12%) By fold prediction of the WfeD sequence, it seems to possess a very different 3-D fold (WecG-TagA) which is likely to be very different from GT-A or GT-B folds. This indicates that the WfeD could have a structural fold completely distinct from that of bovine milk $\beta_{4}$GalT1, but possess a similar catalytic mechanism (DxD) and similar substrate recognition. Further enzyme characterization of the WfeD protein has become a research goal of our lab in order to reach a better understanding of its role in O-antigen biosynthesis.
1.6 Project overview

1.6.1 Project rationale

*Shigellae* are the major cause of bacillary dysentery, and disease treatments have been the focus of studies in many clinical laboratories. It is important to develop new anti-bacterial strategies against the cell surface O-antigens. We propose that studying the WfeD protein can provide feasible assays and new substrates for characterizing more O-antigen synthetic enzymes. The results from these studies will help to identify other GTs in *Shigella* B14 or in other bacteria that also recognize sugar-pyrophosphate-lipids as the substrate. Furthermore, determining the biochemical functions for the putative GTs and studying the mechanism of these novel enzymes will help to understand the complex mechanisms of O-antigen synthesis. This may eventually complement the anti-bacterial strategies either by finding inhibitors for the biosynthetic enzymes to interfere with O-antigen synthesis, or by establishing a technology to synthesize complete O-antigens as vaccines.

1.6.2 Project hypothesis

The *wfeD* gene encodes a GaIT that catalyzes the Galβ1-4GlcNAc linkage in *S. bodyii* B14 O-antigen synthesis, and the recombinantly expressed WfeD is active in its soluble form. Additionally, we propose that the DYEIE sequence of WfeD is a signature motif that is important for divalent metal ion-supported catalysis, and that WfeD is highly specific for GlcNAc-PP-lipid (an analog of the natural substrate GlcNAc-PP-UndP), which could be due to the involvement of the RKR sequence (or other positively charged amino acid clusters) in substrate binding to the enzyme.
1.6.3 Objectives

- His<sub>6</sub>-WfeD fusion protein will be expressed in *E. coli*, and the GT activity of the
total cell lysate will be determined. This will establish the basic activity of the
unpurified enzyme and help define suitable substrates.

- WfeD will be purified by Ni<sup>2+</sup>-NTA chromatography and the membrane
requirement for WfeD activity will be studied.

- The total cell lysate and purified enzymes will be characterized by studying the
kinetics, stability and requirement for other cellular cofactors (i.e. metal ions,
membrane association).

- The structure of the enzyme reaction product will be analyzed to prove the type of
new linkage synthesized.

- Substrate specificity and cofactor requirement will be studied on purified WfeD to
better illustrate the type of reaction, and to explore the common feature(s)
between WfeD and previously characterized GTs. These studies are the
prerequisites for the development of specific substrate analog inhibitors.

- Mutants with changes in proposed critical amino acids will be analyzed. The
ultimate goal will be the discovery of catalytically important amino acid residues
in WfeD.

These studies will yield important information about the enzyme-substrate
interactions and the reaction mechanism of WfeD.
CHAPTER 2 MATERIALS AND METHODS

Materials. All reagents were purchased from Sigma unless indicated otherwise. Radioactive nucleotide sugars were purchased from Perkin-Elmer. The GlcNAc-R acceptor substrates were synthesized as reported previously\textsuperscript{37}. The acceptor substrate GlcNAc\textsubscript{α}-PO\textsubscript{3}-PO\textsubscript{3}-lipid and its structural analogs were synthesized by Walter Szarek’s group\textsuperscript{29,38}. GlcNAc\textsubscript{α}-PO\textsubscript{3}-PO\textsubscript{3}-(CH\textsubscript{2})\textsubscript{9}-CH\textsubscript{3} was kindly provided by O. Hindsgaul, (Carlsberg Laboratories, Copenhagen, Denmark)\textsuperscript{43}.

2.1 DNA plasmid construction and expression of the wfeD gene

2.1.1 Construction of the mutants

The construction of mutants was done by our collaborator Lei Wang (Nankai University, China). Oligonucleotide primers FM, RM, R1, and F1 (Table 2) were designed based on the known sequence of DNA. For DNA amplification, two PCR reactions were set up: PCR 1 contained template DNA, 10×PCR buffer, Mg\textsuperscript{2+}, dNTPs, primer FM, primer F1, thermostable DNA polymerase; PCR 2 contained Template DNA, 10×PCR buffer, Mg\textsuperscript{2+}, dNTPs, primer RM, primer F2, and thermostable DNA polymerase (following the Stratagene Quikchange protocol). DNA was amplified using the program: 95°C 5min, 95°C 30sec, 50°C 45sec, 72°C 1min, 24 cycles, 72°C 2min. This was followed by gel electrophoresis for purification of DNA. For ligation, products from PCR1 and PCR2 were mixed with 10x amplification buffer, thermostable DNA polymerase, dNTPs, primer F1 and R1, using the program: 95°C 5min, 95°C 1min, 50°C 1min, 72°C 1min, 24 cycles, 72°C 2min.
Table 2. List of forward and reverse primers used in single amino acid mutagenesis.

Subsequent DNA sequencing of the mutant constructs was performed by the Robarts Research Institute (London, Ontario). Mutations produced are: A296C (of mutant D99A), A302C and A303T (of mutant E101), A308C (of mutant E103), A616G and G617C (mutant R210A), A331G, G332C and A333T (of mutant R212A), and A328G, A329C, A330T (of mutant K211A). The nucleotides shaded in grey indicate where point mutations were made.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Primer</th>
<th>Sequence</th>
<th>Site</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FM (Forward)</td>
<td>5'-CGCGGATCCGGTGAATCTCATAAAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RM (Reverse)</td>
<td>5'-CCGCTCGAGTCATTTCTCGTGAATATTT</td>
<td></td>
</tr>
<tr>
<td>D99A</td>
<td>F1 (Forward)</td>
<td>5'-TTAATCGGTTCTAAGGCACTATGA</td>
<td>D99</td>
</tr>
<tr>
<td></td>
<td>R1 (Reverse)</td>
<td>5'-TCATAACCTTGAACACCGGATTAA</td>
<td></td>
</tr>
<tr>
<td>E101A</td>
<td>F2 (Forward)</td>
<td>5'-TCGGTTCTAAGGATTATGCT AT</td>
<td>E101</td>
</tr>
<tr>
<td></td>
<td>R2 (Reverse)</td>
<td>5'-ATACGATAATCCCTAGAAACCAGA</td>
<td></td>
</tr>
<tr>
<td>E103A</td>
<td>F3 (Forward)</td>
<td>5'-AGGATTATGAAATAGCGCAAGC</td>
<td>E103</td>
</tr>
<tr>
<td></td>
<td>R3 (Reverse)</td>
<td>5'-GCTTGGCTATTTCTATACAAA</td>
<td></td>
</tr>
<tr>
<td>R210A</td>
<td>F4 (Forward)</td>
<td>5'-ATATAAACATGGTTACGAAAAG</td>
<td>R210</td>
</tr>
<tr>
<td></td>
<td>R4 (Reverse)</td>
<td>5'-CTTTTGGAACATGTATAT</td>
<td></td>
</tr>
<tr>
<td>R212A</td>
<td>F5 (Forward)</td>
<td>5'-TAGGAAAAGCTTTTTATCGAT</td>
<td>R212</td>
</tr>
<tr>
<td></td>
<td>R5 (Reverse)</td>
<td>5'-ATCGATAAAAAACGTTCTTA</td>
<td></td>
</tr>
<tr>
<td>K211A</td>
<td>F6 (Forward)</td>
<td>5'-CATGTTAGGCTAGTATTT</td>
<td>K211</td>
</tr>
<tr>
<td></td>
<td>R6 (Reverse)</td>
<td>5'-AAAATCTAGGCCCTACATG</td>
<td></td>
</tr>
</tbody>
</table>
For constructing mutants, four primers and three PCR reactions were used to create a site-specific mutation by overlap extension. One pair of primers is used to amplify DNA that contains the mutation site together with upstream sequences. The second pair of primers is used in a separate PCR to amplify DNA that contains the mutation site together with downstream sequences. The mutation of interest is located in the region of overlap and therefore in both amplified fragments. The overlapping fragments are mixed, denatured, and annealed to generate heteroduplexes that can be extended and, in a third PCR, amplified into a full-length DNA using two primers that bind to the extremes of the two initial fragments. Subsequent sequencing confirmed the correct sites of mutation.

2.1.2 Transformation of plasmid constructs into the expression cell lines

Transformation involved the use of the electro-competent *E.coli* BL21 cells (DE3, codon plus, Kanamycin and Ampicillin resistant) in 100 µl/aliquots (OD$_{600}$=0.6) prepared in John Allingham’s lab. The electroporation apparatus (Eppendorf Electroporator 2510) was set at “1700v, 5ms”. One µl of 80ng/µl DNA was added to each 100 µl cell suspension, and kept on ice for 1 min. Then the mixture was transferred to an electroporation cuvette in the electroporator for transformation. At the end of the process, 500 µl LB medium were added to the cuvette, and 600 µl cell suspension were transferred to a culture tube for 1hr incubation at 37°C. The cell suspension was then kept at 4°C before plating on an agar plate (containing 50 µg/ml kanamycin and 50 µg/ml ampicillin) for overnight growth at 37°C. Single colonies of the wild type and each mutant plasmid construct were picked from the agar plates for small scale or large scale enzyme expression.
2.1.3 Expression of wild type and mutant wfeD in E.coli BL21 cells

For a small scale expression, cells containing the plasmids by transformation were grown overnight (16 hrs) in 6 ml LB broth containing 50 μg/ml of kanamycin and 50 μg/ml of ampicillin, at 37°C with constant shaking. On the second day, the E.coli cells were transferred into 125 ml LB broth containing the antibiotics, and incubated for 90 min at 37°C with constant shaking. For large scale expression, cells were grown overnight in 125 ml LB broth containing 50 μg/ml kanamycin and 50 μg/ml ampicillin, and were transferred to 1 liter LB broth with the same antibiotics for induction using IPTG. When the suspension reached an absorbance at 600 nm of 0.6, the cell growth was stopped by cooling down the cell culture suspension on ice for 15 min. A final concentration of 1 mM IPTG was added to the cell culture with OD₆₀₀ of 0.6, to induce protein expression for 4 hrs at 30°C. Cells were harvested by centrifugation for 10 min at 3145 g (4000 rpm, IEC 21000R Centrifuge). Pellets were washed with 5 ml PBS and resuspended in 10 ml PBS containing 10 % glycerol. Aliquots of bacteria were stored at -20°C for enzyme assays.

2.2 Galactosyltransferase activity and inhibition assays

Bacterial homogenates containing WfeD were prepared by sonication in 0.05 M sucrose buffer as described. The standard assay mixtures for WfeD had a total volume of 40 μl containing acceptor substrate (GlcNAc-R analogs), bacterial homogenates (0.012 mg/ml protein), 75 mM MES buffer (pH 7), 5 mM MnCl₂, and 1 mM donor substrate UDP-[³H]Gal (2000-3500 cpm/nmol). In standard GalT assays, 0.25 mM GlcNAcα-PO₃-PO₃-(CH₂)₁-OPh (GlcNAcα-PP-PhU) was used as the acceptor substrate, while the
control assays lacked acceptor substrate\textsuperscript{32,38}. Due to the low solubility of GlcNAc\(\alpha\)-PP-PhU in water, 10\% MeOH was added to all assays. Preliminary testing showed that there was no significant inhibition of GalT activity up to 20 \% MeOH in the assay. Mixtures were incubated for 10 min at 37°C (reaction time course showed rate linearity up to 20 min). Reactions were stopped by the addition of 200 \(\mu\)l ice-cold water, which abolished the temperature-dependent enzyme activity. For purifying the product, 500 \(\mu\)l water was added and the mixture was passed through a short C18 Sep-Pak column for separation based on hydrophobicity, and the column was washed with 4 ml water (1 ml/fraction). Radioactive product was then eluted with 4 ml MeOH (1 ml/fraction). Scintillation fluid (4.5 ml Ready Safe, Beckman Coulter Inc., USA) was added to each 1 ml fraction, and radioactivity was determined by scintillation counting (LS6500 multi-purpose scintillation counter, Beckman Coulter Inc., USA). Radioactive product in the first 3 ml of MeOH eluates was concentrated by flash evaporation for HPLC analysis. Residues were taken up in MeOH. The standard assay mixtures for bovine \(\beta\)4GalT1 were as described\textsuperscript{37,41}. Kinetic parameters were determined using the EnzFit program\textsuperscript{30,32,37}.

For GalT activity inhibition assay, the standard assay mixtures for WfeD had a total volume of 40 \(\mu\)l containing the same components as the standard activity assay described above plus 0.5 mM potential inhibitor (GlcNAc-R analogs). The standard GalT assay mixture (without any potential inhibitor compound) was used as the positive control. Mixtures were incubated for 10 min at 37°C. Reactions were stopped by the addition of 200 \(\mu\)l ice-cold water. The same product isolation and quantification methods were used as described above.
The enzyme activity assay for bovine milk β4GalT also used UDP-[3H]Gal (2000-3500 cpm/nmol) as donor substrate and GlcNAcβ-Bn as acceptor substrate, followed by AG1 resin anion exchange for product purification as previously described\cite{37}.

### 2.3 Structural analysis of the reaction product

#### 2.3.1 Product identification using mass spectrometry and HPLC

The reaction product was prepared as a radioactive product for HPLC analysis, and also as a non-radioactive product for mass spectrometry (MS) analysis. Separation of substrates and enzyme products was achieved by HPLC using a C18 column and acetonitrile (AN)–water mixtures at a flow rate of 1 ml/min. For a separation of substrate GlcNAc-PP-PhU and product Gal-GlcNAc-PP-PhU, 24 % AN in water was used as the mobile phase. Under these conditions, substrate eluted at 33 min and product at 25 min (Figure 11). Compounds were subsequently analyzed by MS, using Electrospray (ESI)-MS in the negative ion mode, as described\cite{39}.

#### 2.3.2 Product linkage confirmation by galactosidases

The anomeric configuration of the linkage formed in the radioactive product was determined by digestion with linkage-specific galactosidases: Green Coffee Bean (α-specific) α-galactosidase (0.065 U/μl), Jack Bean (β4-specific) β-galactosidase (0.06 U/μl) and bovine testicular (β1-3, 1-4, 1-6-specific) β-galactosidase (0.095 U/μl).

Aliquots of the radioactive reaction product (800 cpm) were treated in a total volume of 100 μl with 25 μL MacIlvaine buffer (0.1 M citric acid/0.2 M Na-phosphate), pH 4.3, 10 μl of 0.1 % bovine serum albumin and 2 to 40 μl galactosidase.
Figure 11. HPLC separation of the purified reaction product.

The product was first purified by Sep-Pak C18 column and concentrated by flash evaporation for HPLC. The flow rate was 1 ml per min, while the fractions were collected by 2 min per fraction. The radioactivity in cpm was obtained by counting every collected 2 ml fraction. An aliquot (1100 cpm) of the synthesized product was injected, and 900 cpm were recovered during the first 30 min. The absorbance at 195 nm is indicated by the dashed line, and cpm counts are indicated by the solid line. Peak 1 was UDP-Gal, Gal or degradation product; peak 2 is disaccharide lipid product, and peak 3 is the remaining acceptor substrate.
The reaction mixtures were incubated for 30 min at 37°C, diluted with 800 µl of water, and applied to 0.4 ml AG1 columns (100-200 mesh, Cl-form, anion-exchange resin, BioRad). Released radioactivity ([3H]Gal) was eluted with 2.8 ml water, while unreacted enzyme product stayed in the AG1 column.

2.3.3 Product analysis by NMR

To prepare large amounts of reaction product for NMR analysis, GalT assays were carried out as follows. The incubation mixtures (a total of 8 ml) contained: 0.2 mg bacterial homogenate in 50 mM sucrose (E. coli BL21 cells complemented with plasmid pET28a), 2.5 µmol GlcNAc-PP-PhU, 10 µmol UDP-Gal, 1 mmol MES buffer, pH 7, and 50 µmol MnCl₂. After incubation for 30 min at 37°C, 8 ml of cold water were added and the mixtures were applied to C18 Sep-Pak columns. Each column was washed with 4 ml water, and the product was eluted with 4 ml MeOH. The methanol fractions of Sep-Pak columns were pooled, flash evaporated, and redissolved in 500 µl MeOH. Aliquots of the methanol-dissolved product were purified by HPLC, using a C18 column and AN/water (24:76) as the mobile phase. The reaction product was dried, exchanged three times with 99.96 % D₂O, dissolved in CD₃OD, and analyzed by 600 MHz 1D ¹H spectroscopy, 2D COSY and ¹H-¹³C HSQC (by Dr. Francoise Sauriol at the Chemistry Department, Queen's University) as described³³.

2.4 Enzyme purification and analysis of the protein by SDS-PAGE and mass spectrometry

For purification of the His₆–WfeD fusion protein, bacteria were sonicated in 0.2 g/ml of the lysis buffer (Tris pH 8.0, 500 mM NaCl, 10 mM imidazole and 1 mM EGTA), for
six cycles of 1-min sonication and 1-min waiting time (Misonix Sonicator 3000, Program 1). The homogenate was then centrifuged at 13000 g for 30 min. The His<sub>6</sub>–WfeD fusion protein in the supernatant was purified by affinity chromatography using a Ni<sup>2+</sup>-NTA (nickel-Nitrilotriacetic acid) Sepharose Fast Flow column (BioRad). 50 ml of supernatant was loaded onto the column and then washed with 50 ml of washing buffer (10mM imidazole, Tris-HCl pH 8.0, 500 mM NaCl and 1 mM EGTA), and their run-through solutions were collected as “flow-through” and “wash-through” fractions. The protein was eluted with a linear gradient of imidazole (from 0 to 600 mM) in Tris-HCl pH 8.0, 500 mM NaCl and 1 mM EGTA. In total, 60 fractions of 2 ml were collected as the “elution fractions”. WfeD eluted between 200 to 400 mM imidazole buffer. Eluted protein fractions were dialyzed against 2 L of 50 mM Tris pH 8.0, 300 mM NaCl and 1 mM EGTA for one day. The protein concentrations of the cell lysate, supernatant, and eluted fractions were measured by Bradford reagent (Bio-Rad) using bovine serum albumin (BSA) as the standard. The purified fusion protein was analyzed by SDS–PAGE (15 % gel), western blotting, and by our GalT activity assays. The gel slices of the suspected protein band at 27~28 kDa were analyzed by peptide mass fingerprinting at the Protein Function Discovery Facility, using MALDI-TOF (MS). The experimental data as a list of peptide mass values from the enzymatic digest of the band slices were fed to the online search engine Mascot (www.matrixscience.com) to identify the protein from the primary sequence databases.

2.5 Detergent test on enzyme activity and separation of protein and membrane aggregate pellet
To test whether the detergents affect the enzyme activity, 4 μL of the three different detergents 1% (v/v) Triton X-100, Tween-80, and NP-40 (nonyl phenoxypolyethoxyl-ethanol), as well as varying concentrations of Triton X-100, were included in standard assay mixtures of 40 μl (as described in section 2.2). The effect of those detergents on enzyme activity was compared to the control (without any detergent).

To test whether the inclusion of detergent increases the solubility of WfeD in aqueous solution after cell homogenization and centrifugation, 0.1% Triton X-100 and 0.1 % Tween-20 were separately included in the sonication buffer. Additionally, two concentrations of Tween-20 (0.1 % and 1 %) were included in the sonication buffer to test if increased detergent concentration would lead to higher protein solubility. After the same sonication and centrifugation procedures as described in section 2.4, the supernatant was separated from the pellet. The supernatants obtained after sonication with different concentrations of detergents were subjected to SDS-PAGE, to determine whether better separation of protein from the pellet was achieved.

2.6 Western blot analysis of purified WfeD (wild type and mutants)

Western blot analysis was performed with rabbit antibody against the His-tag as primary antibody (kindly provided by P. Davies, Queen’s University), and HRP-conjugated anti-rabbit IgG as the secondary antibody (Promega). Twenty μL of the diluted cell homogenates of the wild type and mutant proteins (150~250 μg/ml) were diluted 5:1 in 4 μL SDS loading buffer consisting of 1 mM Tris–HCl, pH 6.8, 1 % SDS, 10 % glycerol, and 5 % β-mercaptoethanol, and run on SDS-PAGE (15 % gel). The non-His-tagged bovine β4 GalT was used as negative control. Proteins were
electrophoretically transferred to nitrocellulose, and then probed sequentially with 10 ml of 0.5 mg/ml anti-His antibody (1:5000) and 10 ml of anti-rabbit HRP-linked IgG (1:10000) (Promega). Labeling was visualized with a 1:1 mixture of luminol reagent and peroxide solution of 10 ml in total (Millipore). Prestained protein standards (Fermentas) were used to calibrate the gels.

For quantifying the approximate band density, the ImageJ1.43 software was used to calculate the optical density of the protein bands on the film\textsuperscript{44-46}. The amount of anti-His-tag stained protein in the band was represented as the integrated density (IntDen) in pixels. The relative anti-His-tag protein amount in protein solution was obtained by dividing the net IntDen (=IntDen of protein band-IntDen of negative control) by the total amount of protein being loaded to each well of the western blot.

\[
\frac{\text{IntDen of mutant – IntDen of negative control}}{\text{Loaded mutant protein amount}} \quad \frac{\text{IntDen of wild type – IntDen of negative control}}{\text{Loaded wild type protein amount}}
\]

2.7 Testing enzyme activity with amino acid modifying reagents

To evaluate the role of amino acids in activity, prior to the assay, the crude and purified enzyme preparations were pre-incubated with amino acid modifying reagents in the reaction buffer (MES and lysis buffer) for 10 min at room temperature. Reaction was initiated by the addition of UDP-Gal, GlcNAc-PP-PhU and MnCl\textsubscript{2}. Diethyl pyrocarbonate (DEPC, reacts with His), iodoacetic acid (IAA, reacts with Cys), \(p\)-hydroxyphenylglyoxal (HPG, reacts with Arg), or 5,5'-dithio-bis-2-nitrobenzoic acid (DTNB, reacts with Cys) were used at 0.2 mM concentration in the assay. Disulfide bond reducing agent dithiothreitol (DTT) was used at the concentrations of 0.2 to 3 mM in the assay, and \(\beta\)-mercaptoethanol was used at concentrations of 1 to 30 \(\mu\)M.
CHAPTER 3  RESULTS

3.1 Enzyme characterization of crude and purified WfeD protein

3.1.1 Sequence hydrophobicity analysis for the potentially membrane-associated region

In the proposed O-antigen synthesis pathway of *E.coli* and *Shigella*, a GT that catalyzes the transfer of the first few sugars of the O-chain is expected to be close to the inner cell membrane. To assess whether WfeD may be a membrane-associated enzyme, the hydrophobicity of the amino acid sequence was analyzed to indicate the possible membrane associating regions. The hydropathy profile (Figure 12) of WfeD indicates the presence of a possible transmembrane region of this enzyme. In general, the amino acid residues within a membrane-bound region usually have a hydrophobic score of >1\(^47\). For WfeD, there are five regions of varying lengths containing this hydrophobicity (shown by the arrows). The most likely membrane-bound region seems to be the hydrophobic stretch of Tyr47 to Thr65, which has a hydrophobicity score (>1.5) and 19 residues that may form a stable 3-D structure inside the membrane. The remaining regions with hydrophobic score of >1 may also associate with the membrane. Based on this finding, more experiments were conducted to explore the membrane requirement of WfeD, including the enzyme activity test in the presence of detergents.

3.1.2 Detergent test on enzyme activity and the separation of protein and membrane-aggregate pellet

To obtain a pure and active WfeD protein, the protein needs to be isolated from the lipid membrane and should retain its activity in an aqueous solution without membrane
Figure 12. Kyte & Doolittle hydropathy profile of the WfeD sequence (as shown in Figure 10A).

The profile was created by inserting the amino acid sequence of WfeD into the online program ProtScale (Expasy) for computation. The x-axis represents the amino acid residue numbers, and y-axis represents the hydrophobicity scores. The sequence has a minimal score of -2.267 and a maximal score of 2.189. Scores greater than zero indicate that the residues are hydrophobic. Scores greater than one are the common scores of amino acids in membrane-bound regions. The arrows indicate the amino acids stretches with the score of >1.
environment. Different detergents of the same concentration (0.1 %) were added to the reaction mixtures to test whether the enzyme activity would be affected. By treating the protein in cell homogenates with mild detergents, such as Triton X-100, Tween 80, and NP-40, we observed that the inclusion of detergents did not reduce the enzyme activity (Figure 13A). To show whether there is a concentration effect of the detergent on enzyme activity, the reaction mixtures were treated with varying concentrations of Triton X-100. As the concentration of Triton X-100 increased to 6 % in the reaction assay, the enzyme activity decreased but was maintained at 65 % of the original activity (Figure 13B).

Once we confirmed that the addition of detergents maintained the enzyme activity, we proceeded to test whether the inclusion of a detergent in the cell sonication or lysis buffer would help increase the solubility of the enzyme protein in aqueous solution. It was thought that lysing the cells with detergent would provide a lipid-like environment for easier separation of a membrane-associated protein that utilizes a hydrophobic substrate. Two types of detergents, Triton X-100 (0.1%) and Tween-20 (0.1 %) were further used separately in the sonication, but neither of them increased the amount of WfeD protein in the supernatant. This may indicate that the hydrophobic sequence in WfeD does not anchor the enzyme into the membrane. The SDS-PAGE analysis of the cell supernatants obtained by including three concentrations of Tween-20 (0 %, 0.1 % and 1 %) in sonication are shown in Figure 14. As the detergent concentration increased from 0 to 1 %, no obvious evidence of an increase of WfeD in the supernatants was observed. It is still possible that Tween-20 could not solubilise the protein sufficiently. However, this does not preclude the likelihood of successfully increasing solubility of the protein by another detergent or lipid and by ultracentrifugation.
Figure 13. Detergent effect on WfeD enzyme activity.

A) Effects of three detergents on WfeD enzyme activity. The three detergents were included in the 40μL reaction assay at 0.1% (v/v) using WfeD protein in cell homogenate. A standard assay without any detergent being added (‘no detergent’) was used as a control. All the conditions were assayed in duplicates, and bars indicate the difference between the duplicates.

B) Effect of increasing concentration of Triton X-100 on enzyme activity. The activity was obtained by measuring the radioactivity of the reaction product after incorporation of UDP-[3H]Gal into GlcNAc-PP-PhU.

- 38 -
Figure 14. SDS-PAGE using Coomassie Blue staining: detergent effect on the separation of protein and membrane aggregate pellet.

From left to right: Lane 1 is the molecular weight marker; Lane 2 is the supernatant of the cell homogenate without IPTG for induction in cell growth, which was treated with 0 % Tween-20 in sonication; Lane 3 to 5 are the supernatants of cell homogenates with IPTG for induction in cell growth, and each of them had a different lysis buffer (containing 0 %, 0.1 % or 1 % Tween-20) for sonication. The supernatants from the cell lysates were obtained by centrifugation and subjected to SDS-PAGE (15% gel) analysis. The molecular weights with corresponding migrating distances are indicated on the left.
It seemed likely that the protein is not anchored to the membrane or is loosely associated with membrane, and thus using detergent in sonication would not affect solubility. Moreover, it is likely that an insoluble portion of the total WfeD proteins formed a pellet with the membrane and other cellular organelles due to its mis-folding, which could have resulted from detergent disruption of protein 3-D folding. Since the WfeD protein was partially soluble in aqueous solution, further purification could be achieved.

3.1.3 Enzyme purification by Ni\(^{2+}\)-NTA chromatography

Since WfeD was expressed with a His\(_6\) tag at the N-terminus, Ni\(^{2+}\)-affinity chromatography was used to purify WfeD from the sonicated cell homogenate. The His-tag within the protein allowed WfeD to bind to the Ni\(^{2+}\)-linked agarose beads as the stationary phase, and to be eluted with imidazole buffer of increasing concentrations (0 to 600 mM, elution buffer). Eleven out of sixty fractions collected from the elution buffer were analyzed by SDS-PAGE. The enzyme molecular weight of the expressed WfeD with His-tag was calculated to be 28 kDa (Protein molecular weight calculator, Science Gateway). The specific GalT activity of the protein in eluted fractions and cell homogenate was calculated by dividing the reaction rate (\(\mu\)mol/hr) by the amount of total protein (mg) to assess the overall purity and activity of the purified enzyme.

Fractions 15 to 25 showed a band at 27-28 kDa of relatively high intensity which is the expected molecular weight for WfeD, while fractions 30 to 50 showed the same 27~28 kDa band of a relatively high purity (Figure 15A). The specific activity of fractions 14 to 26 and fractions 30 to 50 were calculated. The purified protein in fractions
30 to 50 had a specific activity 16-fold higher than that of the total cell lysate (Figure 15B), meaning that there was 16 times more active GalT enzyme in the purified fractions than in the cell homogenate before purification. The fact that the GalT activity in the purified fractions 30 to 50 was 16-fold higher than that in the cell homogenate and that the amount of the 27~28kDa protein appeared to be much less than 1/16 of the total protein in the cell supernatant (indicated by the “supernatant” lane in Figure 15), suggested that a large proportion of the expressed WfeD protein was still insoluble in aqueous solution.

The 27~28 kDa band in the “wash-through” lane indicated that the target protein had been partially washed through with 10 mM imidazole washing buffer. The Flow-through fraction, which was expected to have much less WfeD protein than the soluble fraction of cell lysate, showed an equally intense band (27~28 kDa) as the “supernatant”. That suggested a low binding affinity for the Ni²⁺ column. Although the anti-His-tag antibody is highly specific, the identity of the suspected target band at 27~28 kDa needed to be confirmed later by MS and western blot.

3.1.4 Product confirmation by mass spectrometry and western blot

In the SDS-PAGE analysis of 250 mM imidazole-eluted supernatant fractions (Figure 16A), the most abundant protein was the one at 27~28 kDa. To confirm the identity of the band, MALDI peptide mass fingerprinting was performed. The band in Figure 16A was excised from the gel, digested with trypsin, and analyzed by MALDI-TOF MS. A list of the obtained peptide masses was fed to the online Mascot search, to match with known protein sequences in the database. WfeD (gi| 187880540) was obtained as the sole ‘hit’, with a protein score of 196 (scores greater than 83 are significant (p<0.05)) (Figure 16B).
Figure 15. Analysis of the Ni\textsuperscript{2+}-NTA purified WfeD with linear gradient of imidazole.  

A) SDS-PAGE of cell supernatant fractions purified by a Ni\textsuperscript{2+}-NTA column and eluted with increasing imidazole concentration. Ten \( \mu \)l of the supernatant, “flow-through”, “wash-through”, and column “eluted fractions” were subjected to SDS-PAGE analysis and stained with Coomassie Blue. The protein standards used are shown on the left. 

B) The specific activity of Ni\textsuperscript{2+}-NTA eluted protein fractions (combined fractions of #14-26 and #30-50) and cell homogenate. The specific activities are 13 and 33 \( \mu \)mol/hr/mg for #14-26 and #30-50, respectively. Bars indicate the difference between the duplicates.
The matching of query peptides with the WfeD glycosyltransferase has a sequence coverage of 52% and the peptide match of 15/21, meaning that 52% of the WfeD sequence was accounted for in the mass analysis and that 15 out of 21 peptides were able to match the WfeD sequence, thereby confirming the 27~28 kDa band as WfeD.

Western blot analysis using anti-His-tag antibody was performed to show the expression level of WfeD. Approximately 5 μg protein of cell lysate and 0.5 μg protein of Ni\textsuperscript{2+}-NTA eluted protein solution were run on SDS-PAGE (15% gel). A much more intense band appeared with the WfeD from the purified protein solution compared to the one in the cell lysate (Figure 16C). Using ImageJ software for densitometry, the optical density of the protein bands in three lanes (the negative control, the cell lysate and the Ni\textsuperscript{2+}-NTA purified protein) can be calculated as an indication of the relative His-tagged protein amount in cell homogenate and in the solution of purified protein\textsuperscript{44-46}. The ImageJ output data are shown in Table 3. The net band density of the purified protein is 2.2-fold of the crude protein in the cell lysate, meaning that the His-tagged protein was 2.2 times more abundant in the purified protein solution than in the cell homogenate. In the absence of the loading control, the western blotting was performed in duplicates with a variation of <10% for the calculated band intensity. Since the total amount of the protein in the cell lysate was 10-fold greater than that of the purified protein solution, approximately 4.5 % of the total protein in cell lysate was His-tagged protein (presumably the WfeD protein). Therefore, among all the proteins being expressed in the \textit{E.coli} cells, WfeD protein took up approximately 4.5 % of the total protein amount.
**Figure 16. Analysis of small scale Ni\textsuperscript{2+}-NTA purified WfeD.**

A) SDS-PAGE separation of Ni\textsuperscript{2+}-NTA purified WfeD for MS analysis, using Coomassie Blue staining. Band #1 at 25-27 kDa was excised and subjected to MS analysis. **B) Mascot score histogram of the 27\textasciitilde28 kDa band** (Figure 16A), by online MS peptide fingerprinting (Matrixscience), for the identification of query protein. X-axis represents the protein score, and protein score is \(-10^\text{Log(P)}\) where P is the probability that the observed match is a random event. Y-axis represents number of different protein hits that have certain degree of sequence match with the query peptides. The highest protein score hit was WfeD protein (gi | 187880540) with protein score of 196. **C) Western blot analysis of crude WfeD in the cell lysate and purified WfeD.** Proteins were stained using anti-His-tag antibody, with bovine \(\beta\)4 GalT (33 kDa) as a negative control. There is only one band at 27\textasciitilde28 kDa present in cell lysate and Ni\textsuperscript{2+}-NTA purified protein solution.
Table 3. The pixel values of the bands on the western blot film (Figure 16C), and calculated relative band density.

For an 8-bit image, all the measured pixel values are still under the upper limit of 255. Area, area of the protein bands. Max. and Min., the highest and lowest optical density values among the smallest area units. Mean, the average optical density value among the smallest area units. NC, negative control. IntDen, integrated density. The net density was calculated by

\[ \text{IntDen of band in cell lysate (or purified protein) – IntDen of band in NC} \]

<table>
<thead>
<tr>
<th></th>
<th>Area (mm²)</th>
<th>Mean (pixel)</th>
<th>Min (pixel)</th>
<th>Max (pixel)</th>
<th>IntDen (pixel)</th>
<th>Net Density (pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell lysate</td>
<td>1548</td>
<td>49.251</td>
<td>18</td>
<td>89</td>
<td>70240</td>
<td>63153</td>
</tr>
<tr>
<td>Ni²⁺-NTA purified protein</td>
<td>1557</td>
<td>22.72</td>
<td>7</td>
<td>39</td>
<td>35375</td>
<td>28288</td>
</tr>
<tr>
<td>Negative control</td>
<td>1524</td>
<td>4.902</td>
<td>0</td>
<td>31</td>
<td>7087</td>
<td></td>
</tr>
</tbody>
</table>

Net Density of purified Protein / Net Density of crude Protein = 2.2
3.1.5 *WfeD* kinetics and storage stability of crude and purified enzyme forms

The reaction rate was proportional to enzyme protein concentration up to 0.016 mg/ml. Untransformed control BL21 cells without inserted pET28a plasmid, showed a background of <5% radioactivity compared to transformed BL21 cells, suggesting a high specific activity in the crude enzyme preparation. The enzyme kinetics data were calculated by EnzFit program. For the crude cell lysate, the apparent $K_M$ for UDP-Gal was 0.1 mM and the $V_{max}$ was 11 μmol/hr/mg, measured at 1 mM GlcNAc-PP-PhU. The apparent $K_M$ for the acceptor substrate GlcNAc-PP-PhU was 0.38 mM with a $V_{max}$ of 12 μmol/hr/mg, measured at 1 mM UDP-Gal (Figure 17A, B). For the purified enzyme, the apparent $K_M$ for UDP-Gal was 0.25 mM and the $V_{max}$ was 40 μmol/hr/mg, measured at 1 mM GlcNAc-PP-PhU. The apparent $K_M$ for the acceptor substrate GlcNAc-PP-PhU was 0.1 mM with a $V_{max}$ of 42 μmol/hr/mg, measured at 1 mM UDP-Gal (Figure 18A, B). The kinetic parameters were determined by two independent repeats with a difference of <10%. The purified enzyme has much higher $V_{max}$ values for UDP-Gal and GlcNAc-PP-PhU than those of the crude cell lysate, which could be explained by the removal of non-specific proteins by purification. The variation between the $K_M$ of the crude and purified enzymes is small. The enzyme stability test under different storage temperatures was conducted to see whether this enzyme has to remain frozen to maintain its activity, and if purification destabilizes the enzyme. The crude enzyme in the cell lysate was stored at -20°C and 4°C. Storage of the crude enzyme in cell homogenate (in 50 mM sucrose) at 4°C maintained 60% of the original activity up to 9 weeks (Figure 19A).
Figure 17. Kinetics of crude WfeD enzyme.

A) The enzyme reaction rate vs UDP-Gal concentration curve, using 1 mM GlcNAc-PP-PhU, and varying UDP-Gal concentrations (0.1 to 1.5 mM). B) The enzyme reaction rate vs GlcNAc-PP-PhU concentration curve, using 1 mM UDP-Gal, and varying GlcNAc-PP-PhU concentrations (0.05 to 2 mM). The parameters were determined by the EnzFit program.
Figure 18. Kinetics of purified WfeD enzyme.

A) The enzyme reaction rate vs UDP-Gal concentration curve, using 1mM GlcNAc-PP-PhU, and varying UDP-Gal concentrations (0.025 to 2 mM). B) The enzyme reaction rate vs GlcNAc-PP-PhU concentration curve, using 1 mM UDP-Gal concentration, and varying GlcNAc-PP-PhU concentrations (0.05 to 0.8 mM). The parameters were determined by the EnzFit program.
Figure 19. Stability of GalT activity of WfeD.

A) Activity of enzyme in bacterial homogenates. Bacteria were homogenized in 50 mM sucrose and stored at +4°C and -20°C.  B) Activity of purified WfeD enzyme at +4°C. The Ni²⁺-NTA purified enzyme protein was stored at +4°C in dialysis buffer of 50 mM Tris, pH 8, 300 mM NaCl and 1 mM EGTA. Enzyme solutions were made up in aliquots and stored at low temperatures. The GalT activity of each aliquot was measured at each time point. All the activity data points were obtained by duplicate measurements, and bars indicate the difference between duplicates.
The decay of the activity by storing at -20°C was not much different from that at 4°C. The Ni²⁺-NTA purified protein was also stored at 4°C (Figure 19B) in 50 mM Tris and 300 mM NaCl, and maintained about 25% of the original activity after storage for 5 weeks.

3.2 Confirmation of Gal transfer and newly synthesized linkage of Galβ1-4GlcNAc

3.2.1 Molecular weight confirmation of the reaction product

The addition of a Gal residue and the β1,4 linkage formed in the product needs to be confirmed to elucidate the type of reaction that the enzyme catalyzes. For the crude cell homogenate catalyzed reaction, the reaction mixture was subjected to Sep-Pak (C18) purification, which separated the hydrophilic donor substrate UDP-Gal (in the H₂O fraction) from the more hydrophobic acceptor substrate GlcNAc-PP-PhU and product Gal-GlcNAc-PP-PhU (in the MeOH fraction). Electrospray MS was performed on the concentrated (by flash evaporation) acceptor substrate and product fraction by Sep-Pak elution. The spectrum (Figure 20), showed a peak of m/z 788 for the product Gal-GlcNAc-PP-PhU and m/z 626 for the substrate GlcNAc-PP-PhU. Since the MS was conducted at negative ion mode of [M-H]⁻ (which means that the compounds are missing one proton and carry a negative one charge), all of the peaks represented [M-1] for the molecular weight. The difference between the two peaks was 162, which corresponded to the molecular weight of one Gal minus water. No peak with the molecular weight of (788 + 162n) was observed in the spectrum, indicating only one Gal (m/z 162) had been added to the substrate. The next sugar in the O-chain is also a Gal residue.
Figure 20. Electrospray MS (negative ion mode) of the isolate reaction product Gal-GlcNAc-PP-PhU (MW. 789), containing substrate GlcNAc-PP-PhU (MW. 627).

The disaccharide lipid product was synthesized by crude WfeD enzyme-catalyzed reaction, purified by Sep-Pak C18 column and subjected to MS analysis.
But the results clearly showed that WfeD has no dual activity and cannot utilize the Gal-GlcNAc-PP-PhU product as a substrate to add another Gal residue.

### 3.2.2 Confirmation of the β-linkage in the reaction product

Galactosidases release terminal Gal residues of a certain linkage from an oligosaccharide. Three different galactosidases were used to determine the linkage between Gal and GlcNAc in the \[^3\text{H}]\text{Gal-GlcNAc-PP-PhU}\) product. The released free \[^3\text{H}]\text{Gal}\) was separated from the reaction mixture by an AG1 column that binds the uncleaved product. We showed that Gal was β-linked in the reaction product, and likely in a 1-4 linkage. The cleavage of the Gal-GlcNAc linkage catalyzed by Jack Bean (β1,4 specific) β-galactosidase and bovine testicular (β1-3,-4,-6-specific) β-galactosidase were 52 % and 60 % of the untreated control Gal-GlcNAc-PP-PhU. No cleavage of Gal-GlcNAc-PP-PhU by Green Coffee Bean α-galactosidase was observed (Figure 21), confirming the β-linkage in the reaction product.

### 3.2.3 Linkage identification of the product by NMR

The \(^1\text{H}-\text{NMR}\) spectrum of the reaction product (in CD\(_3\)OD), as compared to that of the substrate GlcNAc-PP-PhU, showed a new doublet signal at 4.345 ppm with a coupling constant of 7.6 Hz indicative of H-1 of Gal in β-linkage (Table 4). In the spectrum of the reaction product of β3GalT WbbD (Galβ1-3GlcNAc-PP-PhU), the H-1 signal of the terminal Gal was also at 4.35 ppm\(^{39}\). The chemical shifts of the other \(^1\text{H}\) signals of GlcNAc in the substrate (H-1, H-2, and H-3) were only slightly different in the product (Table 4).
Figure 21. Galactose cleavage of the galactosidase-treated reaction product.

The percentage of free radioactive Gal released from WfeD product [\(^{3}H\)Gal-GlcNAc-PP-PhU is shown. Enzyme product [\(^{3}H\)Gal-GlcNAc-PP-PhU was synthesized using purified WfeD enzyme. Galactosidases were used as described in Section 2.3.2. Bovine testicular \(\beta\)-galactosidase is specific for \(\beta\)-1-3, \(\beta\)-1-4, and \(\beta\)-1-6 linkages; Green coffee bean \(\alpha\)-galactosidase is specific for \(\alpha\)-linkages, and Jack bean \(\beta\)-galactosidase is specific for \(\beta\) 1-4 linkages. All galactosidase conditions were tested in duplicates; and bars indicate the difference between the duplicates.
C) A section of the enlarged 1D $^1$H NMR spectrum from 3.4 to 4.6 ppm. B) A section of the enlarged 2D $^1$H-$^1$H COSY spectrum from 3.4 to 4.5 ppm. C) A section of the enlarged 2D $^1$H-$^{13}$C HSQC spectrum, from $^1$H 3.5 to 4.1 ppm and from $^{13}$C 52 to 80 ppm. The product (‘Gal $\beta$ 1-4GlcNAc$\alpha$ -PP-PhU) of WfeD-catalyzed reaction was isolated by C18 Sep-Pak column and C18 reverse phase HPLC. ‘H-1 of ‘Gal shows a chemical shift of 4.355 ppm indicating a $\beta$-configuration for the Gal-GlcNAc linkage, and C-4 of GlcNAc shows a chemical shift of 79.5 indicating a glycosylation modification of C-4. A total amount of 500 $\mu$mol of reaction product was obtained, dried by flash evaporation, exchanged three times with 99.96 % D$_2$O, and dissolved in CD$_3$OD for 600-MHz NMR analysis.

Figure 22. 600 MHz proton NMR spectrum of the WfeD product.
Table 4. $^1$H and $^{13}$C NMR parameters established from 1D $^1$H NMR, 2D COSY and $^1$H-$^{13}$C HSQC.

Comparison between NMR parameters of substrates GlcNAc$\alpha$-PP-PhU (Montoya et al.$^{29}$) and WfeD product 'Gal$\beta$1-4GlcNAc$\alpha$-PP-PhU (established from Figure 22).

<table>
<thead>
<tr>
<th></th>
<th>H</th>
<th>Shift (ppm)</th>
<th>J coupling (Hz)</th>
<th>C</th>
<th>Shift (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAc$\alpha$-PP-PhU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNAc</td>
<td>H-1</td>
<td>5.53</td>
<td></td>
<td>C-1</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>H-2</td>
<td>3.97</td>
<td></td>
<td>C-2</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>H-3</td>
<td>3.75</td>
<td></td>
<td>C-3</td>
<td>72.9</td>
</tr>
<tr>
<td></td>
<td>H-4</td>
<td>3.33</td>
<td></td>
<td>C-4</td>
<td>71.9</td>
</tr>
<tr>
<td>'Gal$\beta$1-4GlcNAc$\alpha$-PP-PhU (WfeD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNAc</td>
<td>H-1</td>
<td>5.561</td>
<td></td>
<td>C-1</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td>H-2</td>
<td>4.027</td>
<td></td>
<td>C-2</td>
<td>53.1</td>
</tr>
<tr>
<td></td>
<td>H-3</td>
<td>3.846</td>
<td></td>
<td>C-3</td>
<td>69.4</td>
</tr>
<tr>
<td></td>
<td>H-4</td>
<td>3.626</td>
<td></td>
<td>C-4</td>
<td>79.1</td>
</tr>
<tr>
<td>'Gal</td>
<td>'H-1</td>
<td>4.355</td>
<td>7.6</td>
<td>'C-1</td>
<td>103.4</td>
</tr>
<tr>
<td></td>
<td>'H-2</td>
<td>3.538</td>
<td>6.9, 10.1</td>
<td>'C-2</td>
<td>70.8</td>
</tr>
<tr>
<td></td>
<td>'H-3</td>
<td>3.465</td>
<td>2.9, 9.5</td>
<td>'C-3</td>
<td>73.1</td>
</tr>
<tr>
<td></td>
<td>'H-4</td>
<td>3.792</td>
<td>2.5</td>
<td>'C-4</td>
<td>68.8</td>
</tr>
</tbody>
</table>
1D $^1$H NMR (Figure 22A), 2D COSY (Figure 22B) and 2D $^1$H-$^{13}$C HSQC (Figure 22C) identified 70% of the carbon and proton shifts of the disaccharide moiety of the Gal-GlcNAc-R product. There was a large downfield shift of GlcNAc H-4 from 3.33 ppm in the substrate to 3.626 ppm in the product and a large shift of the C4 signal of 71.9 in the substrate to 79.1 in the product (Table 4). This indicated that Gal was linked to the 4-position of GlcNAc and a Galβ1-4GlcNAc linkage in the enzyme product. Therefore, these results proved that WfeD is a UDP-Gal: GlcNAc-R β1,4-Gal-transferase. These chemical shifts are consistent with the ones being reported in previous study.\textsuperscript{29,48}

### 3.3 Substrate specificity, enzyme inhibition and cofactor requirement of WfeD

#### 3.3.1 Donor substrate specificity

To determine whether WfeD is only specific to UDP-Gal as the donor substrate and if any non-specific transferases that utilize other nucleotide sugars exist in the cell, six other nucleotide sugars were tested using the crude cell homogenate. Those nucleotide sugars were chosen for their structural resemblance to UDP-Gal. The result of the donor substrate specificity study of the crude cell homogenate (Figure 23) showed that none of the nucleotide sugars yielded greater than 2% of the activity observed with UDP-Gal. This indicated that WfeD is highly specific for UDP-Gal as the donor substrate. Although the 4-epimerase present in \textit{E.coli} cells usually acts to convert UDP-Glc to UDP-Gal, the crude enzyme in the cell homogenate did not show any activity with UDP-Glc. Therefore, the result not only indicated a strict specificity of WfeD for UDP-Gal as the donor substrate, but also showed that the 4-epimerase does not convert any UDP-Glc to UDP-Gal within the 10 min incubation time.
Figure 23. Donor specificity of the crude enzyme, using seven nucleotide sugars at the indicated concentrations.

The activities are shown as $\mu$mol/hr/mg and the relative activities as percentage of the positive control using UDP-Gal (100%). Each substrate was tested in duplicates; and bars indicate the difference between the duplicates. CMP-SA stands for CMP-sialic acid.
3.3.2 Role of pyrophosphate in acceptor substrate specificity

In order to define the requirements of WfeD for the acceptor substrate, two categories of GlcNAc analogs (the neutral group and the negatively charged group) were tested as substrates (Table 5). Using the AG1 anion exchange method to isolate product, none of the neutral compounds (with either hydrophobic or hydrophilic aglycone substituent) showed activity higher than 5% of the standard substrate. The GlcNαc-\((CH_2)_{11}\)-O-phenyl compound which had only one phosphate group showed 11% activity compared to the standard substrate. The charged compounds which have a pyrophosphate group but different lipid moieties, showed various levels of high activity (C18 Sep-Pak). This is consistent with the structure of the natural substrate GlcNAc-PP-Und. The results indicated that the lipid moiety is not a critical determinant of the acceptor substrate specificity, although a hydrophobic group may be required. However, WfeD has an absolute requirement for minimally one phosphate group, and requires the pyrophosphate group in the acceptor for full activity.

3.3.3 Screening of compounds as potential inhibitors

Potential inhibitors with the structure GlcNAc-R were selected that had an activity of < 5% compared to GlcNAc-PP-PhU, or were previously shown to be inhibitors of mammalian GalT, i.e. GlcNαcβ-(2-naphthyl) and GlcNBuβ-S-(2-naphthyl). The enzyme activity in the presence of both substrate 580 (GlcNAc-PP-PhU) and potential inhibitor compounds were recorded in \(\mu\)mol/hr/mg (Table 6), and the percentage of inhibition was calculated by

\[
\frac{(Activity_{without\ inhibitor} - Activity_{with\ inhibitor}) \times 100}{Activity_{without\ inhibitor}}
\]
Table 5. WfeD GalT acceptor substrate specificity.

Synthetic compounds were used as acceptor substrates at 0.25 mM concentration in the GalT assay. The reaction rate is represented as the percentage of the activity in control assays using GlcNAcα-PO₃-PO₃-(CH₂)₁₁-OPh, with an activity of 20 μmol/hr/mg (100%). Assays were performed as described in the Methods section. Each substrate was tested in at least duplicate determinations.
<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Chemical Structure</th>
<th>Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAcα-PO₃-PO₃-(CH₂)₁₁-OPh</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td>100</td>
</tr>
<tr>
<td>GlcNAcα-PO₃-PO₃-(CH₂)₆-OPh</td>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>39</td>
</tr>
<tr>
<td>GlcNAcα-PO₃-PO₃-(CH₂)₁₆-OPh</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td>92</td>
</tr>
<tr>
<td>GlcNAcα-PO₃-(CH₂)₁₁-OPh</td>
<td><img src="image4" alt="Chemical Structure" /></td>
<td>11</td>
</tr>
<tr>
<td>GlcNAcα-PO₃-PO₃-(CH₂)₉CH₃</td>
<td><img src="image5" alt="Chemical Structure" /></td>
<td>74</td>
</tr>
<tr>
<td>GlcNAcα-(CH₂)₁₁-OPh</td>
<td><img src="image6" alt="Chemical Structure" /></td>
<td>&lt;5</td>
</tr>
<tr>
<td>Chemical Structure</td>
<td>Concentration</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-(CH$<em>2$)$</em>{11}$-OPh</td>
<td>&lt;5</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-CO-CH$_2$-PO$_3$-(CH$<em>2$)$</em>{11}$-OPh</td>
<td>&lt;5</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-CO-CH$_2$-CO-O-(CH$<em>2$)$</em>{11}$-OPh</td>
<td>&lt;5</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-(3-isoquinolinyl)</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNBuβ-(2-naphthyl)</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-(2-naphthyl)</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNBuβ-S-(2-naphthyl)</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ-Bn</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNAcα-Bn</td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ</td>
<td><img src="image" alt="GlcNAcβ" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>----</td>
</tr>
<tr>
<td>GlcNAcβ-S-phenyl</td>
<td><img src="image" alt="GlcNAcβ-S-phenyl" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>4-deoxy-GlcNAcα-Bn</td>
<td><img src="image" alt="4-deoxy-GlcNAcα-Bn" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>4-deoxy-GlcNAcα</td>
<td><img src="image" alt="4-deoxy-GlcNAcα" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>6-deoxy-GlcNAcα</td>
<td><img src="image" alt="6-deoxy-GlcNAcα" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>6-thio-GlcNBuβ-(2-naphthyl)</td>
<td><img src="image" alt="6-thio-GlcNBuβ-(2-naphthyl)" /></td>
<td>&lt;1</td>
</tr>
<tr>
<td>GlcNBuβ-NH-(2-naphthyl)</td>
<td><img src="image" alt="GlcNBuβ-NH-(2-naphthyl)" /></td>
<td>&lt;1</td>
</tr>
</tbody>
</table>
Table 6. Screening of potential inhibitors of WfeD GaIT.

The synthetic compounds were selected from those being inactive as substrates. Each potential inhibitor was used at 0.5 mM concentration along with 0.25 mM GlcNAcα-PO₃-PO₃-(CH₂)₁₁-Oph in the assay. The positive control assay contained the acceptor substrate only, and the activity of 20 μmol/hr/mg was used as the standard (100 %). Assays were carried out as described in the Methods section. Each compound was tested in duplicate determinations.

<table>
<thead>
<tr>
<th>Compound Name</th>
<th>Chemical Structure</th>
<th>Activity Inhibition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAcα-CO-CH₂-PO₃-(CH₂)₁₁-Oph</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
<tr>
<td>GlcNAcβ-(3-isoquinolinylyl)</td>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
<tr>
<td>GlcNAcβ-(2-naphthyl)</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
<tr>
<td>GlcNAcβ-Bn</td>
<td><img src="image4" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
<tr>
<td>GlcNAcα-Bn</td>
<td><img src="image5" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
<tr>
<td>1-Thio-N-butrylGlcNAc-β-(2-naphthyl)</td>
<td><img src="image6" alt="Chemical Structure" /></td>
<td>0</td>
</tr>
</tbody>
</table>
Having a concentration of twice the substrate concentration in the assay solution, none of these compounds showed inhibition of the enzyme activity in the presence of GlcNAc-PP-PhU (Table 6). This suggested that an acceptor substrate analog inhibitor also has to have the pyrophosphate group or a close structural analog for binding to the enzyme.

3.3.4 Divalent metal ions enhance WfeD enzyme activity

Mammalian GT activities often have a requirement for divalent metal ions. To examine the cofactor requirement of WfeD, salt solutions containing Mn\(^{2+}\), Mg\(^{2+}\), Cu\(^{2+}\), Ca\(^{2+}\), Pb\(^{2+}\) or Ni\(^{2+}\) were added to the assay. EDTA and H\(_2\)O alone (without any cofactor ion) were used as negative controls. Mn\(^{2+}\) showed the highest stimulation of enzyme activity in both crude and purified enzyme activity assays (Figure 24). Ca\(^{2+}\) and Mg\(^{2+}\) were also effective. Surprisingly, the Pb\(^{2+}\) ions, often inhibitory to enzymes, and the Ni\(^{2+}\) ions, which are not usually used as GT cofactors due to their toxicity, were both shown to induce high enzyme activity in the cell homogenate as well as in the purified enzyme (Figure 24)\(^9\).

A further test showed that the effect of Pb\(^{2+}\), Ni\(^{2+}\), and Mn\(^{2+}\) on enzyme activity was concentration-dependent (Figure 25A). The ionic radius of these divalent metal ions (Figure 25B), which was thought to be a major factor in metal ion—enzyme—substrate interaction, shows no clear correlation between the size of the ionic radius and the enzyme activity\(^50\). HPLC analysis confirmed the identity of the enzyme products formed in the presence of these metal ions. This finding may uncover a novel role of Pb\(^{2+}\) in bacterial metabolism, and may help to define the tolerance to Pb\(^{2+}\) ions in physiological conditions.
Figure 24. Cofactor (divalent metal ion) requirement of WfeD.

**A) Cofactor requirement of the crude WfeD enzyme.** Freshly homogenized *E.coli* cells were used as crude enzyme. Cu$^{2+}$ ions did not stimulate the activity. **B) Cofactor requirement of the purified WfeD enzyme.** The purified WfeD solution stored for 3 weeks was used for the assay. The reaction rate is represented by activity (µmol/hr/mg). Each ion was tested by duplicates, and bars indicate the difference between duplicates. The products obtained by using Mn$^{2+}$, Pb$^{2+}$ and Ni$^{2+}$ ions were confirmed by HPLC (C18, 24% / 76% of AN/H$_2$O).
Figure 25. Effect of metal ion cofactor on WfeD GalT activity.

A) Cofactor concentration effect on WfeD GalT activity. MnCl₂ (solid curve), Pb(OAc)₂ (evenly dotted curve), and NiSO₄ (unevenly dotted curve) were present in the GalT assays using 0.012 mg/ml crude cell homogenate. B) A line chart of the ionic radius and stimulated enzyme activity correlation of six divalent metal ions. The six metal ions listed from left to right are in the order of their atomic numbers. The activity values are represented by squares and the ionic radius values are represented by diamonds.
The Pb\(^{2+}\) effect on enzyme activity was further tested on another enzyme to examine whether this is a general phenomenon for GTs. The bovine milk β4GalT was chosen, since its activity was also enhanced by Mn\(^{2+}\) as a cofactor. Thus Pb(OAc)\(_2\) was tested with the bovine milk β4GalT using H\(_2\)O and MnCl\(_2\) as the controls. As a result (Figure 26A), bovine β4GalT activity was stimulated by Pb\(^{2+}\) ions, but much less than WfeD. This is an unexpected and unusual effect of a toxic metal on both mammalian and bacterial GalTs. The WfeD activity was also tested with different acetate salts, such as NaOAc and Pb(OAc)\(_2\), to see whether the acetate anion had an effect on the enzyme activity. NaOAc did not show a significantly higher activity than the background (the negative control without any cofactor ion being added), while Pb(OAc)\(_2\) showed a high stimulation relative to MnCl\(_2\) (Figure 26B).

3.4 Role of amino acid residues in enzyme activity

3.4.1 Effect of amino acid modifying reagents on WfeD activity

In order to examine the involvement of specific amino acids in catalysis, we employed a number of amino acid modifying reagents as a route to facilitate identification of residues that are of particular importance to WfeD function. Specifically, reagents were chosen that could react covalently with amino acids and alter the interaction between the amino acid and its environment, or break disulfide bonds. The inclusion of 0.2 mM DEPC (diethylpyrocarbonate) in the assay showed a 31 % inhibition of activity, while 0.2 mM HPG (p-hydroxyphenylglyoxal) showed a 50 % inhibition (Figure 27). This suggested a potential role of His and Arg in protein structure or catalysis.
Figure 26. Further test of Pb{$^{2+}$} effect on enzyme activity.

A) Pb{$^{2+}$} ion effect on the activity of bovine β4GalT. Effect of MnCl$_2$ and Pb(OAc)$_2$ on bovine milk β4 GalT enzyme. Each reaction was carried out using the condition of 0.017 mg/ml GaLT, 0.5 mM GlcNAcβBn, 5 mM divalent ion, 1 % BSA, 75 mM MES and 1 mM UDP-Gal. The reaction product was purified by AG1X8 resin ion exchange chromatography. B) Acetate anion effect on the activity of the WfeD enzyme. All of the assay results were obtained in duplicates, and bars indicate the difference between the duplicates.
Different amino acid reagents were pre-incubated with purified enzyme for 10 min prior to the assays. Percentages of activation or inhibition were obtained by comparing to the positive control (without any amino acid reagent). DEPC—diethyl pyrocarbonate, HPG—p-Hydroxyphenylglyoxal, DTNB—5,5′-Dithio-bis (2-nitrobenzoic Acid), IAA—iodoacetic acid. Each amino acid reagent was tested in duplicates, and bars indicate the difference between the duplicates.
The alkylating agent IAA (iodoacetic acid) showed a 20% inhibition, indicating that one or both of the Cys residues may be involved. The most drastic effect was seen with DTNB (5,5′-dithio-bis-2-nitrobenzoic acid) which binds to the reduced –SH groups of Cys and inhibited the activity by 85%. Dithiothreitol (DTT), which reduces disulfide bonds, stimulated the enzyme activity 5-fold (Figure 28A), while β-mercaptoethanol also increased the activity (Figure 28B). This suggested that one or both of the Cys residues may be required to be reduced for full activity, and that the enzyme is activated by reducing the oxidized Cys residues. The roles of both Cys in the catalytic mechanism or enzyme stability remains to be further investigated. As compared to the cellular organelles and the periplasmic space, the cytosol is relatively free of oxidizing agents that would greatly impair the activity of WfeD, therefore, it would make more biological sense for WfeD to be localized in the cytosol than any other cellular compartment.

### 3.4.2 Sequence comparison of WfeD and putative glycosyltransferases for the selection of crucial amino acids

Most GalTs known to utilize the DxD motif in their catalytic mechanism are members of GT-A fold family, which have low overall sequence similarity and poor sequence alignment. By sequence similarity, WfeD is classified as a member of GT26 family, with most of its enzymes proposed to have the GT-B fold. Although the WfeD protein contains a DxD resembling motif, it possesses a fold (WecG/TagA) different from GT-A or GT-B. By BLAST search (blastp, NCBI), several GT26 family proteins (ZP0525486 and ZP05257042) and GT WecG/TagA folded proteins (YP737451 and ZP02177500) were shown to have highest sequence similarity (>30%) to WfeD (Figure 29), but none of them
Figure 28. Effect of disulfide reducing agents on WfeD activity.

A) Effect of varying concentrations of DTT on WfeD activity. Crude WfeD GaIT activity (with crude cell homogenate stored for 3 months) was tested with varying concentrations (0 to 3 mM) of DTT. B) Effect of varying concentrations of $\beta$-mercaptoethanol on WfeD activity. Crude WfeD GaIT activity (with crude cell homogenate stored for 3 months) was tested with varying concentrations (0 to 30 $\mu$M) of $\beta$-mercaptoethanol.
Figure 29. Alignment of the complete WfeD sequence with other putative GTs that have high sequence similarity to WfeD.

Proteins having similarity (>30%) to WfeD (by BLAST, NCBI) have been aligned by T-COFFEE. The light grey shaded areas indicate amino acid similarities, and light green shaded sequences resemble the DxD motif. The top line represents the WfeD sequence. Although the DxD motif exists in various forms, they are found in all of these proteins. Most proteins have multiple Cys (highlighted in pink) but only Cys148 (in WfeD) is conserved. The RKR sequence (highlighted in yellow) is found in all of these proteins.
CLUSTAL FORMAT for T-COFFEE Version_8.06

SCORE=71, Nseq=18, Len=321

WfeD
YP_130857 teichoic acid biosyn protein
YP_737451 GT WecG/TagA/CpsF Shewanella
YP_02177500 GTWecG/TagA/CpsF Hydrogenivir
ZP_0525486 BacteroidesGT26
ZP_0525486 BacteroidesGT26

WfeD
YP_130857 teichoic acid biosyn protein
YP_737451 GT WecG/TagA/CpsF Shewanella
YP_02177500 GTWecG/TagA/CpsF Hydrogenivir
YP_130857 teichoic acid biosyn protein
YP_737451 GT WecG/TagA/CpsF Shewanella
YP_02177500 GTWecG/TagA/CpsF Hydrogenivir
YP_0525486 BacteroidesGT26

WfeD
YP_130857 teichoic acid biosyn protein
YP_737451 GT WecG/TagA/CpsF Shewanella
YP_02177500 GTWecG/TagA/CpsF Hydrogenivir
YP_0525486 BacteroidesGT26

WfeD
YP_130857 teichoic acid biosyn protein
YP_737451 GT WecG/TagA/CpsF Shewanella
YP_02177500 GTWecG/TagA/CpsF Hydrogenivir
has been functionally identified. This means that the functional analysis of WfeD might reveal the characters of a group of undiscovered enzyme proteins.

Sequence alignment (Figure 29) of these putative GTs with relatively high sequence similarity (>30%) to WfeD revealed some highly conserved (or potentially important) amino acid residues, such as the RKR sequence and Cys residues. WfeD has a DYEIE sequence resembling a DxD motif found in virtually all putative GTs (Figure 29). Therefore, three single amino acid mutants were created by mutating Asp99, Glu101 and Glu103 to Ala. Moreover, sequence alignment showed a high conservation of an RKR motif among WfeD and other putative GTs which corresponds to the effects of amino acid specific reagents that suggested that one or more Arg residues are crucial for enzyme activity. Therefore, we proposed that positively charged amino acids in the cluster RKR may be important for substrate binding which could be tested in mutants R210A, R212A and K211A.

3.4.3 Expression confirmation of WfeD mutant proteins

Lei Wang in China made mutants of WfeD encoding plasmids (including D99A, E101A, E103A, R210A, K211A and R212A) by site-direct mutagenesis. The success of mutagenesis was confirmed by subsequent DNA sequencing. The mutant plasmids were transformed into BL21 E.coli cells by electroporation and grown overnight on agar plates. Single colonies were then picked for small scale cell growth and expression testing in LB. The expression of mutant proteins was induced by IPTG, similar to the procedure for the expression of the wild type WfeD. Small scale Ni²⁺-NTA purification was applied to help identify and quantitate the amount of each mutant protein. The protein concentrations of the purified mutants were measured prior to SDS-PAGE (15% gel) analysis (Figure 30).
Figure 30. SDS-PAGE of Ni\(^{2+}\)-NTA (small scale) purified mutants using Coomassie Blue staining

From left to right, Lane 1 shows the supernatant of D99A mutant cells that was not treated with IPTG in cell growth; Lane 2 to Lane 7 are small scale purified protein solutions of the six mutants that were treated with IPTG in cell growth. The molecular weight markers are on the right. 15 \(\mu\)l of each elution solution was loaded into the well, with the amount of proteins varying from 5 to 10 \(\mu\)g. A band at 27~28 kDa appeared in every lane of the six mutants, and was absent in the pre-induction supernatant of mutant D99A.
Figure 30 showed that the bands at 27~28 kDa in each mutant were the only protein showing enhanced expression after IPTG induction, indicating their identities as the WfeD variants. The protein expression of those mutants was further confirmed by western blotting.

Western blotting was performed on purified mutants following the same procedure as the one for the wild type WfeD (in Section 3.1.4). Dilutions were made to the purified protein solutions to adjust protein amounts within the sensitivity range of the antibody detection. The western blot (Figure 31) showed that all of the WfeD mutants were expressed and appeared to have a molecular weight of 29~30 kDa, and that none of these proteins appeared to be degraded. Some non-specific light bands shown in the lane of “wild type” might be due to partial degradation of WfeD or in complete wash of the western blot membrane. With the help of ImageJ for densitometry, the relative amount (using wild type as the standard) of all mutants and the wild type proteins were calculated (Table 7)\textsuperscript{44-46}. In the absence of the loading control, the western blotting was performed in duplicates with a variation of <10% for the calculated band intensity. Following the western blot analysis, we proceeded to test the activity of all mutant constructs. Since all mutant proteins were expressed, the western blot analysis proved that any great reduction among the activities of the mutant and wild type enzymes should not be due to insufficient expression of protein.
Figure 31. Western blot of Ni\textsuperscript{2+}-NTA (small scale) purified wild type and six mutant proteins of WfeD.

From left to right are small scale Ni\textsuperscript{2+}-NTA (250 mM imidazole in elution buffer) purified wild type and mutant WfeD elution fractions, and a bovine \(\beta_4\text{GalT} \) (33 kDa) without His-tag as the negative control. The amounts of each mutant loaded onto the SDS gel were 3.6 \(\mu\text{g} \) (D99A), 2.5 \(\mu\text{g} \) (E101A), 2.5 \(\mu\text{g} \) (E103A), 3.3 \(\mu\text{g} \) (R210A), 2.7 \(\mu\text{g} \) (K211A), 1.5 \(\mu\text{g} \) (R212) and 2.7 \(\mu\text{g} \) (wild type). An anti-His-tag antibody was used against the WfeD protein. The molecular weights with corresponding migrating distances are indicated on the left. A band at 28-30 kDa was shown in all mutants and wild type lanes.
**Table 7. Western blot (Figure 31) band density calculated by ImageJ and the estimated relative amount of His-tagged protein in Ni\(^{2+}\)-NTA purified WfeD mutants**

For an 8-bit image, the maximal pixel values are still under the upper limit of 255. 

**IntDen**- integrated density, **WT**- wild type, **NC**- negative control protein.

The **Net IntDen** was calculated by

\[
\text{Net IntDen} = \text{IntDen of WT(or Mutant)} - \text{IntDen of NC}
\]

The **Protein Amount Specific Band Density** was calculated by

\[
\frac{\text{Net IntDen}}{\text{Loaded protein amount}}
\]

The **Relative Amount of His-tagged Protein** was calculated by

\[
\frac{\text{Protein Amount Specific Band Density of Mutant Protein}}{\text{Protein Amount Specific Band Density of WT Protein}}
\]

<table>
<thead>
<tr>
<th></th>
<th>Area (mm(^2))</th>
<th>Mean (pixel)</th>
<th>Max (pixel)</th>
<th>IntDen (pixel)</th>
<th>Net IntDen (pixel)</th>
<th>Loaded protein (µg)</th>
<th>Protein Amount Specific Band Density (pixel/µg)</th>
<th>Relative Band Density</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wild Type (WT)</strong></td>
<td>2257</td>
<td>53</td>
<td>135</td>
<td>119651</td>
<td>112364</td>
<td>2.7</td>
<td>41616</td>
<td>1</td>
</tr>
<tr>
<td>D99A</td>
<td>2646</td>
<td>78</td>
<td>130</td>
<td>206594</td>
<td>199307</td>
<td>3.6</td>
<td>55363</td>
<td>1.3</td>
</tr>
<tr>
<td>E101A</td>
<td>2257</td>
<td>49</td>
<td>105</td>
<td>111079</td>
<td>103791</td>
<td>2.5</td>
<td>41517</td>
<td>1</td>
</tr>
<tr>
<td>E103A</td>
<td>2439</td>
<td>25</td>
<td>43</td>
<td>62088</td>
<td>54801</td>
<td>2.5</td>
<td>21920</td>
<td>0.5</td>
</tr>
<tr>
<td>R210A</td>
<td>2240</td>
<td>32</td>
<td>70</td>
<td>75634</td>
<td>68347</td>
<td>3.3</td>
<td>20711</td>
<td>0.5</td>
</tr>
<tr>
<td>K211A</td>
<td>2826</td>
<td>19</td>
<td>40</td>
<td>52608</td>
<td>45321</td>
<td>2.7</td>
<td>16786</td>
<td>0.4</td>
</tr>
<tr>
<td>R212A</td>
<td>2288</td>
<td>11</td>
<td>22</td>
<td>24108</td>
<td>16820</td>
<td>1.5</td>
<td>9894</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>NC</strong></td>
<td>2496</td>
<td>6</td>
<td>31</td>
<td>7287</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4.4 GalT activity of the WfeD mutants

The cell lysates for each WfeD mutant and wild type were assayed for GalT activity using the standard assay condition. The D99A and E103A mutants were highly active (equally high as the wild type), but the E101A mutant exhibited minimal activity (Figure 32A). R210A and R212A mutants were active while the K211A mutant also lacked activity (Figure 32A).

The GalT activity of each mutant was then divided by the relative amount of His-tagged protein (Table 7), to obtain the normalized GalT activity (Figure 32B). Although the western blot analysis of all the mutant and wild type WfeD has confirmed the expression of each protein, it is still necessary to clarify that the variations among the activities are mainly due to the altered protein activity and not the difference in protein expression. Therefore, the normalized activity was used to estimate the GalT activity of all mutants based on the quantified variation among their His-tagged protein expression levels seen in the western blot, and also based on the assumption that the relative His-tagged protein expression levels of wild type and mutant WfeD in the cell lysate were reflected by the relative band intensities in the western blot. The limited activity observed for the Glu101 and Lys211 mutants relative to wild type indicated an important role of these residues in substrate binding, catalysis, or protein structure, possibly by forming interaction between the charged functional groups of the amino acids and the substrates or, alternatively, with other residues within the enzyme. E101 is the third residue in the DYEIE sequence of WfeD, which resembles the DxD motif found in many GTs32,33,36,37. Therefore, the DYEIE sequence in WfeD may indeed function as a DxD motif and Glu101 is necessary for catalysis. Likewise, the K211 residue from the RKR sequence
could be involved in the substrate and enzyme interaction, since the positively charged amino acid cluster could facilitate substrate binding by ionic interaction with the pyrophosphate group. Although the mutants R210A and R212A maintained their GalT activity, the loss of activity by HPG (Arg modifying reagent) indicated that some of the other Arg residue(s) are involved in enzyme activity.

The GalT activity of mutants was also tested in the presence of Pb\(^{2+}\) as the metal ion cofactor (Figure 33), and product formation was confirmed by HPLC. Replacement of E101 and K211 with Ala greatly impaired enzyme activity in the presence of either Mn\(^{2+}\) or Pb\(^{2+}\), although a higher activity was seen in the K211A mutant in the presence of Pb\(^{2+}\). The comparative test of Mn\(^{2+}\) and Pb\(^{2+}\) ions showed that the divalent metal ions are vital for the activity (or catalysis). However, the Pb\(^{2+}\) ions seemed to play more than one role, and may also be involved in maintaining the activity in the absence of K211, possibly by interacting with the acceptor substrate. It remains to be shown if indeed Pb\(^{2+}\) can substitute for the positively charged amino acid to promote acceptor substrate binding.

The conformation of the wild type and mutant enzymes may also differ. A Met344His mutant of mammalian \(\beta 4\)GalT1 prefers to stay in the conformation that allows acceptor binding and has different metal ion binding properties\(^{40,41}\). The K211 mutant of WfeD may also have an acceptor binding conformation where only the large Pb\(^{2+}\) ion facilitates catalysis. This could be shown by crystallography of the wild type and mutants, together with either Mn\(^{2+}\) or Pb\(^{2+}\) ions.
Figure 32. Activity assay of WfeD mutant enzymes. A) Activity of WfeD wild type and mutant enzymes. GalT activity of WfeD mutants K211A, R212A, R210A, E103A, E101A and D99A were assayed using 1 mM UDP-Gal and 0.25 mM GlcNAc-PP-PhU acceptor substrate, and 5 mM MnCl$_2$. The reaction rates are shown as activity in $\mu$mol/hr/mg. B) Normalized activity of WfeD wild type and mutant enzymes. Normalized activity is the specific activity of each mutant and wild type (from Figure 32A) divided by the relative His-tagged protein amount in each strain. Each WfeD strain was tested in duplicates, and bars indicate the difference between the duplicates.
Figure 33. Activity assay of two WfeD mutants after Pb$^{2+}$ and Mn$^{2+}$ stimulation. A) Activity of WfeD wild type and two mutants by Pb$^{2+}$ and Mn$^{2+}$ stimulation. B) Normalized activity of WfeD wild type and two mutants by Pb$^{2+}$ and Mn$^{2+}$ stimulation. The normalized activity was obtained by the same method used for obtaining the normalized activity in Figure 32B. This activity assay of wfeD wild type and mutants K211A, E101A used 5 mM MnCl$_2$ (grey histogram) and 5 mM Pb(OAc)$_2$ (white histogram). Each mutant was tested in duplicates, and bars indicate the difference between the duplicates.
CHAPTER 4  GENERAL DISCUSSION AND FUTURE DIRECTIONS

4.1 General discussion

4.1.1 Membrane requirement of WfeD enzyme activity

The sequence hydrophobicity analysis showed that there is a long hydrophobic stretch of 19 amino acids at the N-terminus of WfeD that might mediate the association with the cell membrane. Since the enzyme activity was still maintained after treating the cell homogenate with detergents and solubilising the enzyme, the membrane association may not be a critical factor for in vitro activity. In vivo, the association may be necessary to hold the detergent-like acceptor substrate and the enzyme in the right position that favours Gal transfer. Enzymes that assemble O-antigens are thought to be loosely associated with the membrane while mammalian GalT are imbedded into the Golgi membrane with their N-terminal transmembrane region. For the in vivo activity, a number of factors have to be considered such as the relative location of membrane associating sequence compared to the active site of the enzyme. In vitro, disruption of membrane-protein association by detergent solubilisation or cell homogenization might affect the folding or conformation of the enzyme active site or the access of enzyme to its acceptor substrate. Membrane components as well as detergents can form micelles and thus alter the accessibility of the enzyme active site or the substrate. Both the natural and the synthetic acceptors have detergent-like properties. For more accurately identify the cellular localization of WfeD, cell fractionation studies should be conducted.
4.1.2 Enzyme purification efficacy and enzyme stability

The specific activity comparison of the Ni$^{2+}$-NTA column-purified WfeD and the WfeD in the cell lysate indicated that the former has 16-fold more concentrated active GalT protein than the latter (Section 3.1.3). Meanwhile, the western blot analysis of the cell lysate and purified WfeD showed that the amount of WfeD protein is approximately 1/20 of the total protein in the cell lysate (Section 3.1.4). Therefore, protein solubilisation and the purification procedure have been successful. Previous attempts to purify a bacterial β3GalT did not result in significant activity of the purified protein fraction, possibly because the enzyme required the membrane association for activity$^{39}$. A western blot analysis of the WfeD amount in cell homogenate and supernatant could have been done to identify the proportion of soluble WfeD in all of the expressed WfeD protein. The SDS-PAGE of Ni$^{2+}$-NTA eluted fractions (Figure 14A) indicated that an imidazole concentration of 300 mM (Fraction #30) is desirable for obtaining a relatively pure WfeD. Although the dialysis process was relatively standard for removing the imidazole ions from the protein solution, a trace amount of imidazole could still have remained. A control study of imidazole concentration effect on the enzyme activity of WfeD should be done, to determine whether the activity was altered by the presence of imidazole in the purified solution. However, there are still experimental problems that need to be addressed. While loading the supernatant onto the Ni$^{2+}$-NTA column, a large portion of WfeD seemed to pass through with the mobile phase (Figure 14A). Possibly the His-tag on the protein was not properly exposed for Ni$^{2+}$ column binding due to the protein folding. It may be feasible to first denature the freshly expressed WfeD and purify it by Ni$^{2+}$-NTA before refolding into the active conformation$^{53-55}$. By doing so, the His$_6$ tag can
be exposed more for immobilization to the column. However, the nature of the protein determines whether it can be refolded to the native conformation.

Replacing the His$_6$ tag with other tags may also help to increase the solubility of WfeD. Protein tags such as GST (Glutathione S-transferase) and MBP (maltose binding protein) have been widely used as affinity tags for purifying recombinant proteins. It was frequently observed that these fusion proteins over-express better and exhibit enhanced solubility and sample stability compared to their untagged counterparts.$^{56}$ However, the relative size of protein tag to the target protein has to be considered, which could affect the synthesis and folding of a protein. There is a need for large amounts of highly pure protein for structural analysis using the X-ray crystallography method, meaning the least possible amount of non-specific proteins should remain in the purified protein solution while the WfeD enzyme activity and protein solubility were maintained. Therefore, more efficient expression and purification methods should be developed for this enzyme protein.

4.1.3 Enzyme kinetics and stability

The kinetic properties of WfeD β1,4-Gal-transferase resemble those of other bacterial and mammalian GTs that have similar activities but very different sequences.$^{32,33,37,40}$ The non-purified protein in the homogenate containing 50 mM sucrose appeared to be more stable at low temperatures than the purified protein in dialysis buffer (Figure 18). The main reason for such a distinct stability between the purified and the non-purified forms may be that sucrose and the cell lysate act as a protective and stabilizing agent for the enzyme. In contrast, the purified WfeD may be more susceptible to oxidation in the
dialysis buffer containing only Tris and sodium chloride. The activation of stored enzyme with disulfide reducing agents DTT and β-mercaptoethanol supports the idea that the enzyme is active in the reduced form and less active upon oxidation (possibly by disulfide bonding). The role of Cys-SH could involve substrate binding, catalysis or protein conformation.

4.1.4 GlcNAcα-PP-PhU, acts as a good natural substrate analog in vitro

The synthetic acceptor substrate GlcNAcα-PP-PhU has amphipathic properties similar to those of the endogenous substrate GlcNAcα-PP-Und that acts as a lipid carrier in O-antigen biosynthesis\(^1,^2\). We have shown that GlcNAcα-PP-PhU is an excellent in vitro acceptor substrate for WfeD GalT. The advantage of GlcNAcα-PP-PhU is that the substrate and the enzymatic reaction product can be readily detected and separated by HPLC (C18, reverse phase). The compound has been shown to be an excellent substrate for β1,3-GalT WbbD from \textit{E. coli} strain VW187, β1,3-GlcT WfgD from \textit{E. coli} serotype O152, and β1,3-GlcT WfaP from \textit{E. coli} O56\(^3,^33\). There are more than fifty different serotypes of bacteria (including \textit{E. coli}, \textit{Shigella}, \textit{Salmonella}, \textit{Yersinia} and \textit{Vibrio enterica}) having GlcNAc as their first sugar of the O-chain, which indicates that our assays may be used to characterize many more GTs using GlcNAcα-PP-PhU as the acceptor substrate\(^3,^38\).

4.1.5 Insights into inhibitor development

The compounds that were tested as potential inhibitors of WfeD did not contain the phosphate group, which was found to be essential in the substrate. Thus, none of these compounds bound to and inhibited WfeD. We therefore propose to design a GlcNAc-PP-
derived compound as an inhibitor that resembles the acceptor substrate and may bind to the acceptor binding site of WfeD. Bovine milk β4GalT has several properties resembling those of WfeD. It binds to UDP-Gal, is activated by Mn$^{2+}$, binds to GlcNAc-acceptors and as an inverting transferase catalyzes the transfer of Gal in β1-4 linkage to GlcNAc. An inhibitor of the bovine milk GalT is 1-Thio-N-butylglucosamineβ-(2-naphthyl) which is an acceptor substrate analog but this inhibitor is inactive with WfeD. This suggests that the acceptor binding sites are distinctly different between the two enzymes, although the binding site of the inhibitor in bovine milk GalT has not yet been determined.

4.1.6 Activity stimulation by divalent metal ions

The divalent metal ions in the reaction mechanism of WfeD may interact with the pyrophosphate group of UDP-Gal, and facilitate the nucleophilic attack of GlcNAc created by the Glu101 residue. There seemed to be no absolute requirement of the cation size (or the ionic radius). It is very surprising that other metal ions such as Pb$^{2+}$ can also stimulate the WfeD activity, since Pb$^{2+}$ is highly toxic, and is used as an enzyme inhibitor in organisms. The Pb$^{2+}$ effect was not an artefact; it was Pb$^{2+}$ concentration-dependent, and the enzyme product was demonstrated by HPLC. In contrast, the bovine milk β4GalT which also utilizes divalent metal ion in its catalysis, was much less stimulated by Pb$^{2+}$. Pb$^{2+}$ was found to activate a brain protein kinase C (PKC) at very low concentrations. For this diacylglycerol-activated calcium and phospholipid dependent protein kinase, lead acetate, lead chloride and lead citrate at picomolar concentrations stimulated its activity over 10-fold, compared to calcium chloride. It is possible that the
UDP-Gal binding site of WfeD can accommodate the large (in ionic radius) Pb$^{2+}$ ion as a replacement of Mn$^{2+}$. This novel structural feature and the interactions of the binding site could be explained by X-ray analysis of the WfeD protein.

4.1.7 Involvement of the DxD motif in WfeD activity

The DxD motif, which is found in many GTs, has been shown to be involved in divalent metal ion-stimulated enzyme catalysis$^{27,36}$. Bovine β4GalT, E.coli WbbD β3GalT, N. meningitidis LgtC β4GalT, E.coli O56 WfaP β3 GlcT and E.coli O152 WfgD β3 GlcT all have a DxD motif and are stimulated by Mn$^{2+}$ (Table 1)$^{27,32,33}$. The predicted folding of these enzymes indicates that they are in the GT-A superfamily, although their sequence identity and similarity are very low. The DxD motif has been delineated to interact with the sugar substrate in the Mn$^{2+}$ stimulated reaction of LgtC β1,4GalT from N. meningitides$^{27}$. Among those enzymes identified to contain a DxD motif, their overall sequences can vary greatly, however, the neighbouring sequences are usually similar in overall properties (DxD are flanked by apolar amino acids on both sides) and only one of two Asp/Glu residues is essential for enzyme activity$^{27,36}$. The WfeD protein has a DYEIE sequence (Figure 9A), flanked by several apolar residues on both sides. We showed that only the third residue (Glu101) within the DXD resembling sequence is essential for activity. A similar amino acid sequence arrangement was also found in rabbit GlcNAc transferase I which allowed for the accommodation of the cofactor ion with both substrates complexed in the catalytic site$^{36}$. 
4.1.8 Involvement of RKR in WfeD activity

The mutations of the RKR motif were made based on the assumption that it may interact with the pyrophosphate group of the acceptor substrate. The activity loss of Lys211 (with Mn$^{2+}$ catalysis) suggests that it might be involved in catalysis or be part of a functional structure of WfeD. The sequence comparison of WfeD and other putative GTs showed that the RKR sequence is conserved (Figure 28). Many of the enzymes involved in O-antigen synthesis bind to pyrophosphate substrates and have clusters of positively charged amino acids Lys and Arg in their sequences. However, our study is the first one performing mutagenesis on the positively charged residue cluster, and further investigations are necessary to determine the exact functions of these amino acids. An activity rescue study could be conducted on the K211A mutant by replacing Lys211 with Arg or His, to determine if the activity can be restored in the presence of another positively charged residue. Since Pb$^{2+}$ was able to activate the K211A mutant, it appears that this mutant is properly folded and capable of catalysis. X-ray structure analysis may reveal whether Pb$^{2+}$ has replaced the role of K211A in supporting the enzyme activity.

4.1.9 Role of Cys residue(s) in enzyme activity or stability

The rescue of WfeD activity with DTT and β-mercaptoethanol suggested that the oxidation of Cys residues or disulfide bond formation was responsible for the loss of activity over long time storage. Since the Cys reducing reagents restored the activity, and reagents that covalently modify Cys-SH inhibited the activity, we expect that at least one of the two Cys-SH-groups is essential for activity by either participating in the catalysis, by binding the substrate or being involved in the folding of the protein. There is only one
highly conserved Cys residue (Cys148 in WfeD) (Figure 28). The roles of Cys residues could be further demonstrated by mutagenesis, X-ray structure determination or by locating the oxidized Cys or the disulfide bond using MS. Oxidation of Cys residues and formation of disulfide bonds have been shown in previous studies to cause the loss of enzyme activity among other GTs\textsuperscript{61,62}. There might be a distortion of structure resulting from an intramolecular or intermolecular disulfide bond which could change the conformation or accessibility of the active site.

4.2 Future directions

4.2.1 Further examination of Glu101 and Lys211 involvement in WfeD activity

Although the E101A and K211A were shown to have greatly impaired activity as compared to the wild type, further mutagenesis could be conducted to determine if the activities can be rescued. For example, substituting the Ala residues with the original amino acid residues could be performed using overlap extension method same as applied for creating the mutant plasmids of E101A and K211A. If the activities of the two mutants can be restored, it would confirm the involvement of Glu101 and K211 in enzyme activity. Substituting the Glu and Lys residues with similar amino acids, such as Asp and Arg respectively, could be another way of verifying the role of those two amino acids. Since Asp exhibits the same charge as Glu, supposedly it may carry the same function as Glu and the E101D mutant may show an equally high enzyme activity, which indicates the importance of the negative charge on Glu for the activity. Likewise, substituting Lys with Arg could indicate whether the positive charge on Lys is related to the enzyme activity.
4.2.2 Further improvements of protein solubility and purity

Since a large portion of WfeD protein still remained in the membrane aggregate pellet, increasing protein solubility would be critical for retrieving more WfeD protein from the pellet. Although the bacterial LPS and peptidoglycan synthesizing GTs are thought to be membrane-associated proteins (such as MurG), some of them can be solubilised at high concentrations in an active form, in the presence of lipid vesicles derived from vesicular intracellular membranes\textsuperscript{57,58}. Therefore, the inclusion of lipid molecules may help solubilising the protein. The protein tags also play a role of altering protein solubility, and thus have been applied for enhancing protein solubility as well as stability. More individual specific protein tags, called solubility-enhancement tags have been shown to be highly effective in overcoming solubility and stability problems\textsuperscript{56}.

To improve the Ni\textsuperscript{2+}-NTA purification efficiency of WfeD, a possible way is to denature the protein in the cell pellet before purification, and refold the purified protein afterwards, which may increase the Ni\textsuperscript{2+}-column binding affinity of the protein\textsuperscript{53,54}. However the success of the refolding depends on the intricate structure properties of the protein. In a purification study of recombinant Cancer/Testis antigen BJ-HCC-2 (a hydrophobic protein with two transmembrane regions) expressed in \textit{E. coli} M15 cells, refolding the protein (following a Ni\textsuperscript{2+}-NTA chromatography of the denatured protein in the pellet) in a simple dialysis buffer was very successful\textsuperscript{53}. The denaturing agent (N-lauroyl sarcosine) was superior to Guanidine-HCl, NP-40 and glycerol, and was able to interact with hydrophobic residues to reduce hydrophobicity without impairing the Ni\textsuperscript{2+}-NTA purification and protein refolding. This method has been proposed to be fairly effective in purifying hydrophobic proteins that are difficult to purify by other methods\textsuperscript{55}. 

- 93 -
Using the highly purified protein, crystallization of the WfeD can be attempted to solve the 3D structure of the protein in its apo form as well as in complex with its substrates and cofactors. This could provide more information about the catalytic mechanism and the functional involvement of specific amino acids and divalent cations.

4.2.3 Further investigations on the role of the two Cys residues in WfeD activity

The effects of Cys reducing reagents (DTT and β-ME) and amino acid specific reagents IAA and DTNB indicated a role of Cys residues in enzyme activity. Therefore, Cys148 or Cys174 should be mutated to Ala, combined with GalT assays to confirm which Cys is involved in activity. Another question is whether there is an intra- or inter-molecular disulfide bond being formed that affects the enzyme stability or activity. This could be determined by mutagenesis combined with non-reducing protein gel electrophoresis or ultracentrifugation. The analytical ultracentrifugation can separate subunits of a complex mixture and leave the disulfide bond intact, which could be used to separate the monomer from the dimer (if it exists) based on their different molecular size. On the other hand, MS analysis of trypsin-digested disulfide bonded protein might also reveal which Cys residues are linked to each other, and could be applied as an alternative method to mutagenesis. Among the MS-based strategies, MALDI-TOF MS has been successfully used for the assignment of disulfide crosslinkages. Isotope-labelled free Cys reactive reagent could be used in the MS analysis following tryptic digestion. The oxidized Cys residue would be shielded and only the free Cys would be alkylated with the isotope labelling. The position of the isotope-labelled Cys will be revealed by MS, demonstrating which Cys is or is not in the oxidized state.
4.2.4 *WfeD* cellular localization determination

The intra-cellular localization of WfeD can be determined by cell fractionation or fluorescent labelling. For cell fractionation, the His\(_6\) tag fused WfeD protein should be expressed in *E. coli* cells. The cell fractionation technique involves homogenizing cells in hypotonic buffer (which causes the cells to swell so that they can be efficiently broken open for homogenization) and centrifuging the cell homogenate at low speed of approximately 200g (which separates the cytoplasmic fraction and cell organelle fraction into supernatant and pellet, respectively\(^6\)). The use of anti-His-tag antibody and western blotting will help to detect the amount of WfeD protein in the two components. Fluorescent labelling of WfeD will demonstrate a more precise cellular localization of the protein by microscopy examination\(^7\). The green fluorescent protein (GFP) needs to be co-expressed with WfeD in *E. coli*. The GFP acts as the reporter protein that can be visualized as bright green fluorescence with the exposure to blue light. There, the localization of WfeD in certain cellular compartment can be determined using fluorescence microscopy.

4.2.5 Significance for intestinal disease

The characterization of WfeD enzyme is significant not only since a new enzyme has been discovered with a rare folding (WecG/TagA) and a novel feature of stimulation by Pb\(^{2+}\), but also in establishing the materials and technology to characterize other novel GTs. The technology acquired in this study could be used to characterize other GT of *Shigella* B14 but also of other bacteria that have similar O-antigen structures. The functions of the remaining genes in the O-antigen cluster of *Shigella* B14 can now be unravelled.
The study on the GTs involved in O-antigen synthesis of *Shigella* is a basis for the development of O-antigen vaccine by enzymatic synthesis of O polysaccharide-carrier protein conjugates. It also gives an insight into methods to control O-antigen synthesis to alter the pathogenicity of specific bacteria, which could lead to the development of a new generation of adjuvant antibiotics.
REFERENCES


16. Hale TL, Keusch GT. 1996. ‘Shigella’: structure, classification, and antigenic types. Medical Microbiology, 4e. eds Baron, The University of Texas Medical Branch at Galveston: Ch22.


