

**Accessing Pseudaminic Acid (Pse5Ac7Ac)  
containing Glycosides through the  
Characterisation of Pse5Ac7Ac  
Processing Enzymes**

**Emily Kate Pedley Flack**

**PhD**

**University of York**

**Chemistry**

**August 2019**

## Abstract

Cell-surface carbohydrate pseudaminic acid (Pse5Ac7Ac) is known to contribute to the virulence of several multi-drug resistant bacterial pathogens.<sup>1</sup> Pse5Ac7Ac and its derivatives are not commercially available in appreciable quantities and chemical synthesis of these molecules has proved to be challenging.<sup>2-9</sup> Access to Pse5Ac7Ac and activated CMP-Pse5Ac7Ac has been a hindrance in studies into the biological significance of Pse5Ac7Ac, including Pse5Ac7Ac-processing enzymes, which may be novel therapeutic targets.<sup>1</sup> This project aimed to characterise enzymes which process pseudaminic acid and to chemoenzymatically synthesise glycosides which contain pseudaminic acid.

Firstly, nucleotide-activated pseudaminic acid (CMP-Pse5Ac7Ac) was produced via a chemoenzymatic synthesis route. Six recombinant biosynthetic enzymes which are encoded in *Campylobacter jejuni* and *Aeromonas caviae* were purified for use in this reaction. Purification and characterisation of the resultant CMP-Pse5Ac7Ac confirmed the role of *Aeromonas caviae* PseF as an  $\alpha$ -CMP-Pse5Ac7Ac synthetase.

With CMP-Pse5Ac7Ac in-hand, a library of bacterial sialyltransferases were assayed for activity with CMP-Pse5Ac7Ac as donor. Success from this initial screen led to the synthesis of glycosides containing  $\beta$ -linked Pse5Ac7Ac, mediated by promiscuous sialyltransferases.

Efforts were made to recombinantly produce five putative glycosyltransferases which were predicted to use CMP-Pse5Ac7Ac or a derivative as their natural donor (pseudaminyltransferases) however, all proteins were initially insoluble. *Acinetobacter baumannii* retaining pseudaminyltransferase was solubilised through the construction of an Im9-fusion protein. Activity studies monitored by Liquid Chromatography – Mass Spectrometry confirmed that Im9-KpsS1 could utilise CMP-Pse5Ac7Ac as a donor. Finally, Im9-KpsS1 was used in a seven enzyme one-pot chemoenzymatic synthesis to produce  $\alpha$ -2,6-Pse5Ac7Ac-*p*NP- $\beta$ -D-Glc, confirming that KpsS1 functions as a retaining pseudaminyltransferase. To our knowledge the work presented herein details the first examples of chemoenzymatic synthesis of glycosides containing Pse5Ac7Ac and the first *in vivo* study of a pseudaminyltransferases to provide unequivocal functional characterisation of this novel class of enzyme.

# Contents

|  |           |
|--|-----------|
| 1. <b>Abstract</b> .....   | <b>2</b>  |
| 2. <b>Contents</b> .....   | <b>3</b>  |
| 3. <b>Acknowledgements</b> .....   | <b>7</b>  |
| 4. <b>Author's Declaration</b> .....   | <b>9</b>  |
| 5. <b>Chapter 1: Introduction</b> .....  | <b>10</b> |
| 1.1 Carbohydrates in Biology .....   | 10        |
| 1.1.1 Carbohydrate Structures.....   | 10        |
| 1.1.1 Biological Significance of Glycosylation and Glycans.....                            | 12        |
| 1.2 Glycosyltransferases .....   | 16        |
| 1.2.1 Glycosyltransferase classification .....   | 16        |
| 1.2.2 Glycosyltransferase folds .....  | 17        |
| 1.2.3 Glycosyltransferase function .....   | 17        |
| 1.2.3 Mechanisms of glycosyltransferases.....  | 18        |
| 1.3 Sialic Acids.....  | 23        |
| 1.3.1 Sialic Acid Structures and Function in Eukaryotes.....                               | 23        |
| 1.3.2 Sialic acid biosynthesis and scavenging in bacteria .....                            | 24        |
| 1.3.3 Sialic acid in bacterial glycoconjugates .....                                       | 25        |
| 1.4 Pseudaminic Acids .....  | 26        |
| 1.4.1 Occurrence and Biological Significance of Pseudaminic Acids .....                    | 26        |
| 1.4.2 Biosynthesis of Nucleotide-Activated Pseudaminic Acids .....                         | 28        |
| 1.4.3 Chemical synthesis towards Pseudaminic Acids and Pseudaminic Acid<br>glycosides..... | 33        |
| 1.4.4 Chemoenzymatic synthesis of Pseudaminic Acids.....                                   | 44        |
| 1.4.4 Introduction to Pseudaminic Acid Processing Enzymes .....                            | 47        |
| 1.5 Project Outline.....   | 49        |
| 6. <b>Chapter 2: Chemoenzymatic Synthesis of CMP-Pseudaminic Acid</b> .....                | <b>50</b> |

|   |    |
|---|----|
| 2.1 Introduction .....  | 50 |
| 2.2 Experimental .....  | 54 |
| 2.2.1 Overexpression and Purification of PseB, PseC, PseH, PseG and Psel .....                              | 54 |
| 2.2.2 Enzymatic synthesis of Pse5Ac7Ac with Acetyl-thiocholine Iodide .....                                 | 56 |
| 2.2.3 Small-molecule screen for potential inhibition of PseB and Psel .....                                 | 56 |
| 2.2.4 Recombinant Expression and purification attempt of <i>C. jejuni</i> PseF in <i>E. coli</i> ..         | 58 |
| 2.2.5 Expression and purification of <i>A. caviae</i> PseF .....  | 60 |
| 2.2.6 Activity assay of <i>A. caviae</i> PseF with Pse5Ac7Ac .....  | 61 |
| 2.2.7 Purification of <i>Aeromonas caviae</i> PseF for Crystallisation trials .....                         | 61 |
| 2.2.8 <i>A. caviae</i> PseF Protein Identification by Mass Spectrometry .....                               | 62 |
| 2.2.9 <i>A. caviae</i> PseF Circular Dichorism .....  | 62 |
| 2.2.10 <i>A. caviae</i> PseF Size Exclusion Chromatography - Multi-Angle Laser Light<br>Scattering .....    | 63 |
| 2.2.11 Crystallisation trials for <i>A. caviae</i> PseF .....   | 63 |
| 2.2.12 Large scale synthesis and purification of CMP-Pse5Ac7Ac, via a one-pot six<br>enzyme synthesis ..... | 64 |
| 2.3 Results .....   | 66 |
| 2.3.1 Overexpression and Purification of PseB, PseC, PseH, PseG and Psel .....                              | 66 |
| 2.2.3 Enzymatic synthesis of Pse5Ac7Ac with Acetyl-thiocholine Iodide .....                                 | 67 |
| 2.3.3 Small-molecule screen for potential inhibition of PseB and Psel .....                                 | 68 |
| 2.3.4 Recombinant Expression and purification attempt of <i>C. jejuni</i> PseF in <i>E. coli</i> ..         | 72 |
| 2.3.5 Expression and purification of <i>A. caviae</i> PseF .....  | 73 |
| 2.3.6 Activity assay of <i>A. caviae</i> PseF with Pse5Ac7Ac .....  | 74 |
| 2.3.7 Purification of <i>A. caviae</i> PseF for Crystallisation trials .....                                | 76 |
| 2.3.8 <i>A. caviae</i> PseF Protein Identification by Mass Spectrometry .....                               | 77 |
| 2.3.9 <i>A. caviae</i> PseF Circular Dichorism .....  | 78 |
| 2.3.10 <i>A. caviae</i> PseF Size Exclusion Chromatography - Multi-Angle Laser Light<br>Scattering .....    | 79 |
| 2.3.11 Crystallisation trials for <i>Aeromonas caviae</i> PseF .....  | 79 |

|   |            |
|---|------------|
| 2.3.12 Large scale synthesis and purification of CMP-Pse5Ac7Ac, via a one-pot six enzyme synthesis .....                              | 80         |
| 2.4 Discussion.....   | 83         |
| 2.5 Conclusions and Future Work .....   | 90         |
| <b>7. Chapter 3: Sialyltransferase mediated synthesis of glycosides containing Pseudaminic Acid .....</b>                             | <b>91</b>  |
| 3.1 Introduction .....  | 91         |
| 3.2 Experimental .....  | 94         |
| 3.2.1 Initial Activity assays of sialyltransferase library with CMP-Pse5Ac7Ac donor..   | 94         |
| 3.2.2 PmST acceptor screen with CMP-Pse5Ac7Ac as donor .....  | 96         |
| 3.2.3 Large Scale synthesis of $\beta$ -Pse5Ac7Ac-2,3- <i>p</i> NP- $\beta$ -D-Galp using PmST .....                                  | 96         |
| 3.3 Results .....   | 98         |
| 3.3.1 Initial Activity assays of sialyltransferase library with CMP-Pse5Ac7Ac donor..   | 98         |
| 3.3.2 PmST acceptor screen with CMP-Pse5Ac7Ac as donor .....  | 100        |
| 3.3.3 Large Scale synthesis of $\beta$ -Pse5Ac7Ac-2,3- <i>p</i> NP- $\beta$ -D-Galp using <i>Pasteurella multocida</i> tPm0188Ph..... | 102        |
| 3.4 Discussion.....   | 104        |
| 3.5 Conclusions and Future Work .....   | 107        |
| <b>8. Chapter 4: Characterisation of Putative Pseudaminyltransferases .....</b>   | <b>108</b> |
| 4.1 Introduction .....  | 108        |
| 4.2 Experimental .....  | 121        |
| 4.3.1 Sequence analysis of Putative Pseudaminyltransferases.....  | 121        |
| 4.3.2 Homology Modeling of Putative Inverting Pseudaminyltransferases .....   | 121        |
| 4.2.3 Putative Pseudaminyltransferase Expression Trials.....  | 121        |
| 4.2.4 Solubilisation trials of Putative Pseudaminyltransferases Lst .....   | 123        |
| 4.2.5 Detergent Screen for Solubilisation of Putative Pseudaminyltransferases Lst and KpsS1 .....                                     | 124        |
| 4.2.6 Cloning of KpsS1 Fusion Protein Plasmids .....  | 125        |
| 4.2.7 KpsS1 Fusion Protein Expression Trials .....  | 127        |
| 4.2.8 Im9-KpsS1 Large Scale Expression.....   | 127        |

|   |            |
|---|------------|
| 4.2.9 Im9-KpsS1 Purification .....  | 127        |
| 4.2.10 Im9-KpsS1 Protein Identification Mass Spectrometry.....  | 128        |
| 4.2.11 Im9-KpsS1 Activity Screen.....   | 128        |
| 4.2.12 Chemoenzymatic synthesis of glycoside containing Pse5Ac7Ac, <i>via</i> a one-pot,<br>seven-enzyme reaction using Im9-KpsS1 ..... | 129        |
| 4.3 Results .....   | 131        |
| 4.3.1 Sequence analysis of Putative Pseudaminyltransferases .....   | 131        |
| 4.3.2 Further Sequence analysis and Homology Modeling of Putative Inverting<br>Pseudaminyltransferases.....                             | 131        |
| 4.3.3 Initial Expression trials of Putative Pseudaminyltransferases .....   | 134        |
| 4.3.4 Expression trials of Putative Pseudaminyltransferases, Lst and KpsS1 in <i>E. coli</i><br>Tuner (DE3).....                        | 139        |
| 4.3.5 Solubilisation trials of Putative Pseudaminyltransferase Lst .....  | 141        |
| 4.3.6 Detergent Screen for Solubilisation of Putative Pseudaminyltransferases Lst<br>and KpsS1 .....                                    | 142        |
| 4.3.7 KpsS1 Fusion Plasmids .....   | 147        |
| 4.3.8 KpsS1 Fusion Proteins Expression Trials .....   | 147        |
| 4.3.9 Im9-KpsS1 Purification .....  | 150        |
| 4.3.10 Im9-KpsS1 Protein Identification Mass Spectrometry.....  | 150        |
| 4.3.11 Im9-KpsS1 Activity Screen.....   | 151        |
| 4.3.12 Chemoenzymatic synthesis of glycoside containing Pse5Ac7Ac, <i>via</i> a one-pot,<br>seven-enzyme reaction using Im9-KpsS1 ..... | 157        |
| 4.4 Discussion.....   | 161        |
| 4.5 Conclusions and Future Work .....   | 165        |
| <b>9. Chapter 5: Concluding Remarks and Future Prospectives.....</b>  | <b>167</b> |
| General methods .....   | 206        |
| <b>10. Abbreviations .....</b>  | <b>209</b> |
| <b>11. References .....</b>   | <b>214</b> |

## Acknowledgements

First and foremost, I would like to thank my supervisor's Dr Martin Fascione and Professor Gavin Thomas for giving me the opportunity to start this PhD and for your on-going patience and encouragement that enabled me to complete it. Martin, thank you for your efforts in teaching me chemistry and for somehow making it less daunting than it used to be. Gavin, thank you for reminding me how fun biology is. I am grateful to you both for allowing me to experience the best parts of both worlds throughout this PhD. Additional thanks are owed to my TAP member Professor Jen Potts for advice on this research during meetings.

Thank you to the technical staff in the departments of Biology and Chemistry for their assistance in this project. I would specifically like to thank Dr Ed Bergström, for always ensuring that a working LC-MS was available to the Fascione group.

I would like to thank Dr Kun Huang for your help with sialyltransferase work at MIB. Thanks to Tasha for sugar synthesis and patiently helping me with chemistry. Reyme, thank you for introducing me to crystallography. Thank you to the wonderful Darshita, Tess and Clare for answering my stupid questions and for always being willing to help. "Lovely Darshita" thank you for sneakily making that disaccharide. Also, thank you for your tough-love approach. It was just what I needed. Julia, thank you for sharing your scientific wisdom and for keeping the place running. I'm sure the fridges, freezers and gel room will be a lot emptier once I've finished. (Sorry about that.) We do not deserve you!

Thank you to the Thomas lab for ensuring there was always somewhere I could go to discuss interesting biology and for allowing me to raid the lab for supplies. Huge thanks are owed to Martin, Alison, Lianne and Julia for their hard work in creating a brilliant lab to work in and finding great team members to work here. To all of the Parkin-Fascione-Willems lab members (B Block babes), thanks to each of you for making this such a weird, yet wonderful place to work! Please do not change. Thanks for the tea-breaks (usually instigated by Mark), numerous pub lunches (Nick, you are to blame for most of these) and many other bizarre distractions, that helped make days at work so great.

I would like to thank the other three quarters of the Fascione Four: Richard, Harriet and Robin, for five years of friendship, understanding and advice. Rich, thanks for cheering me up with chats of home. Harriet, thank you for being the better half of “Hamily”/Team Pse. Robin, you have been the greatest PhD buddy I could have wished to be stuck with! (I guess we can finally stop asking one another “can we quit yet?”). This would have been significantly harder and much less fun without you guys.

To my family and closest friends, thank you for your continued love and support. Mum and Eddy- thank you for subtly asking Matt for an update, whenever I was being evasive. Also, thanks for knowing that the only proper way to get me back to Baildon from York, is to drive us over the moors! I’d like to acknowledge my Granny and Grandad for their influence in my decision to spend seven years studying science in York. If only we’d known on our many day trips here, that I’d end up calling it home. Finally, Matt, thank you for being my very best friend and for always being on my team.

## Author's Declaration

I declare that I am the sole author of this thesis. The work presented here was neither published before nor used previously to obtain a degree at this or another university. The work presented was carried out by me with the exception of:

### Chapter 2:

- Compounds screened as PseB and PseI inhibitors were synthesised by Joe Ferner, University of Sheffield.
- *A. caviae* PseB and *A. caviae* PseI was purified by Joe Ferner, University of Sheffield.
- SNAc was synthesised by Dr Harriet Chidwick
- SEC-MALLS data was obtained and analysed by Dr Andrew Leech for the University of York Technology Facility.
- PseF crystals were fished by Reyme Herman and diffraction data was collected by Sam Hart.

### Chapter 3:

- Initial sialyltransferase activity screen with LacNAc-ITag was performed with Dr Kun Huang (KH). Mass spectrometry data for these samples was collected by KH at University of Manchester.

### Chapter 4:

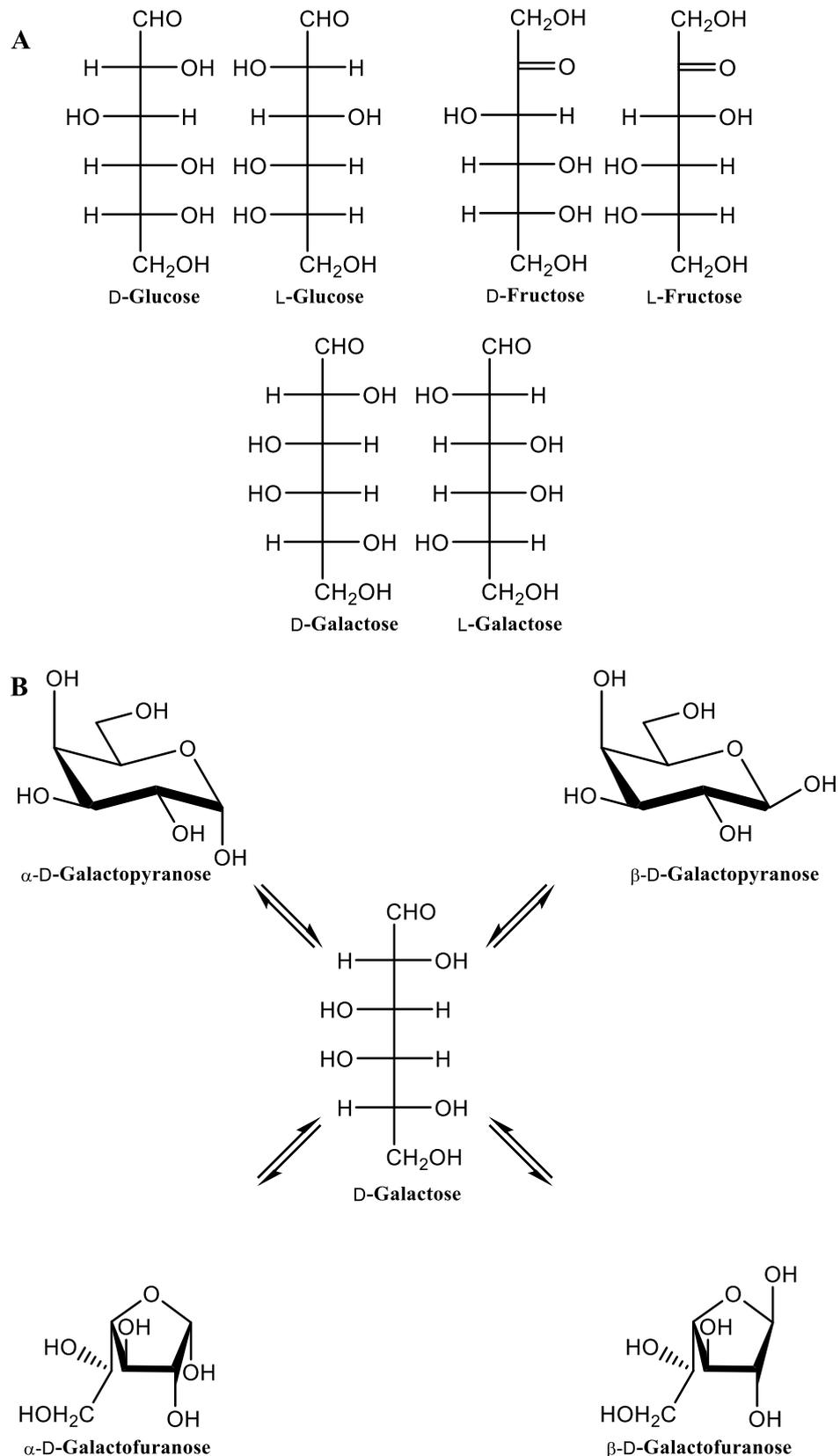
- Synthesis and characterisation of  $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe was performed by Dr Darshita Budhadev and Natasha Hatton.
- Trypsin digest and MS analysis of Im9-KpsS1 was performed by Rachel Bates for the University of York Technology Facility

# Chapter 1: Introduction

## 1.1 Carbohydrates in Biology

### 1.1.1 Carbohydrate Structures

All cell surfaces, without known exception, are covered with carbohydrates.<sup>10</sup> Within biological systems carbohydrates may be covalently linked to one-another to form oligo- or polysaccharides (glycans) or linked to other classes of biomolecules to form glycoconjugates, including glycolipids and glycoproteins, the latter generated by post-translational, for instance modification of serine and threonine residues. The complexity of carbohydrate structures must be considered in order to appreciate the challenges in studying this large class of biomolecules. Beginning with monosaccharides, these units may exist in aldose or ketose form and may be ring open or closed (Figure 1.1). Carbohydrates may be defined as “reducing” or “non-reducing” due to the presence of an aldehyde or ketone moiety respectively in ring-open form. Carbohydrates contain multiple additional stereocentres, therefore each hydroxyl group may epimerise, generating enantiomers and diastereomers for a carbohydrate of any given number of carbons (Figure 1.1 A). The carbon of the carbonyl group in the ring open form carbohydrate, forms a chiral anomeric centre when the molecule is ring closed, giving the carbohydrate either  $\alpha$  or  $\beta$  stereochemistry (Figure 1.1 B). Additionally, structural diversity is seen as some hexoses can interconvert between five and six-membered furanose or pyranose rings (Figure 1.1 B). Further to this, carbohydrates may deviate from their  $C_n(H_2O)_n$  empirical formula, as hydroxyl groups may be substituted for alternative functional groups.



**Figure 1.1: A: Fischer projections of hexose monosaccharides: L- and D-Glucose, an example of enantiomeric aldose carbohydrates; L- and D-Fructose, a pair of ketose enantiomers; L- and D-Galactose, diastereomers of L- and D-Glucose. B: Ring open D-Galactose can form ring closed  $\alpha$ - and  $\beta$ -Galactopyranose and  $\alpha$ - and  $\beta$ -Galactofuranose (six and five membered rings respectively).**

Carbohydrates polymerise to form glycans which are the most structurally diverse major class of molecules. During polymerisation, glycosidic bonds are formed via condensation reactions, usually involving the anomeric carbon. The linkage can be defined as  $\alpha$  or  $\beta$  in stereochemistry and may join the carbohydrate to its neighbouring carbohydrate at one of several positions. Carbohydrates which form oligo- or polysaccharide may be linear or branched structures.<sup>10</sup>

In calculating the number of possible structures for a polysaccharide of any given length, the following factors must be accounted for: epimeric identity of each monomer, D or L sugars, anomeric configuration of each monomer, position of glycosidic linkage, possible branching, ring size and reducing terminal attachment. There are 1056 possible isomers for a tetrasaccharide, increasing the oligosaccharide size to a hexasaccharide gives  $\sim 1.05 \times 10^{12}$  possible structures.<sup>11,12</sup> Whilst many of these potential structures are not known to occur in nature, this vast number of possible isomers adds a level of complexity when trying to determine the structure of a chemically synthesised or biologically extracted oligosaccharide.<sup>12</sup> Technical difficulties in studying glycans and glycoconjugates can in-part, account for why studies of these molecules have lagged behind other biomolecular classes.

### **1.1.1 Biological Significance of Glycosylation and Glycans**

At the beginning of the millennium, carbohydrates were termed the “Cinderella” molecules of biology, as the neglected, over-looked sister of the nucleic acids and proteins.<sup>13</sup> Recently, the ubiquity of glycosylation and glycans has become apparent and it is now accepted that carbohydrate play significant roles beyond that of an energy source. Glycans are as universal as nucleic acids, proteins, and lipids and essential for all living organisms.<sup>14,15</sup> The assembly, modification and interaction of these four biomolecular groups is crucial for cell development and survival.<sup>14</sup> Our understanding of the biological significance of glycoconjugates and glycans has been predominantly gained from studies undertaken in the past three decades.<sup>16,17</sup>

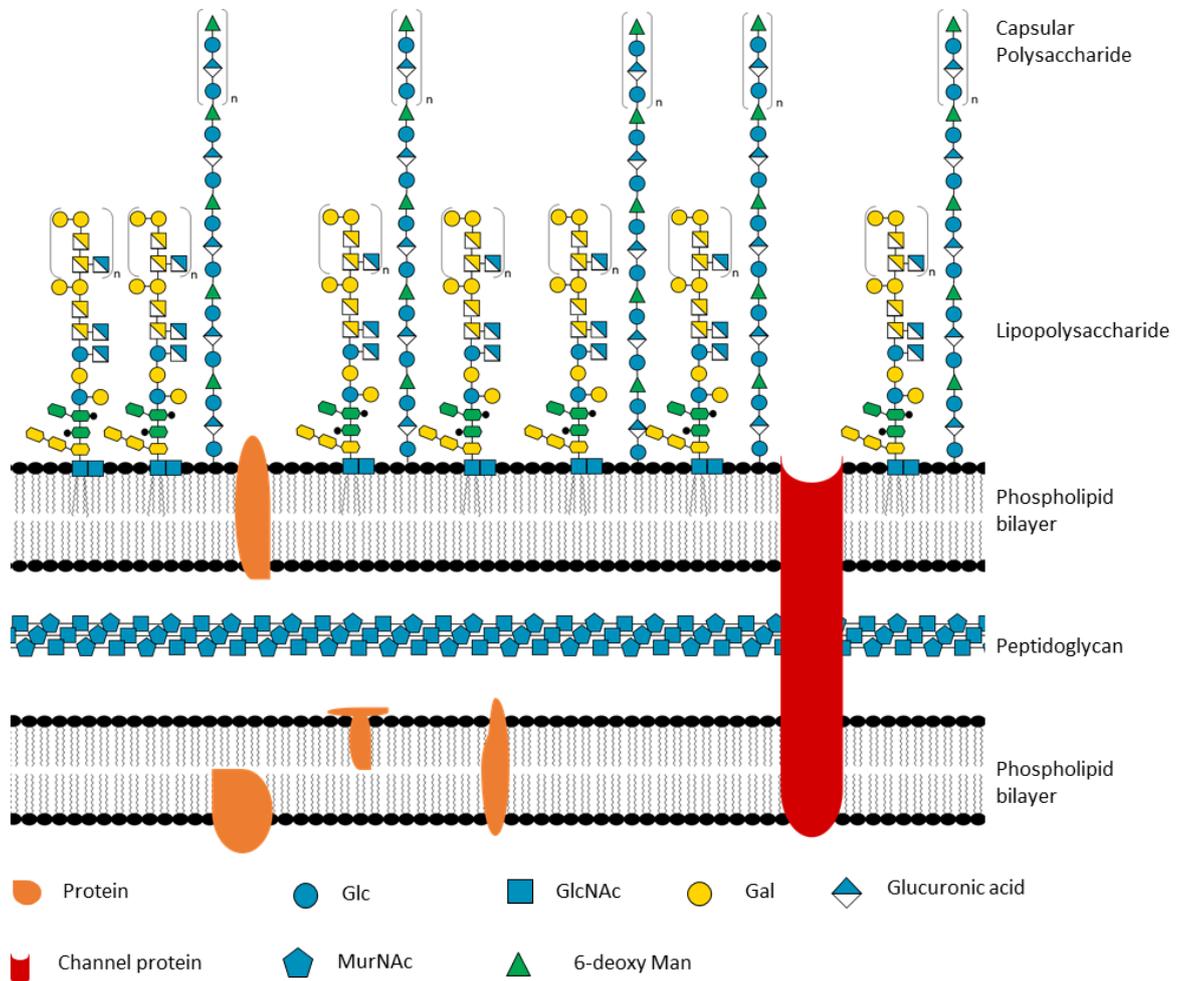
The functions of glycans, or glycosylation, may be classified into three broad groups: 1) structural and modulatory roles, including nutrient storage; 2) recognition by other molecules (as either extrinsic/interspecies recognition or intrinsic/intraspecies recognition); and finally, 3) molecular mimicry of host carbohydrates.<sup>10,17</sup>

In addition to structural complexity, a further hindrance in the study of glycans can be attributed to a lack of template for their biosynthesis, unlike the nucleic acids and proteins, for which well-understood, largely universal set of instructions exist for their biosynthesis.<sup>10</sup> A plethora of rare human diseases, collectively termed Congenital Disorders of Glycosylation (CDG) are known to be caused by glycosylation defects. However, heterogeneity in glycosylation and glycans makes understanding the molecular mechanism of each CDG particularly challenging. Whilst the genetic basis of many CDG's were known, they were often overlooked as glycosylation patterns in laboratory-based tissue culture, do not represent the glycosylation pattern *in vivo* and therefore the physiological manifestation of the defects is altered. This not only highlights the biological significance of eukaryotic glycosylation but indicates that many important functions of glycans are only apparent in whole organisms.<sup>10</sup>

Despite recent advances, glycan structure and function remains unexplored in many forms of life including bacteria, archaea, fungi and algae.<sup>17</sup> Many classes of glycosylation were once believed to be reserved as a eukaryotic post-translational modification. The discovery of the Pgl pathway with *Campylobacter jejuni* was the first example of a general N-linked glycosylation pathway in prokaryotic cells.<sup>18</sup> This was a significant breakthrough in the field of bacterial glycobiology. The types of monosaccharide found in bacteria show greater diversity than eukaryotic monosaccharides and within bacteria carbohydrates often undergo an extensive range of chemical modifications (including alkyl, acyl, aminoacyl, phosphoryl, and nucleosides).<sup>19</sup> Key cell-surface glycoconjugates in Gram-negative bacteria include peptidoglycan, lipopolysaccharide (LPS), lipooligosaccharide (LOS), extracellular polysaccharide (EPS), capsular polysaccharide (CPS) and glycoproteins (some of which are depicted in Figure 1.2). Gram-positive bacteria do not contain LPS or LOS but do have glycosylated lipoteichoic acids and wall teichoic acids (depicted in Figure 1.3).<sup>19</sup>

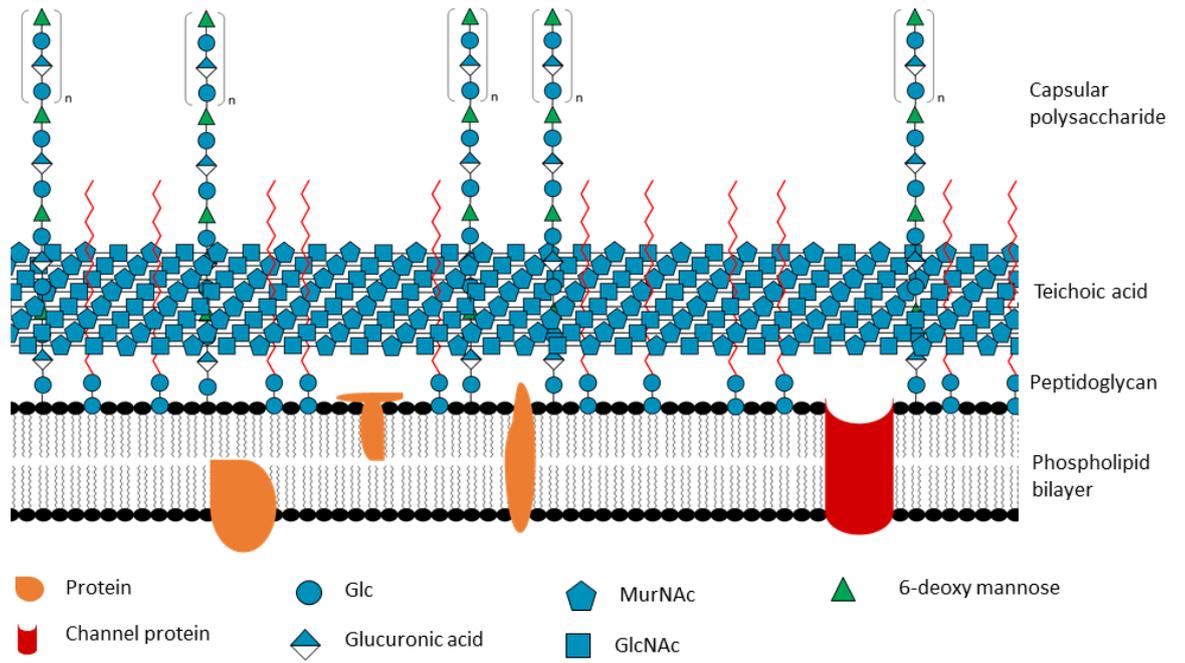
Perhaps one of the most intriguing biological roles of carbohydrates is the aforementioned “molecular mimicry” that microbial pathogens engage in, by covering their surfaces with glycans that are similar to host glycans, with the aim of evading host immune response or increasing host tolerance towards the pathogen.<sup>10</sup> This is observed with the nonulosonic acids (NulOs), a group of nine-carbon carbohydrates, some of which are found as terminal residues in human glycoconjugates and often in bacterial glycoconjugates.<sup>20</sup> Interestingly,

phylogenetic analysis of NulO biosynthesis and N-glycosylation genes revealed that NulOs and N-glycosylation pathways may have originated in bacteria and archaea.<sup>21,22</sup>



**Figure 1.2: Depiction of Gram-Negative cell wall and examples of cell surface glycoconjugates, LPS and CPS.\***

\* Adapted from figure originally produced by Dr Harriet Chidwick



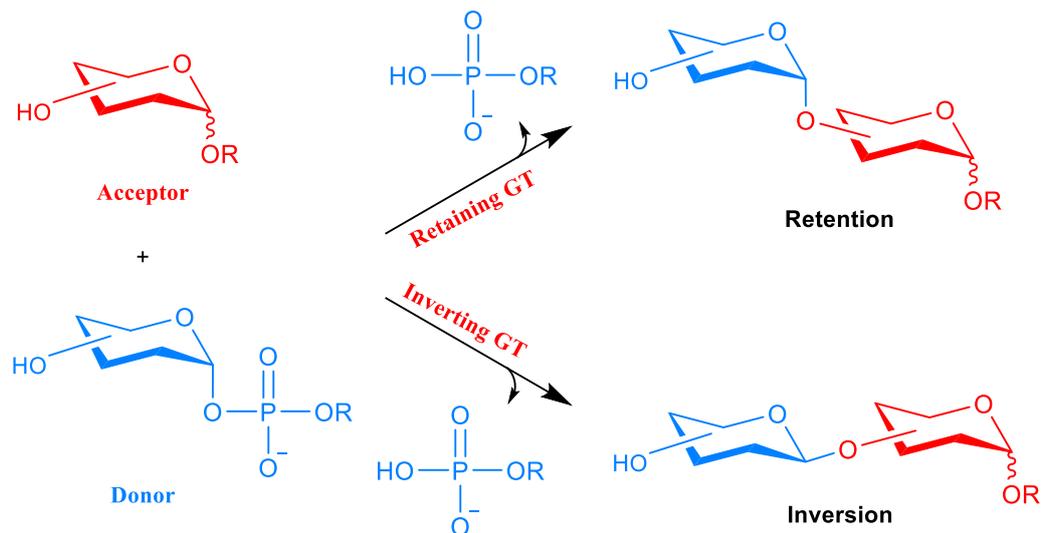
**Figure 1.3: Depiction of Gram-Positive cell wall and examples of cell surface glycoconjugate and CPS.<sup>†</sup>**

<sup>†</sup> Adapted from figure originally produced by Dr Harriet Chidwick

## 1.2 Glycosyltransferases

### 1.2.1 Glycosyltransferase classification

Around two thirds of all biological carbon occurs in the form of carbohydrates.<sup>23</sup> Taking in to consideration the range of potential structures, described above, it can be reasoned that a vast number of enzymes exist to build and degrade carbohydrates.<sup>23</sup> Structural and kinetic studies of these enzymes are important for gaining insights into their function. A molecular level understanding of their function can help to identify enzymes involved in disease, or that could be engineered to synthesize biologically or therapeutically relevant molecules.<sup>24</sup> Glycosyltransferases (GTs) are a class of enzymes which transfer an activated donor, to an acceptor moiety, forming a glycosidic bond with either inverted or retained stereochemistry at the anomeric centre.<sup>23</sup> Around 65% of all donor carbohydrates are activated with a nucleotide.<sup>25</sup> The identity of the acceptor can include other carbohydrates, proteins, and lipids (Figure 1.4 depicts the mechanism of a GT transferring an activated carbohydrate to a general carbohydrate donor). It should be noted that whilst glycosyltransfer usually occurs at the oxygen of a hydroxyl, it can also occur to nitrogen (e.g. N-linked glycoproteins), sulfur (e.g. the formation of thioglycosides), and carbon (e.g. C-glycoside antibiotics).<sup>26</sup>



**Figure 1.4: General mechanism of glycosyltransferases. Stereochemistry of the anomeric centre may be either inverted or retained with respect to that of the activated donor carbohydrate when a GT forms a glycosidic bond between the donor and acceptor moieties.**

In recent years, there has been an exponential growth in the number of gene sequences available that encode GTs.<sup>24</sup> The Carbohydrate-Active EnZymes (CAZy) database is curated to describe the abundance of known GTs based on their sequences, structures and functions. CAZy classifies GTs into sequence-based families, of which there are currently 107 (July 2019).<sup>27</sup> GTs may take one of three overall three-dimensional folds and a GT family generally contain proteins with shared fold.<sup>26</sup> The polyspecificity seen within some GT families means that this classification alone can make it difficult to predict GT function and therefore empirical evidence is valuable in correctly assigning the function of novel GTs.<sup>24,27,28</sup>

### **1.2.2 Glycosyltransferase folds**

Structurally characterised GTs typically display one of three overall folds. All structurally characterised GTs which used nucleotide-activated donor substrates have either GT-A or GT-B folds. The catalytic domain of a GT-A structure contains a sequence of around 120 residues which resembles the Rossmann fold of nucleotide-binding proteins. This region of the GT interacts with the nucleotide of the activated donor sugar. Almost all GT-A fold GTs contain a characteristic DXD motif, which co-ordinates a metal-ion (significance described below in 1.2.2). GTs classed as GT-B fold in architecture, have two distinct domains both of which are similar to the Rossmann type fold and the active-site is found in a cleft between domains. The C-terminal domain interacts with the nucleotide moiety of the donor substrate. These enzymes lack a DXD motif and are metal-ion independent.<sup>10</sup> Unlike the GT-A and GT-B fold GTs, GT-C enzymes use lipid-linked donor substrates. GT-C is a more recently defined GT fold, with only a few structures available. GT-C fold enzymes are integral membrane proteins and it is predicted that GT-C members are related due to similarities in transmembrane region, may not have similar catalytic domain structure.<sup>10,26</sup>

### **1.2.3 Glycosyltransferase function**

An initial challenge in studying GTs, is often their insolubility. The nature of GT function means that they are often membrane bound or membrane associated, therefore production of high-yielding, recombinant, pure protein can be challenging.<sup>29</sup> To combat this, GT activity may be assayed from crude cell lysate, to eliminate the requirement for pure protein.<sup>30</sup> Secondly, their functional characterisation requires identification of both

acceptor and donor molecules, which may not be commercially available and can be challenging molecules to synthesise.<sup>24</sup> Given the instability of nucleotide-activated sugars, their *in-situ* generation is a routine approach when studying GT function.<sup>30–32</sup> Additionally, substrate analogues may be used if native substrates cannot be accessed with ease.<sup>30,33</sup>

Once the GT and potential substrates, or analogues are in hand, several assays may be used to study their function. Initially, thin layer chromatography (TLC) may be used to monitor the depletion of substrates and the formation of products.<sup>28,30</sup> Beyond this more quantitative assays such as spectrophotometric and fluorescence based methods, may be used to measure the depletion of the acceptor or the nucleotide donor or the accumulation of the free nucleotide or the glycosylated product can be monitored. Similarly, radiochemical assays may be used quantitative monitoring of GT activity if radiolabelled substrates can be accessed. These assays can be now used in small volumes, be applied in high throughput and assay membrane bound GTs.<sup>28,34</sup> Immunological approaches may be used, particularly in assays of bacterial GTs, where antibodies or lectins may identify the reaction products.<sup>35</sup>

More sophisticated chromatography techniques such as high-performance liquid chromatography (HPLC) may be used to monitor GT reactions, where quantitative data can be obtained.<sup>36–40</sup> Alternatively with mass spectrometry (MS) is frequently used or chromatography and MS techniques may be paired, therefore two readouts are obtained for each sample.<sup>36,37,41,42</sup> Finally, chemical analysis of products, through techniques such as NMR, is often required to confirm the identity, stereochemistry and linkage of GTs, such that their function may be unequivocally assigned.<sup>21,29,32,33</sup>

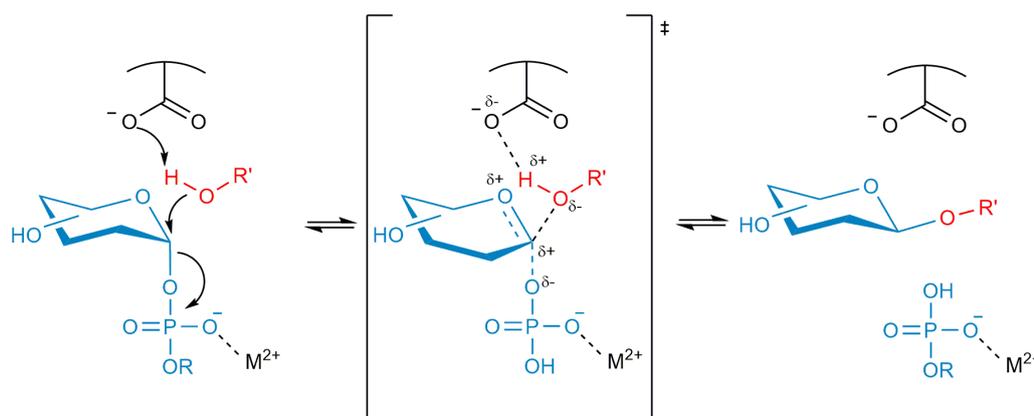
### **1.2.3 Mechanisms of glycosyltransferases**

As depicted above, glycosyltransferases catalyse the formation of glycosidic bonds in which the resulting configuration of the anomeric centre in the product is either inverted or retained in comparison to that of the donor carbohydrate (Figure 1.4).<sup>23</sup> In order to rationally design inhibitors of GTs as therapeutics, the precise mechanism by which these enzymes perform these nucleophilic substitution reactions must be understood.<sup>24</sup> As mentioned above, the insolubility of GTs may preclude the production of high-yielding recombinant protein, which is required for structural studies.<sup>29</sup> Beyond these initial

hurdles, GTs may be difficult to crystallise as they are often multidomain proteins which undergo significant conformational changes in the presence of substrates. Probing the mechanisms of these enzymes not only relies upon being able to obtain structures of these proteins, it also requires site-direct mutagenesis and chemical probes, which again can be challenging to synthesise.<sup>24</sup>

### Mechanisms of inverting glycosyltransferases

Experimental data has revealed that inverting GTs use a single-displacement,  $S_N2$  mechanism. A catalytic base residue within the active-site, deprotonates the nucleophilic acceptor moiety, thus enabling an  $S_N2$ -like displacement of leaving group on the activated donor. This reaction proceeds via an oxocarbenium ion-like transition state (Figure 1.5). A divalent cation, typically  $Mg^{2+}$  or  $Mn^{2+}$ , which is coordinated by the conserved DXD motif, aids departure of the leaving group by stabilising the negative charge of the transition state.<sup>26</sup> However, structural data of some inverting GTs suggests that an alternative mechanism must be employed, due to the lack of a catalytic base within the active site.<sup>24</sup>



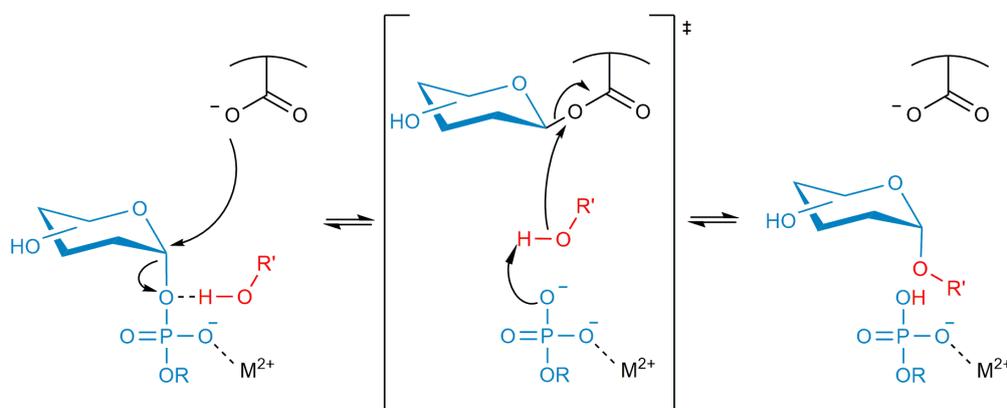
**Figure 1.5:  $S_N2$  mechanism of inverting GTs. Blue molecule: activated donor moiety; Red molecule: acceptor moiety; R: a nucleoside, nucleoside monophosphate, lipid phosphate, or phosphate; R': carbohydrate or protein.**

### Proposed mechanisms of retaining glycosyltransferases

There is more uncertainty surrounding the mechanisms used by retaining GTs. Much of the knowledge on the mechanisms on retaining GTs stems from an understanding of the mechanism used by retaining glycosylhydrolases (GHs). Koshland noted that in biomolecules where stereochemistry was retained it was likely that enzymatic reaction proceeded via a double-displacement reaction, requiring an enzymatic nucleophile.<sup>44</sup> In the first step, a nucleophile active-site residue forms a covalent-enzyme intermediate with the

anomeric carbon of a substrate, with inversion of stereochemistry in the intermediate. A second inverting displacement occurs when a second nucleophile, H<sub>2</sub>O for GHs, performs a back-side attack of the anomeric carbon, resulting in overall retention of the original configuration. This mechanism is widely accepted as it proceeds with lower free-energy than the alternative S<sub>N</sub>1-like mechanism, in which it was proposed that the enzyme shields the opposite face of the substrate, such that it may not be attacked from the shielded face, after a covalent-enzyme intermediate has formed.<sup>26</sup>

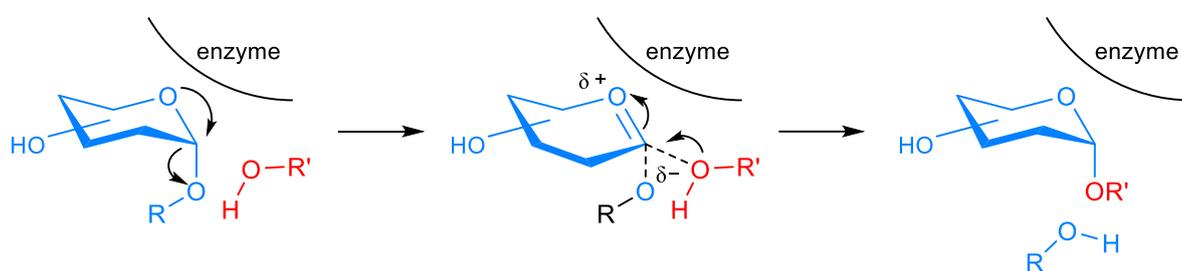
In the absence of experimental data, it was assumed from studies of GHs that retaining GTs use a double-displacement mechanism, forming covalent-enzyme intermediate, initiated by enzymatic nucleophilic attack of the anomeric carbon (Figure 1.6). In comparison to inverting GTs, divalent metal cations (coordinated by aforementioned enzymatic DXD motifs) or helix-dipoles, stabilise negative charge of the leaving group. Additionally, the phosphate moiety of the leaving group is proposed to act as the base to deprotonate the acceptor moiety.<sup>26</sup> In retaining GTs, the deprotonated acceptor acts as the non-catalytic nucleophile, in place of H<sub>2</sub>O, which is used in GHs. However, of the structurally characterised GTs only the GT6 family has a suitably placed residue, which may act as the catalytic nucleophile residue.<sup>45</sup>



**Figure 1.6: Proposed double-displacement mechanism of retaining GT. Blue molecule: activated donor moiety; Red molecule: acceptor moiety; R: a nucleoside, nucleoside monophosphate, lipid phosphate, or phosphate; R': carbohydrate or protein.**

In response to these experimental findings, an alternative mechanism was proposed, termed S<sub>N</sub>i-like (substitution internal nucleophilic or internal-return mechanism). A front-face S<sub>N</sub>i mechanism is seen in solution based glycosyltransfer and has been confirmed to occur during solvolysis of glycosyl fluorides in ethanol and chloroethanol solution, using

kinetic isotope effect measurements.<sup>45</sup> In  $S_{Ni}$ -like reactions facilitated by retaining GTs, the donor carbohydrate acts as the nucleophile as the leaving group decomposes. The leaving group and acceptor are held as an ion pair on the same face of the donor. This intermediate breaks down and nucleophilic attack occurs, so that the resulting product has retained stereochemistry at the anomeric centre (Figure 1.7).<sup>46,47</sup> Structural data obtained for some GTs now supports an  $S_{Ni}$ -like mechanism.<sup>37,48</sup> However, for many retaining GTs,  $S_{Ni}$ -like mechanisms have been proposed due to a lack of experimental evidence to support an alternative mechanism, for example when an enzymatic nucleophile has not been identified.<sup>26</sup>



**Figure 1.7:  $S_{Ni}$  mechanism of inverting GTs. Blue molecule: activated donor moiety; Red molecule: acceptor moiety; R: a nucleoside, nucleoside monophosphate, lipid phosphate, or phosphate; R': carbohydrate or protein.**

When studying retaining GHs, substrate analogues may be used, in which hydroxyls are substituted with fluorine at the C2 or C5 position in pyranose sugars, to capture the intermediate formed during a double displacement mechanism.<sup>26</sup> The increased electronegativity of the fluoro-sugar destabilises the oxocarbenium ion that forms during the double displacement mechanism, resulting in a decreased reaction rate. Secondly, a good leaving group can be used in the substrate analogue, so that the enzyme can perform step one of the hydrolysis mechanism, but the glycosyl-enzyme intermediate will accumulate. This approach allows the long-lived intermediate to be captured via analytical techniques such as X-ray crystallography.<sup>49</sup> However, implementing similar approaches for the examination of retaining GTs, proves challenging due to the enzymes specificity for leaving groups, where phosphates or nucleotides are required. Activated-donor sugars are metabolically expensive to biosynthesise; therefore, it can be anticipated that the cleavage of the activating group from the carbohydrate is rate limiting, preventing the accumulation of an intermediate. Additionally, for GHs substrate analogues may be used along with mutagenesis of the predicted acid/base catalyst within the active site to decrease reaction

rate and capture intermediate. However, for GTs the leaving group is predicted to act as a base, in place of an enzymatic base residue, therefore the suggested mutagenesis strategy cannot be implemented. In the absence of strategies to manipulate the rate of intermediate formation, structural studies must be used to predict the mechanism of these enzymes.<sup>26</sup>

A challenge with inferring mechanism from apo-structure alone is that this does not account for conformational changes in the enzyme that may occur when substrate binds.<sup>26</sup> Similarly, in structures with substrate or substrate analogue in the active site, a residue in suitable proximity to the anomeric centre cannot be assumed to be the enzymatic nucleophile. Site-directed mutagenesis can be used to probe the function of active site residues.<sup>50-52</sup>

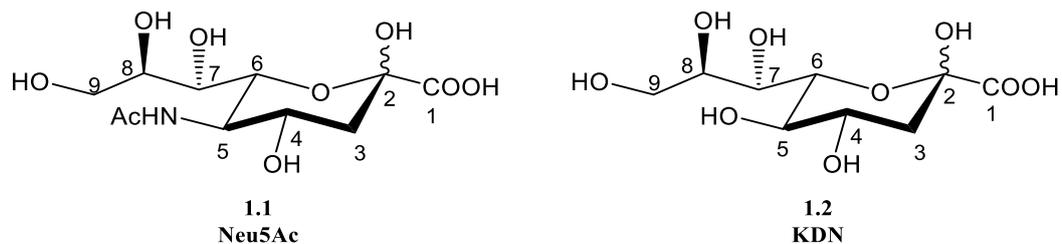
Due to a lack of conclusive experimental data on retaining GT mechanisms, theoretical approaches have been used.<sup>26,45</sup> Early work was focused on modelling the active site of retaining GTs. The modelling suggested that  $S_Ni$  reactions in these enzymes were feasible if the nucleophile and leaving group were both on one face of the sugar, with the mechanism using a single transition state. However, a similar model indicated that a double-displacement mechanism would occur if a nucleophile was near the anomeric carbon. A major caveat of both studies was that only the active site residues were modelled. Therefore, the protein environment and dynamics could not be accounted for.<sup>26</sup> More recently, models of full retaining GTs have been used.<sup>45</sup>

GTs are clearly challenging proteins to study. However, with perseverance, the functional structural and mechanistic characterisation of novel GTs is able to guide further studies such as the design of inhibitors for clinically relevant GTs, or the engineering of these enzymes, so that a greater range of biological relevant glycosides may be accessed via enzymatic methods.<sup>24,53,54</sup>

## 1.3 Sialic Acids

### 1.3.1 Sialic Acid Structures and Function in Eukaryotes

The term nonulosonic acid (NuLO) refers to a class of carbohydrates defined as nine carbon  $\alpha$ -keto acid sugars.<sup>1</sup> A well-studied sub-class of NuLOs are the sialic acids (Figure 1.6), of which over 50 naturally occurring derivatives have been identified.<sup>55</sup> Sialic acids can be found in all three domains of life, and include ubiquitous -keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid (N-acetyl-neuraminic acid, Neu5Ac **1.1**), as well as those found exclusively in one species.<sup>56,57</sup> All sialic acids have D-glycero-D-galacto stereochemistry and most sialic acids are derivatives of Neu5Ac **1.1** or 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN **1.2**).<sup>58</sup> The most common derivatives are O-acetylated at carbon positions 4, 7, 8, or 9 (Figure 1.8 depicts carbon numbering).<sup>59</sup>



**Figure 1.8: Structures of nonulosonic acids Neu5Ac 1.1 and KDN 1.2**

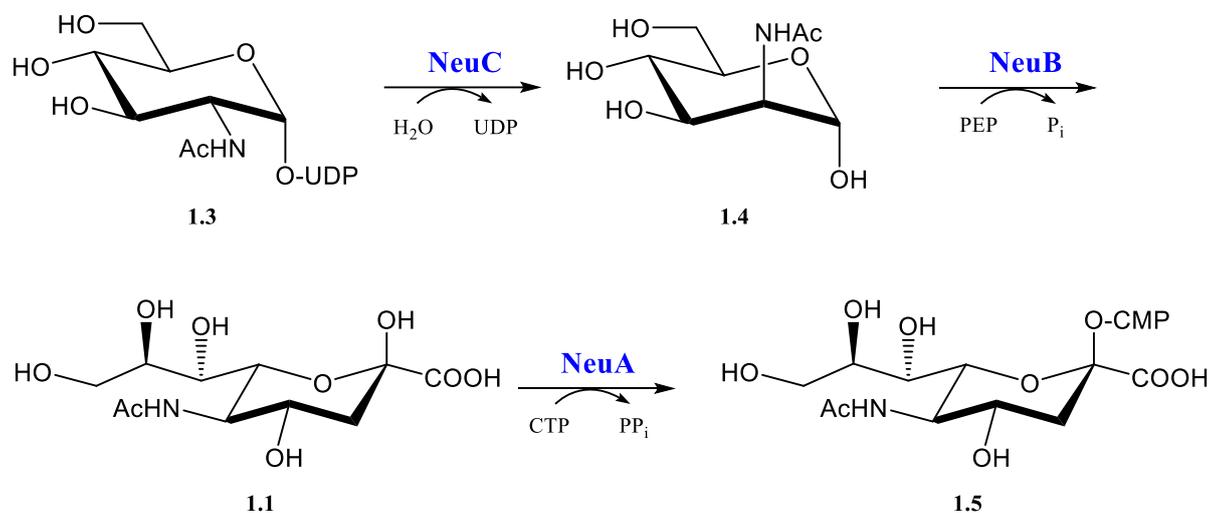
Within mammals, sialic acids are prevalent in cell-surface glycoconjugates of epithelial cells throughout the gastrointestinal tract and mucosal surfaces. Sialylated glycans in humans range from di- to polysaccharides, where they may be linked to various other carbohydrates and sialylglycoconjugates including a range of glycoproteins and glycolipids.<sup>60</sup> Studies in mice have demonstrated that sialic acids are essential for embryonic development, where embryonic death occurred in those with mutations in sialic acid biosynthesis genes.<sup>61</sup> As the most abundant terminal sugar on eukaryotic surface glycoconjugates, sialic acids are important in many cell-cell interactions and recognition processes.<sup>61,62</sup>

In eukaryotes, sialic acids may perform the following (but not exhaustive) functions: stabilisation of glycoconjugates and cell membranes due to charge-charge repulsion; mediating cell-cell regulation; act as chemical messengers; regulation of transmembrane receptor function; affecting membrane transport; controlling the half-lives of cells.<sup>63</sup>

Additionally, sialic acids may be involved in communication between host and pathogen and in the immune response.<sup>64</sup>

### 1.3.2 Sialic acid biosynthesis and scavenging in bacteria

Sialic acids may be synthesised *de novo* within bacteria (Figure 1.9), including *Neisseria meningitidis* and *Campylobacter jejuni*, where the biosynthesis pathway begins with uridine 5'-diphosphate-N-acetyl-D-glucosamine (UDP-GlcNAc **1.3**). The dual function enzyme, NeuC, hydrolyses this nucleotide-activated sugar and epimerises carbon 2 to yield, N-acetyl-D-mannosamine (ManNAc **1.4**). The resultant hexose is then converted into the Neu5Ac, via a condensation reaction with phosphoenol pyruvate (PEP), facilitated by sialic acid synthase (NeuB). CMP-sialic acid synthetase (NeuA) then nucleotide-activates Neu5Ac **1.1** using cytidine-5'-triphosphate (CTP) to produce using cytidine-5'-monophosphate-Neu5Ac (CMP-Neu5Ac **1.5**) (Figure 1.9) which may be utilised by CMP-sialic acid glycosyltransferases (sialyltransferases, (SiaTs)) to produce a range of sialylated glycoconjugates.<sup>58</sup>



**Figure 1.9: CMP-Neu5Ac 1.5 biosynthesis pathway in bacteria.**

Alternatively, sialic acid may be scavenged by bacteria from their environment. In the case of pathogenic bacteria, including *Pasteurella multocida*, *Haemophilus influenzae* and *Neisseria gonorrhoeae*, sialic acids may be scavenged from their host.<sup>65-67</sup> Some scavengers secrete a glycosylhydrolase (sialidase) to cleave sialic acid from host glycoconjugates. Bacteria must then uptake environmental sialic acid, through a range of sialic acid transporters.<sup>68</sup> However, some bacteria are unable to synthesise sialic acids, yet do not

secrete sialidases. Therefore, they are reliant on free sialic acids, released by the sialidases of other pathogens or cleaved by the host. These pathogens typically possess high-affinity sialic acid transporters.<sup>69</sup> Alternatively, free CMP-sialic acids, found with the host, can be used by cell-surface SiaTs to directly sialylate the bacterial cell-surface.<sup>70</sup>

Even though sialic acids are an important virulence factor for many pathogenic bacteria, some scavengers can use environmental sialic acid in catabolic as well as anabolic pathways.<sup>65</sup> Sialic acid can be used as a carbon and nitrogen source, where a series of enzymes convert the NuO into fructose-6-phosphate which can be used in glycolysis, in addition to pyruvate and ammonia.<sup>69</sup> A second function of some NeuA proteins is that they may de-acetylate intracellular sialic acid monomers.<sup>67</sup> This deacetylase function may be crucial in enabling scavenged sialic acids to be used in catabolism.<sup>67</sup> Alternatively, scavenged sialic acids may be used as a source of amino-carbohydrate for cell wall biosynthesis.<sup>63</sup> There exist bacteria that can scavenge sialic acid, yet do not use sialic acid in glycoconjugates, therefore, it appears that sialic acid is solely utilised in catabolic pathways within these cells.<sup>60</sup> Experimental evidence using a mouse model confirmed that catabolising scavenged host sialic acid, liberated by the sialidase activity of commensal bacteria, enhanced colonization of bacterial pathogens *Salmonella typhimurium* and *Clostridium difficile*.<sup>71</sup>

### **1.3.3 Sialic acid in bacterial glycoconjugates**

Sialic acids are present on the cell surfaces of both commensal and pathogenic bacteria.<sup>63</sup> Due to their abundance on cell surfaces of vertebrates, sialic acids are often used in pathogenic bacteria as an aforementioned form of “molecular mimicry”, in order to evade host immune systems.<sup>65,72,73</sup> Sialic acids enable pathogens to dampen the innate immune response, via interaction with the sialic acid-binding lectins (siglecs).<sup>74,75</sup> Sialic acids also provide bacterial pathogens a mask to escape the adaptive immune system.<sup>76</sup>

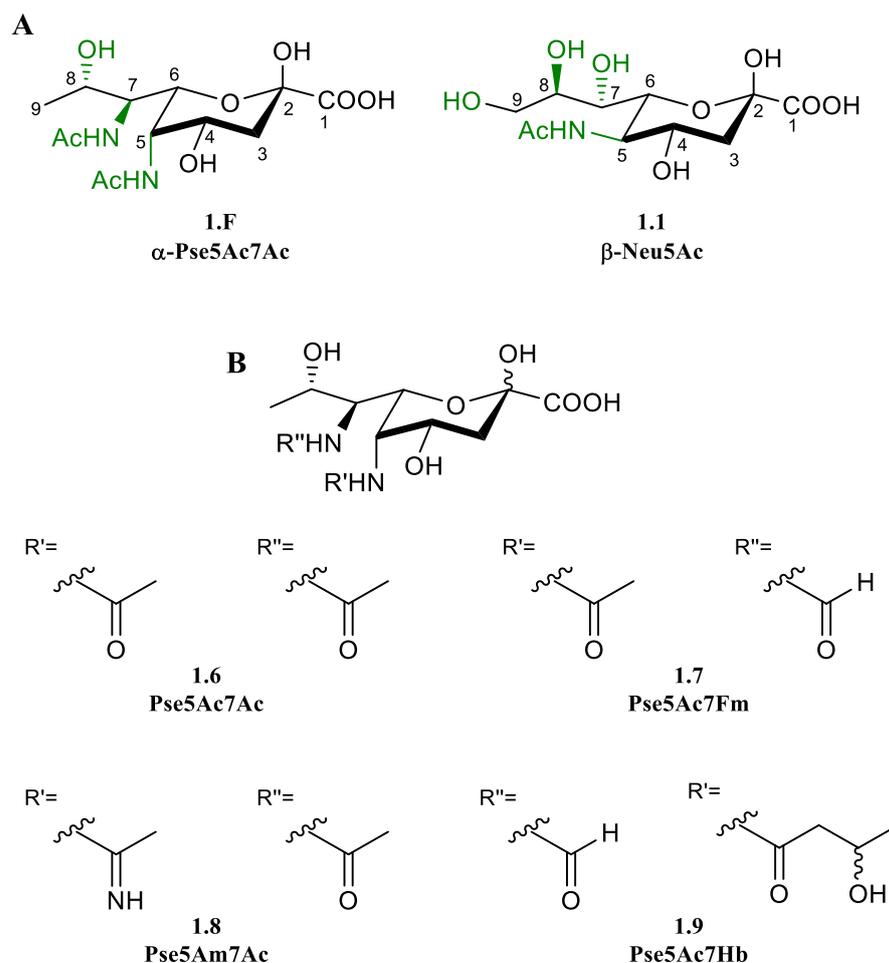
Whether scavenged or biosynthesised, sialic acid must be nucleotide-activated through the action of NeuA, for use in anabolic pathways, to build glycoconjugates. It may be incorporated into bacterial capsular polysaccharide which mimics host cells molecules, conferring low immunogenicity to these bacteria.<sup>63</sup> In *E. coli* O-acetyltransferases may modify Neu5Ac, at carbon 7 or 9, which is incorporated into the polysialic acid CPS.<sup>67</sup>

Additionally, sialic acid may be incorporated into the LPS O-antigen repeat unit, or lipooligosaccharide of a number of bacteria.<sup>77</sup> The sialic acid in some LPS may also be O-acetylated. Therefore, a diverse range glycan structures containing sialic acid can be found on bacterial cell surfaces. Several bacterial SiaTs are well-characterised and have been shown to be able to biosynthesise glycans containing sialic acid that is attached to several different carbohydrates, through various different linkages.<sup>41,43,78–80</sup>

## 1.4 Pseudaminic Acids

### 1.4.1 Occurrence and Biological Significance of Pseudaminic Acids

A less common, more recently discovered, group of NuOs which are of particular interest are the sugars generally termed pseudaminic acids as these “sialic acid-like” structures are not present in mammalian cells, but have been found in many bacterial species.<sup>81</sup> The general term “pseudaminic acid” (Pse) arose from the initial discovery of  $\alpha$ -5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno-non-2-ulosonic acid (Pse5Ac7Ac **1.6**) (Figure 1.10 A), within the LPS O-antigen of *Pseudomonas aeruginosa* O7 and O9.<sup>81</sup> Pse5Ac7Ac varies from Neu5Ac in stereochemistry at carbons 5, 8 and 9 and functionality at several positions (Figure 1.10 A). Derivatives of Pse5Ac7Ac have since been characterised in which *N*-linked substituents at C5 and C7 can show wide diversity (Figure 1.8 B).<sup>1</sup> The Bacterial Carbohydrate Structure Database currently lists 149 structures in which Pse5Ac7Ac or a derivative is present (<http://csdb.glycoscience.ru/bacterial/>) (May 2019).



**Figure 1.10: A: Structures of nonulosonic acids  $\alpha$ -Pse5Ac7Ac 1.F and  $\beta$ -Neu5Ac 1.1. Green highlights differing stereochemistry and/or functionality. B: Structures of prevalent pseudaminic acids: Pse5Ac7Ac 1.6, Pse5Ac7Fm 1.7, Pse5Am7Ac 1.8, Pse5Ac7Hb 1.9.<sup>1</sup>**

Pse glycoconjugates have also been identified in human gut archaea, *Methanobrevibacter smithii*, where Pse5Ac7Ac is present in the CPS.<sup>82,83</sup> To date this remains the sole reported example of Pse in archaea.<sup>21,82</sup> Although putative nonulosonic acid biosynthesis genes have been found in several other species of archaea, complete pathways within these genomes are yet to be assigned.<sup>82</sup>

Pse and its derivatives are widespread in a range of bacterial cell-surface glycans, where they are known to be present in both Gram-negative and Gram-positive species. Unlike sialic acids which are most commonly terminal glycan residues, Pse is predominantly found as an internal component of a glycoconjugate, although there are examples of terminal Pse residues.<sup>64,84</sup> In addition to Pse being found in LPS O-antigen and CPS of bacteria, serine and threonine residues of flagella and pili may be *O*-glycosylated with Pse-derivatives, present in both  $\alpha$  and  $\beta$  forms.<sup>81,85–87</sup>

Most notable is the presence of Pse-containing glycoconjugates within pathogenic bacteria, where Pse-derivatives are known to play a crucial role in the virulence of many clinically relevant, drug resistant strains.<sup>88–90</sup> For example, *Helicobacter pylori* flagella are exclusively glycosylated with Pse5Ac7Ac and experimental evidence showed that a disruption to flagellin glycosylation resulted in a loss of bacterial motility, due to a lack of detectable flagella.<sup>91</sup> *Campylobacter jejuni* flagellin has been suggested to be an immunoprotective antigen of this human pathogen.<sup>92</sup> In *Campylobacter jejuni*, a pathogen which is the leading cause of gastroenteritis, the FlaA1 flagellin protein is glycosylated with Pse5Ac7Ac or a derivative at multiple residues, where all but one are surface-exposed in the assembled flagellin.<sup>93</sup> Mutations to the *C. jejuni* flagellin serine residues which are glycosylated with Pse5Ac7Ac resulted in truncated flagella and reduced motility.<sup>94,95</sup> Motility has previously been shown to be a key factor in the ability of these two bacteria to colonize the host, therefore novel therapeutics could target Pse biosynthesis or flagellin glycosylation.<sup>96,97</sup>

It is generally believed that terminal sequences, rare structures and modifications of glycosides may mediate more specific biological functions.<sup>10</sup> Therefore, as Pse5Ac7Ac and derivatives are relatively rare carbohydrates, which are biochemically expensive for cells to produce, it can be assumed that they play significant biological roles. Although the biological relevance of Pse-derivatives within pathogens is not fully understood, based upon structural similarities to eukaryotic sialic acids, it has been noted that immune responses to bacteria may be dampened by the presence of cell surface Pse-based moieties.<sup>98</sup> Pse derivatives on the flagella of *C. jejuni* have been shown to bind to the human immune-modulatory receptor siglec-10, resulting in increased murine dendritic cell interleukin-10 expression. This evidence demonstrates the direct role that Pse-derivatives play in host-pathogen interaction and immune response to infection.<sup>20</sup>

#### **1.4.2 Biosynthesis of Nucleotide-Activated Pseudaminic Acids**

Given the importance of Pse-derivatives in several multidrug resistant pathogens the biosynthetic pathways of nucleotide-activated  $\alpha$ -CMP-Pse5Ac7Ac **1.10** and its derivatives have been proposed in multiple organisms (all pathways described in this section are depicted in Figure 1.11).<sup>86,99,100</sup> These organisms include the aforementioned *C. jejuni* and *H. pylori*, another gastric pathogen and the only known bacterial species associated with

cancer, where the  $\alpha$ -CMP-Pse5Ac7Ac biosynthetic pathways have been well characterised, *in vitro* (Figure 1.11).<sup>99,101,102</sup> The biosynthesis of Pse has a number of common features that are seen in the biosynthesis of all CMP-NuOs. Firstly, an activated hexose is utilised in all CMP-NuO biosynthesis pathways, in Pse biosynthesis UDP-GlcNAc **1.3**. Varying numbers of steps, including hydrolysis of the nucleotide are required to convert the activated hexose into the desired hexose, before a condensation with a three-carbon substrate, phosphoenol pyruvate (PEP **1.11**) for **1.6**, yields a NuO. Finally, all NuOs are nucleotide activated, usually CMP-activated, as is seen in **1.6** biosynthesis, such that they may be incorporated into glycoconjugates.<sup>1,58</sup>

Typically, within Gram-negative bacteria, biosynthesis of  $\alpha$ -CMP-Pse5Ac7Ac and its derivatives is performed in six enzymatic steps.<sup>99</sup> In *C. jejuni* and *H. pylori*, **1.3** is dehydrated and epimerised by PseB to give UDP-2-acetamido-2,6-dideoxy- $\beta$ -L-arabino-hexos-4-ulose **1.12**, the substrate for L-glutamine dependent aminotransferase PseC, which produces UDP-4-amino-4,6-dideoxy- $\beta$ -L-AltNAc **1.13**.<sup>103</sup> N-acetyltransferase, PseH, then uses **1.13** and acetyl-coenzyme A (Ac-CoA) to yield UDP-2,4-diacetamido-2,4,6-trideoxy- $\beta$ -L-altropyranose **1.14**, which undergoes hydrolysis of UDP, facilitated by nucleotidase PseG, producing 2,4-diacetamido-2,4,6-trideoxy-L-altropyranose (6-deoxy-AltdiNAc **1.15**). Pseudaminic acid synthase, PseI, utilises PEP to convert **1.15** to the nonusonic acid Pse5Ac7Ac **1.6**.<sup>103</sup> The sixth enzyme installs the CMP-moiety at the C1 hydroxyl, using CTP in produce CMP-Pse5Ac7Ac **1.10** (Figure 1.11).<sup>1</sup> This varies in the human pathogen *Aeromonas caviae*, where only five enzymes are required due to the dual-functionality of FlmD which performs the acetyl-transferase and UDP-hydrolase roles which are carried out by PseH and PseG, respectively within *C. jejuni* and *H. pylori*.<sup>104</sup>

Within the Gram-positive bacterium *Bacillus thuringiensis*, the biosynthesis of  $\alpha$ -CMP-Pse5Ac7Ac is carried out by seven enzymatic reactions, where the UDP-GlcNAc 5-epimerase and 4,6-dehydratase roles of Gram-positive PseB are performed by the enzymes "Pen" and "Pal" respectively.<sup>99,105</sup> This yields the PseC or Pam substrate **1.12**, from which point the remainder of the pathway is analogous to that in *C. jejuni* and *H. pylori* (Figure 1.11). These seven enzymes are encoded in one operon.

The putatively assigned  $\alpha$ -CMP-5-acetamido-7-N-(3-hydroxybutanyl)-3,5,7,9-tetradecyloxy-L-glycero-L-manno-non-2-ulosonic acid (CMP-Pse5Ac7Hb **1.16**) biosynthetic pathway can be

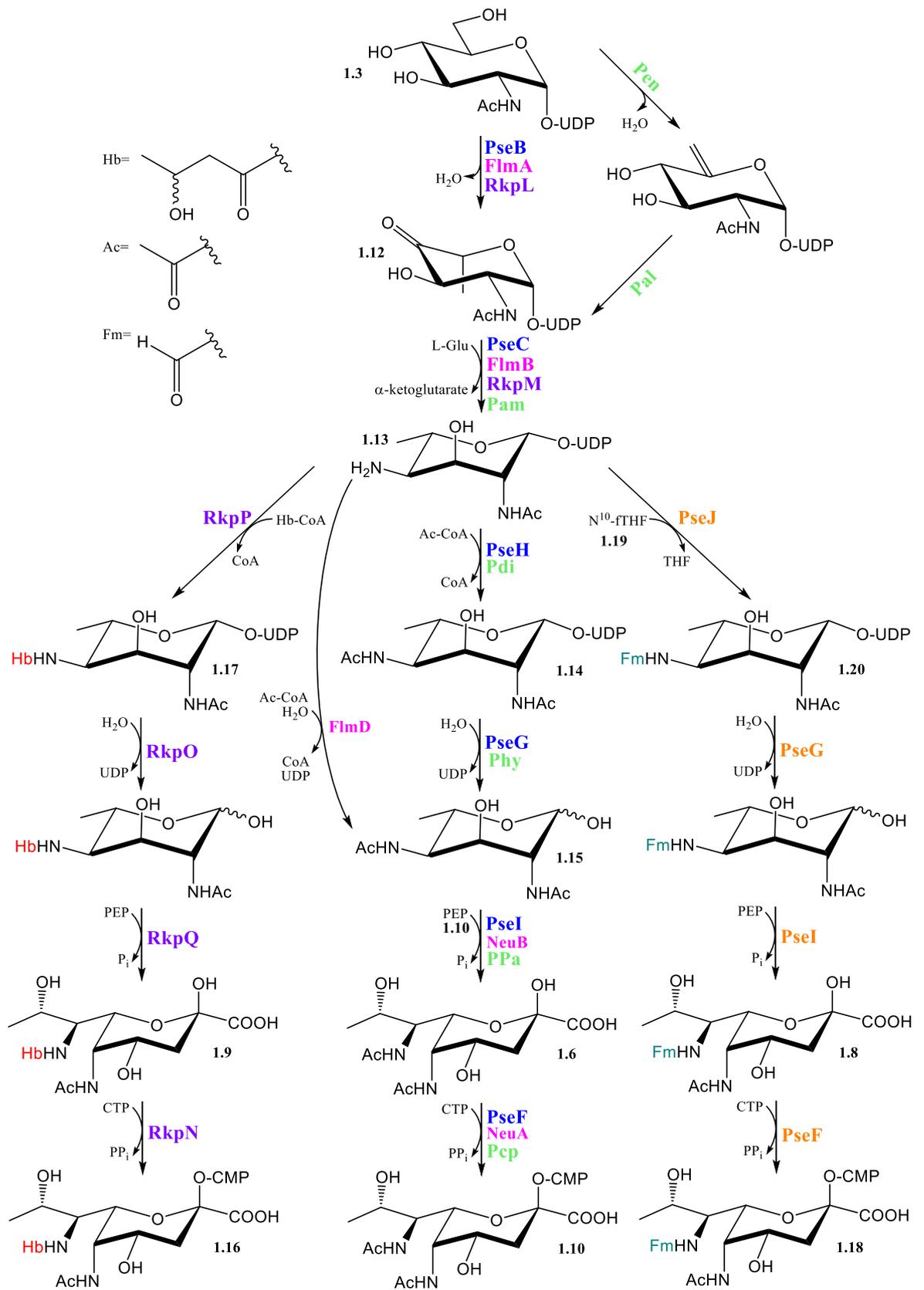
found within *Sinorhizobium fredii* HH103, where 5-acetamido-7-N-(3-hydroxybutanyl)-3,5,7,9-tetra-deoxy-L-glycero-L-manno-non-2-ulosonic acid (Pse5Ac7Hb **1.9**) makes up a homopolysaccharide K-antigen.<sup>106</sup> The biosynthetic pathway begins with two enzymes, homologous to that of *C. jejuni* PseB and PseC. The pathways diverge at the third enzyme, where in *S. fredii* the primary amine of C7 has a hydroxybutanol group installed **1.17**, by the transferase RkpP, in place of an N-acetylation that occurs in Pse5Ac7Ac biosynthesis. The remainder of the CMP-Pse5Ac7Hb pathway, comprises of three subsequent enzymes, homologous to those of the CMP-Pse5Ac7Ac pathway.<sup>99,100</sup>

Most recently the  $\alpha$ -CMP-5-acetamido-7-N-formyl-3,5,7,9-tetra-deoxy-L-glycero-L-manno-non-2-ulosonic acid (CMP-Pse5Ac7Fm **1.18**) biosynthesis pathway from Gram-positive *Anoxybacillus kamchatkensis* has been partially elucidated *in vitro* (Figure 1.9).<sup>107</sup> This pathway requires six-enzyme, demonstrating that a six-enzyme pathway is not an exclusive feature of Gram-negative species. The activities of the first three enzymes, PseB, PseC and PseJ were confirmed *in vitro*. PseB and PseC were shown to function similarly to their Gram-negative homologues to produce **1.14** from **1.3**. In a one-pot three enzyme synthesis, the formyl donor N<sup>10</sup>-formyltetrahydrofolate (N<sup>10</sup>-fTHF, **1.19**) was used to formylate the C4 amine of **1.14** to produce UDP-4,6-dideoxy-4-formamido-L-AltNAc **1.20**.<sup>107</sup> Three downstream enzymes have been putatively assigned as those required to produce CMP-Pse5Ac7Fm from **1.18** in steps which are analogous to those used in *C. jejuni* CMP-Pse5Ac7Ac **1.10** biosynthesis.<sup>107</sup> Currently there are no other examples of *in vitro* characterisation of enzymes involved in biosynthesis of Pse-derivatives.

In addition to Pse5Ac7Ac, the derivative 5-acetamido-7-acetamido-3,5,7,9-tetra-deoxy L-glycero-L-manno-non-2-ulosonic acid CMP-Pse5Ac7Am **1.20**, is found on the flagella of *C. jejuni*. It is currently unclear at which point in the pathway PseA converts the acetyl group to an acetamido group and it has been suggested that Pse or CMP-Pse5Ac7Ac is the substrate.<sup>108</sup> CMP-Pse5Ac7Am was isolated for *C. jejuni* lysate, indicating that the modification is made prior to glycosylation of the flagellin protein.<sup>108</sup>

The elucidation of the CMP-Pse5Ac7Ac biosynthetic pathway has been complemented by structural characterisation some of the individual enzymes from *H. pylori* and *C. jejuni*. As Pse can be considered a key virulence factor for many pathogens, it follows that these enzymes are therefore valid therapeutic targets for inhibitor development

studies.<sup>99,103,109,110</sup> The absence of Pse in eukaryotes increases their suitability as targets for novel therapeutic compounds.

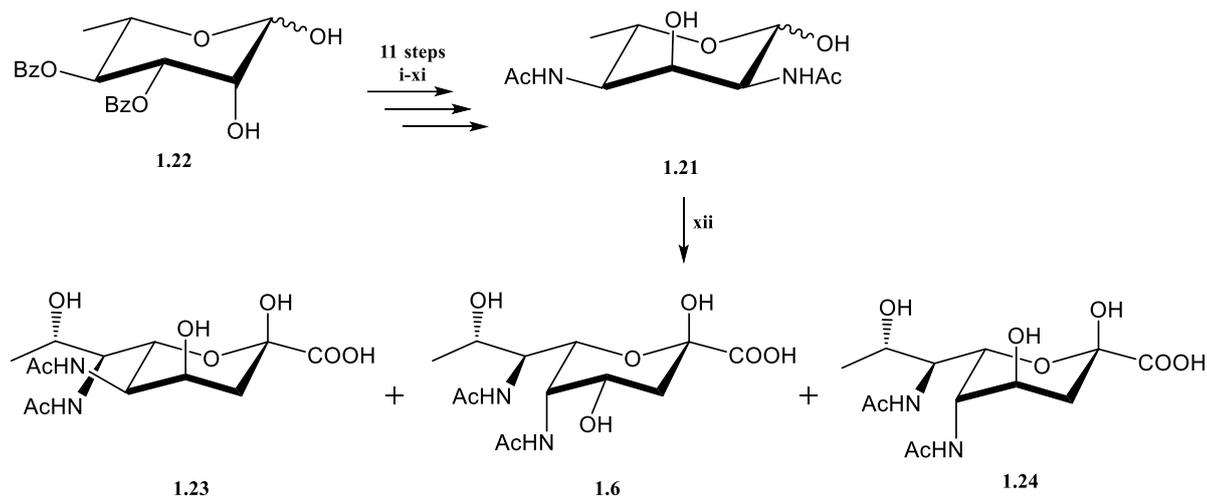


**Figure 1.11: Biosynthesis pathways for CMP-Pse5Ac7Hb 1.16, CMP-Pse5Ac7Ac 1.10 and CMP-Pse5Ac7Fm 1.18. *C. jejuni* and *H. pylori* enzymes in blue, *S. fredii* HH103 enzymes in purple, *B. thuringiensis* enzymes in green, *A. caviae* enzymes in pink and *A. kamchatkensis* enzymes in orange.**

### 1.4.3 Chemical synthesis towards Pseudaminic Acids and Pseudaminic Acid glycosides

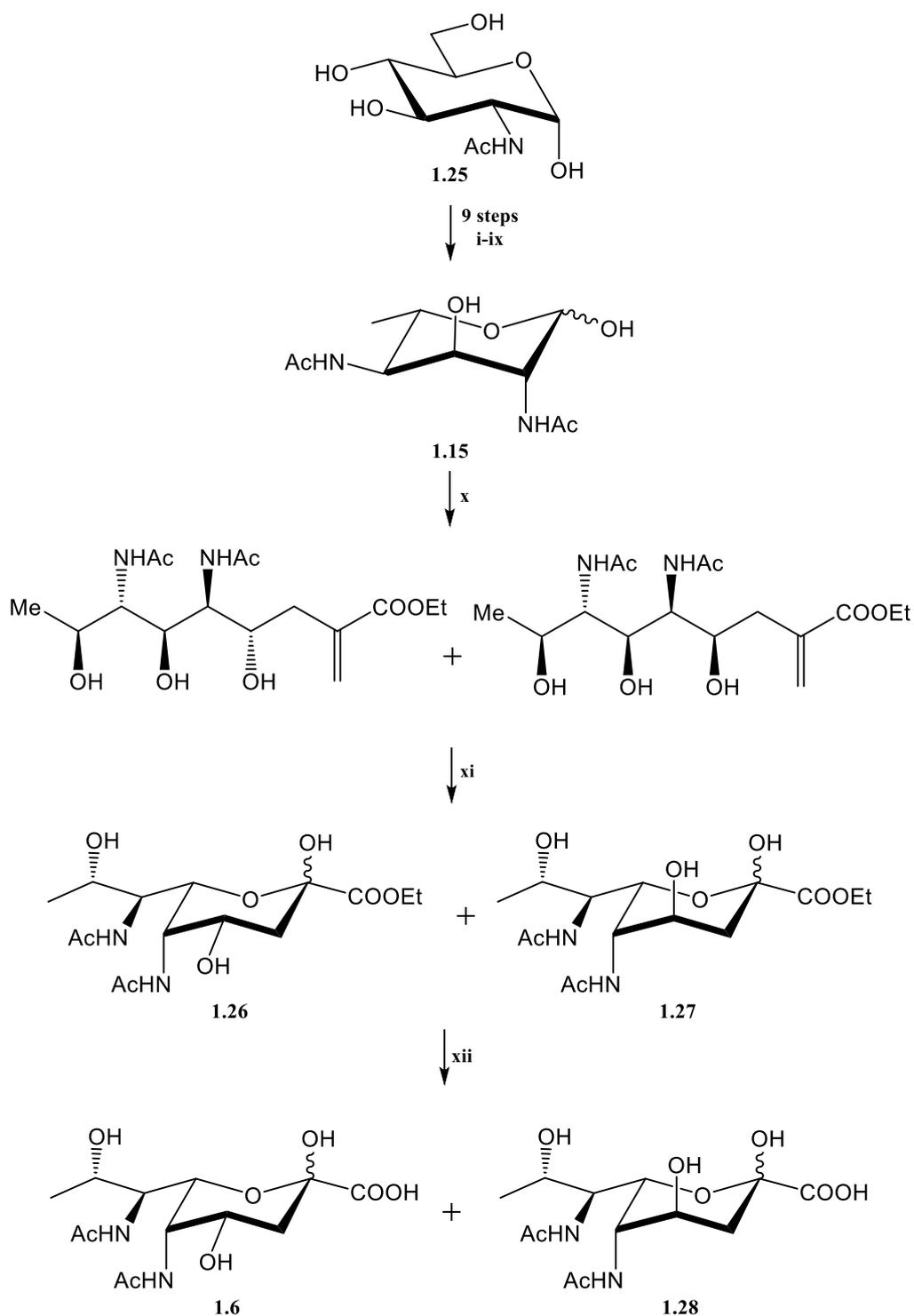
Many studies on the biological significance of Pse are reliant on a supply of Pse, Pse-derivatives and Pse-based chemical probes. Currently Pse-derivatives are not commercially available in appreciable quantities. Whilst Pse5Ac7Ac **1.6** is structurally similar to commercially available Neu5Ac **1.1**, synthesis of Pse5Ac7Ac has been much more demanding and very low yielding, and indeed limitations can be found in each of the published methods.<sup>2-9</sup> Many of the complications in synthesis of Pse-based molecules can be attributed to the epimeric stereochemistry at the C5 position and the propyl chain (C7-9) and different desired functionalities at several positions. Generally, all current literature describes one of three broad approaches: synthesis begins with a NuLO acid, which requires inversion of several stereocentres, achieved with varying levels of success; alternatively, a four-carbon amino acid is used as starting material; or thirdly, synthesis of the biosynthetic precursor 6-deoxy-AltdiNAc **1.14** (or similar) and subsequent condensation reaction to attach the functionalised propyl chain, via chemical or biological routes.<sup>2-9</sup> Whilst routes to Pse5Ac7Ac and derivatives have required much optimisation, more recent reports also use Pse-based donors in chemical glycosylation, yielding Pse-based glycosides.<sup>3,6,7,9</sup>

The initial publication reporting the synthesis of Pse5Ac7Ac came from Knirel and co-workers in 2001, 17 years after his reported discovery of Pse5Ac7Ac.<sup>2,81</sup> They used an approach that was analogous to the synthesis of the sialic acid KDN, where a six-carbon carbohydrate, undergoes a condensation reaction with oxaloacetic acid, followed by decarboxylation, to yield nine-carbon product.<sup>111</sup> In the synthesis employed by Knirel and colleagues, 2,4-diacetamido-2,4,6-trideoxy-L-allose **1.21** was used as the six-carbon intermediate, an epimer of naturally occurring biosynthetic intermediate **1.15** (Scheme 1.1).<sup>2</sup> A 12 step synthesis beginning with 3,4-di-O-benzoyl-β-L-rhamnopyranoside **1.22** was required to produce **1.21**. The condensation step produced three nonulosonic acids, **1.23**, **1.6** and **1.24**, in 8%, 3% and 1% yields respectively. Knirel *et al.* remarked that the stereoselectivity of the condensation reaction was unexpected. As Pse5Ac7Ac was a very low-yielding by-product, in a multistep synthesis there was clear scope for improvement of synthetic routes to Pse5Ac7Ac.<sup>2</sup>



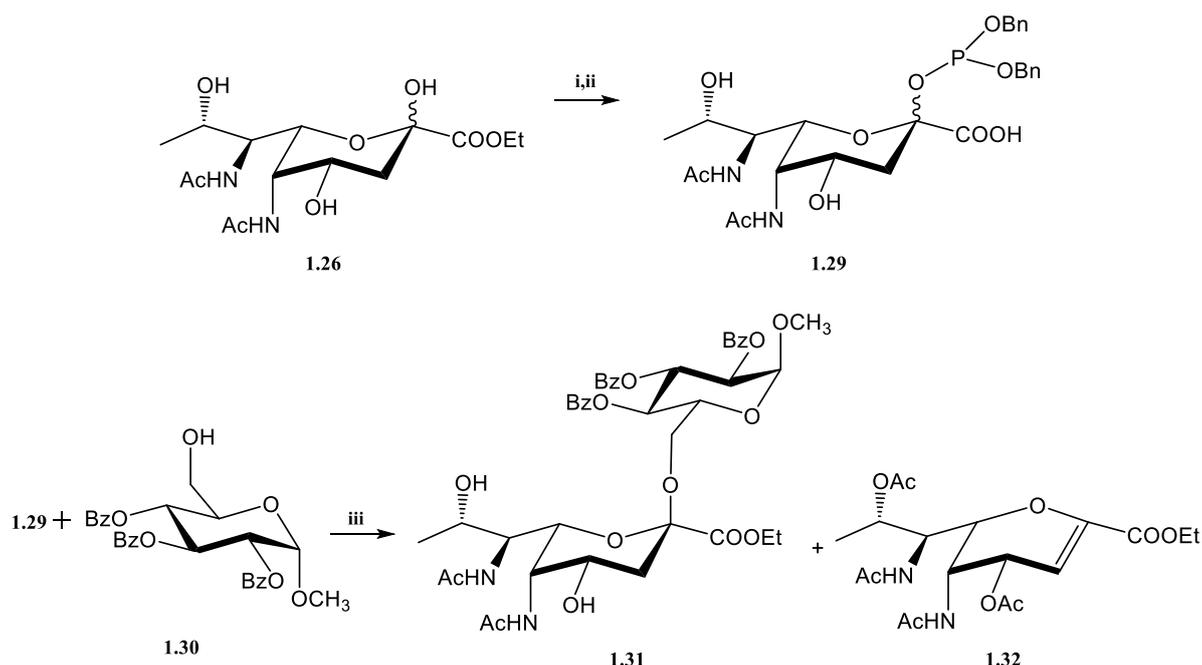
**Scheme 1.1: Synthesis of nonulosonic acids 1.23, 1.6 and 1.24 via biosynthetic intermediate 1.21, as reported by Knirel and co-workers.<sup>2</sup> Reagents and conditions: i: Bu<sub>2</sub>SnO, benzene, reflux; ii: BnBr, Bu<sub>4</sub>NBr, benzene, reflux; iii: MeONa, MeOH; iv: trimethyl orthoacetate, PTSA, MeCN, rt; v: Ac<sub>2</sub>O, pyridine, rt; vi: 80% aq AcOH, rt; vii: Tf<sub>2</sub>O, pyridine, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C; viii: NaN<sub>3</sub>, DMF, rt; ix: NaN<sub>3</sub>, NH<sub>4</sub>Cl, aq EtOH, reflux; x: H<sub>2</sub>, Pd(OH)<sub>2</sub>/C, MeOH, 30°C; xi: Ac<sub>2</sub>O, MeOH, rt; xii: oxaloacetic acid, Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub>, pH 10.5.**

Ito and co-workers reported synthesis of Pse5Ac7Ac via production of biosynthetic intermediate 6-deoxy-L-AltdiNac **1.15** (the naturally occurring product of PseG enzyme) using an inexpensive starting material, N-acetyl-glucosamine (GlcNAc **1.25**) starting material. The route to **1.15** was not trivial and required nine steps (Scheme 1.2), as the initial strategy for introducing a C4 amino moiety to **1.15** precursor was unsuccessful. Therefore, an alternative four step strategy was employed to obtain **1.15** in 15% yield, which was converted to the desired Pse5Ac7Ac over an additional three steps (Scheme 1.2). Conversion of **1.15** to **1.6** involved the synthesis of the ethyl-ester **1.26**,  $\alpha:\beta = 5:1$ . However, an undesired C4 epimer **1.27** was produced, lowering the potential yield of Pse5Ac7Ac. **1.15** and **1.27** were saponified to produce 4-epi-Pse5Ac7Ac **1.28** and Pse5Ac7Ac **1.6**, in a low overall yield of 4%.



**Scheme 1.2: Synthesis of Pse5Ac7Ac 1.6 and 4-*epi*-Pse5Ac7Ac 1.28 using 1.15 as reported by Ito and coworkers.<sup>3</sup> Reagents and conditions: i: as reported by Sharma 1990<sup>112</sup>; ii: I<sub>2</sub>, PPh<sub>3</sub>, imidazole, THF, 0 °C, 2 h, 88%; iii: TIPSOTf, 2,6-lutidine, CH<sub>2</sub>Cl<sub>2</sub>, 12 h, 92%; iv: t-BuOK, THF, 70 °C, 10 h, then TBAF, THF, 2 h, 96%; v: H<sub>2</sub>, RhCl(PPh<sub>3</sub>)<sub>3</sub>, benzene, EtOH, 3 h, 84%; vi: Dess–Martin periodinane, NaHCO<sub>3</sub>, CH<sub>2</sub>Cl<sub>2</sub>, 17 h, 94%; vii: MeONH<sub>2</sub>·HCl, NaHCO<sub>3</sub>, MeOH, 65 °C, 17 h, 95%, major/minor = 2:1; viii: Sml<sub>2</sub>, MeOH, THF, 12 h, then Ac<sub>2</sub>O, pyridine, 6 h, 66%; ix: H<sub>2</sub>, Pd(OH)<sub>2</sub>, EtOH, 4 h, 94%, α:β = 2:1; x: In, 0.1 N HCl–EtOH (1:6), 40 °C, 12 h, 77%; xi: O<sub>3</sub>, MeOH, -78 °C, 30 min, followed by 30% H<sub>2</sub>O<sub>2</sub>, H<sub>2</sub>O, HCO<sub>2</sub>H, 90 min, for 1.28: 85%, α:β = 5:1; for 1.31: 86%, α:β = 3:1; xii: Et<sub>3</sub>N–H<sub>2</sub>O (1:3), 0 °C, 2 h, for 1.6: 96%, for 1.32: 92%**

Aware of several examples of equatorial ( $\alpha$  for Neu5Ac,  $\beta$  for Pse5Ac7Ac) Neu5Ac glycosides being synthesised using Neu5Ac thioglycoside donors, Ito and co-workers used established conditions with the aim of producing a Pse5Ac7Ac-thio glycoside but described the reaction as “extremely sluggish” and therefore dibenzyl phosphite ethyl-ester **1.29** was instead used as the Pse5Ac7Ac-based donor (Scheme 1.3 below). Glycosylation of **1.30** in  $\text{CH}_3\text{CN}$  unexpectedly produced  $\alpha$ -glycoside **1.31** ( $\alpha:\beta = 10:1$ ) (combined yield of 28%). However, the major product was undesired glycal **1.32**. The reaction was performed in dichloromethane, yielding 35% Pse5Ac7Ac-based glycoside which was exclusively the  $\alpha$  anomer and a 62% yield of **1.31**. However, whilst this was the first report of a glycosylation using a Pse-based donor, the product was not deprotected to afford a Pse5Ac7Ac glycoside.<sup>3</sup> Overall, this work demonstrated that alternative approaches to those used in synthesis of Neu5Ac glycosides are required to produce Pse-based glycosides, therefore adding additional challenges to already complex syntheses.

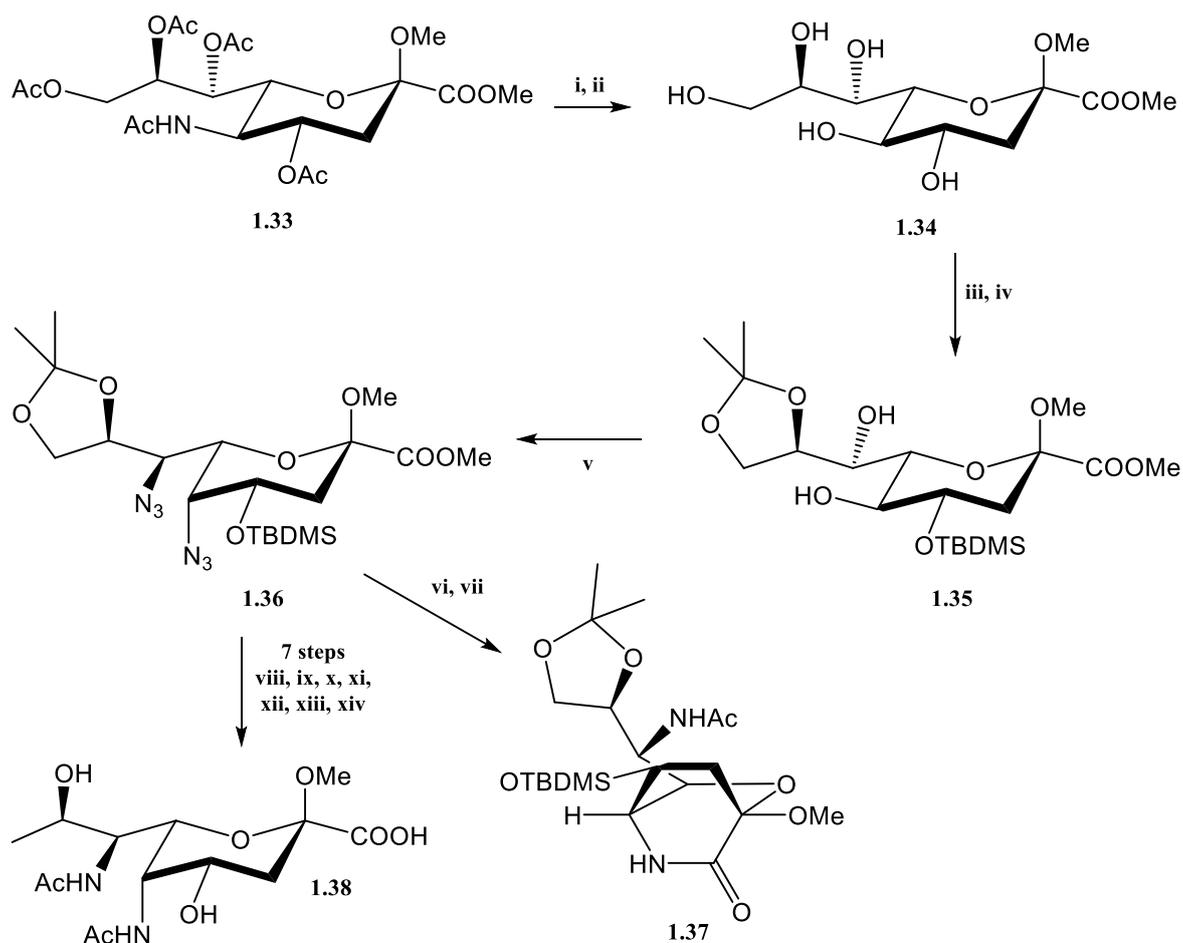


**Scheme 1.3: Synthesis of  $\alpha$ -ethyl-Pse5Ac7Ac glycoside **1.31**, as reported by Ito and coworkers.<sup>3</sup> Reagents and conditions: i:  $\text{Ac}_2\text{O}$ , pyridine,  $0^\circ\text{C}$ , 12 h, 73%,  $\alpha:\beta = 8:1$ ; ii:  $\text{Et}_2\text{NP}(\text{OBn})_2$ , 1*H*-tetrazole, THF, 2 h, 94%; iii: TMSOTf,  $\text{CH}_3\text{CN}$ ,  $0^\circ\text{C}$ , 3 h [1.31: 28% ( $\alpha:\beta = 10:1$ ), 1.32: 47%] or  $\text{CH}_2\text{Cl}_2$ , rt, 6 h [1.31: 35%, 1.32: 62%].**

In 2014 Kiefel and co-workers reported the synthesis of a Pse5Ac7Ac analogue over 14 steps. Their initial retrosynthetic analysis suggested nonulosonic acid KDN **1.2** as starting material, however as this compound is highly costly, this precluded its use in preparative

scale synthesis. They also reported that attempts to use KDN had resulted in both pyranose and furanose products with both  $\alpha$  and  $\beta$  anomers and therefore deemed this wasteful. Alternatively, methyl ester  $\beta$ -methyl glycoside of Neu5Ac **1.33** was used to produce  $\beta$ -methyl glycoside methyl ester of KDN **1.34** in 57% yield, an improvement on previously reported yields. Installation of the N-acetyl groups was problematic and required many optimisations (Scheme 1.4). A 5,7-diol **1.35** was produced on route to installing the key 5,7-di-N-acetyl functionalities of Pse5Ac7Ac. Desired **1.36** was isolated in 80% yield after diols were converted to azido groups, hydrogenated and acetylated. However, initial hydrogenation and acetylation attempts produced 1,5-lactam **1.37**. Serendipitously, this undesired compound may allow for selective acylation of the two amine groups, therefore potentially enabling access to naturally occurring Pse-derivatives.<sup>4</sup> However, no further work on this route has been reported.

Further steps were required to produce a deprotected Pse5Ac7Ac analogue **1.38**. However, in the final compound the stereochemistry of C8 matches that of the sialic acids, therefore **1.38** is termed an 8-*epi*-Pse5Ac7Ac analogue (Scheme 1.4).<sup>113</sup>

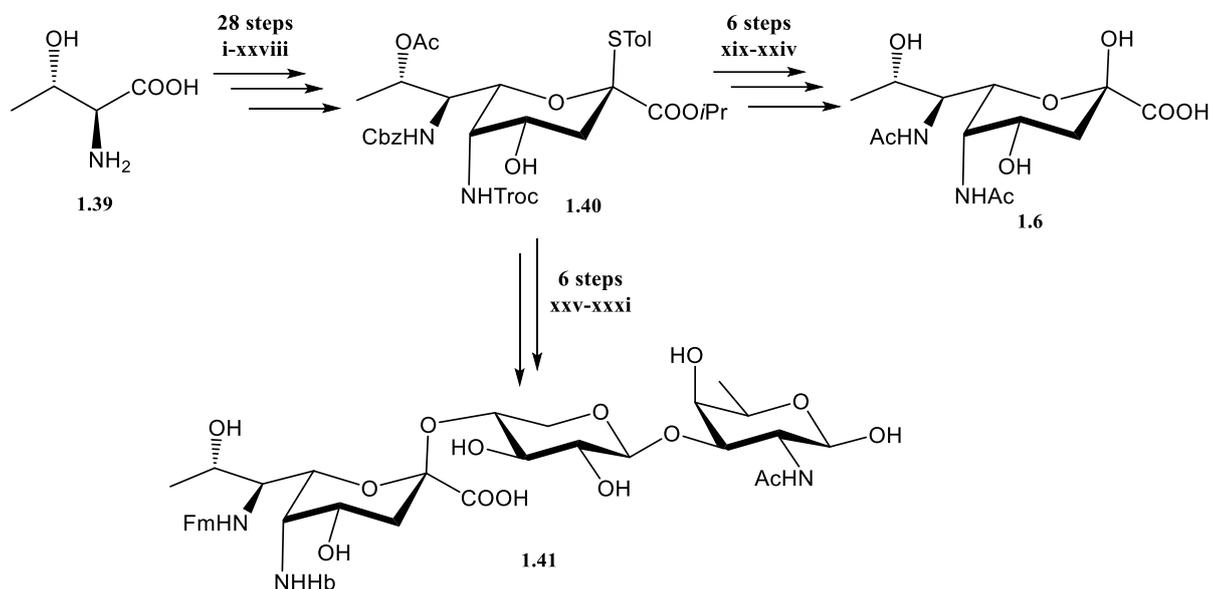


**Scheme 1.4: Synthesis of 8-*epi*-Pse5Ac7Ac **1.38**, reported by Kiefel co-workers.<sup>113</sup> Reagents and conditions: i: NaNO<sub>2</sub>, Ac<sub>2</sub>O-HAcOH (2:1), 0 °C, 1 hr, then 50 °C, 6 hrs; ii: MeONa, MeOH; iii: 2,2-dimethoxypropane, PTSA, acetone, r.t.; iv: imidazole, TBDMS-Cl, DMF, r.t, 16 hrs; v: NaN<sub>3</sub>, DMF, 4 °C; vi: H<sub>2</sub> Pd on C, MeOH; vii: Ac<sub>2</sub>O py; viii: *p*-TSA.H<sub>2</sub>O, Pd(OH)<sub>2</sub>/C, H<sub>2</sub>, MeOH, r.t, 2 hrs; ix: Ac<sub>2</sub>O, pyridine, r.t, 12 hrs; x: TFA, THF-H<sub>2</sub>O (4:1), r.t, 30 mins; xi: NaOMe, MeOH, r.t, 2 hrs. xii I<sub>2</sub>, PPh<sub>3</sub>, imidazole, THF, 60 °C, 2 hrs. xiii *i*Pr<sub>2</sub>EtN, Pd(OH)<sub>2</sub>/C, H<sub>2</sub>, MeOH, r.t, 16 hrs; xiv: aq. NaOH (1 M), 40 °C, 1 hr; xv: Dowex-50WX8(H<sup>+</sup>), 80 °C, 36 hrs.**

2 years after reporting the synthesis of 8-*epi*-Pse5Ac7Ac **1.38** Kiefel, Payne and colleagues were able to produce Pse5Ac7Ac in a 17-step synthesis. This was achieved using the aforementioned intermediate **1.34**, where the stereochemistry of the C8 hydroxyl was inverted using an oxidation then subsequent reduction.<sup>5</sup>

Xuechen and colleagues began synthesis of Pse5Ac7Ac **1.6** with the cheap starting material, L-threonine **1.39** which required a 28-step synthesis to yield a ring closed sugar **1.40** with the desired stereochemistry which was converted to Pse5Ac7Ac **1.6** over five additional steps (Scheme 1.5). This work is a significant improvement in the synthesis of Pse5Ac7Ac **1.6** due higher yields, lower cost of materials and higher control of stereochemistry.<sup>6</sup> Xuechen and colleagues not only produced Pse5Ac7Ac but also, published the first reported

total synthesis Pse-containing trisaccharide **1.41**. This highly impressive strategy was used to produce *Pseudomonas aeruginosa* 1244 pilin glycan,  $\alpha$ -Pse5Hb7Fm-(2 $\rightarrow$ 4)- $\beta$ -Xyl-(1 $\rightarrow$ 3)-FucNAc **1.41**.



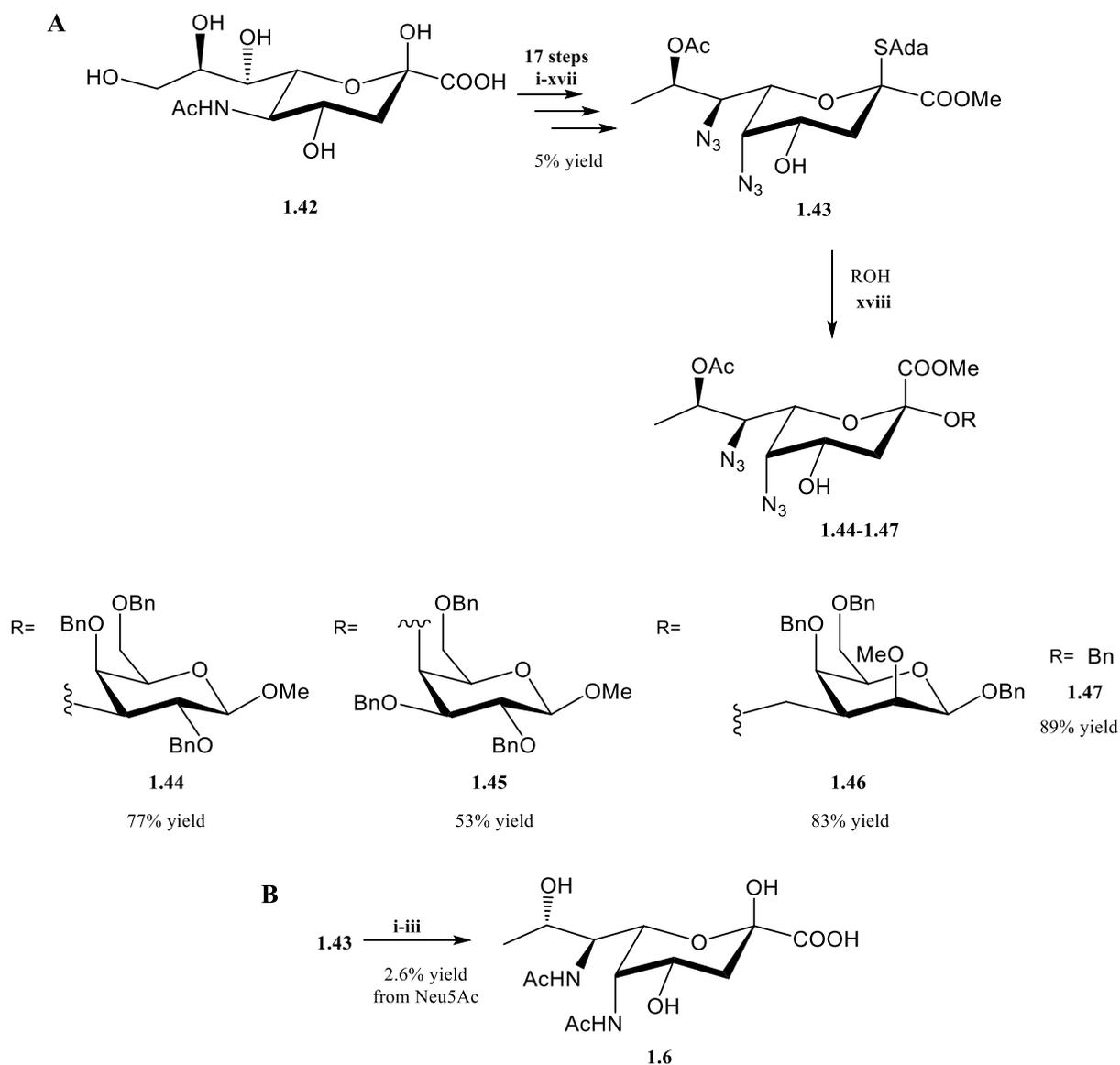
**Scheme 1.5: Synthesis of Pse5Ac7Ac 1.6 and *P. aeruginosa* pilin trisaccharide 1.41, as reported by Xuechen and coworkers.<sup>6</sup>** Reagents and conditions: i: SOCl<sub>2</sub>, MeOH, reflux; ii: AcCl, Et<sub>3</sub>N, CH<sub>2</sub>Cl<sub>2</sub>; iii: SOCl; iv: 10 % aq. HCl, reflux; v: CbzCl, Na<sub>2</sub>CO<sub>3</sub>, H<sub>2</sub>O; vi: MeI, KHCO<sub>3</sub>, DMF; vii: DMP, BF<sub>3</sub>·OEt<sub>2</sub>, DCM, 12 hrs; viii: NaBH<sub>4</sub>, CaCl<sub>2</sub>, EtOH-THF, 24 hrs; ix: BAIB, TEMPO, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C-r.t, 10 hrs; x: LiOTf, iPr<sub>2</sub>NEt, DCE-DMF, r.t, 3 hrs; xi THF-H<sub>2</sub>O, reflux, 10 hrs; xii: Et<sub>3</sub>SiH, Pd on C, THF, 3 hrs; xiii: Indium powder, NH<sub>4</sub>Cl, EtOH, 2 hrs; xiv: Dess-Martin periodinane, CHCl<sub>2</sub>, 0 °C, 2 hrs; xv: TBAF, HOAc, THF, 1 hr; xvi: NaBH(OAc)<sub>3</sub>, HOAc, MeCN, -40 to -20 °C, 10 hrs; xvii: HOAc, H<sub>2</sub>O, 50 °C, 20 hrs; xviii: 3 % aq. HCl in MeOH, 0 °C to r.t, 8 hrs; xix: TrocCl, 0.5 M Na<sub>2</sub>CO<sub>3</sub>, MeCN, 2hrs; xx: O<sub>3</sub>, CHCl<sub>2</sub>, -78 °C, 30 mins, then Me<sub>2</sub>S. xxiii: Ac<sub>2</sub>O, pyridine, DMAP; xxi: TolSH, BF<sub>3</sub>·OEt<sub>2</sub>, CHCl<sub>2</sub>, 16 hrs; xxii: BnOH, TolSCL, AgOTf, AW-300 MS, -78 °C; xxiii: Zn(s), Ac<sub>2</sub>O, HOAc, 40 °C, 3 hrs; xxiv: Pd on C, H<sub>2</sub>, NH<sub>4</sub>OAc, MeOH-CH<sub>2</sub>Cl<sub>2</sub>, 1 hr; xxv: aq. LiOH, MeOH-THF (4:1), r.t, 24 hrs; xxvi: Pd/C, H<sub>2</sub>, MeOH-H<sub>2</sub>O, 12 hrs. xxvii: HCl(aq), acetone-H<sub>2</sub>O, reflux, 16 hrs; xxviii: (*R*)-3-benzyloxybutanoic anhydride, pyridine, DMAP, 12 h; xxviii: (compound 46 in ref. 6), TolSCL, AgOTf, DCM, AW-300 molecular sieves, -78 °C, 2 h; xxix: TBAF, HOAc, THF, 40 °C, 48 h; xxx: NiCl<sub>2</sub>·6H<sub>2</sub>O, NaBH<sub>4</sub>, MeOH, 0 °C, 1 h; then Ac<sub>2</sub>O, 0 °C, 1 h; xxxi: LiOH, THF, H<sub>2</sub>O; xxxii: H<sub>2</sub>, Pd/C, HOAc-H<sub>2</sub>O, 48 h; viii: formic anhydride (in Et<sub>2</sub>O), Et<sub>3</sub>N, MeOH, -20 °C, 5 h.

The Pse-based donor **1.40** underwent six further steps to produce **1.41**. Again, different protecting groups were installed at nitrogen atoms at C5 and C7 including the selective installation of N-3-hydroxyl butanol at C5 and N-formyl at C7. However, many optimisations

were required to introduce the desired group at the nitrogen of C5 before the monosaccharide was used in glycosylation. The trisaccharide was fully deprotected to obtain the desired **1.41** (Scheme 1.5).

The reducing carbohydrate at the end of the trisaccharide gives the possibility for installing chemical handles, via established methodologies, which could be used as reporters or in the synthesis of glycoconjugates. Access to **1.41** is perhaps the most significant achievement in chemical synthesis of Pse-glycosides, and this naturally occurring product may be used in studies the biological function and significance of Pse glycosylation in *P. aeruginosa* and in pathogenesis and development of novel therapeutics.<sup>6</sup> An additional benefit of this overall scheme is that it allows for the incorporation of various N-linked functionalities at C5 and C7, where naturally occurring Pses vary in functionality.<sup>1,6</sup>

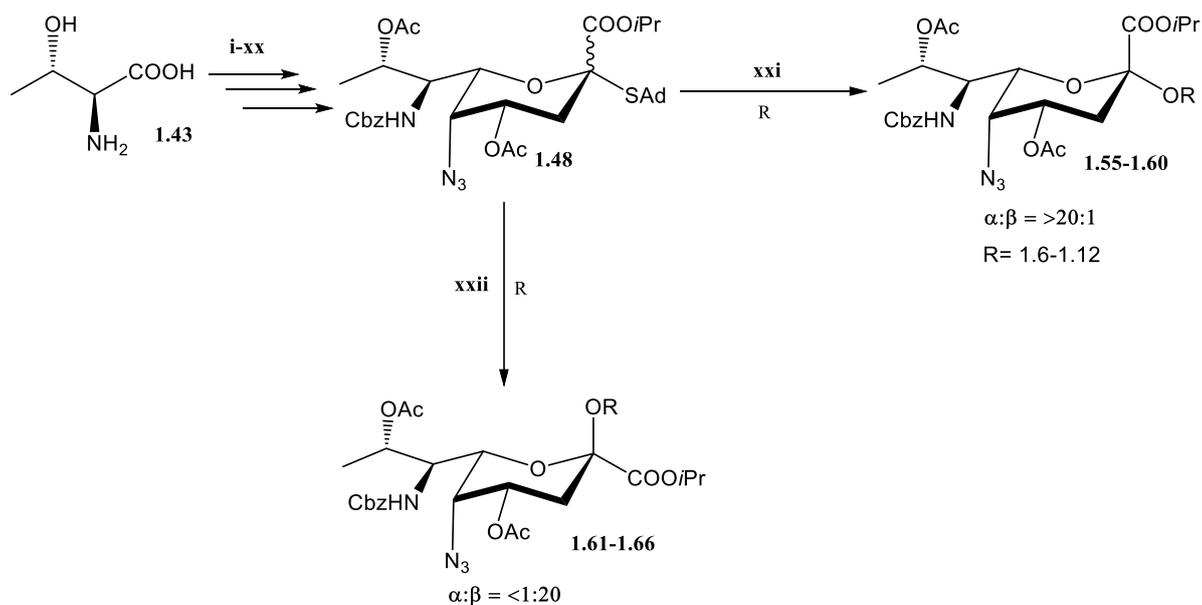
Crich and co-workers recently described both the synthesis of a Pse-donor **1.42** (Scheme 1.6 A) and Pse5Ac7Ac **1.6** (Scheme 1.6 B). Beginning with Neu5Ac **1.1** the Pse-donor **1.6** was synthesised in 5% yield over 18 steps. In contrast to the  $\alpha$ -Pse-based disaccharide produced by Ito *et al.* this work reports the total synthesis of  $\beta$ -Pse-based disaccharides. Di-azido donor **1.43** was used to obtain disaccharides, which were regioselectively Boc-protected at either the C5 or C7 azide, allowing selective conversion of the remaining azide to N-acetyl. Crich noted that this strategy would allow for installation of different N-functionalities at C5 and C7, using a single glycosyl-donor, producing **1.44-1.47**. The  $\beta$ -donor (equatorial OH at C2) was remarked to have “exquisite equatorial selectivity”, predicted to be due to the stronger electron withdrawing effects of the conformation of the C7-9 side chain in the Pse-based molecule vs that of the Neu5Ac side chain. Synthesis of these  $\beta$ -Pse-based glycosides is impressive and selective installation of desired functional groups at N5 and N7, where many Pse5Ac7Ac derivatives vary from that of the parent molecule, would greatly benefit studies into the biological significance of naturally occurring  $\beta$ -Pse-glycosides.<sup>7</sup> However, a reoccurring theme in the Pse total synthesis literature, is that after the challenges of producing the Pse-skeleton have been overcome, the final deprotection steps are not reported.<sup>3,7</sup> Additionally, Pse5Ac7Ac was synthesised was synthesised from **1.47** in three steps, as mixture of  $\alpha$ - and  $\beta$ - anomers, in an overall yield of 2.6%.



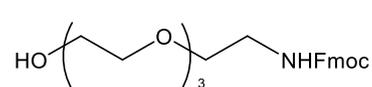
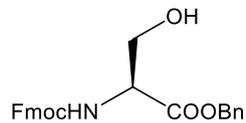
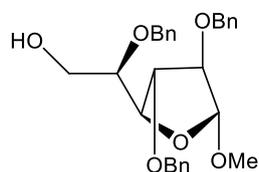
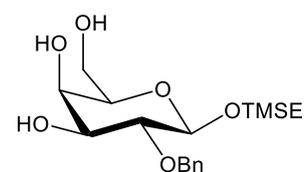
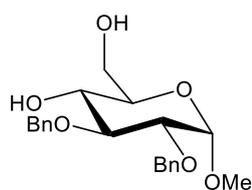
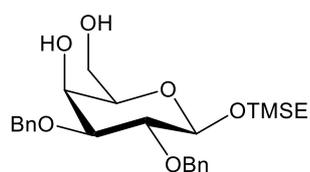
**Scheme 1.6: A: Synthesis of Pse-based glycosylation donor 1.43 and production of glycosides 1.44-1.47. B: Synthesis of Pse5Ac7Ac using 1.43. A and B described by Crich and co-workers.<sup>7</sup> Reagents and conditions: Reagents and conditions: A: i-iii: as reported<sup>114,115</sup>; iv: NaOMe; v: Me<sub>2</sub>C(OMe)<sub>2</sub>, CSA; vi: AcCl, py; vii: TBSOTf, Et<sub>3</sub>N; NaH, NaBr, DMF; viii: NOBF<sub>4</sub>, py; ix: TFA, H<sub>2</sub>O, TPSCl, Et<sub>3</sub>N, Bu<sub>2</sub>SnO; x: Tf<sub>2</sub>O, py; xi: Bu<sub>4</sub>NNO<sub>2</sub>, xii: NaI, Me<sub>2</sub>CO; H<sub>2</sub>, Pd/C, EtOAc, Et<sub>3</sub>N; xiii: Ac<sub>2</sub>O, DMAP; xiv: DDQ; xv: NH<sub>2</sub>NH<sub>2</sub>·H<sub>2</sub>O, AcOH, py; xvi: Tf<sub>2</sub>O, py; xvii: NaN<sub>3</sub>, DMF; xviii: 4Å AW molecular sieves, CH<sub>2</sub>Cl<sub>2</sub>:MeCN 2:1, NIS, TfOH. B: i: AcSH, py; ii: Ba(OH)<sub>2</sub>; iii: H<sub>2</sub>, Pd on C, H<sub>2</sub>O, dioxane.**

Most recently, Xuechen and co-workers expanded on their initial work on Pse synthesis, reporting a total synthesis strategy for stereocontrolled  $\alpha$ - and  $\beta$ -glycosylation using a single glycosyl donor **1.48**, over 16 steps (Scheme 1.7). The donor was prepared on gram scale and contains orthogonal N5 and N7 protecting groups, allowing for selective introduction of N-acyl groups at each position, an analogous feature to the donor prepared by Crich.<sup>7,9</sup> Acceptors were selected for screening based on the known structures of Pse-

containing oligosaccharides (**1.49-1.52**) and glycopeptides (**1.53**), or due to its potential for use in glycoconjugate preparation (**1.54**). Protected Pse-based glycosides **1.55-1.66** were produced with each acceptor in yields from 56-80%. Under different glycosylation conditions the donor showed predominantly  $\alpha$ - or  $\beta$ -selectivity, regardless of the chemistry of the acceptor. None of the glycosides were synthesised with complete anomeric selectivity however, the synthesis of these products is a great advance on previous efforts towards Pse-based biologically relevant glycosides. Efforts were made to produce a  $\beta$ -linked Pse-based trisaccharide **1.67**, which has the same the same skeleton as the repeating unit of LPS O-antigen from *Pseudomonas chlororaphis* UCM B-106. Whilst this is a great achievement, further deprotection and modification of N5 and N7 is required to yield the native LPS O-antigen structure.<sup>9</sup>



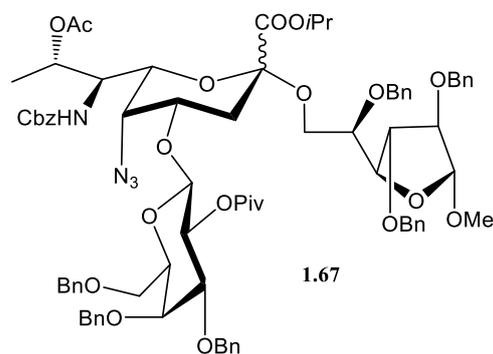
R:



1.52

1.53

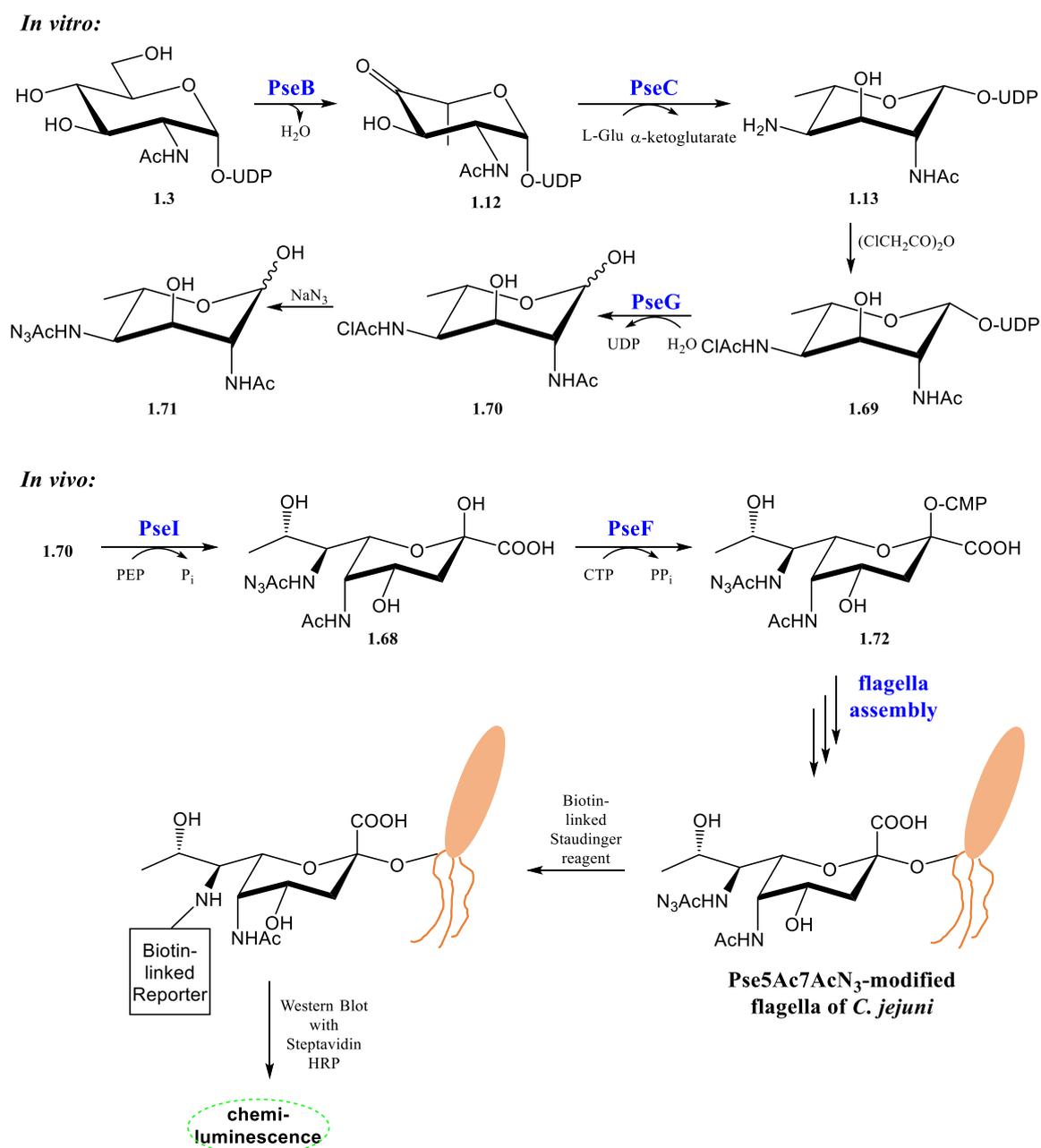
1.54



**Scheme 1.7: Synthesis of  $\alpha$ - and  $\beta$ -Pse-based glycosides using donor 1.48, with acceptors 1.49-1.67 and synthesis of 1.67 trisaccharide as reported by Xuechen and co-workers.<sup>9</sup> i-xi equivalent to i-xv in scheme 1.5; xvi: 3% HCl in MeOH, 0 °C to rt, 8 hr; xvii TfN<sub>3</sub>, 0.5 M Na<sub>2</sub>CO<sub>3</sub>, MeCN, 2 hrs; xviii: O<sub>3</sub>, DCM, -78 °C, 0.5 hr; then Me<sub>2</sub>S; xix: Ac<sub>2</sub>O, Py, DMAP, 0 °C to rt, 95%; xx: AdSH, BF<sub>3</sub>·OEt<sub>2</sub>, DCM, 16 hrs; xxi: NIS/TfOH (2.4/0.1 equiv.), AW-300 molecular sieves, DCM/MeCN (5 equiv.), -78 °C; xxii: NIS/TfOH (2.4/0.1 equiv.) AW-300 MS DCM/DMF (5 equiv.) -40 °C.**

#### 1.4.4 Chemoenzymatic synthesis of Pseudaminic Acids

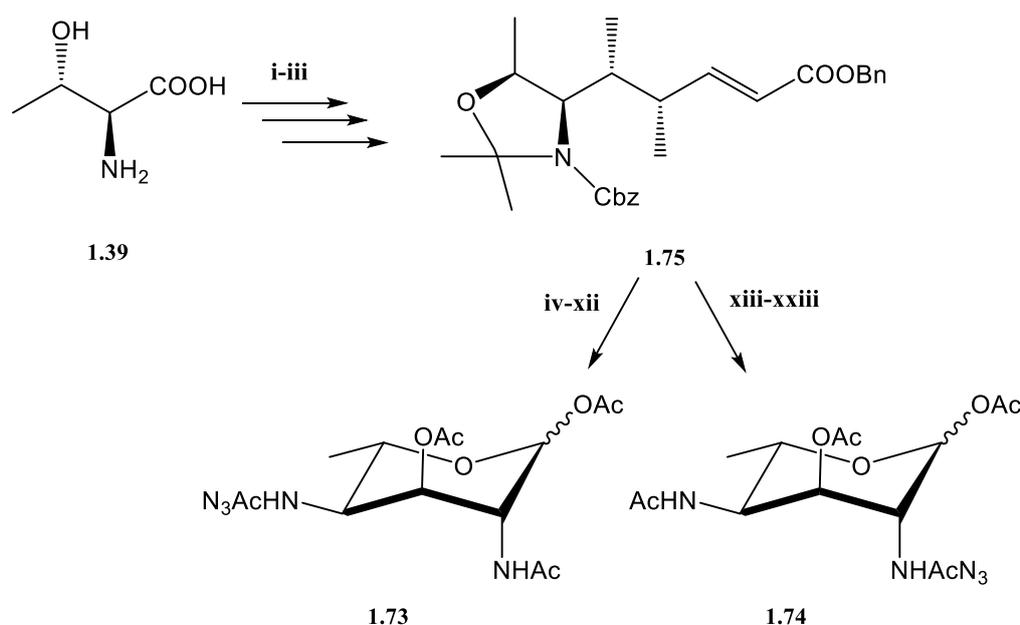
Notably, several of the total synthesis schemes involved installation of azido groups.<sup>5,7,113</sup> These functional groups may be used as bioorthogonal handles on Pse-derivatives, as demonstrated by the work of Tanner and colleagues who used a chemoenzymatic approach to azido-modify *C. jejuni* flagella with Pse5Ac7NHAcN<sub>3</sub> **1.68** (Scheme 1.8).<sup>116</sup> Enzymes PseB and PseC were used *in vitro* to produce **1.13** from UDP-GlcNAc **1.3**. **1.13** was chemically acetylated with chloroacetic anhydride, to produce PseH substrate analogue, UDP-2-acetamido-4-chloroacetamido-2,4,6-trideoxy- $\beta$ -L-Alt (UDP-6-deoxy-AltNAc4NAcCl) **1.69**. The UDP group was hydrolysed by PseG yielding **1.70** and treated with NaN<sub>3</sub> to yield 2-acetamido-4-azidoacetamido-2,4,6-trideoxy- $\beta$ -L-Alt (6-deoxy-AltNAc4NAcN<sub>3</sub> **1.71**) which *C. jejuni* then transported into their cells from the growth media. A biotin-linked Staudinger reagent was used to detect Pse5Ac7NHAcN<sub>3</sub> on the flagella of *C. jejuni*. This indicated that PseI produced **1.70**, PseF nucleotide-activated to yield **1.72**, that served as the donor for inverting glycosyltransferase to glycosylate flagellin subunits.<sup>116</sup>



**Scheme 1.8:** *In vitro* chemoenzymatic synthesis of 6-deoxy-AltNAc<sub>4</sub>NAcN<sub>3</sub> **1.71**, for *in vivo* production of Pse5Ac7AcN<sub>3</sub>-modified flagella of *C. jejuni*. Biotin-linked Staudinger reagent reacts with azido-moiety of Pse5Ac7AcN<sub>3</sub> **1.68** which is detected by Western Blot analysis with Streptavidin HRP. As reported by Tanner and co-workers.<sup>116</sup>

Xuechen and co-workers built upon methodology used in their 2017 total synthesis paper, by using *L*-allo-threonine **1.39** as the starting material for synthesis 1,3-di-O-acetyl-6-deoxy-AltNAc<sub>4</sub>NAcN<sub>3</sub> **1.73** and 1,3-di-O-acetyl-6-deoxy-AltNAc<sub>2</sub>NAcN<sub>3</sub> **1.74**, via intermediate **1.75**, with the aim of producing Pse5Ac7AcN<sub>3</sub> **1.77** and Pse5AcN<sub>3</sub>7Ac **1.78** *in vivo* (Scheme 1.9). Rich growth media for *P. aeruginosa*, *A. baumannii* and *Vibrio vulnificus* was supplemented with **1.73** and **1.74**. Following a 48 hr incubation period cells underwent a

Cu-free click reaction to fluorescently label any resulting cell-surface azido-modified Pse, via the azide moiety of Pse. Studies confirmed that within some of the *P. aeruginosa* studied, the LPS had been labelled, indicating that a Pse-derivative was present in the O-antigen.<sup>8</sup> Further examination confirmed that only Pse5Ac7AcN<sub>3</sub> **1.70** the downstream product of **1.73**, could be labelled in all organisms tested. This suggests that within the *P. aeruginosa*, *A. baumannii* and *V. vulnificus* strains and serotypes tested, an enzyme downstream of PseG is unable to utilise **1.74**. In the case of *P. aeruginosa* 1244 it was suggested that it may be Psel which cannot turnover **1.74** as the native substrate contains an N-hydroxybutanol group at C2. However, further studies are required to find the enzymatic bottleneck in each of the organisms studied.<sup>8</sup> This demonstrates that whilst metabolic engineering can provide valuable information in the biological relevance of Pse, the strategy is not infallible.



**Scheme 1.9: Chemical synthesis of 1.73 and 1.74, from L-allo-threonine 1.39. Synthesised with the intended use in *in vitro* synthesis of Pse5AcN<sub>3</sub>7Ac and Pse5Ac7AcN<sub>3</sub>, as reported by Xuechen and colleagues.<sup>8</sup> Reagents and conditions: i: equivalent to steps i-xi in Scheme 1.X; ii: Pd/C, Et<sub>3</sub>SiH, THF; iii: Ph<sub>3</sub>P=CHCOOBn, DCM; iv: TFA/H<sub>2</sub>O 10/1, 0 °C, 30 min; v: 3% HCl (aq) in MeOH, 0 °C to r.t., 8 hrs; vi: Ac<sub>2</sub>O, Na<sub>2</sub>CO<sub>3</sub>, MeCN/H<sub>2</sub>O; vii: O<sub>3</sub>, DCM, -78 °C, then Me<sub>2</sub>S; viii: TBSOTf, 2,6-lutidine, DCM, ix: Pd/C, H<sub>2</sub>, MeOH; x: 2-azidoacetic acid, HATU, *i*Pr<sub>2</sub>NEt, DMF; xi: TBAF, HOAc, THF; xii: Ac<sub>2</sub>O, pyridine, DMAP; xiii: TFA/H<sub>2</sub>O 10/1, 0 °C, 30 min; xiv: 3% HCl (aq) in MeOH, 0 °C to r.t., 8 hrs; xv: Ac<sub>2</sub>O, Na<sub>2</sub>CO<sub>3</sub>, MeCN/H<sub>2</sub>O; xvi: FmocCl, *i*Pr<sub>2</sub>NEt, DCM; xvii: O<sub>3</sub>, DCM, -78 °C, then Me<sub>2</sub>S; xviii: TBSOTf, 2,6-lutidine, DCM; xix Pd/C, H<sub>2</sub>, EtOAc; xx: Ac<sub>2</sub>O, pyridine; Et<sub>2</sub>NH, DCM, DMF; xxi: 2-azidoacetic acid, HATU, *i*Pr<sub>2</sub>NEt, DMF; xxii: TBAF, HOAc, THF; xxiii: Ac<sub>2</sub>O, pyridine, DMAP.**

The work of both Tanner *et al.* and Xuechen *et al.* demonstrates the benefits of utilising enzymes in synthesis due to their inherent regio- and stereoselectivity, factors which are challenging to control using traditional chemical methods for carbohydrate synthesis.<sup>8,116</sup> It also demonstrates that a degree of enzyme promiscuity may be exploited, allowing substrate analogues to be converted in to desired products. A caveat to this methodology was made apparent when results suggested that **1.74** was not utilised *in vivo* to produce the desired Pse5AcN<sub>3</sub>7Ac-containing glycoconjugates.<sup>8</sup> This work may have benefitted from the precautionary step taking by Tanner *et al.* where the promiscuity of enzymes was tested *in vitro* with substrate analogues before the analogues were used for *in vitro* experiments.<sup>116</sup> However, both studies have shown that chemoenzymatic synthesis has the potential to access products ranging from naturally occurring Pse5Ac7Ac derivatives, to Pse-derivative chemical probes, which may be used to further the biological relevance of Pse and the mechanisms used by Pse processing enzymes.<sup>8,116</sup>

#### 1.4.4 Introduction to Pseudaminic Acid Processing Enzymes

There is currently very limited knowledge of GTs which utilise CMP-Pse (Pseudaminyltransferases, PseTs).<sup>117</sup> A few genes had putatively been assigned as PseTs based upon sequence homology to SiaTs, or other structurally related  $\alpha$ -keto sugars. However, none of the proteins encoded by the putative PseTs had been functionally characterised *in vitro*. Pse-glycoconjugates can be found in both  $\alpha$  and  $\beta$  linkages, as opposed to sialic acid glycoconjugates which exclusively contain  $\alpha$ -anomers, due to the inverting mechanism of SiaTs.<sup>55</sup> Therefore, unlike SiaTs, both inverting and retaining PseTs exist, which utilise CMP-Pse5Ac7Ac **1.10** or a similar CMP-Pse-derivative.<sup>1,55</sup>

In 2018 the first crystal structure of a putative PseT, Maf, from *Magnetospirillum magneticum* AMB-1 was solved in an apo form.<sup>117</sup> Whilst the recombinant expression, purification and successful crystallisation of this enzyme, from a potentially novel GT class, was an impressive feat, the identity of the Pse-derivative from *M. magneticum* AMB-1 could not be determined. Additionally, lack of access to CMP-Pse and Pse-derivatives precluded functional studies of Maf required to confirm its identity as a PseT.<sup>117</sup>

An additional class of Pse processing enzymes that are yet to be studied are glycosylhydrolases. In general, this class of enzyme hydrolyses glycosidic bonds from

glycoconjugates, to yield free carbohydrates. To date, the sole example of a putative Pse GH (pseudaminidase) is PA2794 from *P. aeruginosa*.<sup>118</sup> Studies showed that in mice infected with pa2794 knockout *P. aeruginosa* PA01 ( $\Delta$ 2794), there was a decrease in biofilm formation, in comparison to infection with wild type *P. aeruginosa* PA01 and that  $\Delta$ 2794 strains were unable to colonize the mouse respiratory tract.<sup>119</sup> PA2794 was originally assigned as a neuraminidase, however, *in vitro* studies showed that the enzyme did not cleave Neu5Ac.<sup>118</sup> The crystallographic structure suggest that the active site is unable to accept Neu5Ac as a substrate and *in-silico* docking of Pse5Ac7Ac suggested that this may be the product of the hydrolysis reaction. However, lack of access to a Pse5Ac7Ac-based probe precluded further functional studies of PA2794.<sup>118</sup>

*P. aeruginosa* is the most prevalent cause of lung infection in patients with cystic fibrosis, where PA2794 is highly overexpressed.<sup>119</sup> The role of PA2794 in biofilm formation makes it an attractive therapeutic target. Additionally, if PA2794 is shown to utilise Pse substrates then this enhances its suitability as a therapeutic target as Pse-based inhibitors could be designed which could avoid off-target effects on human neuraminidases.<sup>118</sup>

The biological mechanism behind the increased biofilm formation and ability of *P. aeruginosa* strains expressing *pa2794* to colonize a host remains curious. It is currently unclear where the proposed PA2794 Pse5Ac7Ac-substrate is found. Whilst the full composition of PA01 biofilm is unknown, Pse5Ac7Ac is known to be present in the glycoconjugates of many *P. aeruginosa* strains.<sup>120,121</sup> Based upon knowledge of sialic acid scavenging by bacterial pathogens, it may be speculated that following hydrolysis of Pse5Ac7Ac, *P. aeruginosa* could utilise Pse5Ac7Ac as either a carbon source, or to build Pse5Ac7Ac-based glycoconjugates. As with sialic acid scavenging, this would require PA2794 to contain a transporter, currently no putative Pse transporters have been identified.<sup>69</sup> A desire to examine genomes for putative pseudaminidases, and potential transporters may be initiated by improved access to Pse-based molecules.

## 1.5 Project Outline

This project aimed to characterise enzymes which process pseudaminic acid and to chemoenzymatically synthesise glycosides which contain pseudaminic acid.

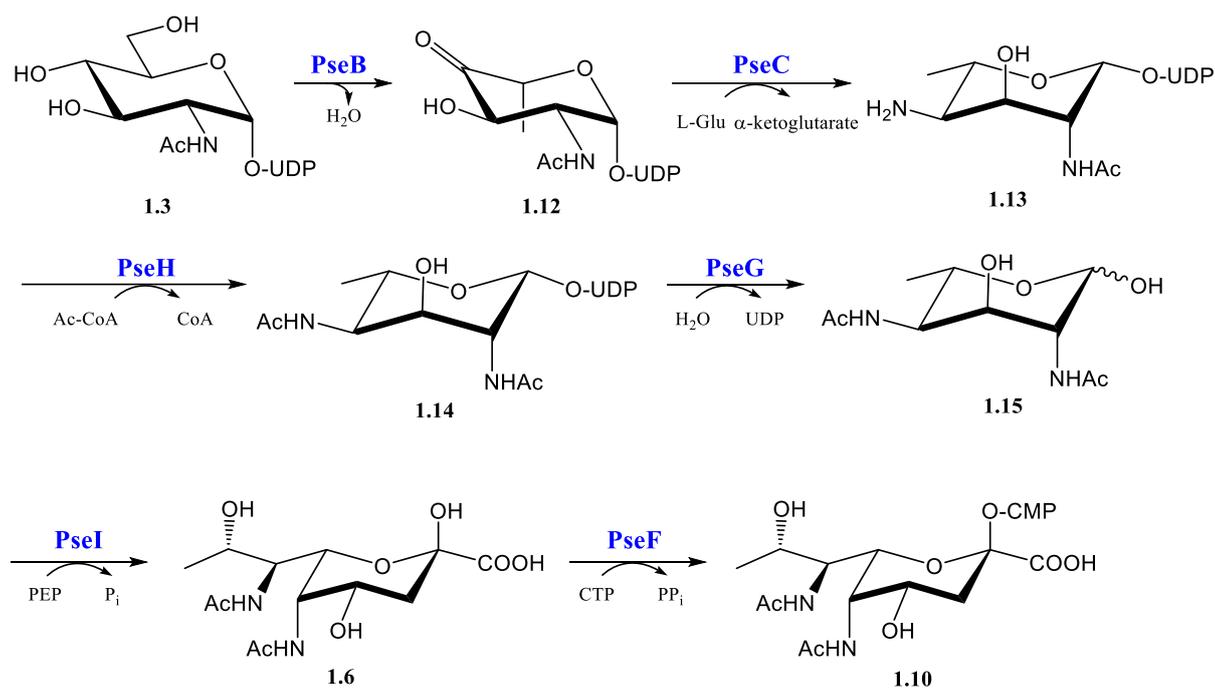
This was achieved through three specific aims:

1. Chemoenzymatic synthesis, purification and characterisation of nucleotide-activated pseudaminic acid, CMP-Pse5Ac7Ac (described in Chapter 2).
2. Synthesis of glycosides containing Pse5Ac7Ac, mediated by promiscuous sialyltransferases (described in Chapter 3).
3. *In vitro* functional characterisation of putative glycosyltransferases for pseudaminic acid (pseudaminyltransferases) (described in Chapter 4).

# Chapter 2: Chemoenzymatic Synthesis of CMP-Pseudaminic Acid

## 2.1 Introduction

In 2006 Schoenhofen *et al.* reported the first enzymatic synthesis of  $\alpha$ -cytidine-5'-monophosphate-Pse5Ac7Ac (CMP-Pse5Ac7Ac **1.10**) facilitated by elucidation of the enzymatic pathway found in within *Helicobacter pylori* and *Campylobacter jejuni* (Figure 2.1).<sup>99</sup> This pathway begins with uridine diphosphate-N-acetyl-glucosamine (UDP-GlcNAc **1.3**) and the first of six enzymes, dehydratase/epimerase PseB, producing UDP-2-acetamido-2,6-dideoxy- $\beta$ -L-*arabino*-hexos-4-ulose **1.12**. **1.12** is the substrate for the L-glutamine dependent aminotransferase PseC, which produces UDP-4-amino-4,6-dideoxy- $\beta$ -L-AltNAc **1.13**.<sup>103</sup> N-acetyltransferase, PseH, then uses **1.13** and acetyl-coenzyme A (Ac-CoA) to yield UDP-2,4-diacetamido-2,4,6-trideoxy- $\beta$ -L-altropyranose **1.14**, which undergoes hydrolysis of UDP, facilitated by nucleotidase PseG, producing 2,4-diacetamido-2,4,6-trideoxy-L-altropyranose (6-deoxy-AltDiNAc **1.15**). Pseudaminic acid synthase, PseI, utilises phosphoenol pyruvate (PEP) to convert **1.15** to the nonulosonic acid Pse5Ac7Ac **1.6**.<sup>103</sup> This condensation reaction between a hexose and PEP is analogous to that seen in the biosynthesis of Neu5Ac.<sup>58</sup> Finally, CMP-Pse5Ac7Ac synthetase, PseF, nucleotide activates Pse5Ac7Ac using cytidine-5'-triphosphate (CTP) to produce CMP-Pse5Ac7Ac **1.10**. **1.10** was found to be a competitive inhibitor of the first enzyme in the pathway, PseB.<sup>103</sup> Therefore, whilst the enzymatic synthesis may be performed in an elegant one-pot system where purification of reaction intermediates is not required, PseF must not be added to the reaction mixture until sufficient conversion of UDP-GlcNAc to **1.12** has been achieved.<sup>103</sup>



**Figure 2.1: CMP-Pse5Ac7Ac biosynthesis pathway in *C. jejuni* and *H. pylori*, starting with UDP-GlcNAc, using enzymes PseB, PseC, PseH, PseG, PseI and PseF.**

Whilst impressive, the enzymatic synthesis of CMP-Pse5Ac7Ac described by Schoenhofen *et al.* was not conducive to large-scale (~100mg) production. Firstly, as is true for many nucleotide activated carbohydrates, the starting material UDP-GlcNAc is costly (£913 per gram, Sigma). Secondly, the third step of the pathway performed by acetyltransferase PseH, requires 1.5 equivalents of co-factor Ac-CoA. The quantity of acetyl-CoA required for a reaction with 1g UDP-GlcNAc currently costs £8850 (Sigma). Therefore, there was a clear need to optimise this enzymatic synthesis in order to reduce cost.

These enzymes are also potential therapeutic targets as non-eukaryotic Pse5Ac7Ac and derivatives are key virulence factors in several multi-drug resistant bacterial pathogens. The strong negative feedback inhibition of PseB by CMP-Pse5Ac7Ac, indicates concentrations of that CMP-Pse5Ac7Ac are regulated within bacteria, highlighting the potential of these enzymes for therapeutic intervention.<sup>122</sup>

Access to Pse5Ac7Ac biosynthesis enzymes has enabled studies into their inhibition.<sup>103,110</sup> Over 96,000 compounds were screened against Pse5Ac7Ac biosynthesis enzymes from *H. pylori* and *C. jejuni*. The study found 320 compound which each inhibited one or more enzymes of the pathway. Five compounds showed concentration-dependent inhibition of



appreciable quantities of **1.6** and **1.10**. Enzymes that process CMP-Pse5Ac7Ac such as Leloir glycosyltransferases, may also be novel therapeutic targets. Therefore, for these enzymes to be studied as novel drug targets, their function must first be unequivocally characterised. These functional studies are reliant on a source of CMP-Pse5Ac7Ac. Structural studies of these enzymes may also require Pse5Ac7Ac and Pse-based probes.

The overall aim of this chapter was the large-scale production of CMP-Pse5Ac7Ac. Firstly, this required the purification of five recombinant enzymes, (PseB, PseC, PseH, PseG and PseI) for use in a reaction producing Pse5Ac7Ac.

Secondly, recombinant PseF was required to convert Pse5Ac7Ac to CMP-Pse5Ac7Ac. This was achieved using *Aeromonas caviae* PseF which was uncharacterised at the commencement of this study, therefore at this point, biochemical characterisation of the enzyme was undertaken. The reaction to produce CMP-Pse5Ac7Ac was scaled to allow purification and characterisation of **1.10**. The production of **1.10** facilitated the studies described in Chapter 3 and 4.

Additionally, small-molecules were screened as potential inhibitors of PseB and PseI. The screening of compounds as potential PseB inhibitors was reliant upon access to active PseB and commercially available UDP-GlcNAc. The screening of compounds as potential PseI inhibitors, active PseI and its substrate, 6-deoxy-AltdiNAc **1.15** was required, which was produced enzymatically during this study.

## 2.2 Experimental

### 2.2.1 Overexpression and Purification of PseB, PseC, PseH, PseG and PseI

*E. coli* BL21 (DE3) cells were transformed via electroporation with plasmid encoding the *H. pylori* gene for either PseB, PseC, PseH, PseG or PseI (Table 2.1) as follows. 50  $\mu$ L electrocompetent cells thawed on ice, to which 20-50 ng plasmid was added and stored on ice for 2 minutes, cells were placed in an ice cold electroporation cuvette and electroporated (Biorad GENEPULSER II), 1 mL SOC media was added and cells placed in a 15 mL tube then incubated at 37 °C, 180 revolutions per minute (r.p.m.) then plated on LB<sub>kan</sub> agar and incubated at 37 °C for 18 hours.

**Table 2.1: Pse5Ac7Ac biosynthesis enzyme plasmid details and conditions for overexpression of recombinant *H. pylori* genes in *E. coli***

| Enzyme | Vector  | Antibiotic resistance | IPTG concentration / mM | Post-induction incubation temperature | Post-induction incubation time |
|--------|---------|-----------------------|-------------------------|---------------------------------------|--------------------------------|
| PseB   | pET30   | Kan                   | 0.1                     | 37 °C                                 | 4 hours                        |
| PseC   | pET-15b | Amp                   | 0.1                     | 16 °C                                 | 18 hours                       |
| PseH   | pET-15b | Amp                   | 0.1                     | 16 °C                                 | 18 hours                       |
| PseG   | pET-15b | Amp                   | 0.5                     | 37 °C                                 | 4 hours                        |
| PseI   | pET-15b | Amp                   | 0.5                     | 37 °C                                 | 18 hours                       |

Cells were plated on LB agar containing either 50  $\mu$ g mL<sup>-1</sup> kanamycin or 100  $\mu$ g mL<sup>-1</sup> ampicillin. For the expression of each plasmid 50 mL LB (in 250 mL Erlenmeyer flask) supplemented with 50  $\mu$ g mL<sup>-1</sup> kanamycin (LB<sub>kan</sub>) or 100  $\mu$ g mL<sup>-1</sup> ampicillin (LB<sub>amp</sub>) was inoculated with a single colony of transformed cells and was cultured overnight at 37 °C, 180 r.p.m. LB<sub>kan</sub> or LB<sub>amp</sub> (1 L in 2L baffled Erlenmeyer flask) was inoculated with a 1 in 100 dilution of overnight culture and incubated at 37 °C, 180 r.p.m. OD<sub>600</sub> readings of the cultures were taken using a spectrophotometer (DeNovix) and overexpression of genes was induced with IPTG once the desired OD<sub>600</sub> value had been reached. Cells were then

incubated at 16 or 37 °C, 180 r.p.m. for 4 or 18 hours (Table 2.1). Cell pellets were harvested via centrifugation (6000 ×g, 30 min, 4 °C) and stored at -80 °C.

Cell pellets from 2 L cultures lysed on ice resuspended in 50 mL lysis buffer (50 mM sodium phosphate, 400 mM NaCl, 10 mM β-mercaptoethanol, pH 7.4) supplemented with protease inhibitor, lysozyme, and Benzonase. Cells were lysed via sonication in ice (30 seconds, 30 second pause, 25 repeats) and centrifuged (37000 ×g, 30 min, 4 °C). Supernatant was applied to a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with lysis buffer, the column was washed with 10 column volumes (C.V.), lysis buffer and protein was eluted with a gradient of elution buffer (lysis buffer + 490 mM imidazole, pH 7.4) over 40 C.V. as 4 mL fractions were collected. 15 μL samples of insoluble material, soluble material, flowthrough wash and fractions were analysed via SDS-PAGE.

SDS-PAGE analysis was performed using 10 or 12% polyacrylamide gels. Samples were mixed with 5x reducing buffer (2% SDS, 2 mM 2-mercaptoethanol, 4% glycerol, 40 mM Tris-HCl pH 6.8, 0.01% bromophenol blue) and were heated to 95 °C for 4 minutes. 6 μL PageRuler Plus Prestained Protein Ladder (Thermo Scientific) molecular weight marker was used per gel. Gels were run at 200 volts for 45-75 min. For Coomassie stain experiments, the gel was washed with fixing solution (40% (v/v) MeOH, 10% AcOH), stained with 0.1% (w/v) Coomassie Brilliant Blue R-250 (50% (v/v) MeOH, 10% AcOH), and repeatedly washed with destain solution (50% (v/v) MeOH, 10% AcOH).

Following SDS-PAGE analysis fractions containing the desired protein were pooled and applied to desalting column (HiPrep 26/10 desalting column, GE Healthcare) to exchange into reaction buffer (50 mM sodium phosphate, 25 mM NaCl, pH 7.4). Protein was collected as it eluted from the column and concentration of protein was recorded using a nanodrop spectrophotometer. Protein was aliquoted in 500 μL aliquots, glycerol was added to a final concentration of 25% (v/v) and samples were flash frozen and stored at – 80 °C. In a later optimisation of this protocol, no glycerol was added prior to flash freezing the enzymes.

## 2.2.2 Enzymatic synthesis of Pse5Ac7Ac with Acetyl-thiocholine iodide

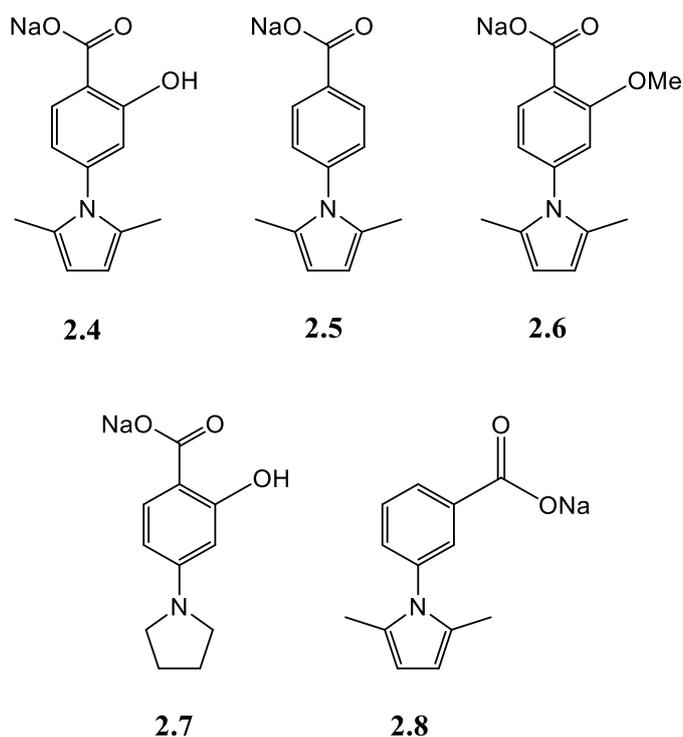
1 mM UDP-GlcNAc **1.3** (200 mg), 0.15 mM coenzyme-A, 2 mM pyridoxal 5'-phosphate, 10 mM L-glutamic acid, 1.5 mM phosphoenolpyruvate, 60 mM acetylthiocholine iodide, 0.1 mg mL<sup>-1</sup> PseB, 0.1 mg mL<sup>-1</sup> PseC, 0.1 mg mL<sup>-1</sup> PseH, 0.1 mg mL<sup>-1</sup> PseG and 0.1 mg mL<sup>-1</sup> PseI (enzymes purified as described in 2.3.1, containing glycerol), 50 mM 25 mM sodium phosphate pH 7.4 in 33 10 mL aliquots. The reaction mixtures were incubated at 37 °C for 6 hours and were analysed via –ESI LC-MS, to monitor the production of Pse5Ac7Ac. Reactions were pooled, enzymes were removed from the reaction mixtures using molecular weight cut off filters (MWCO) (30 kDa and then the subsequent filtrate was passed through a 10 kDa MWCO). The filtrate was lyophilised and resuspended in 50 mL dH<sub>2</sub>O and 5 mL was applied to a 500 mL column packed with LH20 resin (GE Healthcare) in HPLC-grade H<sub>2</sub>O. 3 mL fractions were collected for 40 hours and analysed via LC-MS for the presence of Pse5Ac7Ac. The column was repeated with the remaining lyophilised material and all fractions containing Pse5Ac7Ac were pooled and lyophilised and analysed by NMR (D<sub>2</sub>O, 500MHz). The resulting material was purified by anion exchange (Dowex), with a 0-2 M gradient of formic acid. TLC plates were used to analyse fractions for the presence of carbohydrates (TLC plate soaked in sugar stain: MeOH, 5% H<sub>2</sub>SO<sub>4</sub> and charring) and fractions which appeared to contain carbohydrate was analysed by LC-MS (details in Appendix-General methods.) Fractions containing Pse5Ac7Ac were pooled, lyophilised and analysed by NMR (D<sub>2</sub>O, 500MHz).

## 2.2.3 Small-molecule screen for potential inhibition of PseB and PseI

A library of small-molecules which were synthesised as potential inhibitors of PseB and PseI were made available to this study (Figure 2.3).<sup>‡</sup>

---

<sup>‡</sup> Small-molecule library was synthesised by Joseph Ferner, Shaw Laboratory, University of Sheffield.



**Figure 2.3: Compounds 2.4-2.8 tested for inhibition of *C. jejuni* or *A. caviae* PseI. 2.4 was tested for inhibition of *C. jejuni* PseB.**

To test whether **2.4** is an inhibitor of *H. pylori* PseB the following reaction mixtures containing 0.1 mg mL<sup>-1</sup> PseB, 0.1 mg mL<sup>-1</sup> PseH, 0.1 mg mL<sup>-1</sup> PseG, 0.1 mg mL<sup>-1</sup> PseI (all purified as described in 2.2.1) and 0.2 mg mL<sup>-1</sup> AcPseC, in addition to 1 mM UDP-GlcNAcp **1.3**, 1.5 mM coenzyme-A, 2mM pyridoxal 5'-phosphate, 10 mM L-glutamic acid, 1.5 mM PLP, 60 mM acetylthiocholine iodide, 1 mM PEP, in plus 0 mM 0.1 mM, or 5 mM **2.4** in 50 mM 25 mM sodium phosphate pH 7.4 to a final volume of 50  $\mu$ L. The reactions were performed in duplicate and incubated overnight at 30 °C. Enzyme precipitate was removed by centrifugation (37,000  $\times$ g, 10 min, 4 °C) and supernatant was analysed by -ESI LC-MS, as described in 2.2.2.

To test inhibition of PseI and *Aeromonas caviae* PseI (AcPseI) the substrate 6-deoxyAltDiNAc **1.15** was produced via chemoenzymatic synthesis. To facilitate this the activity of pure, recombinantly expressed *Aeromonas caviae* PseC (AcPseC)<sup>§</sup> was used in combination with *C. jejuni* enzymes PseB, PseH, and PseG, expressed and purified as

<sup>§</sup> AcPseC and AcPseI were expressed and purified by by Joseph Ferner, The University of Sheffield.

described in 2.2.1. A reaction mixture containing 0.114 mg mL<sup>-1</sup> PseB, 0.114 mg mL<sup>-1</sup> PseH, 0.114 mg mL<sup>-1</sup> PseG (all purified as described in 2.2.1) and 0.228 mg mL<sup>-1</sup> AcPseC, in addition to 1.14 mM UDP-GlcNAc **1.3** (2.4 mg), 1.71 mM coenzyme-A, 2.228 mM pyridoxal 5'-phosphate, 11.1 mM L-glutamic acid, 1.71 mM PLP, 68.4 mM acetylthiocholine iodide, 50 mM 25 mM sodium phosphate pH 7.4. The reaction mixture incubated overnight at 30 °C. Enzyme precipitate was removed by centrifugation (37,000 ×g, 10 min, 4 °C). To 87.5 μL of supernatant PseI or AcPseI, PEP and dH<sub>2</sub>O were added to give a final sample volume of 100 μL (such that final concentrations would be equivalent to those in 2.2.2 after PEP and PseI or AcPseI were added) and the samples incubated at 30 °C and incubated for 2 hours. The sample was analysed by LC-MS (details in Appendix-General methods) to check for the presence of Pse5Ac7Ac, to confirm that the initial reaction had produced 6-deoxyAltDiNAc **1.15**.

To test PseI inhibition, 87.5 μL of supernatant, PseI, PEP and an inhibitor compound **2.4-2.8** were added to give a final sample volume of 100 μL (such that final concentrations would be equivalent to those in 2.2.2 after PseI and PEP were added) and inhibitor concentrations of 0 mM, 0.1 mM, 1 mM and 5 mM. The screen was repeated with AcPseI in place of PseI. Reactions were performed in duplicate and incubated at 30 °C and incubated for overnight. The sample was analysed by LC-MS to check for the presence of Pse5Ac7Ac, to assess inhibition of PseI.

#### **2.2.4 Recombinant expression and purification attempt of *C. jejuni* PseF in *E. coli***

Plasmid encoding N-terminal hexahistidine tagged *C jejuni* PseF, codon optimised for *E. coli* (Genscript), in a pET-15b vector (CjPseF) (plasmid details in Appendix) was transformed into *E. coli* BL21 (DE3) cells (via electroporation- as described in 2.2.1). A single colony was used to inoculate 60 mL 2 x yeast tryptone, supplemented with 50 μg mL<sup>-1</sup> kanamycin (2×YT<sub>kan</sub>) that was incubated at 30 °C, 180 rpm, overnight, before being used to inoculate 4 L of 2×YT<sub>kan</sub> (six 2 L baffled Erlenmeyer flasks each contained 1 L LB<sub>kan</sub>). The cultures were incubated at 37 °C, 180 rpm. At OD<sub>600</sub> 0.6, 0.1 mM Isopropyl-β-D 1-thiogalactopyranoside (IPTG) was added to induce expression of CjPseF. Cultures were incubated for 2.75 h post-induction and cells were harvested via centrifugation (6,000 ×g, 30 minutes, 4 °C). Two cell

pellets, each harvested from 2 L culture, were stored at -80 °C. A single cell pellet, corresponding to 2 L culture was thawed on ice. Cells were resuspended in 20 mL lysis buffer (50 mM sodium phosphate, pH 7.3, 400 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 10 mM imidazole, 1 mM MgCl<sub>2</sub>) supplemented with protease inhibitor (1 tablet per 50 mL, Pierce) and Benzonase nuclease (25 units mL<sup>-1</sup>), before being lysed on ice via sonication (30 seconds, 30 second pause, 12 repeats). Lysed cells were centrifuged (17,700  $\times$ g, 35 min, 4 °C) and the resultant supernatant was retained for purification. The supernatant was applied to a 5 mL HisTrap HP column (GE Healthcare), prewashed with lysis buffer. Unbound material was washed through the column with 10 column volumes (C.V.) of lysis buffer. In order to elute bound material from the column, 30 C.V. of 0-50% buffer B (50 mM sodium phosphate, pH 7.3, 400 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 500 mM imidazole) was mixed with lysis buffer, in a linear gradient, followed by 10 C.V. of buffer B. The eluent was collected as 3 mL fractions, which were analysed by 12 % SDS-PAGE (as described in 2.2.1).

Following unsuccessful purification of PseF, due to protein insolubility, expression trials were carried out to find conditions that produced soluble CjPseF. CjPseF plasmid was transformed into *E. coli* Tuner (DE3) cells that were then used in expression trials alongside *E. coli* BL21 (DE3) encoding PseF plasmid. Overnight cultures of both cell strains in 2 $\times$ YT<sub>kan</sub> were incubated at 37 °C, 180 rpm. Each overnight culture was used to inoculate seven 250 mL Erlenmyer flasks, each containing 50 mL 2 $\times$ YT<sub>kan</sub> to give an initial OD<sub>600</sub> of 0.02. Additionally, as a control a pET-15b vector (vector for the CjPseF gene) was cultured in the same conditions. All cultures were incubated at 37 °C, 180 rpm, until expression was induced, at OD<sub>600</sub> measurements of 0.4, 0.6 and 0.8, using either 0.1 mM or 0.5 mM IPTG for each of the three OD<sub>600</sub> values. At this point flasks were incubated at 30 °C, 180 rpm. Three 1 mL samples were taken from each flask at 1 or 2, 4 and 24 hrs post-induction. An uninduced sample was also cultured as a control. The samples were centrifuged (6,000  $\times$ g, 10 minutes, 4 °C) and cells pellets were stored at -20 °C. This process was repeated, with induced cells incubated at 25 °C and 16 °C. Cell pellets were resuspended in 50  $\mu$ L lysis buffer supplemented with 1 mg mL<sup>-1</sup> lysozyme and incubated at 37 °C for 45 minutes to lyse cells. Pellets were centrifuged (6,000  $\times$ g, 10 minutes, 4 °C), following this soluble and insoluble material was separated. The insoluble material was resuspended in 50  $\mu$ L dH<sub>2</sub>O

and 15  $\mu\text{L}$  of both soluble and insoluble fractions, from each sample, were mixed with 3  $\mu\text{L}$  5 $\times$  reducing dye heated to 95  $^{\circ}\text{C}$  for 3 minutes and analysed by 12% SDS-PAGE (as described in 2.2.1).

### 2.2.5 Expression and purification of *A. caviae* PseF

*E. coli* BL21 (DE3) were transformed via electroporation (as described in 2.2.1) with plasmid encoding N-terminal hexahistidine tagged PseF (PseF) from *A. caviae* (pET-28a vector) (PseF)\*\* as follows. Then plated on LB<sub>kan</sub> agar and incubated at 37  $^{\circ}\text{C}$  for 18 hours. A single colony of *E. coli* BL21 (DE3) of transformed cells was used to inoculate 60 mL LB<sub>kan</sub>, that was incubated at 30  $^{\circ}\text{C}$ , 180 rpm, overnight. The cultured cells were used to inoculate seven 250 mL Erlenmyer flasks, each containing 50 mL 2 $\times$  YT<sub>kan</sub> to give an initial OD<sub>600</sub> 0.02. All cultures were incubated at 37  $^{\circ}\text{C}$ , 180 rpm, until expression was induced, at OD<sub>600</sub> measurements of 0.3, 0.4, 0.6 and 0.8, using either 0.1 mM, 0.5 mM or 1 mM IPTG for each of the four OD<sub>600</sub> values. At this point flasks were incubated at 30  $^{\circ}\text{C}$ , 180 rpm. Three 1 mL samples were taken from each flask at 2, 4 and 24 hrs post-induction. The samples were centrifuged (6,000  $\times$  g, 10 minutes, 4  $^{\circ}\text{C}$ ) and cells pellets were stored at -20  $^{\circ}\text{C}$ . This process was repeated, with induced cells incubated at 30  $^{\circ}\text{C}$ , 25  $^{\circ}\text{C}$  and 16  $^{\circ}\text{C}$ . Cell pellets were resuspended in 50  $\mu\text{L}$  Bugbuster supplemented with protease inhibitor (1 tablet per 50 mL, Pierce) plus 1 mg mL<sup>-1</sup> lysozyme and incubated at room temperature for 45 minutes to lyse cells. Pellets were centrifuged (6,000  $\times$ g, 10 minutes, 4  $^{\circ}\text{C}$ ), following this soluble and insoluble material was separated. The insoluble material was resuspended in 50  $\mu\text{L}$  dH<sub>2</sub>O and 15  $\mu\text{L}$  of both soluble and insoluble fractions from each sample were analysed via SDS-PAGE (as described in 2.2.3 and 2.2.1).

A single colony of *E. coli* BL21 (DE3) of transformed cells encoding PseF was used to inoculate 60 mL LB<sub>kan</sub>, that was incubated at 30  $^{\circ}\text{C}$ , 180 rpm, overnight, before being used to inoculate 3 L of LB<sub>kan</sub> (five 2 L baffled Erlenmyer flasks each contained 600 mL LB<sub>kan</sub>). The cultures were incubated at 37  $^{\circ}\text{C}$ , 180 rpm. At OD<sub>600</sub> 0.6, 0.1 mM IPTG was added to induce expression of PseF. Cultures were incubated 30  $^{\circ}\text{C}$ , 180 rpm for 3 hrs post-induction and

---

\*\* Plasmid was gifted by Dr Jonathon Shaw, The University of Sheffield

cells were harvested via centrifugation (6,000 × g, 30 minutes, 4 °C). Two cell pellets were obtained and stored at -80 °C.

Cell pellets were thawed on-ice and resuspended in 50 mL lysis buffer (50 mM sodium phosphate, pH 7.4, 400 mM NaCl, 10 mM β-mercaptoethanol, 1 mM MgCl<sub>2</sub>, 10 mM imidazole and Benzonase nuclease (25 units mL<sup>-1</sup>) (Sigma-Aldrich), before being lysed on ice via sonication (30 seconds, 30 second pause, 10 repeats). Lysed cells were centrifuged (17,700 ×g, 40 min, 4 °C) and the resultant supernatant was retained for purification. The supernatant was filtered (0.45 μm syringe filter) and applied to a 5 mL HisTrap HP column, prewashed with lysis buffer. Unbound material was washed through the column with 10 column volumes (C.V.) of lysis buffer. Bound material was eluted from the column, with a linear gradient of 0-100% elution buffer (50 mM sodium phosphate, pH 7.3, 400 mM NaCl, 10 mM β-mercaptoethanol, 1 mM MgCl<sub>2</sub>, 500 mM imidazole) over 40 C.V. The eluent was collected as 5 mL fractions, which were analysed by 12 % SDS-PAGE. Fractions containing pure PseF were pooled and dialysed (50 mM sodium phosphate, 50 mM NaCl, 1 mM MgCl<sub>2</sub>) overnight at 4 °C. PseF was then either used in enzyme assays or flash frozen and stored at -80 °C (PseF was aliquoted and stored with and without 20% glycerol).

### **2.2.6 Activity assay of *A. caviae* PseF with Pse5Ac7Ac**

100 μL reactions were performed with 130 μg mL<sup>-1</sup> PseF, 0.5 mM Pse5Ac7Ac (Sussex Research), 1.5 mM CTP, 1 mM MgCl<sub>2</sub>, 25 mM sodium phosphate, pH 7.4, 50 mM NaCl and incubated at 25 °C. Methodology was based upon that used by Schoenhofen *et al.* Control reactions were performed with reaction mixture as described without either Pse5Ac7Ac, CTP or PseF. Upon addition of PseF to reaction mixture samples were analysed by -ESI LC-MS (details in Appendix-General methods) every 15 minutes for 2 hours and after 6.5 hours. Freshly purified PseF, thawed PseF containing 20% glycerol and thawed PseF without glycerol was used in these reactions.

### **2.2.7 Purification of *Aeromonas caviae* PseF for Crystallisation trials**

Fractions containing pure PseF (as described in 2.2.4) were dialysed (25 mM Tris, 50 mM NaCl, 2 mM MgCl<sub>2</sub>) overnight at 4 °C.

3 L of *E. coli* BL21 (DE3) encoding BL21 were cultured as described in 2.24. Thawed cell pellet was purified as described with Tris buffers in place of sodium phosphate buffers (Lysis buffer- 50mM Tris, pH 7.65, 400 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 2 mM MgCl<sub>2</sub>, 500 mM imidazole. Elution buffer- 50mM Tris, pH 7.65, 400 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, 2 mM MgCl<sub>2</sub>, 500 mM imidazole.) Fractions containing pure PseF were pooled and dialysed (Tris dialysis buffer- 25 mM Tris, 50 mM NaCl, 2 mM MgCl<sub>2</sub>) overnight at 4 °C. Dialysed protein was centrifuged (6,000  $\times$ g, 40 minutes, 4 °C) to remove precipitant and analysed via 12% SDS-PAGE. Sample was spin concentrated (6,000  $\times$ g, 40 minutes, 4 °C, 10 kDa MWCO tube) and loaded onto a S-200 size-exclusion column (GE Healthcare), pre-equilibrated with Tris dialysis buffer. PseF was eluted from the column in Tris dialysis buffer and collected in 1 mL samples, which were stored on ice. 12% SDS-PAGE was used to analyse the purity of these fractions (as described in 2.2.1).

### **2.2.8 *A. caviae* PseF Protein Identification by Mass Spectrometry**

Protein was extracted from two bands found in SDS-PAGE of PseF following SEC (2.2.7) and subject to trypsin digest. The resultant peptides were analysed by MALDI-MS and MS/MS using a Bruker ultraflex III MALDI-TOF/TOF. Spectral data was compared to the Mascot database to identify the protein contained in each SDS-PAGE band.

### **2.2.9 *A. caviae* PseF Circular Dichroism**

Circular dichroism data was collected for 0.2 mg mL<sup>-1</sup> Pure PseF was dialysed into in 25 mM sodium phosphate buffer pH 7.4, for 16 hours at 4 °C, using 2 x 2L buffer and dialysis tubing (6-8 kDa MWCO, Fisherbrand). Protein concentration was determined using molecular extinction co-efficient (41,940 M<sup>-1</sup> cm<sup>-1</sup>) and nanodrop spectrophotometer (DeNovix) and protein was diluted with dialysis buffer to a final concentration of 0.2 mg mL<sup>-1</sup>. Circular dichroism data was collected at 30 °C from 180-230 nm, using a Jasco J810 CD Spectrophotometer. A quartz cuvette with a path length of 1 mM was used, a 400  $\mu$ L of dialysis buffer was used to obtain a blank spectrum, after which 400  $\mu$ L of PseF was used and a spectrum was recorded. Secondary structure predictions were made from CD data using K2D3 (<http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/>).

### **2.2.10 *A. caviae* PseF Size Exclusion Chromatography - Multi-Angle Laser Light Scattering**

For the collection of size exclusion chromatography - multi-angle laser light scattering (SEC-MALLS) data pure PseF (prepared as described in 2.2.4, without glycerol) was dialysed in to running solvent (20 mM Tris, 50 mM NaCl, 2 mM MgCl<sub>2</sub>, pH 7.8). Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system (SPD-20A UV detector, LC20-AD isocratic pump system, DGU-20A3 degasser and SIL-20A autosampler). Work was conducted at room temperature (20 ±2°C). Running solvent was 0.2 µm filtered and at least 2 column volumes were used to equilibrate Superdex S200 (10/300 GL #0805015 (G.E. Healthcare)) size-exclusion column, running buffer was applied until for UV, light scattering and refractive index detectors were all stable. PseF was supplied at a nominal concentration of 10 mg mL<sup>-1</sup> and was diluted with the running buffer to 4 mg mL<sup>-1</sup> and 1 mg mL<sup>-1</sup>, then microfuged for 3 mins before use. 100 µL samples were applied to a Superdex S200 size-exclusion column (G.E. Healthcare), ran at flow rate 0.5 mL min<sup>-1</sup>.<sup>††</sup> A 2.5 mg/mL BSA sample was run as a standard. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data were analysed using Astra V software.

### **2.2.11 Crystallisation trials for *A. caviae* PseF**

Following purification, PseF was spin concentrated (6,000 ×g, 50 minutes, 4 °C, 10 kDa MWCO tube) resulting in two 200 µL samples at 10 and 6.5 mgmL<sup>-1</sup>. 2 mM DTT was added to 100 µL of each sample. The samples were stored on ice before being used in initial crystallisation screens. 98-well sitting drop plates were set up, using a Mosquito robot, to screen both concentrations of PseF, with and without DTT. Two commercially available screens were used, PEG/ion and PBD minimal set (Hampton research). Above each well a 150 nL drop of PseF was mixed with 150 nL of well mother-liquor, this was repeated with the PseF sample which contained DTT. Trays were stored at room temperature for approximately four weeks and examined for crystals. Regular crystals were fished from the

---

<sup>††</sup> SEC-MALLS data was collected by Dr Andrew Leech, Technology facility, Department of Biology, University of York

solution, flash frozen in N<sub>2</sub> (l) and stored in N<sub>2</sub> (l) prior to diffraction testing. X-ray diffraction was carried out using in-house facilities; a Rigaku MicroMax 007HF generator, RAXIS IV++ imaging plate detector and an Actor robotic sample changer.

Two 24-well hanging drop trays were set up based upon successful conditions found in the initial screen (Conditions in Appendix Table A.1). 0.3% (v/v) Pentaethylene Glycol Monoethyl Ether (C8E5) detergent was added to mother liquor in 24 well trays with the aim of slowing the rate of crystal formation. Trays were stored at room temperature for approximately four weeks and were analysed by ultraviolet (UV) light in the Xtal imager.

### **2.2.12 Large scale synthesis and purification of CMP-Pse5Ac7Ac, via a one-pot six enzyme synthesis**

2 mM UDP-GlcNAc **1.3** (90 mg), 0.3 mM coenzyme-A, 4 mM pyridoxal 5'-phosphate, 20 mM L-glutamic acid, 3 mM phosphoenolpyruvate, 20 mM N-acetyl-S-acetylcysteamine (SNAC)<sup>††</sup>, 0.2 mg mL<sup>-1</sup> PseB, 0.4 mg mL<sup>-1</sup> PseC, 0.2 mg mL<sup>-1</sup> PseH, 0.2 mg mL<sup>-1</sup> PseG and 0.2 mg mL<sup>-1</sup> PseI. The reaction mixture was incubated at 37 °C for 12 hours and was monitored via –ESI LC-MS, to monitor the production of Pse5Ac7Ac. After 12 hours, 0.2 mg mL<sup>-1</sup> PseF, 4 mM CTP and 20 mM MgCl<sub>2</sub> were added to facilitate the production of CMP-Pse5Ac7Ac **1.10**, as confirmed by –ESI LC-MS analysis. The reaction incubated at 37 °C for 4 hours and was analysed via –ESI LC-MS. The mixture was lyophilised and resuspended in 10 mL dH<sub>2</sub>O and 10 mL EtOH and stored at 4 °C for 30 minutes to precipitate enzymes, the solution was centrifuged (38,759 ×g, 1 hours, 4 °C). The supernatant was diluted in dH<sub>2</sub>O to a final volume of 200 mL before lyophilisation. The lyophilised material was resuspended in 7 mL dH<sub>2</sub>O and passed through a 45 μM Millex<sup>®</sup> syringe filter (Merck) to remove any remaining enzymes before being applied to a 500 mL column packed with Bio-Gel<sup>®</sup> P-2 resin (Biorad) in HPLC-grade H<sub>2</sub>O. 4 mL fractions were collected for 24 hours and analysed via LC-MS (details in Appendix-General methods) for the presence **1.10** of and all fractions containing **1.10** were pooled and lyophilised. Following NMR analysis of the resulting material (D<sub>2</sub>O, 500MHz).

---

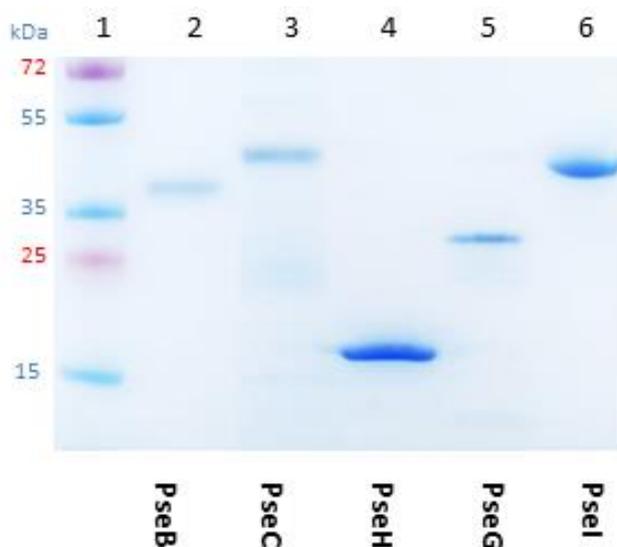
<sup>††</sup> Synthesised and characterised by Matthew Best (MChem student- Fascione laboratory)

$^1\text{H}$  NMR (500 MHz,  $\text{D}_2\text{O}$ )  $\delta$  8.02 (d,  $J = 7.5$ , 1H, H2'), 6.14 (d,  $J = 7.6$  Hz, 1H, H1'), 6.00 (d,  $J = 4.1$  Hz, 1H, H3'), 4.37 – 4.32 (m, 2H, H4', H6'), 4.32 – 4.22 (m, 6H, H4, H5, H6, H5', H7a', H7b'), 4.12 (dd,  $J = 6.5, 5.1$  Hz, 1H, H8), 4.03 (dd,  $J = 9.9, 5.0$  Hz, 1H, H7), 2.23 (dd,  $J = 13.4, 4.5$  Hz, 1H, H3eq), 2.00 (s, 3H, 5NHAc), 1.97 (s, 3H, 7NHAc), 1.61 (ddd,  $J = 13.4, 12.0, 5.2$  Hz, 1H, H3ax), 1.20 (d,  $J = 6.5$  Hz, 3H, 9CH<sub>3</sub>).  $^{13}\text{C}$  NMR (126 MHz,  $\text{D}_2\text{O}$ )  $\delta$  174.5, 173.2, 165.8, 157.3, 141.7, 100.0, 96.5, 89.2, 82.9, 74.1, 72.7, 69.2, 68.6, 64.9, 64.38, 53.7, 48.77, 36.0, 22.0, 21.9, 17.2. HR-MS data was collected with  $-\text{ESI}$  MS: Expected  $[\text{M}-\text{H}]^-$   $m/z$ : 638.1789, measured  $[\text{M}-\text{H}]^-$   $m/z$ : 638.1727. ATR-FTIR  $V_{\text{max}}$ : 3227, 2941, 2888, 1592, 1403, 1075  $\text{cm}^{-1}$

## 2.3 Results

### 2.3.1 Overexpression and Purification of PseB, PseC, PseH, PseG and PseI

To facilitate the enzymatic production of Pse5Ac7Ac *E. coli* BL21 (DE3) were transformed with plasmid containing one of the PseB, PseC, PseH, PseG, or PseI genes from *C. jejuni*. Transformed cells were each cultured on a 4 L scale and overexpression of the gene was induced with IPTG, such that the resultant His-tagged protein could be purified by nickel-affinity chromatography. The conditions for expression and purification has been previously optimised<sup>§§</sup> therefore, as anticipated each purification yielded high levels of protein, which was sufficiently pure for use in chemoenzymatic synthesis (Figure 2.4). The yield for each enzyme were as follows: PseB- 14 mg L<sup>-1</sup>, PseC- 9 mg L<sup>-1</sup>, PseH- 18 mg L<sup>-1</sup>, PseG- 12 mg L<sup>-1</sup>, PseI- 20 mg L<sup>-1</sup>.



**Figure 2.4: 4-20% SDS-PAGE for Pse5Ac7Ac biosynthesis enzymes following nickel-affinity purification. Lanes loaded as follows: 1-PAGE Ruler MW marker, 2- PseB (39.5 kDa), 3- PseC (44.6 kDa), 4- PseH (20.9 kDa), 5-PseG (33.4 kDa), 6- PseI (40.8 kDa).**

Following purification proteins were mixed with 25% glycerol, flash frozen and stored at -80 °C. Glycerol was used as a cryoprotectant for the enzymes, to ensure that proteins retained activity after a freeze-thaw cycle. However, as an aim of this study was to purify

<sup>§§</sup> Optimisation of expression and purification of PseB, PseC, PseG, PseH and PseI was performed by Dr Harriet Chidwick and Emily Flack (BSc project).

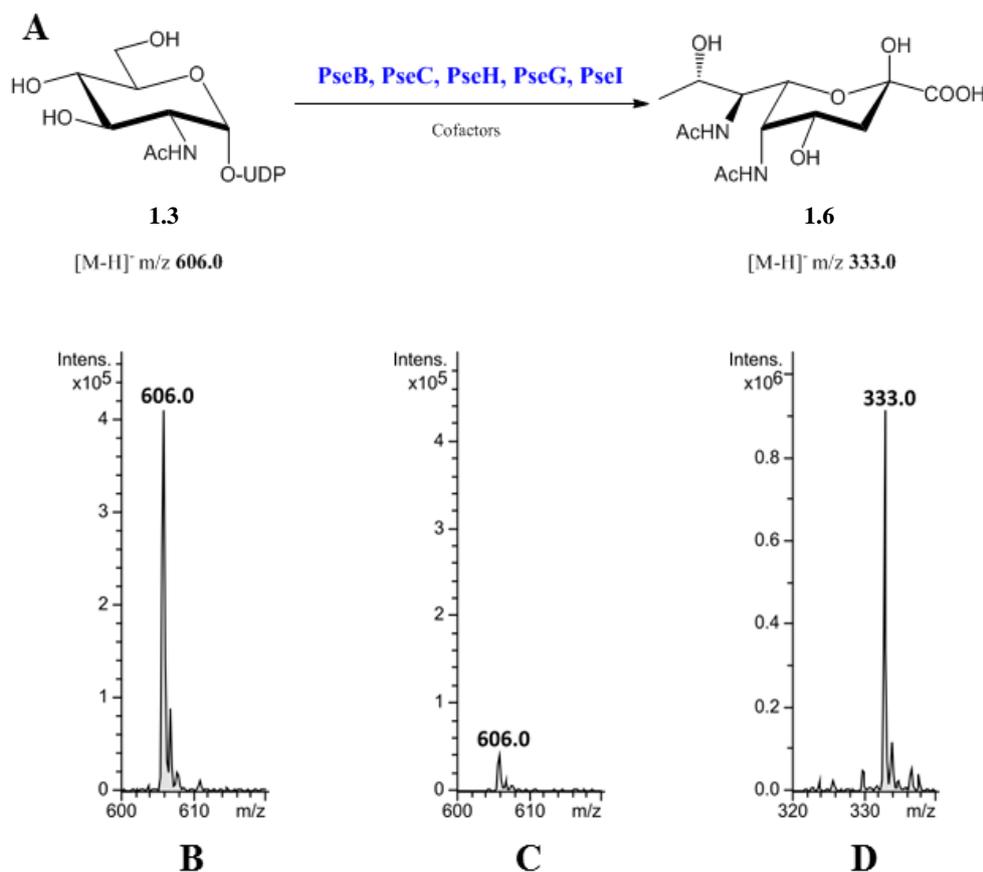
CMP-Pse5Ac7Ac, from a complex reaction mixture, it was decided that the removal of glycerol from enzymes stocks may aid CMP-Pse5Ac7Ac purification. Therefore, expression and purifications were repeated, and enzymes were flash frozen without glycerol.

### 2.2.2 Enzymatic synthesis of Pse5Ac7Ac with Acetyl-thiocholine Iodide

To test the activity of purified enzymes described in 2.3.1 a one pot reaction was performed which contained enzymes PseB, PseC, PseH, PseG and PseI along with UDP-GlcNAc starting material, co-factors and the acetyl-CoA regeneration co-factor acetylthiocholine iodide (reaction summarised in Figure 2.5, A).\*\*\* The reaction incubated at 37 °C and LC-MS analysis showed that over 2 hours there was a decrease in intensity of a peak corresponding to UDP-GlcNAc **1.3** (Figure 2.5, B and C. [M-H]<sup>-</sup>, m/z 606.0). The reaction was successful as a peak corresponding to Pse5Ac7Ac was observed in LC-MS data after 2 hours (Figure 2.5, D. [M-H]<sup>-</sup>, m/z 333.0). As all enzymes were active and this reaction was successful it was determined that efforts should focus on the expression and purification of *C. jejuni* PseF (CjPseF), with the aim of using it to produce CMP-Pse5Ac7Ac **1.10**.

---

\*\*\* Acetylthiocholine iodide was previously shown to be an acetyl-CoA regeneration co-factor by Dr Harriet Chidwick, Fascione Laboratory.



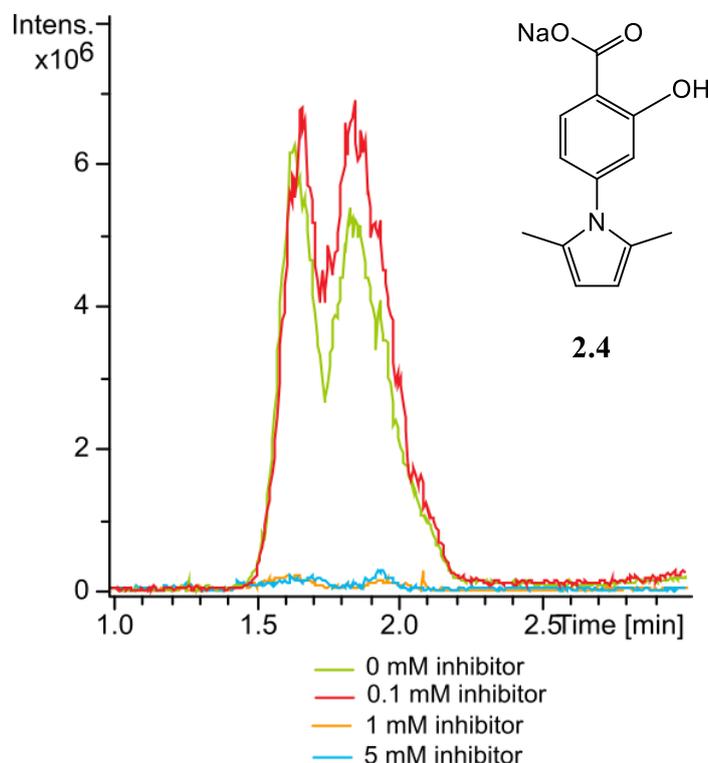
**Figure 2.5: Chemoenzymatic synthesis of Pse5Ac7Ac. A: Summary of reaction-** starting UDP-GlcNAc, **1.3** was mixed with enzymes PseB, PseC, PseH, PseG and PseI, plus co-factors to produce Pse5Ac7Ac **1.6**. **B-** LC-MS peak corresponding to UDP-GlcNAc, 0 hours. **C-** LC-MS peak corresponding to UDP-GlcNAc, 2 hours. **D-** LC-MS peak corresponding to Pse5Ac7Ac, 2 hours.

### 2.3.3 Small-molecule screen for potential inhibition of PseB and PseI

A library of small-molecules synthesised as potential inhibitors for the Pse5Ac7Ac synthase, PseI were made available to this study.<sup>†††</sup> It has previously been shown that **2.4** is an inhibitor of *H. pylori* PseB. A reaction containing recombinantly express *C. jejuni* enzymes PseB, PseC, PseH, PseG, PseI, UDP-GlcNAc **1.3** and all corresponding co-factors, plus 0 mM 0.1 mM, or 5 mM **2.4** incubated overnight and were analysed by -ESI LC-MS. Results showed that there was a decrease in peak corresponding to Pse5Ac7Ac **1.6** molecular ion ([M-H]<sup>-</sup>

<sup>†††</sup> Synthesis, characterisation and purification performed by Joseph Ferner, The University of Sheffield.

m/z 333.2) as the concentration of **2.4** increased, suggesting that there was an inhibitory effect (extracted ion chromatograph (EIC) for m/z 333.2 is shown in Figure 2.6)



**Figure 2.6 Inhibition studies for PseB with compound 2.4. Extracted ion chromatograms for Pse5Ac7Ac, m/z 333.2 [M-H]<sup>-</sup>.**

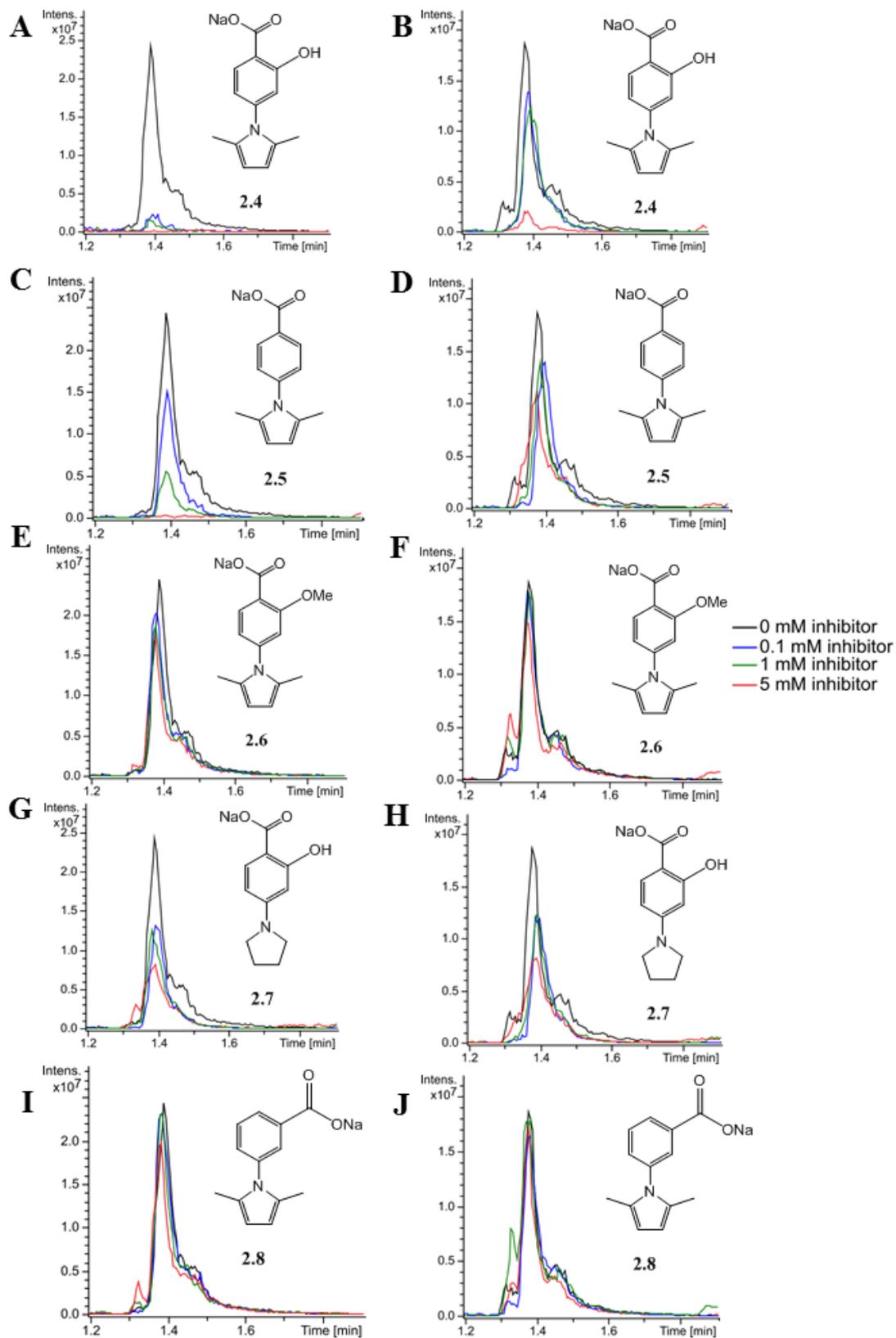
For monitoring inhibition of PseI, the substrate 6-deoxy-AltdiNAc **1.15** and active PseI were required. In addition to *C. jejuni* PseI which was produced as described in 2.2.1, *Aeromonas caviae* PseI (AcPseI) was also tested.<sup>†††</sup> To produce **1.15** a reaction mixture containing enzymes PseB, PseC, PseH, and PseG along with the required starting material **1.3** and co-factors was set up and incubated for 12 hours at 30 °C. LC-MS analysis was used to monitor the reaction. As a peak corresponding to 6-deoxyAltdiNAc was not visible in LC-MS spectra, PseI and PEP were added to a sample of reaction mixture to give final volumes of 100 µL, which incubated for 2 hours at 30 °C, before being analysed via LC-MS for the presence of Pse5Ac7Ac. A peak at 333.2 m/z in LC-MS spectra confirmed that Pse5Ac7Ac was being produced (data in Appendix Figure A.1), therefore it was concluded that the mixture containing enzymes PseB-G had produced 6-deoxyAltdiNAc and that PseI was active.

<sup>†††</sup> AcPseI produced by Joseph Ferner, The University of Sheffield.

Following overnight storage at 4 °C, enzyme precipitate was removed via centrifugation and the resultant supernatant was used in an inhibition assay. Compounds **2.4-2.8** were added at 0.1 mM, 1 mM or 5 mM to reaction mixture along with PEP and Psel or AcPsel to give final sample volumes of 100 µL. Reactions were performed in duplicate and a control without potential inhibitor compound (**2.4-2.8**) was also performed for both Psel and AcPsel. Reactions incubated at 30 °C overnight and were subsequently analysed via LC-MS. EICs for ions with m/z 333.2 ([M-H]<sup>-</sup> Pse5Ac7Ac) for each sample was compared to that of the control, to determine the extent of Psel inhibition (Figure 2.7).

This data showed a clear decrease in the intensity of the EIC (and MS peak) for ions with m/z 333.2, concentration of potential inhibitor increased, in samples containing compounds **2.4**, **2.5** and **2.7** (Figure 2.7 A-D, G and H), suggesting that these compounds cause an inhibitory effect on both Psel and AcPsel. Samples containing **2.4** showed a greater reduction in Pse5Ac7Ac, for both Psel and AcPsel, than equivalent samples containing **2.5** (Figure 2.7 A vs C and B vs D), suggesting that the addition of the hydroxyl moiety contributed towards Psel inhibition. Compound **2.4** (Figure 2.7 A and B) appeared to be most effective concentration-dependent inhibitor of Psel and AcPsel, at 5 mM no Pse5Ac7Ac was detected in the samples containing *C. jejuni* Psel.

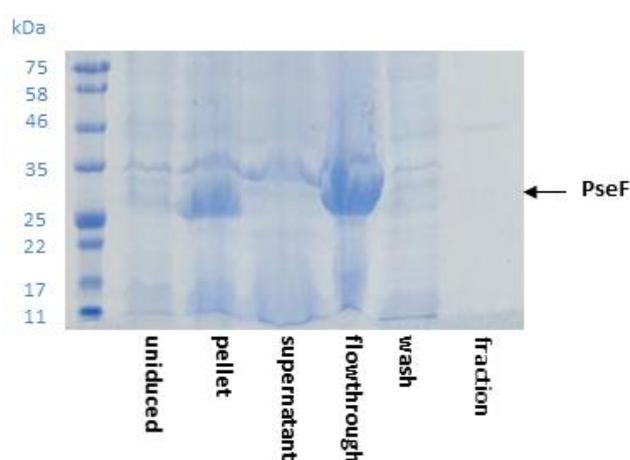
In samples containing compounds **2.6** and **2.8** the EIC peaks for ions with m/z 333.2 also show a decrease in intensity as concentrations of **2.6** and **2.8** increase, suggesting some inhibition of Psel and AcPsel; however, the concentration-dependent effect is less apparent than in other samples. This suggests that that **2.6** and **2.8** may not have as greater inhibitory effects on Psel and AcPsel as compounds **2.4-2.8**. A phosphate release assay may be used as a more quantitative assay for Pse5Ac7Ac production, as phosphate is released when PEP is used by Psel in the condensation reaction with **1.6**. All compounds could be rescreened in the assay, in particular **2.6** and **2.8**, to confirm their function as Psel inhibitors.



**Figure 2.7:** Inhibition studies for PseI and AcPseI with compounds 2.4-2.8. Extracted ion chromatographs for PseAc7Ac,  $m/z$  333.2 [M-H]<sup>-</sup>. A, C, E, G and I- *C. jejuni* PseI, B, D, F, H and J- *A. caviae* PseI. Compounds added to each sample is depicted.

### 2.3.4 Recombinant Expression and purification attempt of *C. jejuni* PseF in *E. coli*

The  $\alpha$ -CMP-Pse5Ac7Ac synthase, CjPseF was recombinantly produced in *E. coli* BL21 (DE3). A cell pellet collected from 2 L of cultured cells was lysed and purification of the His<sub>6</sub>-tagged protein was attempted using nickel-affinity chromatography, following a published protocol outlined by Schoenhofen *et al.* The lack of a peak at A<sub>254</sub> recorded on the chromatogram throughout the procedure suggested that PseF had not been successfully purified (chromatogram in Appendix). Further inspection of the SDS-PAGE showed that the CjPseF was insoluble, accounting for the lack of success in purification (Figure 2.8).

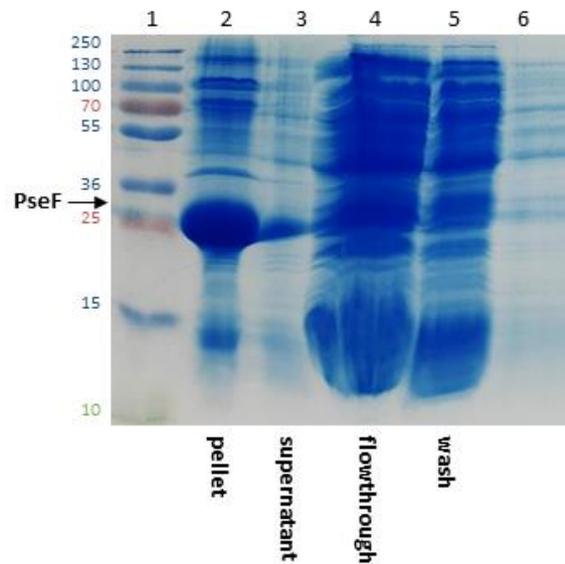


**Figure 2.8: 12% SDS-PAGE for CjPseF purification. Lanes loaded as follows: 1- NEB prestained broad range MW Marker, 2- uninduced control, 3- lysate, 4- supernatant, 5- pellet, 6- flowthrough, 7- wash. PseF (MW 28.0 kDa) can be seen in lysate and pellet samples.**

With the aim of yielding soluble CjPseF a series of expression trials were undertaken. However, SDS-PAGE analysis indicated that soluble CjPseF was not obtained in any of the conditions tested.

The plasmid used in this work was codon optimised for *E. coli* (gene sequence in Appendix- Plasmid information, gene sequences and recombinant protein sequences- CjPseF Codon optimised), unlike that used by Schoenhofen *et al.*<sup>99</sup> Therefore plasmid encoding *CjPseF* which was not codon optimised for *E. coli* (gene sequence in Appendix- Plasmid information, gene sequences and recombinant protein sequences- CjPseF Non-codon optimised) was used in subsequent attempts to recombinantly express *CjPseF* and purify the resultant protein. However, no peak was present at A<sub>254</sub> on the chromatogram recorded as the elution buffer passed through the column (Appendix Figure A.20) the SDS-

PAGE analysis confirmed that CjPseF remained insoluble (Figure 2.8, lane 2) and purification attempt via nickel-affinity chromatography failed to yield pure CjPseF (Figure 2.9, lane 6). It was concluded that studies should focus on a PseF homologue from *A. caviae*, which was yet to be shown to be active *in vitro* as a CMP-Pse5Ac7Ac synthetase.<sup>§§§</sup>

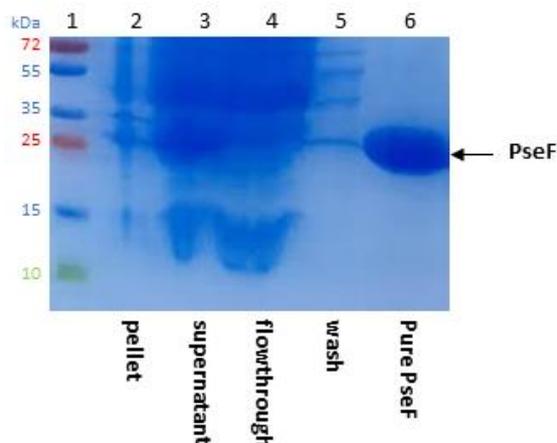


**Figure 2.9: 12% SDS-PAGE for CjPseF purification, from plasmid which was not codon-optimised for *E. coli*. Lanes loaded as follows: 1- PAGE Ruler MW Marker, 2- pellet, 3- supernatant, 4- flowthrough, 5- wash, 6- representative fraction. PseF (MW 28.0 kDa) can be seen in pellet sample.**

### 2.3.5 Expression and purification of *A. caviae* PseF

The putative CMP-Pse5Ac7Ac synthetase, PseF from *A. caviae* (referred to as PseF, throughout), was recombinantly overexpressed in *E. coli* BL21 (DE3) (30 °C, OD<sub>600</sub> 0.6, 0.1 mM IPTG, 3 hr post-induction incubation) based upon the results of small-scale expression trials (data not shown). A cell pellet collected from 3 L of cultured cells was lysed and the His<sub>6</sub>-tagged protein was purified using nickel-affinity chromatography, following a published protocol outlined by Schoenhofen *et al.* SDS-PAGE analysis of fractions obtained from chromatography showed that PseF had been successfully purified (Figure 2.10), yielding approximately 13 mg L<sup>-1</sup> of *E. coli* culture. It was concluded that the protein was sufficiently pure for use in activity assays.

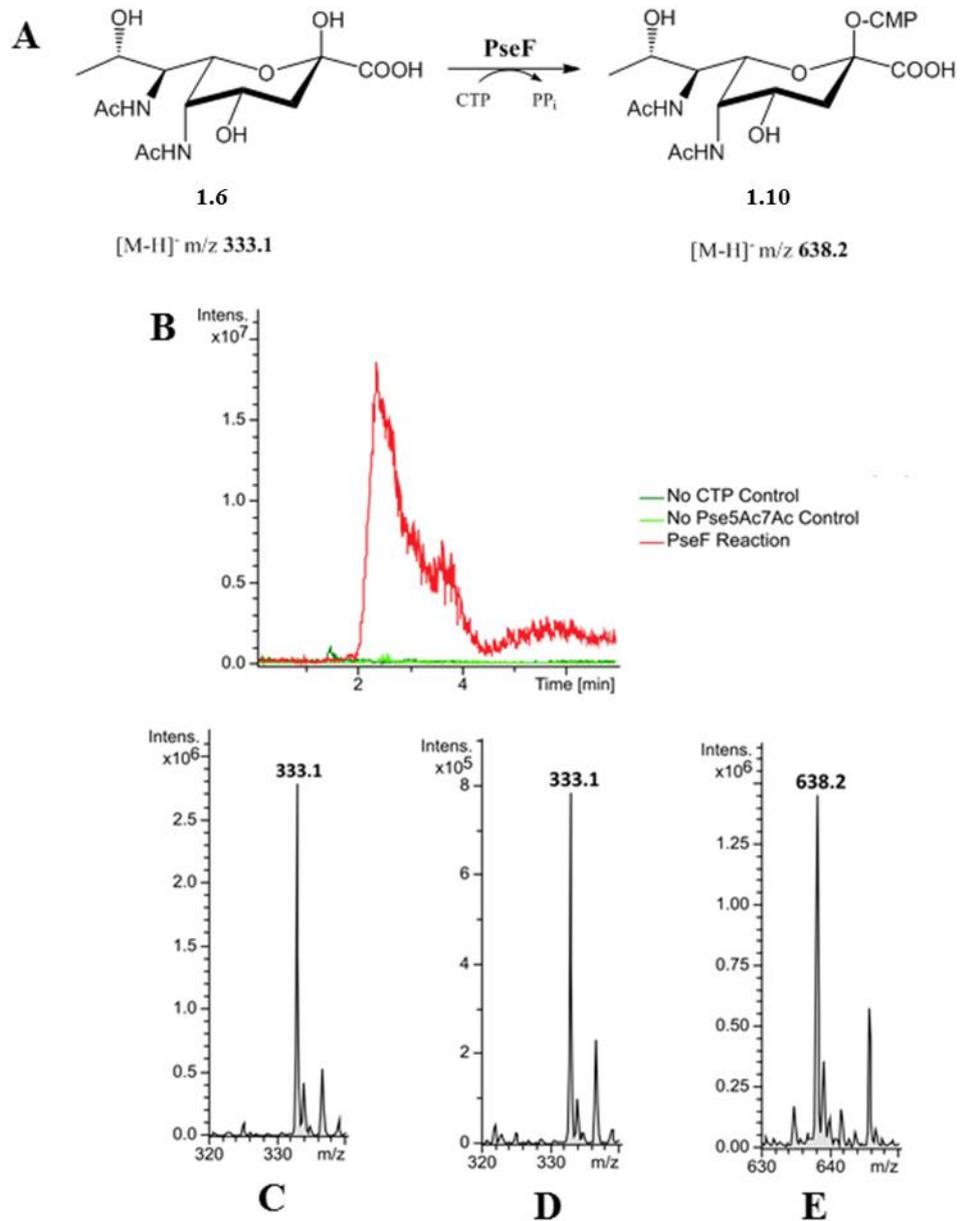
<sup>§§§</sup> *A. caviae* PseF (and resultant protein AcPseF) is referred to as NeuA (and resultant protein NeuA) in previous literature, which is concurrent with sialic acid biosynthesis nomenclature.<sup>104</sup>



**Figure 2.10: 12% SDS-PAGE for PseF purification. Lanes loaded as follows: 1- PAGE Ruler MW Marker, 2- pellet, 3- supernatant, 4- flowthrough, 5- wash, 6- pure PseF (MW 28.0 kDa) can be seen in pellet, supernatant and pure PseF samples.**

### 2.3.6 Activity assay of *A. caviae* PseF with Pse5Ac7Ac

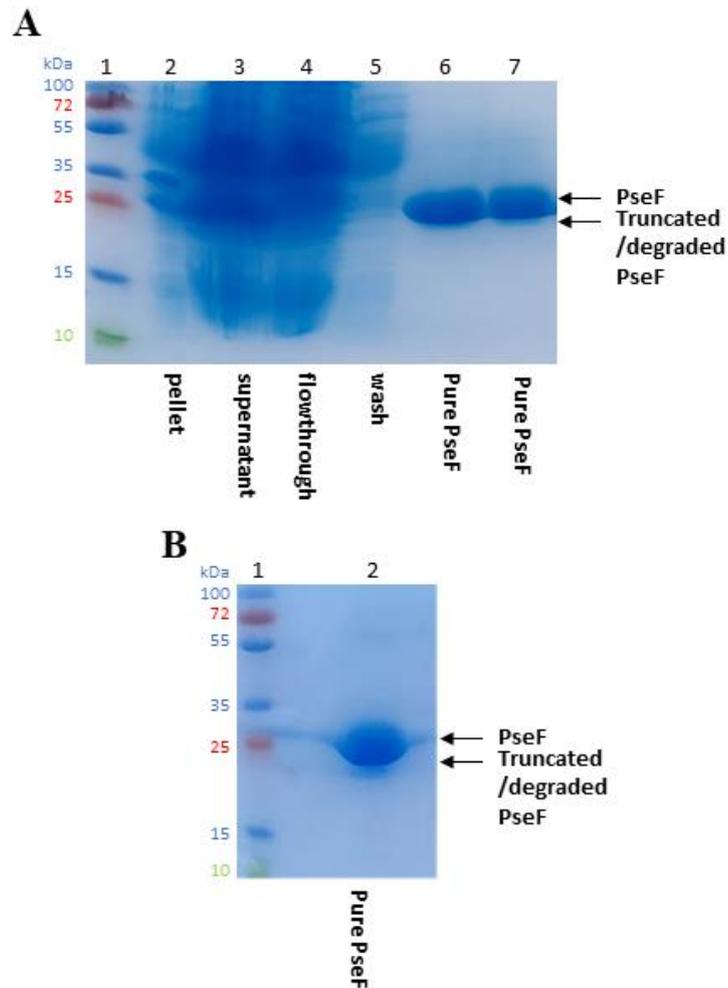
Pure PseF was used in enzymatic synthesis of CMP-Pse5Ac7Ac. Reactions were monitored via negative ESI LC-MS. CMP-Pse5Ac7Ac was produced as indicated by the presence of  $[M-H]^-$  peak, at  $m/z$  638.2 (Figure 2.11, E). In all spectra for reactions  $[M-H]^-$  peaks for both CMP-Pse5Ac7Ac and Pse5Ac7Ac ( $m/z$  333) (Figure 2.11 D shows Pse5Ac7Ac) indicating that whilst the reaction was successful, either complete conversion to product was not occurring or that CMP-Pse5Ac7Ac was hydrolysing to give CMP and Pse5Ac7Ac. The reactions were performed using both fresh and thawed PseF, in both cases PseF was shown to be active. Control reactions were performed with reaction mixture as described without either Pse5Ac7Ac, CTP or PseF. In control experiments peaks corresponding to CMP-Pse5Ac7Ac cannot be seen in mass-spectrometry data (Figure 2.11, B- shows extracted ion count chromatographs for CMP-Pse5Ac7Ac  $[M-H]^-$ ) and a peak corresponding to Pse5Ac7Ac can be seen in the no CTP control can be observed (Figure 2.11, C). This data confirms that *A. caviae* PseF functions as a CMP-Pse5Ac7Ac synthetase.



**Figure 2.11: LC-MS analysis of activity assay for *A. caviae* PseF with Pse5Ac7Ac.** A- Summary of reaction. PseF converts Pse5Ac7Ac to CMP-Pse7Ac5Ac using CTP, releasing pyrophosphate (PP<sub>i</sub>). B- Extracted Ion Chromatographs for m/z 638.2 (CMP-Pse5Ac7Ac [M-H]<sup>-</sup>) Red: PseF reaction, Dark green: No CTP control, Light green: No Pse5Ac7Ac control. C- No CTP control, peak at m/z 333.1 (Pse5Ac7Ac [M-H]<sup>-</sup>), D- PseF reaction, peak at m/z 333.1 (Pse5Ac7Ac [M-H]<sup>-</sup>), E- PseF reaction, peak at m/z 638.2 (CMP-Pse5Ac7Ac [M-H]<sup>-</sup>)

### 2.3.7 Purification of *A. caviae* PseF for Crystallisation trials

To our knowledge, the functional activity study of CjPseF, described by Schoenhofen *et al.* was the sole *in vitro* study on any PseF homologue, before this study.<sup>99</sup> Therefore, it was determined that further studies, including structural studies should be pursued with pure PseF. To screen crystallisation conditions for PseF, purified protein (as described in 2.3.4) was dialysed in Tris dialysis buffer at 4 °C overnight, however protein precipitated during this step. With the aim of combating precipitation, PseF was purified using Tris buffer in place of sodium phosphate-based lysis and elution buffers and fractions were analysed by SDS-PAGE (Figure 2.12 A). This analysis showed the presence of potential contaminant proteins and truncated or degraded PseF, indicated by a band of lower molecular weight in pure PseF fractions (Figure 2.12 A, lanes 6 and 7). With the aim of further purifying PseF by size-exclusion chromatography, fractions containing pure PseF were pooled and dialysed overnight into Tris buffer, where protein precipitation occurred once more. The sample was centrifuged, and supernatant separated from pelleted precipitant and analysed via SDS-PAGE (data not shown). The presence of a strong band of around 28 kDa in SDS-PAGE showed that the supernatant still contained soluble PseF, truncated PseF and contaminant proteins. Therefore, the sample was subjected to size-exclusion chromatography (SEC) where the protein was eluted from the column in Tris dialysis buffer. SDS-PAGE analysis revealed that fractions containing PseF, indicated by strong band at around 28 kDa also still contained a less-prominent band of a slightly lower molecular weight (Figure 2.12 B, lane 2), this may be due to proteolysis of PseF. It was concluded that this protein sample should be used in crystals trials, in order to gather preliminary data which may indicate whether the protein was likely to be amenable to crystallisation (results of crystal trials discussed in 2.3.11).



**Figure 2.12: SDS-PAGE analysis of PseF purification in Tris buffer for crystallisation trials. A: 12% SDS\_PAGE analysis following HisTrap Purification. Gel loaded as follows: 1-PAGE Ruler MW marker, 2- pellet, 3- supernatant, 3- flowthrough, 4- wash, 5 and 6- Pure PseF. Truncated/degraded PseF (28.0 kDa) can be seen in lanes 5 and 6. Pooled PseF fractions (28.0 kDa). B: Purification of PseF using Sephacryl S-400 column and Tris buffer. Gel loaded as follows: 1-PAGE Ruler MW marker, 2-Pooled PseF fractions (truncated/degraded PseF can be seen).**

### 2.3.8 *A. caviae* PseF Protein Identification by Mass Spectrometry

As described in 2.3.6, following SEC of PseF, SDS-PAGE showed the presence of two protein bands at the approximate MW of PseF. As the protein sample was to be used for crystallisation trials, it was crucial to establish the identity of the proteins. Protein was extracted from the bands, subjected to trypsin digest and the resultant peptides were analysed by mass-spectrometry, which identified both bands as PseF. Peptides were found for both proteins which corresponded to the ninth residue from the C-terminus (Figure 2.13 A and B), suggesting that the protein was not truncated during translation and may be of lower MW due to proteolysis. Despite unsuccessful attempts to separate the two species

via SEC, it was concluded that crystallisation trials should be attempted with this protein sample.

**A**

|     |            |            |            |            |            |
|-----|------------|------------|------------|------------|------------|
| 1   | MNIAIIPARG | GSKRIPRKNI | KPFHSPKPIA | WSILAACKAG | CFERIIVSTD |
| 51  | DAEIAAVALE | YGAEVPFTRP | AEIANDYATT | GEVISHAINW | LINQQGQVPE |
| 101 | NVCCLYATAP | FVEPDDLQCG | LELLTFNKEC | QFVFSATRFS | FPIQRAIKLD |
| 151 | ESGWVSMFHP | EYQLTRSQDL | EEAYHDAGQF | YWGKANAWLN | KLPIFAVHTQ |
| 201 | VVLLPSHRVQ | DIDTQDDWLR | AEKLFTLR   |            |            |

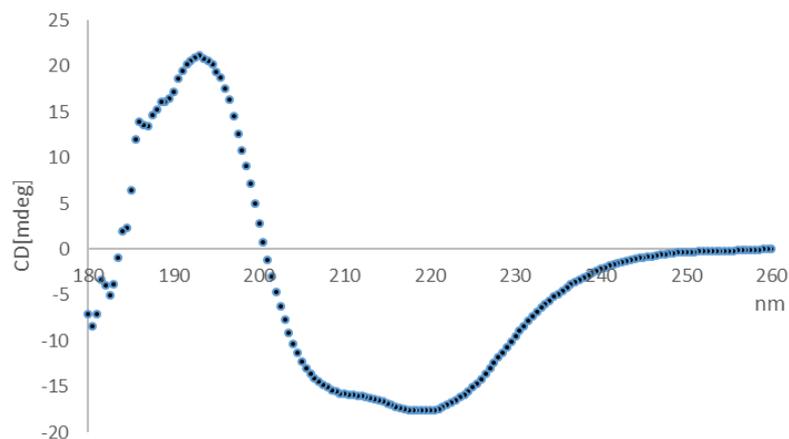
**B**

|     |            |            |            |            |            |
|-----|------------|------------|------------|------------|------------|
| 1   | MNIAIIPARG | GSKRIPRKNI | KPFHSPKPIA | WSILAACKAG | CFERIIVSTD |
| 51  | DAEIAAVALE | YGAEVPFTRP | AEIANDYATT | GEVISHAINW | LINQQGQVPE |
| 101 | NVCCLYATAP | FVEPDDLQCG | LELLTFNKEC | QFVFSATRFS | FPIQRAIKLD |
| 151 | ESGWVSMFHP | EYQLTRSQDL | EEAYHDAGQF | YWGKANAWLN | KLPIFAVHTQ |
| 201 | VVLLPSHRVQ | DIDTQDDWLR | AEKLFTLR   |            |            |

**Figure 2.13: Protein Identification Mass Spectrometry for PseF. A- Analysis following trypsin digest and MS of PseF band seen in figure 2.9, lane 2. B- Analysis following trypsin digest and MS of Truncated/degraded PseF band, seen in figure 2.9, lane 2. Peptides found are shown in blue text.**

### 2.3.9 *A. caviae* PseF Circular Dichroism

Circular dichroism data was collected for 0.2 mg mL<sup>-1</sup> PseF, in 25 mM sodium phosphate buffer pH 7.4 (Figure 2.14). Secondary structure predictions were made using data at 190-240 nm. Results showed that under these conditions PseF was predicted to be 85.21 %  $\alpha$ -helical and 1.33 %  $\beta$ -strands. Therefore 86.54 % of the protein is predicted to have a fixed secondary structure, in the absence of ligands. This suggests that PseF should contain sufficient secondary structure to be amenable to crystallisation.



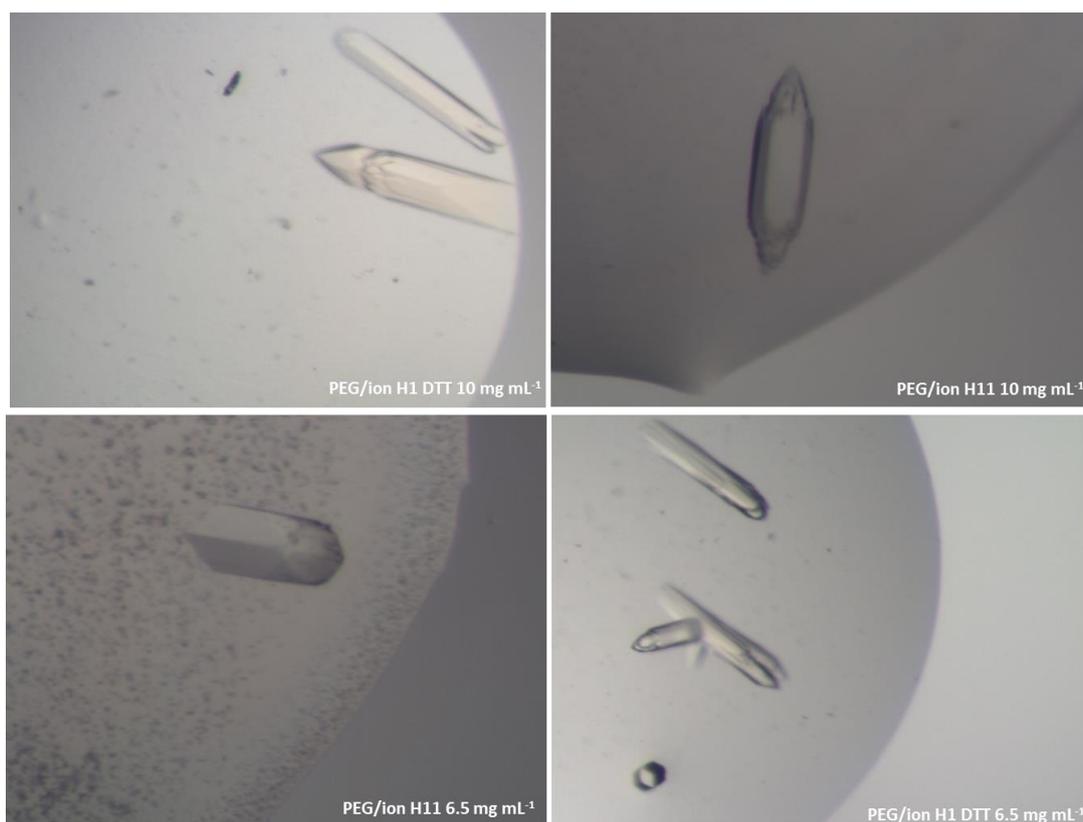
**Figure 2.14: CD spectrum for PseF for 0.2 mg mL<sup>-1</sup> PseF, in 25 mM sodium phosphate buffer pH 7.4.**

### **2.3.10 *A. caviae* PseF Size Exclusion Chromatography - Multi-Angle Laser Light Scattering**

SEC-MALLS data was obtained for PseF to determine the oligomeric state of PseF. PseF was used at a concentration of 4 mg mL<sup>-1</sup> and 1 mg mL<sup>-1</sup>. SEC-MALLS data (Appendix Figure A.3) revealed that a protein of around 54 kDa was present, indicating that 28 kDa PseF was present as a homodimer. This is consistent with literature reports that homologous proteins CMP-sialic acid synthetase and cytidine-monophosphate-3-deoxy-d-manno-oct-2-ulosonic acid (CMP-Kdo) synthetase both exist as homodimers.<sup>126,127</sup>

### **2.3.11 Crystallisation trials for *A. caviae* PseF**

PseF purified as described 2.2.6 in was used in a series of initial crystallisation trials. Two commercially available crystallisation screens, PEG/ion and PBD minimal set (Hampton research) were used to set up 96-well sitting drop plates in which each of the 98 buffer conditions contains two concentrations of PseF, each with and without 1mM DTT. Crystals formed in several of the conditions tested (examples in Figure 2.15), these crystals varied in shape, many of which were hollow needle-like structures. The larger crystals were fished from these trays and diffracted in-house but did not produce high quality diffraction data.



**Figure 2.3: Examples of crystals formed in 96 well tray with PEG ion screen, for PseF**

With the aim of producing larger, higher quality PseF crystals, 24 well hanging drop trays were set up based upon conditions where crystals were found in the initial screen. As some crystals in the initial screen appeared hollow C8E8 detergent was added to mother liquor in 24 well trays with the aim of slowing the rate of crystal formation. There appeared to be crystals present in several wells. However, analysis using the Xtal imager showed that the crystals were not visible under UV light, indicating they were not protein crystals and were likely to be crystallised salts at which point no further attempts were made to crystallise PseF.

### **2.2.3.12 Large scale synthesis and purification of CMP-Pse5Ac7Ac, via a one-pot six enzyme synthesis**

A large scale (90 mg UDP-GlcNAc) chemoenzymatic synthesis of CMP-Pse5Ac7Ac was performed with *C. jejuni* enzymes PseB, PseC, PseH, PseG and PseI to produce Pse5Ac7Ac, followed by *A. caviae* PseF and CTP to afford CMP-Pse5Ac7Ac. This reaction was performed with the aim of purifying CMP-Pse5Ac7Ac, to unequivocally characterise the product of *A.*

*caviae* PseF which had not previously been used *in vitro*. Additionally, the product of this reaction was required to facilitate further work discussed in Chapters 3 and 4.

The reaction mixture containing enzymes PseB-I was incubated at 30 °C and LC-MS was used to monitor the progress of the reaction. After 12 hours CTP and PseF were added to the reaction to facilitate the conversion of Pse5Ac7Ac to CMP-Pse5Ac7Ac. LC-MS analysis of the reaction confirmed that the reaction was successful, observed by a decrease in intensity of peak corresponding to Pse5Ac7Ac ( $[M-H]^-$ ,  $m/z$  333.1) and a peak corresponding to CMP-Pse5Ac7Ac ( $[M-H]^-$ ,  $m/z$  637.6) (Appendix Figure A.2). After 5 hours, the reaction was stopped by precipitating enzymes with EtOH. EtOH also appeared to remove any remaining PLP, which is yellow in appearance. The initially yellow reaction mixture was separated to clear colourless supernatant and yellow pellet containing enzymes and PLP following centrifugation. The enzyme-free supernatant was lyophilised and subsequently resuspended in dH<sub>2</sub>O and applied to a size exclusion column. Fractions were collected and analysed by LC-MS for the presence of CMP-Pse5Ac7Ac. Fractions containing **1.10** were pooled and lyophilised before being analysed by NMR (D<sub>2</sub>O, 500 MHz).

NMR analysis showed that the purification was partially successful in removing some by-products and excess starting materials. Peaks corresponding to both Pse5Ac7Ac and the CMP moiety were assigned (<sup>1</sup>H NMR in Figure 2.16). Crucially, the stereochemistry of the product was assigned. NMR analysis confirmed that the product contained  $\alpha$ -CMP-Pse5Ac7Ac, the stereochemistry was determined by the 0.58 p.p.m. difference between  $\delta$  values for H-3<sub>ax</sub> and H-3<sub>eq</sub> peaks in the <sup>1</sup>H NMR spectrum (reported as ~0.6 p.p.m. for equatorial carboxy group and ~0.9 p.p.m for axial carboxy group).<sup>106,128</sup> However, the NMR data showed that a contaminant remained, and integration of peaks suggested that the mixture was 4:1 CMP-Pse5Ac7Ac to contaminant. A yield of 64% (61 mg) was achieved.

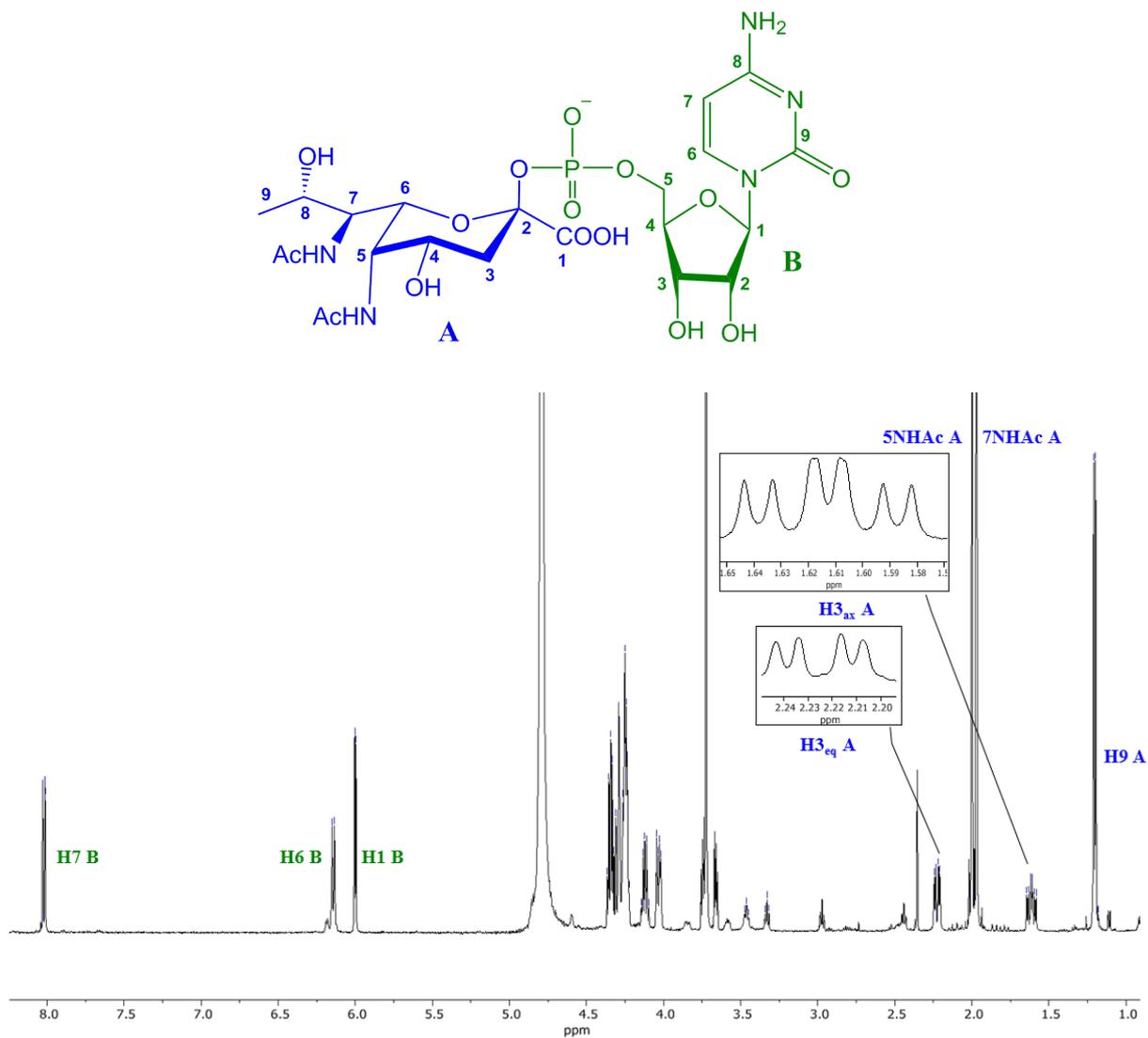


Figure 2.16: 500 MHz  $^1\text{H}$  NMR for  $\alpha$ -CMP-Pse5Ac7Ac 1.10

## 2.4 Discussion

A crucial first step of this work was to successfully purify recombinant PseB-PseI. As conditions for this had been previously optimised, all enzymes were obtained in high yield. With the eventual aim of purifying CMP-Pse5Ac7Ac, it was determined that the enzymes remained activity after being frozen without cryoprotectant.

As demonstrated in Figure 2.2, PseB has a secondary function as a C5 epimerase of its initial product **1.12**.<sup>122</sup> Therefore, to limit the formation of this product it was crucial to use an excess of PseC to drive the conversion of **1.12** to **1.13**. It has recently been demonstrated that *Anoxybacillus kamchatkensis* PseC can function *in vitro* without the addition of PLP to reaction mixtures, indicating that endogenous PLP is sufficient for function.<sup>107</sup> Future enzymatic syntheses which use PseC should omit PLP.

Previous work performed within the Fascione laboratory had optimised the third step in enzymatic synthesis of Pse5Ac7Ac, making the process more economically viable\*\*\*\*. As described in 2.3.2 it was determined that in place of 1.5 equivalents of Ac-CoA which were required for the PseH step in the work of Schoenhofen *et al.* the Ac-CoA could be reduced to 0.01 equivalents when 60 equiv. acetyl-thiocholine iodide was used to regenerate Ac-CoA *in-situ*. Further work performed in the Fascione laboratory focused on an alternative acetyl-donor, SNAc. This compound was used in place of acetyl-thiocholine iodide, in the large-scale synthesis of CMP-Pse5Ac7Ac, described in 2.2.12 and 2.3.12.

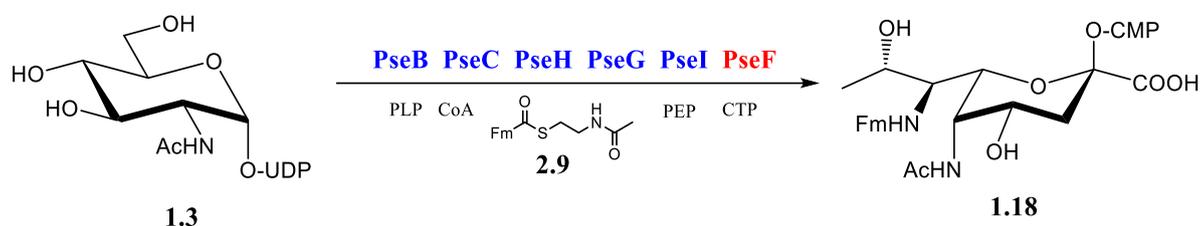
This methodology was further utilised within the Fascione laboratory in the chemoenzymatic synthesis of CMP-5-acetamido-3,5,7,9-tetra-deoxy-7-formyl-L-glycero-L-manno-nonulosonic acid (CMP-Pse5Ac7Fm **1.18**),<sup>††††</sup> a naturally occurring carbohydrate in several bacteria including drug-resistant *P. aeruginosa*.<sup>1,107</sup> This was achieved through the synthesis of N-formyl-cysteamine (SNFm **2.9**) which was used in place of SNAc, and exploited the substrate flexibility of *C. jejuni* PseH, PseG and PseI, in addition to *A. caviae* PseF. However, SNFm appeared to undergo an undesirable reaction with a PseC co-factor. Therefore, to improve yield, a two-step synthesis was performed, beginning with the UDP-GlcNAc, PseB and PseC reaction. Subsequent addition of PseH, PseG and PseI (and co-

---

\*\*\*\* Work performed by Dr Harriet Chidwick

†††† Work performed by Matthew Best, MChem student.

factors including SNFm) produced Pse5Ac7Fm **2.10** from the PseC product **1.13**. Finally, PseF and CTP were added to afford CMP-Pse5Ac7Fm (Figure 2.17). This reaction was performed on a 30 mg scale and the product was successfully purified and characterised, following procedures in line with those described in this chapter.



**Figure 2.17: Summary of reaction for chemoenzymatic synthesis of CMP-Pse5Ac7Fm 1.18.**

This strategy may be used to produce a range of N-R-cysteamine compounds where bioorthogonal groups may be installed as the R group, to be used as chemical handles in Pse5Ac7R and CMP-Pse5Ac7R probes. The success of this strategy is dependent upon substrate promiscuity of PseH and the further three enzymes of the CMP-Pse5Ac7Ac pathway. PseH will need to be able to utilize R-CoA in place of Ac-CoA.<sup>99</sup> In a crystallographic structure of *C. jejuni* PseH with Ac-CoA in the active-site, it appears there is space around the acetyl moiety of Ac-CoA, therefore larger groups may be accommodated.<sup>129</sup> It is also suggested acetyl transfer occurs via nucleophilic attack of Ac-CoA directly from the amino group of the carbohydrate substrate.<sup>129,130</sup> This may provide a mechanistic explanation as to how reactions with Ac-CoA derivatives are possible, as the reaction is initiated without the formation of a covalent Ac-CoA enzyme intermediate. Structural data for the fourth enzyme, PseG, suggests a large binding pocket exists for the carbon 4 group of the substrate **1.15**.<sup>109</sup> Therefore bulky C4 substituents of **1.15** derivatives, which will become carbon 7 in Pse-derivatives should be accommodated by PseG.

Unnatural Pse5Ac7R has previously been synthesized via a similar approach. In the work of Lui *et al.*, PseG and PseI utilized unnatural substrates, to produce azido-modified Pse5Ac7Ac derivative (Pse5Ac7AcN<sub>3</sub>), further highlighting the desired promiscuity of these enzymes.<sup>116</sup> The resulting Pse5Ac7AcN<sub>3</sub>, was used to label and visualize the flagella of *C. jejuni*.<sup>116</sup> Producing a range of derivatives may aid studies on the biological roles of Pse in clinical relevant bacteria.

Following successful synthesis of Pse5Ac7Ac, several small-molecules were screened for their ability to inhibit *C. jejuni* PseB and Psel in addition to *A. caviae* Psel. The results of this study were in line with those described by Ménard *et al.* where **2.4** was found to inhibit *C. jejuni* PseB.<sup>110</sup> PseB is an attractive target in the inhibition of Pse5Ac7Ac biosynthesis for several reasons. Firstly, as stated previously, *C. jejuni* PseB plays a role in both the CMP-Pse5Ac7Ac and pgl pathways. Secondly, PseB has been structurally characterised, therefore the structure-activity relationship of PseB and inhibitors may be examined via *in-silico* docking.<sup>110,131</sup>

The small-molecule screen described in 2.2.3 was a successful initial assay to test whether compounds **2.4-2.8** inhibit Psel and AcPsel. It is crucial to note that a large excess of inhibitor compound (0.1-5 mM concentrations) was used in comparison to that of Psel (2.45  $\mu$ M) and AcPsel (2.3  $\mu$ M), yet in samples where apparent Psel inhibition was observed, Pse5Ac7Ac production was not fully inhibited (in all samples excluding those shown in Figure 2.13 A 5 mM and C 5mM). One of the five compounds screened (**2.4**) has been previously screened against both *H. pylori* and *C. jejuni* Psel and was shown to be inhibitory.<sup>110</sup> The results described in 2.3.3, where *C. jejuni* Psel inhibition was monitored validated the results of Ménard *et al.*<sup>110</sup> Additionally, to my knowledge, this is the first small-molecule screen for AcPsel inhibition where promising results were observed.

As Psel converts 6-deoxyAltDiNAc **1.15** and PEP into Pse5Ac7Ac, phosphate is produced.<sup>99</sup> To obtain more quantitative data on the degree of Psel inhibition caused by compounds **2.4-2.8**, spectrophotometric phosphate release assays may be performed, in which  $\mu$ M phosphate concentrations can be detected. This assay would and was the approach used in the high-throughput screen for inhibitors described by Ménard *et al.*, where HpPsel was screened for inhibition.<sup>110</sup>

Future work on inhibitors of Psel should examine their *in vivo* efficacy in *C. jejuni* and *A. caviae*. A suitable study should analyze Pse5Ac7Ac content and survival of cells treated with these compounds. This assay is crucial as when Ménard *et al.* et al. used **2.4** in a cell-based assay, there was no impact on the growth of *C. jejuni* (monitored by OD<sub>600</sub> values).<sup>110</sup> This compound meets Lipinski's rule of five and has drug-like physicochemical properties for oral bioavailability.<sup>110</sup> However, it was suggested that the negative charge of the carboxylate moiety, resulted in poor membrane permeability.<sup>110</sup> As all the compounds

tested in this screen also contain a carboxylate moiety, similar results may be anticipated from cell-based assays.

PseI is currently not structurally characterized. Structural data for this enzyme could greatly aid inhibition studies, particularly the design of competitive inhibitors which bind to the enzymes active site.<sup>122</sup> Additionally, structural studies may provide insight into the mode of inhibition used by the compounds screen in 2.2.3 and account for variations in the extent of inhibition observed between PseI and AcPseI.

Efforts were made to solubilise CjPseF, using conditions which matched those described by Schoenhofen *et al.* and a range of alternative conditions.<sup>99</sup> However, all attempts failed to obtain soluble, pure CjPseF. Work then focused on AcPseF, which was purified and obtained in a yield of 13 mg L<sup>-1</sup> of *E. coli* culture. Despite success with AcPseF, it remains curious that pure CjPseF was not obtained and that it appeared to be insoluble.

Computational analysis and modelling of *H. pylori* PseF (HpPseF) were reported; the grand average of hydropathy (GRAVY) index of the enzymes amino acid content suggests that this enzyme will also be insoluble.<sup>132</sup> GRAVY calculations for each of the three PseF homologues suggest that *A. caviae* PseF is most likely to be soluble, with the least negative GRAVY index of -0.13 compared to values of -0.23 for CjPseF and -0.24 for HpPseF.<sup>133</sup>

CjPseF and PseF share 47.8% protein identity and HpPseF shares 36.7% sequence identity with PseF (sequences aligned using Clustal Omega). Whilst all three proteins lack apparent trans-membrane regions (sequences analysed using TOPCONS) and are predicted to be cytoplasmic (sequences analysed by PSORTb), it is worth considering that their function as carbohydrate-activating enzymes may suggest that they are membrane associated.<sup>134,135</sup> Pse5Ac7Ac may be activated by PseF in the cytoplasm, near the inner membrane, before being using in glycosylation reactions and transported through the periplasm to the cell surface, where it constitutes cell-surface glycans and glycoproteins.

Serendipitously, the insolubility of CjPseF in this study allowed for the characterisation of the previously putatively assigned PseF from *A. caviae*. In previous studies PseF (referred to as NeuA, throughout this literature) knockout mutant *A. caviae* strains showed a loss of motility, caused by a lack of polar-flagellin and an increase in the LD<sub>50</sub>, attributed to the loss of Pse5Ac7Ac from flagellin proteins and LPS in the mutants compared to wild type *A.*

*caviae*. A loss of Pse5Ac7Ac was observed in the PseF mutant strain.<sup>104</sup> However, unlike enzymes earlier in the biosynthesis pathway PseF, was previously biochemically uncharacterised, perhaps due to a lack of commercially available Pse5Ac7Ac substrate and associated challenges in synthesis.

Analysis of all PseF reactions showed the presence of the substrate, Pse5Ac7Ac and the product, CMP-Pse5Ac7Ac. To determine whether the presence of Pse5Ac7Ac in reaction mixtures is due to lack of conversion to product or due to hydrolysis of the CMP group from the product <sup>18</sup>O-labelled H<sub>2</sub>O could be added to reaction mixtures. In negative ion ESI mass spectrometry an m/z peak at 335 would indicate product is hydrolysed to Pse5Ac7Ac by <sup>18</sup>O-labelled H<sub>2</sub>O.

Additional biochemical studies may be carried out, including quantifying kinetic parameters for the reactions facilitated by this enzyme. As pyrophosphate is released from CTP when Pse5Ac7Ac is nucleotide-activated by PseF, a spectrophotometric assay may be used to quantify the release of pyrophosphate in reactions of PseF, to quantify kinetic parameters of this enzyme. This assay has been used previously the study enzymatic production of CMP-activated carbohydrates which utilize CTP, including CMP-sialic acid production.<sup>127</sup>

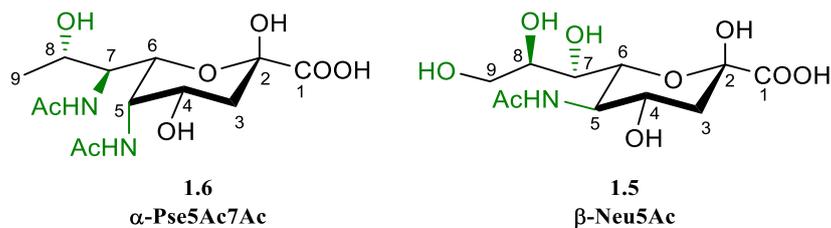
Before commencing structural trials with PseF, CD data was obtained to ensure that the protein was folded and may be amenable to crystallisation. For CD measurements the buffer composition must be carefully considered, as many common buffer components absorb strongly at the wavelength which are generally useful for protein CD (~180- 230 nm). The NaCl and MgCl<sub>2</sub> components of the PseF reaction buffer were omitted from the sample and buffer used for CD as chloride absorbs strongly at X nm. Therefore, the values for predicted secondary structure, may show variation from the overall extent of secondary structure that PseF has in reaction buffer or crystallisation buffer. However, as the data predicted that contained over 86% secondary structure, it is can be assumed that a large proportion of the enzyme maintains secondary structure when buffer is exchanged. This is consistent with computational data obtained for the homologous HpPseF which predicted that this homologue shows  $\alpha/\beta$  fold.<sup>132</sup>

PseF from *A. caviae* shows 27% sequence identity to NmCNS, for which a 2 Å resolution X-ray crystal structure has been solved. Substrate analogue CDP was present in crystals and

the binding of Neu5Ac **1.1** was docked into the structure.<sup>136</sup> Notably, NmCNS was shown to be homodimeric with residues from each monomer comprising the active site.<sup>137</sup> SEC-MALLS data obtained for PseF indicated that this enzyme is also homodimeric.

Amino acid sequences of NmCNS and PseF, along with several other CMP-Neu5Ac **1.5** synthetase homologues and PseF homologues, including CjPseF and HpPseF and also CMP-2-keto-3-deoxy-*manno*-octonic acid (CMP-Kdo) synthetase homologues were aligned. Structural data is available for several of the enzymes included in this alignment.<sup>126,127,137-139</sup> Alignment shows the conservation of several key residues where anticipated, such as those involved in binding the cytosine moiety of CTP. Other conserved residues include those interacting with regions of Neu5Ac where the structure matches that of Pse5Ac7Ac, for example residues interacting with Neu5Ac O4. However, the carbohydrate substrates for these enzymes show several structural differences, such as at the C5 position where the NHAc substituent is equatorial in Neu5Ac and axial in Pse5Ac7Ac (Figure 2.18 Pse5Ac7Ac **1.6** and Neu5Ac **1.1**). NeuA Glu-104 forms a hydrogen bond to the nitrogen of C5 in docked Neu5Ac and also interacts with O8, another position with opposing stereochemistry in Pse5Ac7Ac. Additionally Glu-104 is predicted to discriminate between the C6 propyl chain (C6 to C9) which can vary between sialic acids and is structurally different between Neu5Ac and Pse5Ac7Ac. The Glu-104 residue of NeuA is not well conserved in PseF where Tyr is present.<sup>137</sup> This difference in sequence may in-part account for altered specificity for carbohydrate substrate between these enzymes.

Mutations of key residues in PseF may guide biochemical characterisation and confirm how PseF confers specificity for Pse5Ac7Ac over structurally similar NuLOs, including Neu5Ac, as PseF was not seen to turnover Neu5Ac in this study. Additionally, structural data and mutations may provide insight into the range of Pse5Ac7Ac derivatives that PseF can utilise. Mutagenesis may expand the promiscuity of PseF, allowing for a range of CMP-Pse derivatives to be chemoenzymatically produced, using methodology described in this work.



**Figure 2.18: Structures of nonulosonic acids  $\alpha$ -Pse5Ac7Ac 1.6 and  $\beta$ -Neu5Ac 1.5. Variations in structures are highlighted in green**

NMR analysis of CMP-Neu5Ac synthesised using *E. coli* NeuA shows that the enzyme utilises  $\beta$ -Neu5Ac (equatorial carboxylic acid group) and retains stereochemistry of the anomeric carbon to produce  $\beta$ -CMP-Neu5Ac **1.5**.<sup>138</sup> NMR data confirmed that  $\alpha$ -Pse5Ac7Ac (equatorial carboxylic acid group) is retained by PseF to synthesise  $\alpha$ -CMP-Pse5Ac7Ac. This suggests that there will be similarities in the mechanism used by PseF, compared to that used by NeuA.

Perhaps the most valuable achievement of the work described in this chapter was the large scale chemoenzymatic synthesis of CMP-Pse5Ac7Ac. The synthesis described here shows improvement on that previously described by Schoenhofen *et al.* as the use of SNAc in place of stoichiometric quantities of Ac-CoA to acetylate **1.13**, greatly reduced the cost associated with synthesis of CMP-Pse5Ac7Ac (~£800 Ac-CoA would be required for the 90 mg reaction described in 2.2.12).<sup>99</sup> Further to this, the use of S-cysteamine derivatives provides a facile synthesis for CMP-Pse5Ac7R derivatives.

## 2.5 Conclusions and Future Work

The chemoenzymatic synthesis of CMP-Pse5Ac7Ac described in this chapter was crucial in facilitating further work of this study (Chapter 3 and 4). Access to Pse5Ac7Ac and CMP-Pse5Ac7Ac has been a hindrance in studies into the biological significance of Pse.

This study has provided, to my knowledge the first *in vitro* study of *A. caviae* PseF (previously NeuA), providing unequivocal evidence that this enzyme functions as an  $\alpha$ CMP-Pse5Ac7Ac synthetase. NMR characterisation confirmed the stereochemistry of this carbohydrate, revealing that  $\alpha$ -CMP-PseAc7Ac was produced, this data was crucial for subsequent conclusions of this study (Chapter 3 and 4).

Future work may revisit structural characterisation of PseF, where conditions used to crystallise the homologous NeuA may be screened.<sup>127</sup> The Pse5Ac7Ac and derivatives described in this work may be used in crystallisation of this enzyme. Additionally, PseF mutants may be produced in order to probe the mechanism and substrate selectivity of this enzyme.

Additionally, Psel and AcPsel may be the focus of crystallisation structures to further aid inhibition studies and the substrate selectivity of the enzymes

# Chapter 3: Sialyltransferase mediated synthesis of glycosides containing Pseudaminic Acid

## 3.1 Introduction

The biological roles of diacetamido-3,5,7,9-tetra-deoxy-L-glycero-L-*manno*-non-2-ulosonic acid (Pse5Ac7Ac) **1.6** and its derivatives remain relatively elusive, in part due to lack of access to Pse, glycosides containing Pse and Pse-based probes. Isolation of appreciable quantities of these pure sugars from bacterial isolates is challenging. As discussed in Chapter 1, the production of Pse5Ac7Ac or derivatives by fully synthetic methods is also rare, and although impressively elegant, all reported strategies are low yielding.<sup>5-9</sup> Current literature details one example of a trisaccharide containing Pse5Ac7Ac, produced by total synthesis.<sup>6</sup> When producing glycosides through chemical synthesis, an additional challenge occurs in controlling the stereochemistry of Pse-glycosidic linkages during glycosylation.<sup>141</sup>

Typically in the synthesis of aldose carbohydrates, protecting group chemistry is used at C2 of a glycosyl donor to selectively drive either inversion or retention of stereochemistry at the anomeric centre, during glycosylation.<sup>142</sup> However, ketose sugars such as Neu5Ac and Pse5Ac7Ac are generally glycosylated at the C2 hydroxyl and would therefore rely on the C3 moiety to direct the formation of either  $\alpha$  or  $\beta$  glycosides. However for nonulosonic acids, this is not achievable with traditional protecting groups as C3 is deoxy and therefore lacks neighboring functionality. Further to this, the presence of electron-withdrawing carboxylic acid at the anomeric position makes these donors prone to elimination, resulting in undesirable 2,3-glycal formation. As such, glycosylation of nonulosonic acids usually requires multi-step syntheses, which must be adapted depending on the identity of the acceptor carbohydrate and the desired stereo- and regiochemical outcome.<sup>6,141</sup>

Conversely, the highly stereo- and regioselective nature of enzymatic glycosylation make it an attractive route to Pse-glycosides. Glycosyltransferases (GTs) proposed to utilise cytidine 5'-monophosphate Pse (CMP-Pse) donors (pseudaminyltransferases) have been tentatively identified in a range of bacteria through *in vivo* studies and assignments of

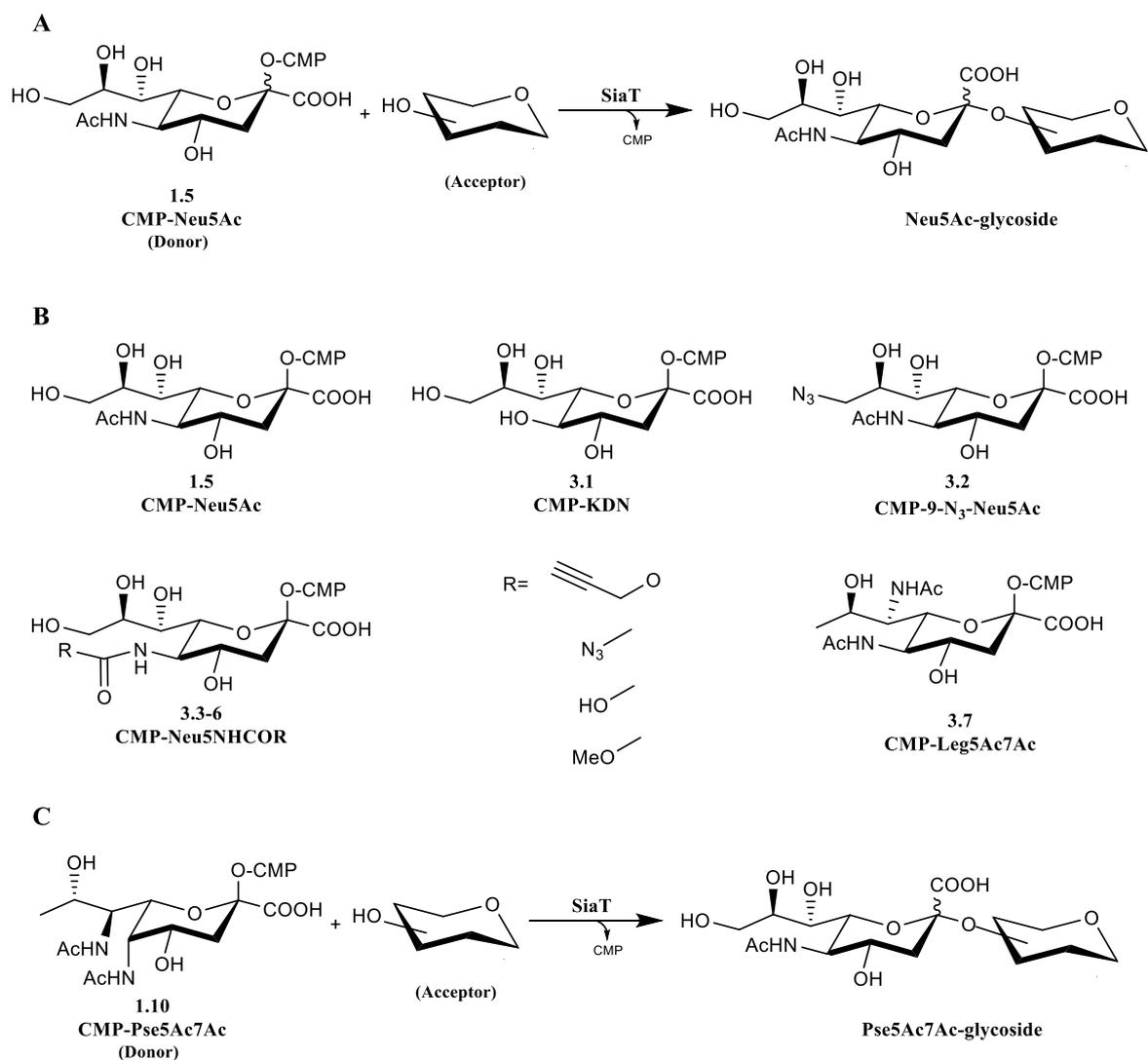
genes based upon homology (see Chapter 4).<sup>89,104,143,144</sup> However, at the commencement of this study, no direct *in vitro* enzymatic transfer of CMP-Pse had been reported.

*In vitro* studies of a number of well-characterised GTs which use CMP-Neu5Ac **1.5** (sialyltransferases (SiaTs)) (general reaction scheme shown in Figure 3.1 A) have been shown to display promiscuity in their utilisation of donor carbohydrates, ranging from CMP-Neu5Ac and a range of natural and bioorthogonal derivatives **3.1-3.6** to bacterial carbohydrate CMP-5,7-diacetamido-3,5,7,9-tetra-deoxy-D-glycero-D-galacto-non-2-ulosonic acid (CMP-legionaminic acid CMP-Leg5Ac7Ac) **3.7** (Figure 3.1 B). Leg5Ac7Ac is an isomer of Pse5Ac7Ac that possesses the D-glycero-D-galacto stereochemistry of Neu5Ac.<sup>81,128</sup>

Given the structural similarities between CMP-Neu5Ac, CMP-Leg5Ac7Ac and CMP-Pse5Ac7Ac (structures all seen in Figure 3.1), we reasoned that SiaTs may be used in the chemoenzymatic synthesis of Pse5Ac7Ac-based glycosides (proposed general reaction scheme shown in Figure 3.1 C). With access to a library of SiaTs, including those which show promiscuity towards donor, and access to CMP-Pse5Ac7Ac (synthesis described in Chapter 2) we were able to address the lack of practical syntheses of Pse-glycosides through chemoenzymatic methodology, concluding with the enzymatic transfer of CMP-Pse5Ac7Ac to a range of glycosyl acceptors using four promiscuous bacterial SiaTs to afford  $\beta$ -2,3/2,6-linked Pse5Ac7Ac di- and trisaccharides.<sup>36,40,145,146\*\*\*\*</sup>

---

\*\*\*\* Library available in Flitsch laboratory, MIB. Originally provided by Prozomix.



**Figure 3.1: A: General reaction scheme for glycosyltransfer facilitated by sialyltransferases. B: CMP-Nonulosonic acids utilised by promiscuous SiaTs. CMP-Neu5Ac 1.5, CMP-KDN 3.7, CMP-9-N<sub>3</sub>-Neu5Ac 3.2, CMP-Neu5NHCOR 3.3-6 and CMP-Leg5Ac7Ac 3.7. C: Proposed general reaction scheme for pseudaminyltransfer facilitated by sialyltransferases, using CMP-Pse5Ac7Ac 1.10 as donor substrate.**

## 3.2 Experimental

### 3.2.1 Initial Activity assays of sialyltransferase library with CMP-Pse5Ac7Ac donor

SiaTs were provided by ProZomix as ammonium sulphate precipitate stocks (Table 3.1 lists enzymes screened). 150  $\mu$ L of enzyme stock was centrifuged (15,000  $\times$  g, 5 min, 4  $^{\circ}$ C), supernatant discarded and pellet resuspended in 40 or 15  $\mu$ L dH<sub>2</sub>O, material was centrifuged (15,000  $\times$  g, 5 min, 4  $^{\circ}$ C) and the resultant supernatant contained SiaT. Concentrations were recorded using a nanodrop spectrophotometer.

For the activity screen of the SiaT library 10  $\mu$ L reactions containing SiaT (Table 1 for concentrations), 1 mM CMP-Pse5Ac7Ac **1.10**, 0.25 mM imidazolium-tagged N-Acetyl-D-lactosamine (LacNAc-ITag) **3.8**, 10 mM MgCl<sub>2</sub> and 0.1 U Pyrophosphatase (ThermoFisher) in 50 mM Tris HCl pH 7.5 were incubated at 37  $^{\circ}$ C for 18 hours and then analysed via matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) for the presence of  $\beta$ -Pse5Ac7Ac-2,3-LacNAc-ITag **3.9** or  $\beta$ -Pse5Ac7Ac-2,6-LacNAc-ITag **3.10** ( $M^+$   $m/z$  984.5).<sup>§§§§</sup> Samples containing *Photobacterium leiognathi*  $\alpha$ -2,6-SiaT (PIST), *Pasteurella dagmatis*  $\alpha$ -2,3-SiaT (PdST), *Pasteurella multocida*  $\alpha$ -2,3-/ $\alpha$ -2,6-SiaT (PmST) and *Photobacterium* sp. JT-ISH-224  $\alpha$ -2,3-SiaT (PJT-ISH-224ST) were subsequently analysed by negative ion mode electrospray ionisation liquid chromatography-mass spectrometry (-ESI LC-MS (LC-MS)).

---

<sup>§§§§</sup> MALDI-TOF MS data was collected by Dr Kun Huang, Manchester institute of Biotechnology, The University of Manchester

**Table 3.1: Sialyltransferases used in initial activity screen with CMP-Pse5Ac7Ac and LacNAc-ITag, known or predicted functions of the enzymes and final concentrations used in reactions**

| Source organism of Sialyltransferase | Function (known or predicted)                           | Final concentration in reaction / mg mL <sup>-1</sup> |
|--------------------------------------|---|---|
| <i>Photobacterium leiognathi</i>     | $\alpha$ -2,6-sialyltransferase <sup>146</sup>          | 0.76  |
| <i>Pasteurella dagmatis</i>          | $\alpha$ 2,3-sialyltransferase                          | 0.86  |
| <i>Pasteurella multocida</i>         | $\alpha$ -2,3-sialyltransferase                         | 0.90  |
|                                      | $\alpha$ -2,6-sialyltransferase <sup>145</sup>          |   |
| <i>Photobacterium</i> sp. JT-ISH-224 | $\alpha$ -2,3-sialyltransferase                         | 2.89  |
| <i>Neisseria meningitidis</i> MC58   | putative $\alpha$ -2,3-sialyltransferase                | 0.260   |
| <i>Photobacterium</i> sp. JT-ISH-224 | $\alpha$ -2,6-sialyltransferase <sup>147</sup>          | 1.86  |
| <i>Haemophilus influenzae</i>        | putative $\alpha$ -2,3-sialyltransferase                | 0.59  |
| <i>Neisseria gonorrhoeae</i>         | $\alpha$ -2,6-sialyltransferase                         | 0.34  |
| <i>Vibrio</i> sp. JT-FAJ-1           | $\alpha$ -2,3-sialyltransferase                         | 0.150   |
| <i>Neisseria lactamica</i>           | putative $\alpha$ -2,3-sialyltransferase                | 0.250   |
| <i>Campylobacter coli</i>            | putative $\alpha$ -2,3-sialyltransferase                | 0.340   |
| <i>Campylobacter jejuni</i>          | putative $\alpha$ -2,3/ $\alpha$ -2,8-sialyltransferase | 0.976   |
| <i>Campylobacter insulaenigrae</i>   | putative $\alpha$ -2,3-sialyltransferase                | 4.47  |

### 3.2.2 PmST acceptor screen with CMP-Pse5Ac7Ac as donor

PmST was provided by Prozomix as ammonium sulphate precipitate and prepared as described in 3.2.1. To screen the activity of PmST with various acceptors and CMP-Pse5Ac7Ac donor, 25  $\mu$ L reactions contained 0.130 mg mL<sup>-1</sup> PmST, 2 mM CMP-Pse5Ac7Ac **1.10** and 8 mM acceptor in 50 mM Tris HCl pH 7.5. One of the following acceptors was added to each reaction: N-acetyl-D-lactosamine (LacNAc) **3.11**,  $\beta$ -D-galactopyranosyl-1,4-D-glucose (lactose (Lac)) **3.12**, D-Galactopyranose (Galactose (Galp)) **3.13**, methyl  $\beta$ -D-galactopyranoside (Me  $\beta$ -D-Galp) **3.14**, 6-deoxy-6-fluoro-D-Galactose (6-F-Galp) **3.15**, 4-nitrophenyl- $\beta$ -D-galactopyranoside (*p*NP- $\beta$ -D-Galp) **3.16**. Reactions were performed in duplicate and control reactions without either PmST, CMP-Pse5Ac7Ac or acceptor were also performed. All samples incubated at 37 °C for 4 hours and were analysed via LC-MS (details in Appendix-General methods).

### 3.2.3 Large Scale synthesis of $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp using PmST

PmST was prepared from ammonium sulphate precipitate as described in 3.2.1. 2 mM CMP-Pse5Ac7Ac (9.5 mg), was added to a solution containing 50 mM Tris HCl pH 7.5, 8 mM *p*NP- $\beta$ -D-Galp, 110  $\mu$ g mL<sup>-1</sup> PmST, 100  $\mu$ g mL<sup>-1</sup> PseF (prepared as described in 2.2.X), 2 mM cytidine-5'-triphosphate (CTP) and 10 mM MgCl<sub>2</sub>, in a final volume of 7.5 mL. The reaction mixture was incubated at 37 °C for 18 hours, and the reaction was monitored by -ESI LC-MS as described for samples in 3.2.2. The reaction mixture was centrifuged (37,000  $\times$  g, 4 °C, 30 minutes) to remove enzyme precipitate and the resultant supernatant was lyophilised and resuspended in 10 mL dH<sub>2</sub>O and cotton wool filtered to remove enzyme precipitate. The filtrate was then mixed with 10 mL EtOH and stored at 4 °C for 30 minutes to precipitate any remaining enzyme. The mixture was cotton wool filtered, and the filtrate was diluted in dH<sub>2</sub>O to a final volume of 200 mL before lyophilisation. The lyophilised material was resuspended in 5 mL dH<sub>2</sub>O and applied to a 500 mL column packed with Bio-Gel® P-2 resin (Biorad) in HPLC-grade H<sub>2</sub>O. 3 mL fractions were collected for 40 hours and analysed via LC-MS for the presence of  $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp **3.17** ([M-H]<sup>-</sup> m/z: 616.2). Fractions containing Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp were pooled and lyophilised. Following NMR analysis of the resulting material (D<sub>2</sub>O, 700MHz), the material was dry loaded onto silica and purified via flash chromatography using EtOH:MeOH:dH<sub>2</sub>O (5:2:1). 2 mL fractions were collected and analysed by TLC (sugar stain: MeOH, 5% H<sub>2</sub>SO<sub>4</sub> and

charring) and LC-MS (details in Appendix-General methods) for the presence of Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp. Solvent was evaporated from desired fractions which were then pooled and analysed by NMR (D<sub>2</sub>O, 700 MHz).

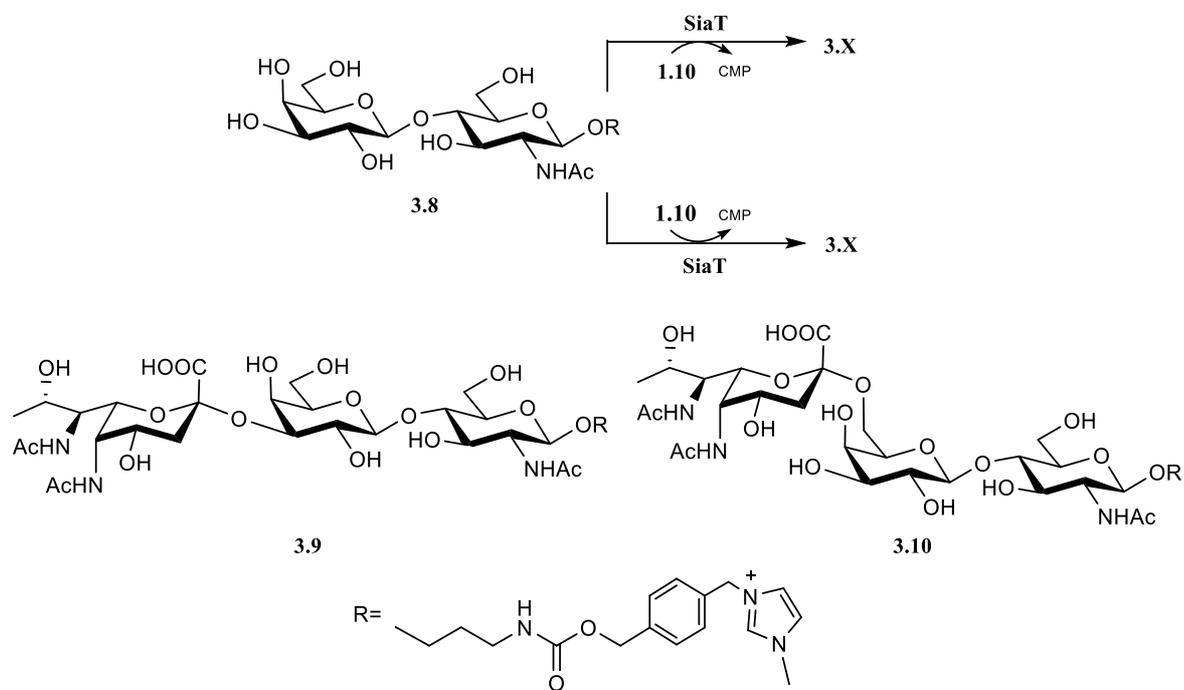
<sup>1</sup>H NMR (700 MHz, Deuterium Oxide)  $\delta$  8.29 (d,  $J$  = 9.3 Hz, 2H, ArH), 7.28 (d,  $J$  = 9.4 Hz, 2H, ArH), 5.32 (d,  $J$  = 7.9 Hz, 1H, H1'), 4.26-4.24 (m, 2H, H3', H6), 4.19-4.18 (m, 1H, H5), 4.12 (t,  $J$  = 3.1 Hz, 1H, H8), 4.05-4.02 (m, 2H, H4', H7), 3.91 (m, 1H, H2'), 3.91-3.90 (m, 1H, H5'), 3.88-3.86 (m, 1H, H4), 3.78 (d,  $J$  = 6.1 Hz, 2H, H6'), 2.58 (dd,  $J$  = 12.6, 4.3 Hz, 2H, H3eq), 2.01 (s, 3H, C5-NHAc), 1.95 (s, 3H, C7-NHAc), 1.69 (t,  $J$  = 12.7 Hz, 1H, H3ax), 1.33 (d,  $J$  = 6.9 Hz, 3H). <sup>13</sup>C NMR (176 MHz, D<sub>2</sub>O)  $\delta$  174.8, 173.6, 173.5, 161.7, 142.6, 126.1, 116.4, 100.1, 99.7, 75.5, 75.1, 73.5, 69.0, 68.8, 68.5, 64.8, 60.7, 53.7, 48.3, 35.85, 22.1, 21.9, 20.0.  $[\alpha]_{D25} = 208.4$  (c 0.08, H<sub>2</sub>O). HR-MS data was collected with -ESI MS: Expected [M-H]<sup>-</sup> m/z: 616.1995, measured [M-H]<sup>-</sup> m/z: 616.2008. ATR-FTIR  $V_{max}$ : 3354, 2971, 2891, 1591, 1406, 1062 cm<sup>-1</sup>. 2.1 mg.  $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp **3.17** was obtained as an off-white solid (23% yield).

## 3.3 Results

### 3.3.1 Initial Activity assays of sialyltransferase library with CMP-Pse5Ac7Ac donor

A library of bacterial SiaTs were screened for their ability to utilise CMP-Pse5Ac7Ac **1.10** as donor, with LacNAc-ITag **3.8** used as the acceptor as the positively charged ITag has previously been used to identify low level product formation in enzymatic reactions monitored by MALDI-TOF-MS (Scheme in Table 3.2).<sup>148</sup> Following overnight incubation the reactions were analysed by MALDI-TOF-MS for the presence of peaks corresponding to  $\beta$ -2,3-Pse5Ac7Ac-LacNAc-ITag **3.9** or  $\beta$ -2,6-Pse5Ac7Ac-LacNAc-ITag **3.10**. It was concluded that four of the 13 SiaTs assayed were active under these conditions, due to the presence of peaks at  $m/z$  985.4 in the MALDI data for these samples, which corresponds to either **3.9** or **3.10** (results summarised in Table 3.2- green text denotes SiaT activity was observed, red text denotes SiaT activity was not observed) (MALDI spectra in Appendix). MALDI spectra for all thirteen samples showed that LacNAc-ITag **3.8** remained in the samples, seen as a positively charged molecular ions peak ( $M^+$ ) at  $m/z$  669.3. As this was the limiting reagent (4:1 equivalents of **1.10** to **3.8**) this suggests that successful reactions did not go to completion. MALDI data also revealed ITag-GlcNAc ( $m/z$  507.2), suggesting that each of the SiaT preparations also contained  $\beta$ -galactosidase, which resulted in the hydrolysis of Gal from LacNAc-ITag. Samples which contained **3.9** or **3.10** were also subjected to  $-ESI$  LC-MS analysis, which revealed that no detectable CMP-Pse5Ac7Ac remained in these samples. Pse5Ac7Ac was present in each sample, which may have been the hydrolysis product of CMP-Pse5Ac7Ac or due to hydrolysis of  $\beta$ -2,3- or  $\beta$ -2,6-Pse5Ac7Ac-LacNAc-ITag. It was concluded that PIST, PdST, PmST and PJT-ISH-224ST were able to function as pseudo-pseudaminyltransferases.

**Table 3.2: Results of Initial Sialyltransferase activity screen with  $\alpha$ -CMP-Pse5Ac7Ac 1.10 and LacNAc-ITag 3.8, for the production of  $\beta$ -2,3-Pse5Ac7Ac-LacNAc-ITag 3.9 or  $\beta$ -2,6-Pse5Ac7Ac-LacNAc-ITag 3.10, as shown in reaction scheme. Green enzymes: active; Red enzymes: no activity observed.**



| Sialyltransferase source organism    | Predicted Product | Reference |
|--------------------------------------|-------------------|-----------|
| <i>Pasteurella multocida</i>         | <b>3.9</b>        | 145       |
| <i>Photobacterium sp. JT-ISH-224</i> | <b>3.9</b>        | 149       |
| <i>Photobacterium leiognathi</i>     | <b>3.10</b>       | 150       |
| <i>Pasteurella dagmatis</i>          | <b>3.10</b>       | 151       |
| <i>Haemophilus influenzae</i>        | <b>3.9</b>        | N/A       |
| <i>Neisseria lactamica</i>           | <b>3.9</b>        | N/A       |
| <i>Campylobacter jejuni</i>          | <b>3.9</b>        | N/A       |
| <i>Campylobacter insulaenigrae</i>   | <b>3.9</b>        | N/A       |
| <i>Neisseria meningitidis MC58</i>   | <b>3.9</b>        | N/A       |
| <i>Vibrio sp. JT-FAJ-1</i>           | <b>3.9</b>        | 39        |
| <i>Campylobacter coli</i>            | <b>3.9</b>        | N/A       |
| <i>Neisseria gonorrhoeae</i>         | <b>3.10</b>       | NA        |
| <i>Photobacterium sp. JT-ISH-224</i> | <b>3.10</b>       | 78        |

### 3.3.2 PmST acceptor screen with CMP-Pse5Ac7Ac as donor

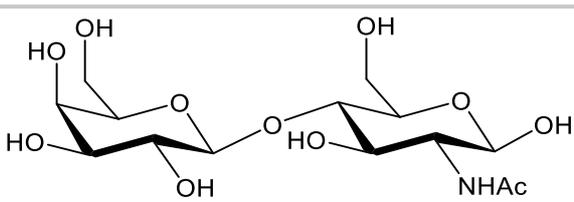
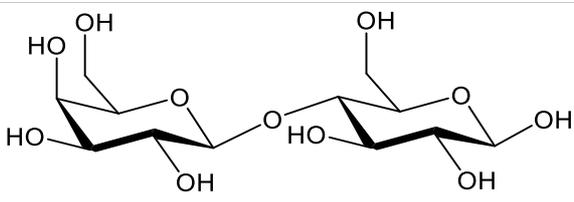
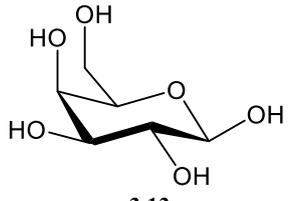
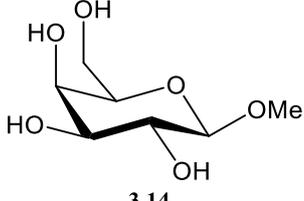
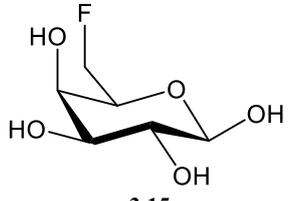
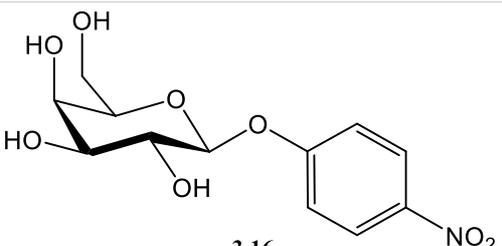
Following the positive results of the initial SiaT activity screen PmST was used with CMP-Pse5Ac7Ac **1.10** to screen a range of galactose derivative acceptors including LacNAc **3.11**, Lac **3.12**, D-Gal **3.13**, Me- $\beta$ -D-Galp **3.14**, 6F-Gal **3.15** or pNP- $\beta$ -D-Galp **3.16**, producing Pse5Ac7Ac galactosides. Reactions at pH 7.5, incubated at 37 °C for 4 hours and were analysed via LC-MS. Standard curves of Pse5Ac7Ac and CMP-Pse5Ac7Ac on -ESI LC-MS and were used to calculate conversion of donor and acceptor to  $\beta$ -2,3-Pse5Ac7Ac-glycoside product. As each reaction contained an initial concentration of 2 mM CMP-Pse5Ac7Ac, the standards curves were used to determine the remaining concentration of CMP-Pse5Ac7Ac and the concentration of any Pse5Ac7Ac, the conversion to Pse5Ac7Ac-glycoside was deduced from these values.

Conversions of CMP-Pse5Ac7Ac to  $\beta$ -2,3-Pse5Ac- $\beta$ -D-galactoside di- or tri-saccharide products ranged from 57%-91% (results summarised in Table 3.3). (Standard concentration curves of CMP-Pse5Ac7Ac and Pse5Ac7Ac, used to calculate conversions can be found in Appendix.) LC-MS analysis showed that, with the exception of the no donor sample, each of the samples contained free Pse5Ac7Ac, observed as a 332.9 m/z ([M-H]<sup>-</sup>). Previous studies of PmST revealed that the enzyme shows sialidase activity towards  $\alpha$ -2,3-sialyl-galactosides, therefore some of the Pse5Ac7Ac found in these samples may be due to hydrolysis of  $\beta$ -2,3-Pse5Ac- $\beta$ -D-galactoside product.<sup>149</sup> However, the reactions were performed above the optimal sialidase activity of PmST (optimal activity at pH 5.0-5.5, reactions were performed at pH 7.5) therefore, it is likely that a large proportion of the Pse5Ac7Ac present in these samples was due to hydrolysis of CMP-Pse5Ac7Ac.<sup>149</sup>

A third function of PmST is that it displays  $\alpha$ -2,6-SiaT activity with CMP-Neu5Ac **1.5** at pH 4.5- 7.0 with less efficiency than the  $\alpha$ -2,3-SiaT activity of the enzyme.<sup>149</sup> As LC-MS analysis reveal that PmST was able to utilise CMP-Pse5Ac7Ac donor and Gal acceptor at pH 7.5, where literature suggests that the enzyme affords 2,3 linked Pse5Ac7Ac galactosides, 6F-Gal was used in this acceptor screen to provide evidence in support of this. No product would be observed in the sample containing 6F-Gal, if PmST was producing 2,6 linked Pse5Ac7Ac galactosides. As anticipated, a peak at 497.0 m/z, corresponding to the molecular ion of Pse5Ac7Ac-6F-Gal was present in LC-MS data, suggesting that C6 hydroxyl

of galactose-based acceptors was not used in glycosidic bond formation under these conditions.

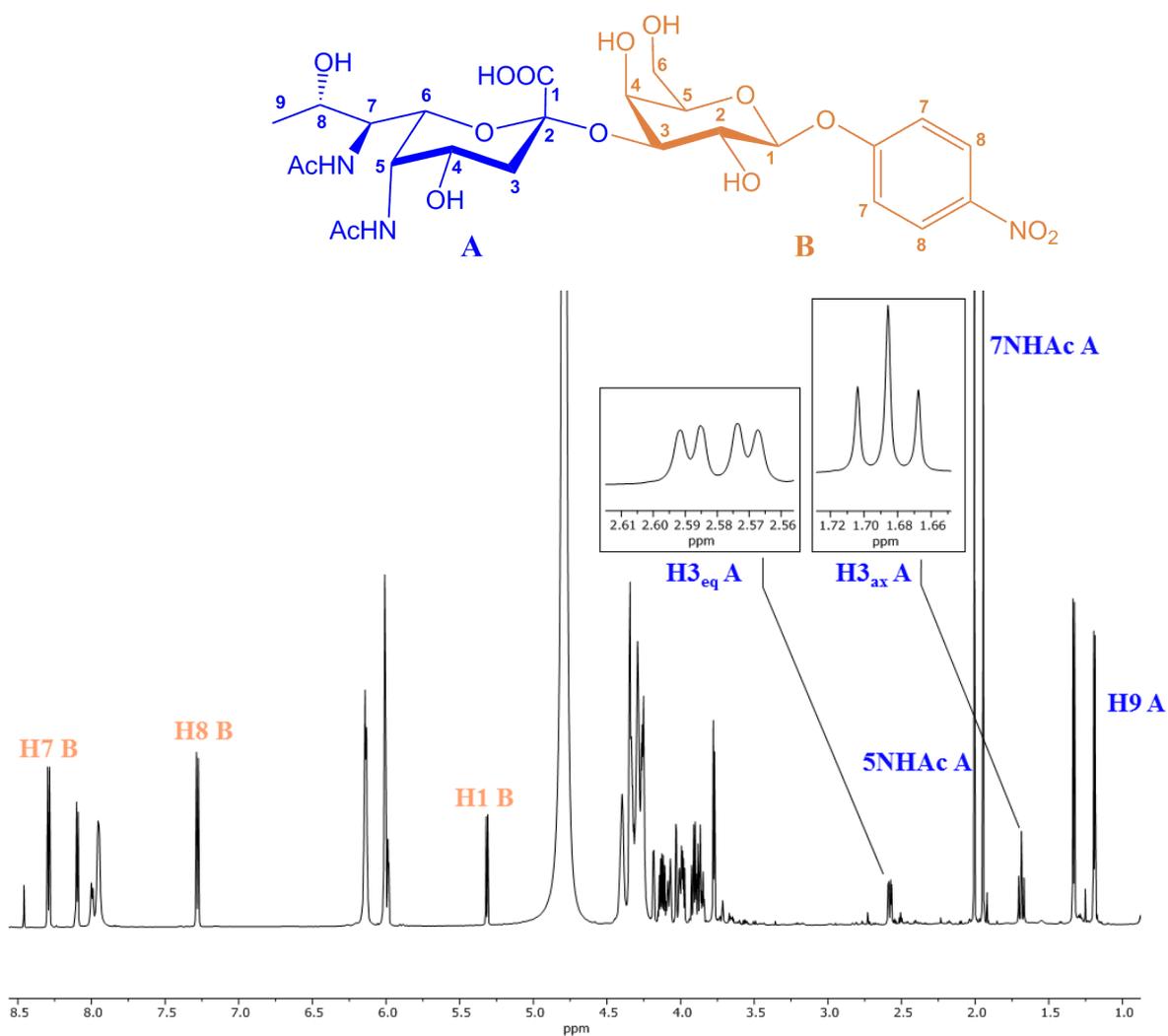
**Table 3.3: Acceptor Screen (compounds 3.11-3.16) of *P. multocida* SiaT (PmST) with CMP-Pse5Ac7Ac as donor. Samples analysed by LC-MS and conversion to  $\beta$ -2,3-Pse5Ac7Ac galactoside calculated using EIC of Pse5Ac7Ac and CMP-Pse5Ac7Ac remaining in samples.**

| Acceptor  | Conversion to $\beta$ -2,3-Pse5Ac7Ac galactoside |
|---|--|
|  <p>3.11</p>   | 91%  |
|  <p>3.12</p>   | 90%  |
|  <p>3.13</p> | 78%  |
|  <p>3.14</p> | 80%  |
|  <p>3.15</p> | 57%  |
|  <p>3.16</p> | 91%  |

### 3.3.3 Large Scale synthesis of $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp using *Pasteurella multocida* SiaT

Following the results of the acceptor screen, in order to unequivocally confirm the stereochemistry and linkage of the products of reactions mediated by PmST, the reaction between CMP-Pse5Ac7Ac **1.10** and *p*NP- $\beta$ -D-Galp **3.16** was performed using 9.5 mg CMP-Pse5Ac7Ac, such that the product could be purified and analysed by NMR. To drive the formation of Pse5Ac7Ac-glycoside product, CMP-Pse5Ac7Ac synthetase, PseF was added to the reaction so that hydrolysed Pse5Ac7Ac could be reactivated with a CMP group. The reaction incubated for 18 hours and progress was monitored via LC-MS. LC-MS analysis showed conversion to  $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp **3.17** was achieved but that CMP-Pse5Ac7Ac remained however, enzyme precipitation was observed so the reaction was stopped at this time. Enzymes were removed from the reaction by lyophilisation, ethanol precipitation and filtration, and the enzyme free material was purified by gel-filtration and subsequent silica flash column chromatography.

$^1\text{H}$ -NMR analysis confirmed that the product did indeed contain a  $\beta$ -Pse5Ac7Ac linkage with the stereochemistry determined by the 0.9 p.p.m. difference between  $\delta$  values for H-3<sub>ax</sub> and H-3<sub>eq</sub> peaks in the  $^1\text{H}$  NMR spectrum (Figure 3.2 -reported as  $\sim 0.6$  p.p.m. for equatorial carboxy group and  $\sim 0.9$  p.p.m for axial carboxy group).<sup>106,128</sup> As PmST is known to function as an inverting GT, this result was anticipated as the CMP-Pse5Ac7Ac starting material shows  $\alpha$ -stereochemistry (Chapter 2.3.9).<sup>145</sup>



**Figure 3.2: 700 MHz  $^1\text{H}$  NMR Spectrum of  $\beta$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp 3.17, in D<sub>2</sub>O.  $^1\text{H}$  peaks for H9, H3<sub>ax</sub>, H3<sub>eq</sub>, 5NHAc, 7NHAc shown (blue labels). *p*NP- $\beta$ -D-Galp H1, H7 and H8 shown (orange labels).**

### 3.4 Discussion

Galan and co-workers previously demonstrated that ionic liquid ITag substrates could be used in sensitive enzymatic reactions and subsequently analysed by MS.<sup>148</sup> On the basis of these results ITag-LacNAc was used with two commercially available SiaTs to produce  $\alpha$ -2,3- and  $\alpha$ -2,6-linked Neu5Ac-LacNAc-ITag trisaccharide.<sup>152</sup> For the initial SiaT screen of this study we used LacNAc-ITag as acceptor as ILs have been shown to give greater peak intensities and lower limits of detection in MS, therefore low level turnover of CMP-Pse5Ac7Ac to Pse5Ac7Ac-LacNAc-ITag product should be detectable in MALDI analysis.<sup>153</sup>

The work of Sittel and Galan had confirmed that PmST and PdST were able to utilise LacNAc-ITag with CMP-Neu5Ac **1.5** as donor.<sup>152</sup> Both of these enzymes showed activity with CMP-Pse5Ac7Ac in the initial screen (3.2.1, results described in 3.3.1). However, during this initial activity screen, no positive control for SiaT activity was collected. Unlike the PmST and PdST SiaTs, some of the enzymes assayed had not been shown to be active with CMP-Neu5Ac (activity listed as putative in Table 3.1), therefore some of the negative hits from this assay should be treated with caution. For the screen described in 3.2.1, CMP-Neu5Ac could have been used as donor to confirm that these enzymes were active and able to utilise LacNAc-ITag as an acceptor. Further to this, the concentrations of enzymes used in this initial screen ranged from 0.15 to 4.47 mg mL<sup>-1</sup>, and activity was seen in samples with 0.76 to 2.89 mg mL<sup>-1</sup> of enzymes. Therefore, a lack of activity in some of the samples may be due to the sample not containing sufficient enzyme to produce a detectable quantity of Pse5Ac7Ac-LacNAc-ITag in the assay. The pH (7.5) and temperature (37 °C) may not have been optimal for some of these SiaTs, therefore this could also be varied if the screen were to be repeated. Additionally, alternative ITagged acceptor carbohydrates may have been screened. However, as four positive hits were obtained with CMP-Pse5Ac7Ac, work with these enzymes was pursued and no further efforts were made to confirm the activity of the remaining SiaTs with CMP-Neu5Ac or CMP-Pse5Ac7Ac under alternative conditions.

CMP-sugars are known to be unstable and rapidly hydrolyse, releasing the CMP group, therefore this can in part account for the free Pse5Ac7Ac observed in samples. However, three of the four SiaTs which utilised CMP-Pse5Ac7Ac have previously been shown to display sialidase activity, cleaving Neu5Ac from glycosides. Pse5Ac7Ac seen in the reaction

containing PIST (performed at pH 7.5) may in part be due to the enzymes sialidase activity, which had previously been shown to display around 20-30% activity at pH 7.0 and pH 8.0.<sup>146</sup> PmST sialidase activity is not reported to have been observed at the pH of this reaction (pH 7.5) with optimal activity reported at pH 5.0-5.5 therefore, the Pse5Ac7Ac found in these samples is therefore likely as a result of CMP-Pse5Ac7Ac hydrolysis.<sup>43</sup> Similarly PdST sialidase is optimal at pH 4.5, given that the reaction with CMP-Pse5Ac7Ac were performed at pH 7.5, the majority of the Pse5Ac7Ac in the sample is again likely due to starting material hydrolysis.<sup>36</sup> Further to this, in the large scale synthesis of  $\beta$ -2,3-Pse5Ac7Ac- $\beta$ -D-Galp mediated by PmST, PseF and CTP were used to drive the conversion of Pse5Ac7Ac to CMP-Pse5Ac7Ac. Whilst hydrolysis of starting material and product may be a caveat of these enzymatic syntheses, this methodology resulted in the synthesis of a number of potentially biologically relevant compounds, where acceptor moiety can easily be altered, unlike in traditional chemical synthesis.<sup>141</sup>

The carbohydrate structure database, currently lists 27 unique, naturally occurring oligosaccharides that contain  $\beta$ -Pse5Ac7Ac or  $\beta$ -Pse-derivative which are linked to galactose via a 2,3- or 2,6- glycosidic bond.<sup>154</sup> These structures range from cell surface glycans such as lipopolysaccharide and exopolysaccharide to the teichuronic acid component of *Kribbella* cell walls.<sup>144,155,156</sup> Cell surface oligosaccharides containing  $\beta$ -Pse5Ac7Ac-2,3-Galp and  $\beta$ -Pse5Ac7Ac-2,6-Galp have been found to occur in drug-resistant bacteria, from clinical isolates.<sup>144,157,158</sup> Therefore, the novel  $\beta$ -Pse5Ac7Ac-2,3-Galp and  $\beta$ -Pse5Ac7Ac-2,6-Galp based carbohydrates synthesised in this study may be valuable tools in work towards new therapeutics.

Sialic acids within bacteria are only known to exist as  $\alpha$ -linked glycosides however, CMP-sialic acids are biosynthesised as  $\beta$ -anomers. All sialic acid GTs SiaTs can be classified as inverting.<sup>69</sup> Each of the SiaTs in this study therefore produces  $\alpha$ -linked Neu5Ac glycosides and where possible,  $\beta$ -linked Pse5Ac7Ac-glycosides (equatorial glycosidic bonds in both Neu5Ac glycosides and Pse5Ac7Ac glycosides). A number of multi-drug resistant bacterial strains from clinical isolates are known to contain  $\alpha$ -Pse-glycosides in cell surface structures, which are crucial for virulence.<sup>159-161</sup> Therefore a more comprehensive toolkit for Pse-glycoside synthesis should ideally be able to produce Pse-glycosides in which the stereochemistry of C1 is retained with respect to CMP-Pse5Ac7Ac, yielding  $\alpha$ -linked Pse

products. This could be achieved through the use of retaining GTs, however to date no retaining SiaTs have been shown to be active *in vitro*. This issue of access to  $\alpha$ -Pse-glycosides, via chemoenzymatic routes is addressed in Chapter 4.

This work has demonstrated the ability of four promiscuous SiaTs to utilise non-natural substrate CMP-Pse5Ac7Ac to generate  $\beta$ -2,3- and  $\beta$ -2,6-Pse5Ac7Ac-glycosides. The products of this screen bear a reactive ITag handle which has been shown to be a useful tool in the studies of the biological roles of carbohydrates.<sup>162</sup> Further to this, the highly promiscuous, well characterised, stable, PmST was shown to utilise a range of acceptors with CMP-Pse5Ac7Ac **1.10**.<sup>43</sup> Structural data is available for three of the four enzymes (PmST, PdST and PJT-ISH-224ST) which displayed activity with CMP-Pse5Ac7Ac.<sup>36,38,147,163</sup> Structural analysis of these enzymes revealed that they all have overall GT-B topology.<sup>36,38,147,163</sup> It has previously been noted that GT-B family enzymes generally have greater promiscuity for substrates than GT-B family transferases.<sup>38</sup>

Crystal structure data of PmST and *Photobacterium* sp. JT-ISH-224  $\alpha$ -2,6-SiaT has previously been used to rationally redesign the active site of PdST such that the enzyme formed  $\alpha$ -Neu5Ac-2,6- $\beta$ -D-galactosides instead of the  $\alpha$ -2,3- wild type activity. This provides precedent for the redesign of activity sites of these well-characterised, highly-stable SiaTs, it is feasible that structural data and site-directed mutagenesis may be used to further engineer these enzymes, enabling a broader range of NuO-glycosides to be enzymatically synthesised.

Given the diversity of CMP-nonulosonic acids that PmST has been previously shown to utilise, it is highly feasible that this enzyme may be used to synthesise a potentially vast range of glycosides incorporating Pse5Ac7Ac or Pse derivatives. Notably, PmST has shown activity for CMP-sialic acids with a range of N-linked functionalities at C5, this is promising for the use of this enzyme with CMP-Pse-derivatives which show a range of N-linked functionalities at this position.<sup>2,145</sup> In addition the crystal structure of this enzyme has been solved with substrate analogues bound in the active-site, this data could potentially be used to engineer the active site to turnover acceptor substrates not currently accommodated by the enzyme or to alter the regioselectivity of the glycosidic bond produced, expanding the library of Pse-based glycosides that can be generated using this approach.<sup>164</sup>

### 3.5 Conclusions and Future Work

This work presented here provides the first examples of chemoenzymatically synthesised Pse5Ac7Ac containing glycosides. Considering that at the commencement of this study no native pseudaminyltransferases had yet been shown to be active *in vitro*, these well-characterised, highly stable, promiscuous SiaTs, which display activity with CMP-Pse5Ac7Ac provide a promising route to synthesis of a range of novel and biologically relevant Pse-glycosides.

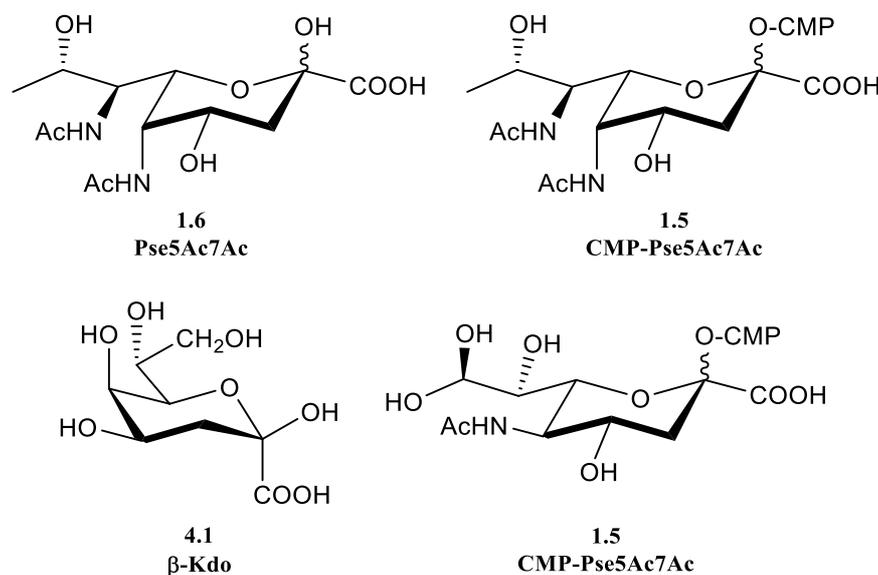
Crystal structure data of PmST with various ligands in the active site are available, including a structure with Lac acceptor and non-hydrolyzable donor substrate analogue, CMP-3-deoxy-3-fluoro-Neu5Ac (CMP-3F-Neu5Ac) both present in the active site.<sup>38,164</sup> Future work could focus on crystallisation of PmST with CMP-3F-Pse5Ac7Ac to gain insight into the molecular basis of donor substrate promiscuity that has been demonstrated in previous studies and expanded on in this work with CMP-Pse5Ac7Ac.

Chapter 2 described a strategy developed in the Fascione laboratory for the chemoenzymatic synthesis of CMP-Pse5Ac7R derivatives. Future work may examine whether any of the SiaT which were shown to be active with CMP-Pse5Ac7Ac are able to utilise these CMP-Pse derivatives as donor substrates, potentially expanding the range of Pse-glycosides which can be synthesised using the approach described in this study.

# Chapter 4: Characterisation of Putative Pseudaminyltransferases

## 4.1 Introduction

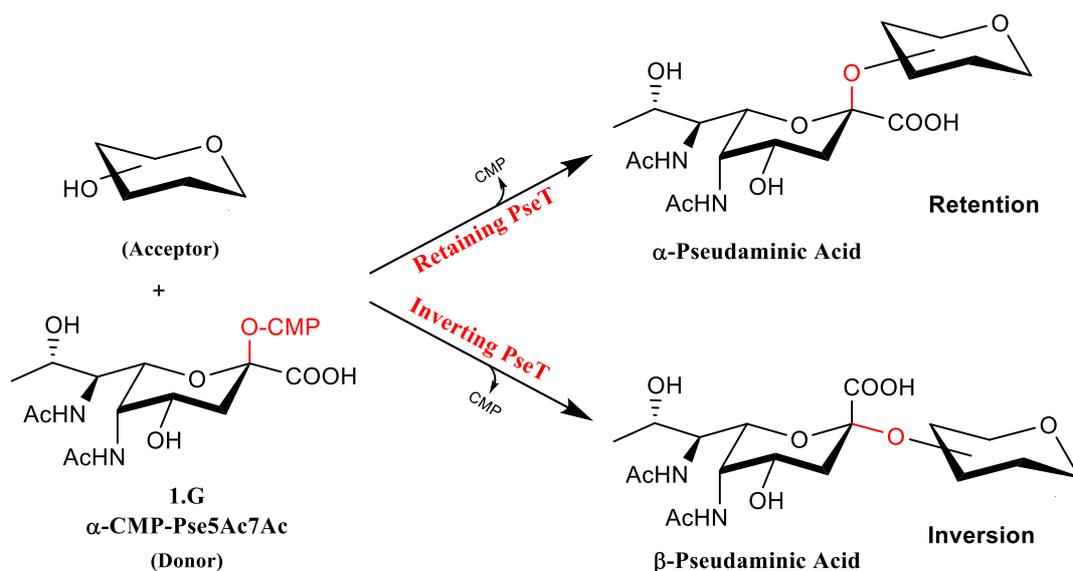
In contrast to the detailed characterisation of PseB, C, G, H, and I enzymes in the biosynthesis of 5,7-diacetamido-3,5,7,9-tetra-deoxy-L-glycero-L-manno-non-2-ulosonic acid (Pse5Ac7Ac **1.6** Figure 4.1), knowledge of pseudaminic acid (Pse) processing enzymes is lacking. Glycosyltransferases (GTs) for which cytidine 5'-monophosphate Pse5Ac7Ac (CMP-Pse5Ac7Ac **1.10** Figure 4.1) or a similar CMP-linked Pse derivative are the substrate, are necessary for the incorporation of Pse into glycoconjugates, but have yet to be unequivocally characterised.<sup>89</sup> The major limitation in the study of pseudaminyltransferases (PseTs) is the lack of available Pse-based molecules, most crucially nucleotide activated CMP-Pse5Ac7Ac and derivatives.



**Figure 4.1: Structures of Ulosonic Acids: Pse5Ac7Ac 1.6, β-Kdo 4.1.**

PseTs are responsible for facilitating the formation of a glycosidic bond between Pse and a range of acceptors. These acceptors include proteins where Pse can be O-glycosylated to serine and threonine residues of cell surface proteins, or carbohydrates, where linkage may be via an oxygen or nitrogen atom of Pse.<sup>1</sup>

When considering the diversity of PseTs, it is also important to account for the differences in the mechanisms of these enzymes. PseTs facilitate the formation of glycosidic bonds formed between Pse and an acceptor with either a retention or inversion of the configuration of the Pse anomeric centre, in comparison to that of the CMP-Pse substrate (Figure 4.2). Whilst a wealth of experimental evidence demonstrates that inverting GTs use a single-displacement  $S_N2$  mechanism, uncertainty remains over the mechanism employed by retaining GTs.<sup>26</sup> Currently, knowledge of the mechanism used by retaining ulosonic acid GTs is limited. The 3-deoxy-d-manno-oct-2-ulosonic acid (Kdo **4.1** Figure 4.1) transferase WbbB, from *Raoultella terrigena* is the first structurally characterised retaining GT to use a ketose donor sugar. Whilst further studies are needed to definitively decipher the mechanism of WbbB<sub>2-401</sub>, structural studies suggest that an internal-return mechanism ( $S_{Ni}$ ) mechanism is not feasible for this enzyme and there is evidence to support an  $S_N2$ -like double displacement mechanism.<sup>31</sup>



**Figure 4.2: Mechanism of retaining and inverting pseudaminyltransferases. A carbohydrate acceptor and the donor  $\alpha$ -CMP-Pse5Ac7Ac 1.10 are used in a PseT catalysed glycosylation reaction where CMP is the leaving group. Anomeric configuration of the donor is either inverted or retained dependent on the mechanism of the PseT.**

To date, the best-studied putative PseTs are those of the motility accessory family (Maf), which are predicted to be involved in O-glycosylation of flagellin proteins in a number of bacterial species.<sup>89,104,108,165,166</sup> Knockout mutational studies of Motility associated factor-1 (Maf1) from the opportunistic pathogen *Aeromonas caviae* have shown that it is likely responsible for the Pse5Ac7Ac glycosylation of flagellin subunits in the cytoplasm. The

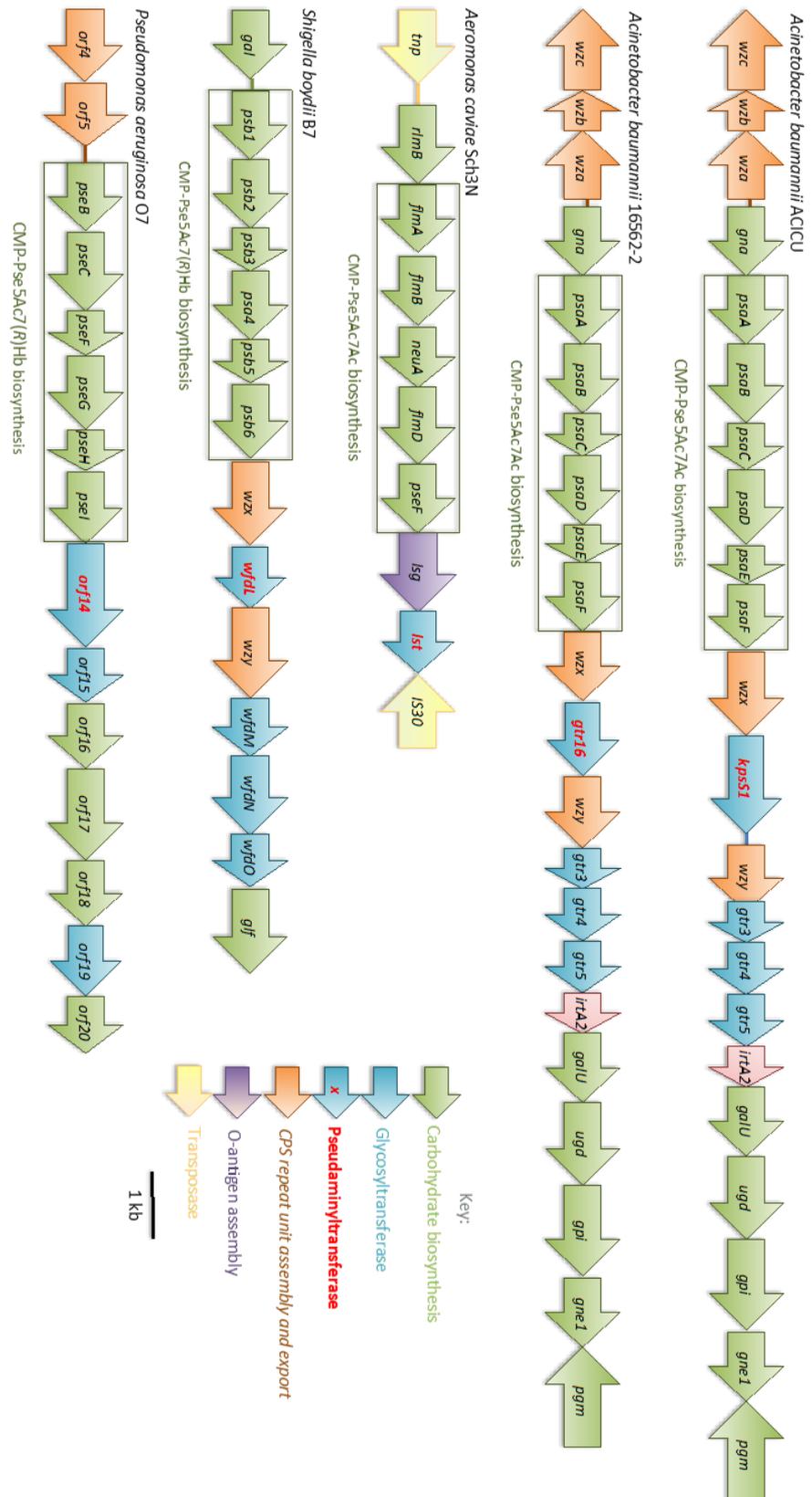
glycosylation of flagellin subunits in *A. caviae* is required for flagella assembly, influencing cell motility.<sup>89,104,165</sup> However, despite this promising initial evidence, no direct biochemical data has been obtained to confirm that Maf1 has been correctly classified as a PseT. *In vitro* characterisation of Maf1 is likely to be hindered by the insolubility of the Maf1 acceptor substrates, FlaA and FlaB flagellin proteins, in addition to lack of access to CMP-Pse5Ac7Ac.<sup>165</sup>

Promisingly, a crystal structure has been solved for Maf from *Magnetospirillum magneticum* AMB-1, putatively assigned as a PseT, which shares 29% sequence identity with *A. caviae* Maf1.<sup>143</sup> MS analysis of the glycans present on the *M. magneticum* flagella revealed that a mass corresponding to Pse5Ac7Ac was present, further to this the genome contains homologues for the six CMP-Pse5Ac7Ac biosynthesis enzymes. Deletion of the *maf* gene resulted in a lack of glycosylation of the flagella, providing further evidence that Maf functions as a putative PseT. Structural analysis showed that Maf shares similarities with inverting sialyltransferases from GT42 and GT29.<sup>166</sup> However, co-crystallisation attempts of Maf with CMP-N-acetyl neuraminic acid (CMP-Neu5Ac **1.5** Figure 4.1) were unsuccessful. Due to a lack of access to CMP-Pse5Ac7Ac and some remaining uncertainty regarding the exact identity of the CMP-nonulosonic acid which this enzyme is predicted to utilise, this enzyme has not been shown to be functional *in vitro*.<sup>166</sup> The literature currently contains no other evidence of biochemical studies of putative PseTs; therefore no member of this enzyme family has yet been unequivocally assigned.

The location of the *maf* gene often distinct to that of the other genes required for flagella formation, however, *maf* genes within *A. caviae* are in close proximity to putative CMP-Pse biosynthesis genes.<sup>166</sup> This association with CMP-Pse biosynthesis genes is also seen for putative PseTs that utilise carbohydrate acceptor moieties, rather than protein acceptors. These potential PseTs are commonly found within genomic islands containing genes for the biosynthesis of capsular polysaccharide (CPS) or O-antigen subunit of the lipopolysaccharide (LPS). Further analysis of these loci revealed the presence of GTs required for repeat unit assembly, including putative PseT genes (Figure 4.3).<sup>104,144,159,160,167,168</sup>

Five putative PseTs that were to be studied in this work (Table 1) are found within four Gram-negative pathogenic bacterial species including the aforementioned *A. caviae*, two

*Acinetobacter baumannii* strains, *Shigella boydii* and *Pseudomonas aeruginosa*.<sup>104,144,159,160,167,168</sup> These transferases are predicted to incorporate Pse5Ac7Ac or a Pse-derivative into either the CPS or LPS O-antigen in the cytoplasm, before subunits are transported across the inner membrane, assembled in the periplasm and presented on the cell surface.<sup>104,144,159,160,167,168</sup> Table 4.1 summarises source organism, Pse-containing glycoconjugate, CMP-Pse substrate and predicted mechanism and GT homologue for each putative PseT.



**Figure 4.3: Genetic islands encoding cell surface carbohydrate synthesis and assembly genes. Islands contain genes for CMP-Pse5Ac7Ac or CMP-Pse5Ac7Hb (green box). Putative pseudaminyltransferase genes are shown (bold red text).**

**Table 4.1: Putative Pseudaminyltransferases used in this study**

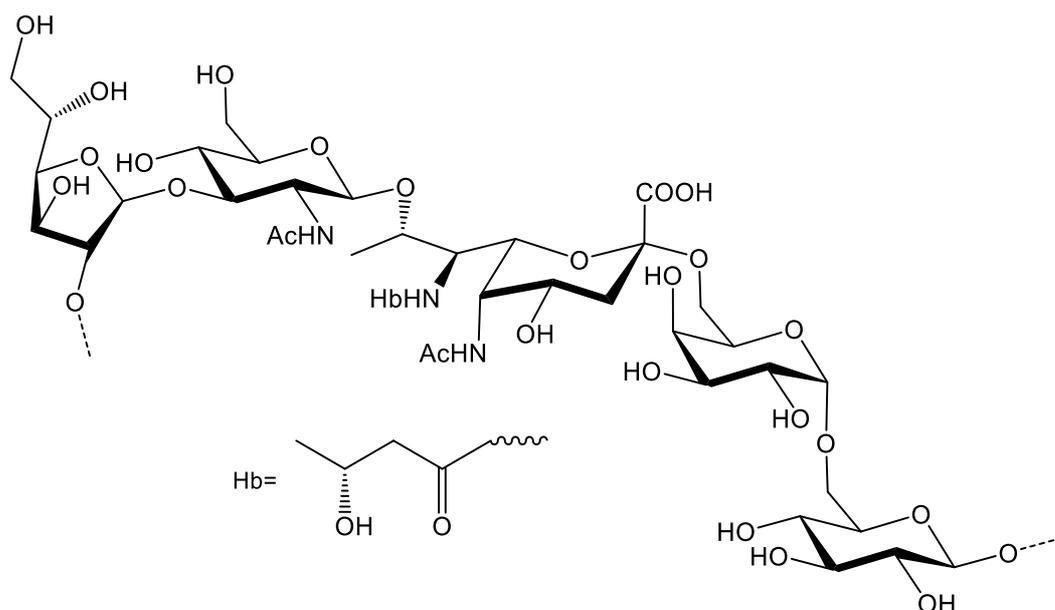
| Putative<br>PseT | Source<br>Organism             | Predicted Pse-<br>containing<br>glycoconjugate | Predicted CMP-Pse<br>substrate | Predicted<br>Mechanism | GT<br>homologue<br>(Sequence<br>identity) |
|------------------|--------------------------------|--|--------------------------------|------------------------|---|
| <b>Lst</b>       | <i>A. caviae</i><br>Sch3N      | LPS O-antigen                                  | $\alpha$ -CMP-Pse5Ac7Ac        | Inverting              | NST (25%)                                 |
| <b>WfdL</b>      | <i>S. boydii</i> B7            | LPS O-antigen (4.2)                            | $\alpha$ -CMP-<br>Pse5Ac7(R)Hb | Inverting              | NST (25%)                                 |
| <b>Gtr16</b>     | <i>A. baumannii</i><br>16562-2 | CPS (4.3)                                      | $\alpha$ -CMP-Pse5Ac7Ac        | Inverting              | NST (25%)                                 |
| <b>KpsS1</b>     | <i>A. baumannii</i><br>ACICU   | CPS (4.4)                                      | $\alpha$ -CMP-Pse5Ac7Ac        | Retaining              | KpsS (30%)                                |
| <b>Orf14</b>     | <i>P. aeruginosa</i><br>1244   | LPS O-antigen (4.5)                            | $\alpha$ -CMP-<br>Pse5(R)Hb7Fm | Retaining              | KpsS (19%)                                |

**Lst:** In addition to *maf1*, the *A. caviae* Sch3N genome also contains *lst*, a putative LPS O-antigen PseT gene.<sup>104</sup> The LPS O-antigen repeat of *A. caviae* Sch3N is known to contain Pse5Ac7Ac, alongside other as yet incompletely characterised hexose sugars **2** (Dr Jon Shaw, personal communication).<sup>104</sup> It is predicted that Lst, the product of the *lst* gene, along with other GTs assemble O-antigen in the cytoplasm, before the Wzx-like flippase, Lsg transports O-antigen repeat units across the cytoplasmic membrane. Mutational studies, carried out alongside characterisation of Maf1, demonstrate that Lst is not involved in flagellin glycosylation. However, mutation of *lst* does result in the loss of O-antigen, as observed by SDS-PAGE analysis of LPS.<sup>104</sup> Although the stereochemistry of the Pse5Ac7Ac linkage in the LPS of *A. caviae* is unknown, it has been shown in this study (Chapter 2.2.12) that *A. caviae* PseF produces  $\alpha$ -CMP-Pse5Ac7Ac, therefore this is the likely substrate for Lst to act as an inverting enzyme, resulting in the synthesis of a  $\beta$ -linked Pse5Ac7Ac glycoconjugate.<sup>104,169</sup> The unknown identity of the acceptor moiety for Lst demonstrates a common hindrance in the study of GTs. Deciphering the identity of two substrates for putative GTs often requires challenging structural analysis of glycans, which may be difficult to perform.<sup>24</sup>

Whilst no direct evidence of the role of Lst has been collected, the functional prediction of Lst as a putative PseT was based upon homology to several inverting sialyltransferases (SiaTs).<sup>104</sup> Notably, studies of Lst may be aided by its homology with *Neisseria meningitidis*

lipooligosaccharide-specific SiaT (NST), which transfers N-acetyl neuraminic acid (Neu5Ac **1.5**) to carbohydrate acceptors.<sup>32</sup> Crystallographic data of NST shows the protein exists as a novel domain swapped homodimer, with a GT-B fold. Typical of GT-B fold enzymes NST operates with a metal ion independent S<sub>N</sub>2-type mechanism, characteristic of inverting GTs.<sup>24,32</sup> The Carbohydrate Active Enzyme (CAZY) database, lists NST and currently 531 other  $\alpha$ -2,3-SiaTs and  $\alpha$ -glucosyltransferases, in the GT52 family (www.cazy.org). CAZY also includes Lst in the GT52 family, based upon sequence homology, however if functional studies confirm that Lst is an inverting PseT this enzyme will likely be the first member of a new GT family.

WfdL: Two additional putative inverting PseTs with homology to NST were to be studied. The first of which is WfdL, a putative PseT predicted to catalyse the incorporation of 5-N-acetyl-7-N-[(R)-3-hydroxybutanoyl]-Pse (Pse5Ac7Hb) into *S. boydii* B7 LPS O-antigen repeat unit **4.2** (Figure 4.4).<sup>168</sup> Liu *et al.* assigned WfdL as a GT based upon sequence homology to the *Salmonella enterica* LPS  $\alpha$ -1,2-glucosyltransferase, WaaH.<sup>168</sup> Further analysis of WfdL shows that it has 25% sequence identity to NST and 27% sequence identity with Lst (observation of this study). We assign WfdL as a putative PseT based upon this homology in addition to its occurrence in the O-antigen gene cluster along with the six *psb* genes needed for  $\alpha$ -CMP-Pse5Ac7(R)Hb biosynthesis, several other glycosyltransferases and transporters (Figure 4.3) required for O-antigen pentasaccharide repeat unit **4.2** (Figure 4.3) assembly and export. Additionally, CAZY currently includes WfdL in the family GT52. The presence of an  $\alpha$ -linked Pse5Ac7Hb at the non-reducing end of the O-antigen suggests that this carbohydrate is likely to play a role in *S. boydii* B7 immunospecificity within the human host.<sup>168,170</sup>



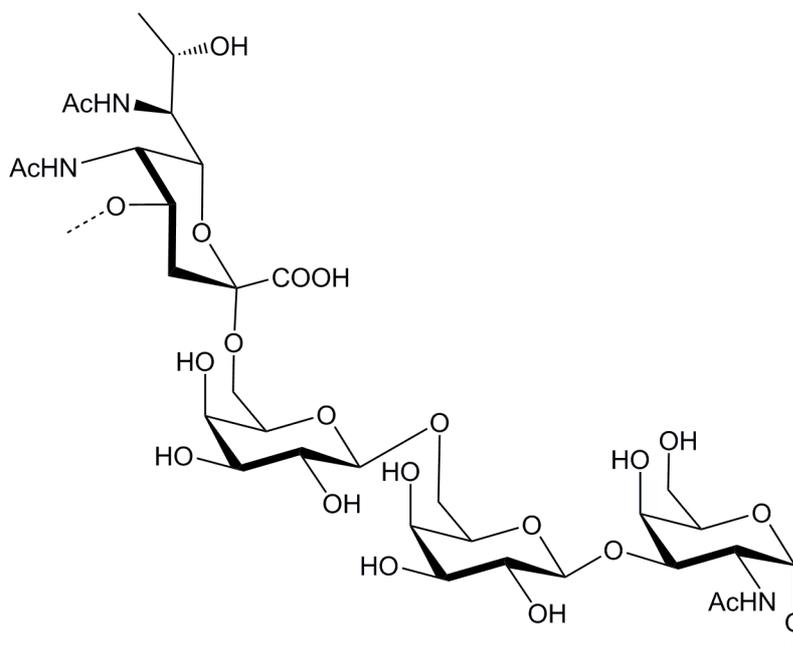
**Figure 4.4: *Shigella boydii* B7 LPS O-antigen repeat unit 4.2:**

**8)- $\beta$ -Pse5Ac7(R)Hb-(2 $\rightarrow$ 6)- $\alpha$ -D-Galp-(1 $\rightarrow$ 6)- $\alpha$ -D-Glcp-(1 $\rightarrow$ 2)- $\beta$ -D-Galf-(1 $\rightarrow$ 3)- $\alpha$ -D-GlcpNAc-(1 $\rightarrow$**

**Putative PseT WfdL is predicted to incorporate  $\beta$ -Pse5Ac7Hb.**

***Gtr16:*** Within clinically relevant, drug-resistant *A. baumannii* genomes, the capsule biosynthesis loci (KL) have been assigned, and contain genes required for carbohydrate biosynthesis, assembly and transport of CPS or LPS. KL6 *A. baumannii* 1656-2 includes putative  $\alpha$ -CMP-Pse5Ac7Ac biosynthetic genes and the putative inverting PseT *Gtr16*, a homolog of NST. Based upon 30-40% sequence identity with NulO GTs, *Gtr16* is predicted to incorporate Pse5Ac7Ac into CPS. The tetrasaccharide CPS repeat unit **4.3** of KL6 *A. baumannii* 1656-2 contain  $\beta$ -Pse5Ac7Ac, linked to both galactopyranose (Galp) and 2-Acetamido-2-deoxy-D-galactopyranose (D-GalNAcp) (Figure 4.5).<sup>144</sup> *Gtr16* is the final putative PseT of this study to show homology to NST. As with *Lst* and *WfdL*, CAZY ([www.cazy.org](http://www.cazy.org)) currently lists *Gtr16* as a GT52 member, however as with *Lst*, functional studies could deem assignment incorrect as inverting PseTs represent a distinct GT family.<sup>171</sup> In addition to KL6, two other *A. baumannii* K loci also encode putative inverting PseTs. The CPS repeat unit produced by the enzymes of the KL16 gene cluster, varies from that of KL6 tetrasaccharide only at the 2,4 linkage of Pse5Ac7Ac to Gal.<sup>172</sup> KL16 encodes *Gtr37* which shares 53% sequence identity with *Gtr16*. Secondly, *A. baumannii* with K93 CPS repeat unit, encode *Gtr167* which is predicted to form a 2,6 linkage between Pse derivative Pse5Ac7Hb and Galp as  $\beta$ -Pse5Ac7Hb is the terminal residue, in the CPS repeat

unit.<sup>173</sup> Gtr167 shares 41% sequence identity with Gtr16, this may be due to difference in the CMP-Pse substrate.



**Figure 4.5: *A. baumannii* 1656-2 CPS repeat unit 4.3:**

**4)- $\beta$ -Pse5Ac7Ac-(2 $\rightarrow$ 6)- $\beta$ -D-Galp-(1 $\rightarrow$ 6)- $\beta$ -D-Galp-(2 $\rightarrow$ 3)- $\alpha$ -D-GalpNAc-(1 $\rightarrow$**

**Putative PseT Gtr16 is predicted to incorporate  $\beta$ -Pse5Ac7Ac.**

*KpsS1*: As highlighted previously, vast differences can be expected in the mechanism used by inverting and retaining PseTs, thus it is of interest to characterize enzymes in both classes. *KpsS1*, the putative retaining PseT from *A. baumannii* ACICU, is found in the KL2 locus that encodes genes for CPS carbohydrate biosynthesis, assembly and transport (Figure 4.3). The CPS repeat unit **4.4** of *A. baumannii* ACICU is made up of Pse5Ac7Ac and three monosaccharides (Figure 4.6), and has also been proposed to constitute a protein glycosylation motif in *A. baumannii* and is also present in the exopolysaccharide of *A. baumannii* 54149 (Ab-54149).<sup>156,167</sup> The CPS repeat unit biosynthesised by enzymes in the *A. baumannii* KL2 locus is structurally similar to that produced by *A. baumannii* KL6 locus **4.3**, by the aforementioned Gtr16 in *A. baumannii*. However structural characterisation suggests that these repeat units vary in the stereochemistry linking Pse5Ac7Ac to a hexose. Inverting PseT Gtr16 facilitates the formation of an 2,6 glycosidic linkage between  $\beta$ -Pse5Ac7Ac and Galp, whereas retaining

KpsS1 is predicted to facilitate formation of a  $\alpha$ -2,6 linkage between Pse5Ac7Ac and glucopyranose (Glc<sub>p</sub> 4.6).<sup>144,167</sup>

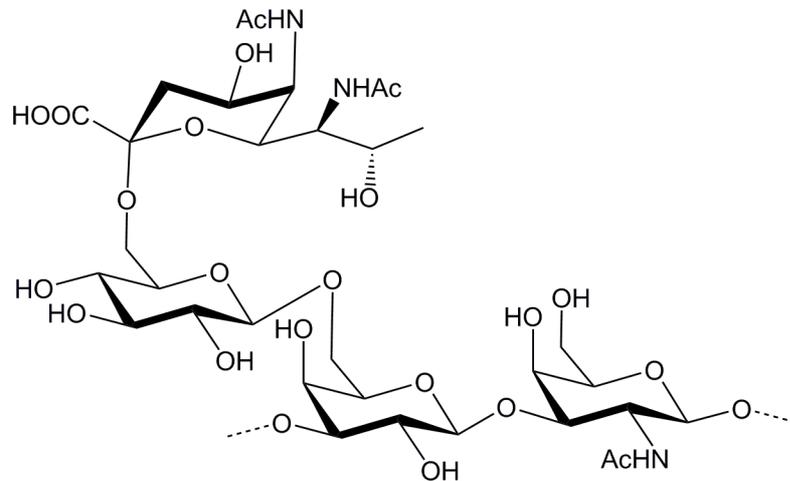
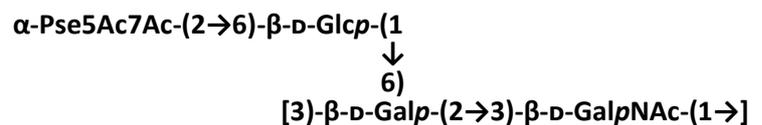


Figure 4.6: *A. baumannii* ACICU CPS repeat unit 4.4:



**Putative PseT KpsS1 is predicted to incorporate  $\alpha$ -Pse5Ac7Ac.**

KpsS1 is named based upon homology to *Escherichia coli* KpsS (30% sequence identity), the first retaining Kdo transferase characterised.<sup>30,160</sup> Biochemically characterised KpsS has been assigned as a member of the GT99 family. Recently WbbB was the first member of this family to be structurally characterized. WbbB is a multidomain  $\beta$ -Kdo transferase that is involved in biosynthesis of the OPS in *Raoultella terrigena*.<sup>31</sup> WbbB<sub>2-401</sub> shows 21% sequence identity to *A. baumannii* KpsS1, likely due to differences in substrate. However, an invariant motif in  $\beta$ -Kdo GTs, which is also highly conserved in CMP-sialic acid GTs, is the active site HP motif, which is also present in KpsS1.<sup>31</sup> Difficulties in complete characterization of GTs are perhaps highlighted by the fact that there is only one structurally characterized protein in the family of KpsS1 homolog, KpsS. The structural characterization of WbbB may facilitate the studies of retaining other GTs that use CMP-activated donors, including KpsS1.

In addition to KL2, several other *A. baumannii* K loci also encode the *psaABCDEF* genes required for the synthesis of CMP-Pse5Ac7Ac or homologous genes, for the biosynthesis of CMP-Pse derivatives.<sup>144,160,174</sup> A GT for the turnover of CMP-Pse5Ac7Ac or a CMP-Pse derivative can also be found within each of these K loci. KL23 encodes a putative GT which

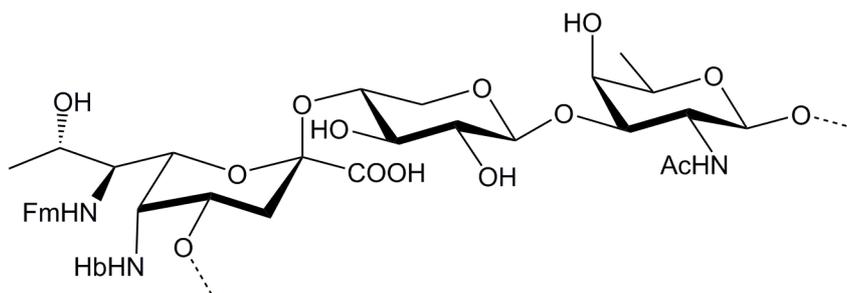
shares 84% sequence identity with KL2 *kpsS1* and has therefore also been named *kpsS1* (*kpsS1*<sub>KL23</sub>).<sup>174</sup> K loci KL33 also contains a putative retaining PseT, which is predicted to form the same  $\alpha$ -Pse5Ac7Ac-1,6- $\beta$ -D-Glcp linkage as KL2 KpsS1. The remainder of the KL33 CPS repeat unit of the KL33 CPS varies from the tetrasaccharide repeat unit in *A. baumannii* containing KL2, which may account for differences in sequence.<sup>161,175</sup> Arbatsky *et al.* annotated this KpsS1 homologue as KpsS2<sub>KL33</sub>.<sup>161</sup> *A. baumannii* LUH5550 KL42 cluster encodes a putative retaining PseT KpsS2<sub>KL42</sub>, required for incorporating Pse5Ac7Hb into the CPS repeat unit. KpsS2<sub>KL42</sub> shows 51% identity to KpsS1 and is a distant *E. coli* KpsS homologue.<sup>176</sup> A more distant KpsS1 homolog occurs in the KL12 locus of *A. baumannii* which contains Grt59, a putative retaining GT which uses CMP-Pse5Ac7Ac epimer CMP-5,7-di-N-acetylacinetaminic acid (CMP-Aci5Ac7Ac) for CPS repeat unit assembly.<sup>177</sup> Structural analysis of *A. baumannii* polysaccharides has aided the annotation of K loci genes, however, biochemical characterisation to confirm these proposed functions of these genes is lacking. Ultimately, unequivocal characterization of KpsS1 and other putative PseTs hinges on the production of CMP-Pse5Ac7Ac.

The *A. baumannii* strains that synthesise **4.4** are antibiotic resistant isolates.<sup>156,167,175</sup> A recent study used bacteriophage  $\Phi$ AB6 tailspike protein which depolymerize Ab-54149 EPS into tetrasaccharide units. This strategy was used in place of alternative chemical methods to access homogenous samples of **4.4**, which was conjugated to a carrier protein for use as a vaccine. Boosted sera from rabbits injected with this glycoconjugate vaccine could recognize both Ab-54149 EPS and the glycoconjugate vaccine. The sera showed bactericidal activity towards Ab-54149, only when the Pse5Ac7Ac present in the carbohydrate moiety of the vaccine and the sera was shown to selectively bind Pse5Ac7Ac over sialic acid. This study has demonstrated the therapeutic potential of **4.4** against an antibiotic resistant *A. baumannii* strain, where Pse5Ac7Ac was shown to be a crucial epitope.<sup>178</sup>

#### Orf14:

*Pseudomonas aeruginosa* is responsible for multidrug resistant infection within immunocompromised hosts. Despite *P. aeruginosa* LPS being considered a virulence factor, crucial in host immune responses, the genes for LPS biosynthesis are relatively under studied.<sup>179</sup> Analysis of the *P. aeruginosa* O7 LPS O-antigen biosynthesis locus (Figure 4.3) shows the presence of genes which may putatively be assigned as those required for  $\alpha$ -CMP-5-N-[(R)-

3-hydroxybutanoyl]-7-N-formyl-Pse (CMP-Pse5Hb7Fm) synthesis. The O7 O-antigen biosynthesis locus also contains the gene annotated *orf14* (accession: AAM27840.1), which we predict encodes the putative retaining PseT.<sup>159</sup> Orf14, which shows sequence homology to KpsS1, is predicted to incorporate  $\alpha$ -Pse5(R)Hb7Fm into the LPS O-antigen repeat unit **4.5** (unpublished observation of this study), a trisaccharide comprised of xylose (Xyl), and *N*-acetyl fucosamine (FucNAc) (Figure 4.7), producing compound **5**.<sup>159</sup>



**Figure 4.7: *P. aeruginosa* O7 O-antigen repeat unit 4.5:**

**4)- $\alpha$ -Pse5(R)Hb7Fm-(2 $\rightarrow$ 4)- $\beta$ -D-Xylp-(1 $\rightarrow$ 3)- $\beta$ -D-FucpNAc-(1 $\rightarrow$**

**Putative PseT KpsS1 is predicted to incorporate  $\alpha$ -Pse5Fm7Hb.**

There is evidence to suggest that the enzymes involved in biosynthesis of this trisaccharide, including Orf14, are also required for pilin glycosylation, as trisaccharide **1.41** (which constitutes one monomer of the trisaccharide repeat unit **4.5**) is present on at the C-terminal Ser148 residue through a  $\beta$ -FucNAc-(1,3)- $\beta$ -Seer linkage, where Pse5Hb7Fm is the terminal carbohydrate.<sup>180</sup> This pilin glycosylation has been shown to play a role in *P. aeruginosa* 1244 virulence. Castaic *et al.* produced a *P. aeruginosa* 1244 strain with a deletion of the *pilO* gene ( $\Delta pilO$ ), encoding the oligosaccharyltransferase that transfers **4.5** to the pilin and confirmed that this mutant strain produced lacked pilin glycosylation. In mouse model studies, wild type *P. aeruginosa* 1244 had greater respiratory pathogenicity, than the  $\Delta pilO$  strain.<sup>181</sup>

In a mouse model study, when injected with pilin fragments glycosylated with **5** mice were able to induce a specific immune response when challenged with *P. aeruginosa* 1244 infection, resulting in higher survival rate. It was shown that the immune response arose from the production of antibodies that targeted **4.5** when present on both the pili and LPS O-antigen. This study showed promise that a glycoconjugate vaccine containing **4.5**, may be used in the treatment of *P. aeruginosa* 1244 infection.<sup>182</sup>

Recently, the total synthesis of the trisaccharide **1.41** has been reported by Lui *et al.*<sup>6</sup> The product of this highly impressive synthesis is not only the first example of a naturally occurring Pse-derivative to be produced by chemical synthesis but is also the first reported synthesis of a Pse-based glycoside. However, as discussed in Chapter 1, the multi-step synthesis required a vast amount of optimisation to yield 8.2 mg of **1.41**.<sup>6</sup>

As discussed in Chapter 2, enzymatic synthesis of Pse-based glycosides may offer a valuable alternative to chemical routes. Access to PseTs will increase the repertoire of Pse-based molecules, which may aid studies into the biology roles of Pse. Additionally, *in vitro* studies of PseTs can provide insight into the functionality and mechanism of these enzymes, enabling them to be studied as new therapeutic targets.

The aims of this chapter were to study of the putative PseTs Lst, WfdL Gtr16, KpsS1 and Orf14, *in vitro*. Therefore, our initial aim was to develop a suitable strategy for the recombinant expression of these targets. Following completion of this, the second aim was to functionally characterise the enzymes, facilitated by access to CMP-Pse5Ac7Ac **1.10** and a strategy for the synthesis of CMP-Pse derivatives (as discussed in Chapter 2).

## 4.2 Experimental

### 4.3.1 Sequence analysis of Putative Pseudaminyltransferases

The sequence of each putative PseT was analysed for the presence of transmembrane helices, using TMHMM Server v. 2.0 ([www.cbs.dtu.dk](http://www.cbs.dtu.dk)). The sequences were further analysed for the presence of secretion signals, using SignalP-5.0 ([www.cbs.dtu.dk](http://www.cbs.dtu.dk)). BLASTp was performed using each putative PseT as the query sequence, to identify homologues. Sequence alignments for Gtr16, WfdL, Lst, NST were performed using Clustal Omega ([www.ebi.ac.uk](http://www.ebi.ac.uk)).

### 4.3.2 Homology Modeling of Putative Inverting Pseudaminyltransferases

A multiple sequence alignment of Nst, Gtr16, Lst and WfdL was produced using Clustal Omega. This was then edited manually to improve the quality of the alignment for use in homology modeling. Models were then produced using MODELER automodel through Discovery Studio 3.5, using the apo structure of NST (PDB entry: 2YK4).

In order to prevent fixed body domain movements, the coordinates of some main chain atoms were copied from the template. The residues chosen, in NST, were Gln60, Tyr92, Tyr139, Tyr212, Gly255, Tyr276 and Tyr318. The actual coordinates in the model vary slightly due to post-building refinement. Some secondary structure restraints were also placed at one of the alpha helices, namely the helix from the conserved Glu260, as there are lots of gaps in the sequence alignment at this region. A restraint was used from Glu210-Ile226 in Gtr16, Glu209-Phe224 in Lst and Glu204-Tyr219 in WfdL. Twenty models were produced for each of the proteins. Both the total PDF energy and the DOPE energy metrics were used in choosing the best quality model.\*

### 4.2.3 Putative Pseudaminyltransferase Expression Trials

Each of the putative PseT encoding genes (*WfdL*, *Lst*, *Gtr16*, *KpsS1* and *Orf14*) were obtained in plasmids constructed from either pET-15b or pET-28a(+) vector, to produce N-terminal His<sub>6</sub> tagged PseT, with genes of interest inducibly expressed using isopropyl β-D-1-thiogalactopyranoside (IPTG) (Table 4.2) (plasmids were obtained from Genscript).

---

\* Modeling was performed by Paul Bond, YSBL, University of York.

Plasmids for the five putative PseTs were initially transformed into chemically competent *E. coli* BL21 (DE3) cells, by heat shock or into electrocompetent *E. coli* BL21 (DE3) cells, by electroporation (as described in 2.2.1). Transformed colonies were selected for using the selectable antibiotic resistant marker, encoded in the pET plasmid of choice (Table 4.2). As controls the empty vectors pET-28a(+) and pET-15b were also transformed into both *E. coli* BL21 (DE3) cells.

**Table 4.2: Putative PseTs Plasmids and Molecular Weights of Recombinant Putative PseTs**

| Putative PseT | Plasmid Vector | Antibiotic Resistance of Plasmid | MW of His <sub>6</sub> PseT (kDa) |
|---------------|----------------|----------------------------------|-----------------------------------|
| <b>Lst</b>    | pET-28a(+)     | Kanamycin                        | 38.3                              |
| <b>WdfL</b>   | pET-15b        | Ampicillin                       | 38.9                              |
| <b>Gtr16</b>  | pET-15b        | Ampicillin                       | 40.4                              |
| <b>KpsS1</b>  | pET-15b        | Ampicillin                       | 59.4                              |
| <b>Orf14</b>  | pET-15b        | Ampicillin                       | 59.5                              |

Transformed *E. coli* cells were then used in expression trials. Throughout these initial trials, the following four parameters were studied in combination: temperature, OD<sub>600</sub> at induction, IPTG concentration and incubation time post-induction. Transformed cells were cultured overnight at 37 °C, grown aerobically at 180 revolutions per minute (r.p.m.) in LB supplemented with 50 µg mL<sup>-1</sup> kanamycin (LB<sub>Kan</sub>) or 100 µg mL<sup>-1</sup> ampicillin (LB<sub>Amp</sub>). For each sample, 50 mL LB<sub>Kan</sub> or LB<sub>Amp</sub> was inoculated with a 1 in 100 dilution of overnight culture of *E. coli* BL21 (DE3), in a 250 mL Erlenmyer flask, and grown at 37°C, 180 r.p.m. until OD<sub>600</sub> reached 0.4 and 0.8 for Lst and 0.4, 0.6 and 0.8 for all other WdfL, Gtr16, KpsS1 and Orf14 at which point cultures were induced with either 0.1 mM or 1 mM IPTG. Post induction cultures were incubated at 30 °C and 37 °C for Gtr16, WdfL and Orf14 and at 25 °C, 30 °C and 37 °C for Lst and KpsS1. 2 mL Samples were taken at either 4 and 24 hours post induction for Lst, WdfL, Gtr16 and Orf14 samples and 1, 4 and 24 hours post induction for KpsS1 samples. The samples were centrifuged (6,000 × g, 10 minutes, 4 °C), supernatant was removed, and cells pellets were stored at -20 °C. An uninduced control was also collected, for each putative PseT, to assess basal expression of the genes and vector only (pET-15b) control sample was taken to analyse alongside KpsS1 samples. To lyse cells, pellets were resuspended in 100 µL BugBuster mix (BugBuster® Protein Extraction Reagent (Merck), supplemented with 1 Pierce Protease Inhibitor Tablet (Thermo Scientific) per 50

mL and 25 units mL<sup>-1</sup> Benzonase® Nuclease (Sigma)), incubated at room temperature for 20 minutes and centrifuged (17,000 ×g, 10 min, 4 °C). Supernatant and insoluble material was separated. The insoluble material was resuspended in 100 µL dH<sub>2</sub>O. 20 µL samples of both the soluble and insoluble material were mixed with 5 µL reducing dye (10% SDS, 10 mM β-mercaptoethanol, 20% glycerol 0.2 M Tris HCl pH 6.8, 0.05 % bromophenol blue) heated to 95 °C (with the exception of Lst which was not heated) then run on 10%, 12% or 4-20% SDS-PAGE (Described in 2.2.1). 5 µL PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific) or 6 µL Blue Prestained Protein Standard, Broad Range marker (New England Biolabs) or both were used as molecular weight markers in SDS-PAGE.

Following these initial expression trials, the plasmids for KpsS1 and Lst were transformed into chemically competent *E. coli* Tuner (DE3) cells. The expression trial was repeated, as described for *E. coli* BL21 (DE3) cells, however, two additional post-induction temperatures of 16 °C and 25 °C were studied. Samples were collected and lysed as described above and analysed via 10 or 12% SDS-PAGE.

#### **4.2.4 Solubilisation trials of Putative Pseudaminyltransferases Lst**

*E. coli* BL21 (DE3) which were transformed with plasmid encoding Lst were cultured overnight at 37 °C, 180 rpm in LB<sub>kan</sub>. The overnight Lst culture was used to inoculate 250 mL Erlenmeyer flasks, each containing 50 mL LB<sub>kan</sub> to give an initial OD<sub>600</sub> 0.02. All cultures were incubated at 37 °C, 180 rpm, until expression was induced, at OD<sub>600</sub> measurements of 0.6, using either 0.1 mM or 1 mM IPTG for each OD<sub>600</sub> value. Post-induction Lst was incubated at 25 °C, 180 rpm. Three 1 mL samples were taken from each flask at 4 and 24 hrs post-induction. The samples were centrifuged (6,000 × g, 10 minutes, 4 °C) and cells pellets were stored at -20 °C.

Cell pellets from each 1 mL cell culture sample were thawed on ice and resuspended in 50 µL Bugbuster mix, supplemented with either 8 M urea, 6 M Guanidine HCl or 1 M NaCl. Cells were then incubated with shaking for 20 minutes at room temperature to facilitate lysis. Lysed cells were then centrifuged (6,000 × g, 10 minutes, 4 °C), and the supernatant was separated from pellet, which was resuspended in 50 µL dH<sub>2</sub>O. 15 µL of each sample were analysed using 12% SDS-PAGE (as described in 2.2.1, with the exception that Lst samples were not heated before being applied to gels).

## 4.2.5 Detergent Screen for Solubilisation of Putative Pseudaminyltransferases

### Lst and KpsS1

A 2 L Erlenmeyer flask, containing 500 mL LB<sub>kan</sub> was inoculated with Lst overnight culture to give an initial OD<sub>600</sub> 0.02 and incubated at 37 °C, 180 r.p.m. Lst expression was induced, at an OD<sub>600</sub> measurement of 0.6, using 0.1 mM IPTG. Cells were centrifuged (6,000 ×g, 10 minutes, 4 °C) and cells pellets were stored at -20 °C. Cells were thawed on ice and resuspended in 25 mL lysis buffer (50 mM sodium phosphate, pH 7.3, 400 mM NaCl, 10 mM β-mercaptoethanol, 10 mM imidazole, 1 mM MgCl<sub>2</sub> supplemented with 1 Pierce Protease Inhibitor Tablet (Thermo Scientific) per 50 mL and 25 units mL<sup>-1</sup> Benzonase® Nuclease (Sigma). 250 μL samples were taken and were each supplemented with one of five detergents, at various concentrations (Table 4.3) and additional lysis buffer to give a final volume of 500 μL. A no detergent control sample was prepared, using lysis buffer only. Cells underwent 10 freeze-thaw cycles to facilitate lysis. Lysed cells were then incubated at 25 °C for 30 minutes, centrifuged (17,000 × g, 30 minutes, 4 °C), and the supernatant was separated from pellet, which was resuspended in 500 μL dH<sub>2</sub>O. 15 μL of each sample, was mixed with 5× SDS reducing buffer and analysed using 12% SDS-PAGE (as described in 2.2.1). This experiment was repeated with KpsS1 from cells cultured at 30 °C, 180 rpm LB<sub>amp</sub>. KpsS1 expression was induced, at an OD<sub>600</sub> measurement of 0.8, using 0.1 mM IPTG. Post-induction KpsS1 was incubated at 30 °C, 180 r.p.m. for 4 hours.

**Table 4.3: Detergents screened for ability to solubilise KpsS1 and Lst**

| Detergent                     | Concentration |
|-------------------------------|---------------|
| n-Dodecyl β-D-maltoside (DDM) | 15 mM         |
|                               | 30 mM         |
|                               | 75 mM         |
| Tween 20                      | 0.06 mM       |
|                               | 0.3 mM        |
|                               | 0.6 mM        |
| CHAPS                         | 6 mM          |
|                               | 16.2 mM (1%)  |
| Triton X-100                  | 0.5%          |
|                               | 2%            |
|                               | 5%            |
| SDS                           | 100 mM        |
|                               | 50 mM         |

## 4.2.6 Cloning of KpsS1 Fusion Protein Plasmids

Three pET-based plasmids each containing commonly used fusion proteins for the solubilisation of recombinantly expressed protein were obtained (University of York, Biology Technology Facility) (plasmid sequences in Appendix) in order to generate fusion plasmids encoding either, glutathione-S-transferase (GST) or colicin E9 immunity protein (Im9), maltose binding protein (MBP) with *A. baumannii* KpsS1 fused to the C-terminal end of each of these proteins. CleR and CleF primers (Table 4.4) were used in gradient PCR to linearize the vectors, in 50  $\mu$ L reactions, using Q5 DNA polymerase (DNAP) (Table 4.5 and 4.6). 5  $\mu$ L of each PCR sample was analysed via gel electrophoresis, using a 0.6% agarose gel where DNA was visualised with SYBR Safe stain. Reaction tubes containing the desired product (Gst: 6050 bp, Im9: 5651 bp) were run on 0.6 % agarose gel and purified via gel extraction (illustra GFX PCR DNA and Gel Band Purification Kit, GE Healthcare). Linearised vector was stored at  $-20^{\circ}\text{C}$ .

**Table 4.4 PCR Primers for KpsS1 Fusion Plasmids**

| Primer        | Sequence                                  |
|---------------|---|
| CleR          | 5'-TTGCTGGTCCCTGGAACAGAACTTCC-3'          |
| CleF          | 5'-CGCGCCTTCTCCTCACATATGGCTAGC-3'         |
| KpsS1 Forward | 5'-TCCAGGGACCAGCAATGAACTTTCTGATCCTGAT- 3' |
| KpsS1 Reverse | 5'-TGAGGAGAAGGCGCGTTACTCAATGTCGGTGT-3'    |

**Table 4.5 PCR Protocol**

| Step                 | Temperature   | Time          |
|----------------------|---|---------------|
| Initial Denaturation | 98 $^{\circ}\text{C}$                                   | 30 seconds    |
| 35 Cycles            | 98 $^{\circ}\text{C}$                                   | 5 seconds     |
|                      | 55–72 $^{\circ}\text{C}$ (gradient over 9 temperatures) | 10–30 seconds |
|                      | 72 $^{\circ}\text{C}$                                   | 30 seconds/kb |
| Final Extension      | 72 $^{\circ}\text{C}$                                   | 2 minutes     |
| Hold                 | 4 $^{\circ}\text{C}$                                    | /             |

The KpsS1 insert was amplified via PCR from the pET-15b KpsS1 (Genscript) plasmid used for initial expression trials. Primers KpsS1 Forward and KpsS1 Reverse (Table 4.4) were used

in PCR to create regions of homology at either end of the KpsS1 gene. PCR was carried out using Q5 DNAP (Table 4.5 and 4.6) as described for vectors, before being analysed via 0.6 % agarose gel electrophoresis and purified by gel extraction and KpsS1 product was stored at  $-20^{\circ}\text{C}$ . A Quick-Fusion cloning kit (Biomake) was used to create two plasmids encoding the KpsS1 fusion proteins from the linearised vector and KpsS1 insert (10  $\mu\text{L}$  reactions were set up containing 1  $\mu\text{L}$  fusion enzyme, 2  $\mu\text{L}$  fusion buffer,  $[0.01 \times \text{bp of vector}] \text{ ng}$  vector, 30 ng KpsS1 insert,  $\text{dH}_2\text{O}$  to 10  $\mu\text{L}$ ). A control reaction using DNA provided in the kit was also performed. Reactions were incubated at  $37^{\circ}\text{C}$  for 30 minutes. 5  $\mu\text{L}$  of each reaction was then transformed into 50  $\mu\text{L}$  XL-1 Blue Supercompetent cells, using heat-shock. Following transformation, the cells were plated on  $\text{LB}_{\text{kan}}$  plates containing a Blue-White Screen for the selection of recombinant cells and were incubated at  $37^{\circ}\text{C}$  overnight.

**Table 4.6 Components for Q5 High-Fidelity DNAP mediated PCR**

| Component                       | 50 $\mu\text{L}$ Reaction    | 15 $\mu\text{L}$ Reaction (Colony PCR)                 | Final Concentration |
|---------------------------------|------------------------------|--|---------------------|
| Q5 High-Fidelity 2X Master Mix  | 25 $\mu\text{L}$             | 7.5 $\mu\text{L}$                                      | 1X                  |
| 10 $\mu\text{M}$ Forward Primer | 2.5 $\mu\text{L}$            | 0.75 $\mu\text{L}$                                     | 0.5 $\mu\text{M}$   |
| 10 $\mu\text{M}$ Reverse Primer | 2.5 $\mu\text{L}$            | 0.75 $\mu\text{L}$                                     | 0.5 $\mu\text{M}$   |
| Template DNA                    | 1 $\mu\text{L}$ pET15b-KpsS1 | 1.6 $\mu\text{L}$ $\text{dH}_2\text{O}$ containing DNA | 20 ng pET-15b KpsS1 |
| Nuclease-Free Water             | to 50 $\mu\text{L}$          | to 15 $\mu\text{L}$                                    |                     |

White colonies were screened for the presence of KpsS1 fusion plasmid using colony PCR. Colonies were selected, spotted onto  $\text{LB}_{\text{kan}}$  agar and placed in 20  $\mu\text{L}$  DNase free  $\text{dH}_2\text{O}$ , which was then heated to  $98^{\circ}\text{C}$  for 5 minutes. 1.6  $\mu\text{L}$  of the  $\text{dH}_2\text{O}$  containing DNA was added to Q5 reaction mixes to generate 15  $\mu\text{L}$  reactions (Table 4.6), PCR was then performed (Tables 4.5 and 4.6, annealing temperature:  $72^{\circ}\text{C}$ ) with KpsS1 Forward and KpsS1 Reverse primers. Gel electrophoresis was used to analyse reactions (0.6 % agarose). Colonies corresponding to positive colony PCR results were incubated in 10 mL  $\text{LB}_{\text{kan}}$  overnight ( $37^{\circ}\text{C}$ , 180 r.p.m.) before plasmid DNA was extracted via miniprep (QIAGEN) and sequenced (GATC) using T7 and T7 reverse primers. DNA encoding each of the desired KpsS1 fusion plasmids was transformed into chemically competent *E. coli* BL21 (DE3) via electroporation (as described in 2.2.1).

#### 4.2.7 KpsS1 Fusion Protein Expression Trials

Following the transformation of either Im9-KpsS1 or GST-KpsS1 plasmid into chemically competent *E. coli* BL21 (DE3) single colonies were picked to inoculate LB<sub>kan</sub>, for overnight cultures incubated at 37 °C, 180 rpm. 24 deep-well blocks were inoculated with the culture to give an initial OD<sub>600</sub> 0.02, in a final volume of 2 mL LB<sub>kan</sub> in each 10 mL well. The cultures were incubated at 37, 30 °C and 16 °C, at 180 r.p.m. and expression of recombinant KpsS1-fusion protein was induced with the addition of either 0.1 mM, 0.5 mM or 1 mM IPTG, at OD<sub>600</sub> values of 0.4 and 0.8, resulting in 2 wells of cells per condition studied. Cells were harvested from each condition 4 hrs and 20 hrs post-induction. OD<sub>600</sub> values of the cultures were recorded at the point of harvesting and cell pellets were obtained (6,000 ×g, 10 minutes, 4 °C) and stored at -20 °C. Pellets were resuspended in Bugbuster mix (without lysozyme) and shaken at room temperature for 45 minutes to lyse cells. Cells were centrifuged (6,000 × g, 10 minutes, 4 °C); following this soluble and insoluble material was separated. The insoluble material was resuspended in an equal volume of dH<sub>2</sub>O and 15 µL of both soluble and insoluble fractions, from each sample, were analysed via 10 % SDS-PAGE (as described in 2.2.1).

#### 4.2.8 Im9-KpsS1 Large Scale Expression

*E. coli* BL21 (DE3) was transformed with Im9-KpsS1 plasmid. LB<sub>kan</sub> (4 x 600mL, in baffled 2 L Erlenmyer flasks) was inoculated with a 1 in 100 dilution overnight of culture of *E. coli* BL21 (DE3) transformed with Im9-KpsS1 plasmid (overnight culture: 50 mL LB<sub>kan</sub>, in 250 mL Erlenmeyer flask, cultured at 37 °C, 180 r.p.m.). Cells were cultured at 37 °C, 180 r.p.m. and protein expression was induced at an OD<sub>600</sub> value of 0.8 with 0.1 mM IPTG. Cells were incubated at 30 °C, 180 r.p.m. for 4 hrs post-induction before cell pellets were harvested via centrifugation (6000 × g, 30 min, 4 °C) and stored at -80 °C.

#### 4.2.9 Im9-KpsS1 Purification

Cell pellet from a 2.4 L culture was resuspended in 40 mL lysis buffer (50 mM sodium phosphate, 400 mM NaCl, 10 mM β-mercaptoethanol, 10 mM MgCl<sub>2</sub>, pH 7.4) supplemented with protease inhibitor, lysozyme, and Benzonase. Cells were lysed via sonication in ice (30 seconds, 30 second pause, 25 repeats) and centrifuged (37000 ×g, 30 min, 4 °C). Supernatant was applied to a 1 mL HisTrap HP column (GE Healthcare) pre-

equilibrated with lysis buffer, the column was washed with 10 column volumes (C.V.), lysis buffer and protein was eluted with a gradient of elution buffer (lysis buffer + 490 mM imidazole, pH 7.4) over 40 C.V. as 1 mL fractions were collected. 15  $\mu$ L samples of insoluble material, soluble material, flowthrough wash and fractions were analysed via 10 % SDS-PAGE. Following SDS-PAGE analysis (as described in 2.2.1) fractions containing Im9-KpsS1 (68.9 kDa, fractions 11-14) were pooled and applied to desalting column (HiPrep 26/10 desalting column, GE Healthcare) to exchange Im9-KpsS1 into reaction buffer (50 mM sodium phosphate, 25 mM NaCl, pH 7.4).

#### 4.2.10 Im9-KpsS1 Protein Identification Mass Spectrometry

Band which appeared to correspond to Im9-KpsS1 in pure fractions of SDS-PAGE (4.2.9) was subjected to protein identification mass spectrometry, as described in 2.2.8 for PseF.

#### 4.2.11 Im9-KpsS1 Activity Screen

40  $\mu$ L reaction mixtures containing 20 mM acceptor (either D-Glucose (D-Glc) **4.6**, methyl  $\beta$ -D-glucopyranoside (Me  $\beta$ -D-Glcp) **4.7**, 4-nitrophenyl  $\beta$ -D-glucopyranoside (*p*NP  $\beta$ -D-Glcp) **4.8**,  $\beta$ -D-glucopyranosyl-2,6-methyl- $\beta$ -D-galactopyranoside ( $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe) **4.9**,\*\*\*\* 4-nitrophenyl  $\beta$ -D-galactopyranoside (*p*NP  $\beta$ -D-Galp) **3.16**, 4-nitrophenyl  $\beta$ -D-xylopyranoside (*p*NP  $\beta$ -D-Xylp) **4.10**, or 4-methoxyphenyl  $\beta$ -D-glucopyranoside (*p*MeP  $\beta$ -D-Glcp) **4.11**), 5 mM CMP-Pse5Ac7Ac **1.10**, 10 mM MgCl<sub>2</sub>, 50 mM sodium phosphate, 25 mM NaCl, pH 7.4 and 12.2  $\mu$ L partially purified Im9-KpsS1 or cell lysate containing Im9-KpsS1 were prepared and incubated at 30 °C for 12-18 hours. Reactions were performed in duplicate. Control reactions were set up as a described above without the addition of one of the following: pure or crude Im9-KpsS1, acceptor or CMP-Pse5Ac7Ac. Reactions were also performed as described above with *p*NP- $\beta$ -D-Glcp acceptor **4.6** and CMP-Neu5Ac **1.5**, in place of CMP-Pse5Ac7Ac and Im9-KpsS1 lysate. Additionally, a reaction was set up with *p*NP- $\beta$ -D-Glcp acceptor, containing both CMP-Neu5Ac and CMP-Pse5Ac7Ac and Im9-KpsS1 lysate. All samples were analysed via negative-ion mode electrospray ionisation liquid-chromatography mass spectrometry (-ESI LC-MS) (details in Appendix-General methods).

---

\*\*\*\* **4.9** was synthesized and characterized by Dr Darshita Budhadev and Natasha Hatton, detailed in Appendix Scheme A.1

#### 4.2.12 Chemoenzymatic synthesis of glycoside containing Pse5Ac7Ac, via a one-pot, seven-enzyme reaction using Im9-KpsS1

2 mM UDP-GlcNAc **1.3** (30 mg), 0.3 mM coenzyme-A, 4 mM pyridoxal 5'-phosphate, 20 mM L-glutamic acid, 3 mM phosphoenolpyruvate, 20 mM N-Acetyl-S-Acetylcysteamine, 0.2 mg mL<sup>-1</sup> PseB, 0.4 mg mL<sup>-1</sup> PseC, 0.2 mg mL<sup>-1</sup> PseH, 0.2 mg mL<sup>-1</sup> PseG and 0.2 mg mL<sup>-1</sup> PseI (enzymes purified as described in Chapter 2.2.1). The reaction mixture was incubated at 37 °C for 6 hours and was monitored via –ESI LC-MS, to monitor the production of Pse5Ac7Ac. Reaction was stored at 4 °C overnight, before enzyme precipitate was removed via centrifugation (37,000 × g, 30 min, 4 °C). Supernatant was retained and 0.2 mg mL<sup>-1</sup> PseF (purified as described in Chapter 2.2.7), 4 mM CTP and 20 mM MgCl<sub>2</sub> were added to facilitate the production of CMP-Pse5Ac7Ac **1.10**, as confirmed by –ESI LC-MS analysis. The reaction incubated at 30 °C for 3 hours. Cell pellet containing Im9-KpsS1 from 1.8 L culture (conditions as described for large scale expression) was lysed as described above. The resulting supernatant was added to the CMP-Pse5Ac7Ac reaction mixture to a final volume of 60 mL to give 1 mM CMP-Pse5Ac7Ac. 5 mM *p*NP-β-D-Glcp was then added to facilitate the formation of Pse5Ac7Ac-*p*NP-β-D-Glcp **4.12**. The reaction incubated at 30 °C for 72 hours and was analysed via –ESI LC-MS. The mixture was lyophilised and resuspended in 10 mL dH<sub>2</sub>O and cotton wool filtered to remove enzyme precipitate. The filtrate was mixed with 10 mL EtOH and stored at 4 °C for 30 minutes to precipitate any remaining enzyme. The mixture was cotton wool filtered, and the filtrate was diluted in dH<sub>2</sub>O to a final volume of 200 mL before lyophilisation. The lyophilised material was resuspended in 10 mL dH<sub>2</sub>O and 5 mL was applied to a 500 mL column packed with Bio-Gel® P-2 resin (Biorad) in HPLC-grade H<sub>2</sub>O. 3 mL fractions were collected for 40 hours and analysed via LC-MS (details in Appendix-General methods) for the presence of Pse5Ac7Ac-*p*NP-β-D-Glcp. The column was repeated with the remaining lyophilised material and all fractions containing Pse5Ac7Ac-*p*NP-β-D-Glcp were pooled and lyophilised. Following NMR analysis of the resulting material (D<sub>2</sub>O, 700MHz), the material was dry loaded onto silica and purified via flash chromatography using EtOH:MeOH:dH<sub>2</sub>O (5:2:1). 2 mL fractions were collected and analysed by TLC (sugar stain: MeOH, 5% H<sub>2</sub>SO<sub>4</sub> and charring) and LC-MS for the presence of Pse5Ac7Ac-*p*NP-β-D-Glcp. Solvent was evaporated from desired fractions that were pooled and analysed by NMR (D<sub>2</sub>O, 700 MHz).

$^1\text{H}$  NMR (700 MHz, Deuterium Oxide)  $\delta$  8.32 – 8.27 (m, 2H, ArH), 7.31 – 7.26 (m, 2H, ArH), 5.31 (d,  $J$  = 7.8 Hz, 1H, H1'), 4.27 (ddd,  $J$  = 12.2, 5.0, 4.0 Hz, 1H, H4), 4.18 (dd,  $J$  = 4.3, 1.8 Hz, 1H, H5), 4.17 – 4.05 (m, 4H, H7, H8, H4' H5'), 4.04 – 3.96 (m, 2H, H6'), 3.84 – 3.80 (m, 2H, H6, H3'), 2.61 – 2.55 (m, 1H, A3<sub>eq</sub>), 2.00 (s, 3H, A5-NHAc), 1.95 (s, 3H, A7-NHAc), 1.69 (t,  $J$  = 12.7 Hz, 1H, A3<sub>ax</sub>), 1.19 (d,  $J$  = 6.4 Hz, 3H, A9).  $^{13}\text{C}$  NMR (176 MHz, D<sub>2</sub>O)  $\delta$  175.5, 174.2, 174.0, 135.0, 126.1, 118.9, 116.2, 100.2, 99.1, 68.6, 68.1, 67.1, 61.5, 60.3, 59.5, 53.4, 48.8, 39.76, 35.8, 30.7, 29.0, 24.4, 24.1, 22.0, 21.9, 20.8, 19.8, 17.9. HR-MS data was collected with – ESI MS: Expected [M-H]<sup>-</sup> m/z: 616.1995, measured [M-H]<sup>-</sup> m/z: 616.2012. ATR-FTIR  $V_{\text{max}}$ : 3236, 2974, 2897, 1588, 1575, 1562, 1407, 1112, 1078 cm<sup>-1</sup>

## 4.3 Results

### 4.3.1 Sequence analysis of Putative Pseudaminyltransferases

To aid the design of plasmids for overexpression, the amino acid sequences of the five putative PseTs (Lst, WfdL, Gtr16, KpsS1 and Orf14) were analysed for the presence of transmembrane helices. As these putative PseTs are predicted to be involved in the synthesis of cell-surface glycoconjugates, there is a likelihood that these enzymes may be membrane bound. As demonstrated with the Lst, WfdL and Gtr16 homologue NST, truncation of proteins to remove membrane helices can aid protein solubility, which is crucial for *in vitro* characterisation.<sup>32</sup> Sequence analysis found that none of the putative PseTs were predicted to contain transmembrane helices. Therefore, plasmids encoding KpsS1, Orf14, Gtr16 and WfdL were all designed with full-length gene sequences. However, a  $\Delta 9$ Lst (referred to as Lst throughout) construct was designed based on NST sequence similarity to the membrane helix of NST, with the aim of producing soluble Lst.<sup>++++</sup> GTs may often be membrane associated and are therefore often insoluble when produced recombinantly.

To determine a suitable expression strategy for the putative PseTs, the protein sequences were analysed for the presence of predicted Gram-negative secretion signals, which may direct the proteins across the cytoplasmic membrane, to the periplasm or out of the cell.<sup>183</sup> None of the target proteins were predicted to contain Sec secretion signals (Appendix-Table A.2) suggesting that these enzymes are cytoplasmic, therefore it was decided that these recombinant proteins should be produced in the cytoplasm, therefore *E. coli* BL21 (DE3) was chosen as the first expression strain to be used in expression trials.

### 4.3.2 Further Sequence analysis and Homology Modeling of Putative Inverting Pseudaminyltransferases

NST catalyses the formation of an  $\alpha 2,3/\alpha 2,6$  glycosidic bond between, the nonulosonic acid (NulO) Neu5Ac **1.1** and a galactosyl containing moiety.<sup>32</sup> NST is the first enzyme of the GT52 family to be structural characterized and has been shown to possess a GT-B fold.<sup>184</sup> Due to

---

<sup>++++</sup> Lst plasmid was designed by Dr Glyn Hemsworth.

a lack of functional and structural characterization, the NST homologs Gtr16 and Lst are currently listed as GT52 family enzymes (cazy.org), based on sequence similarity to this family. Gtr16, WdfL and Lst are the three putative PseTs that show homology to NST and are predicted to transfer a CMP-Pse donor sugar to a hexose acceptor sugar. Therefore, due to similarities in the substrates of these GTs, it can be expected that homology will be seen between the substrate binding sites and active site. The amino acid sequences of NST and the three retaining putative PseTs were aligned (Figure 4.8) and used in combination with NST structural data to build homology models.

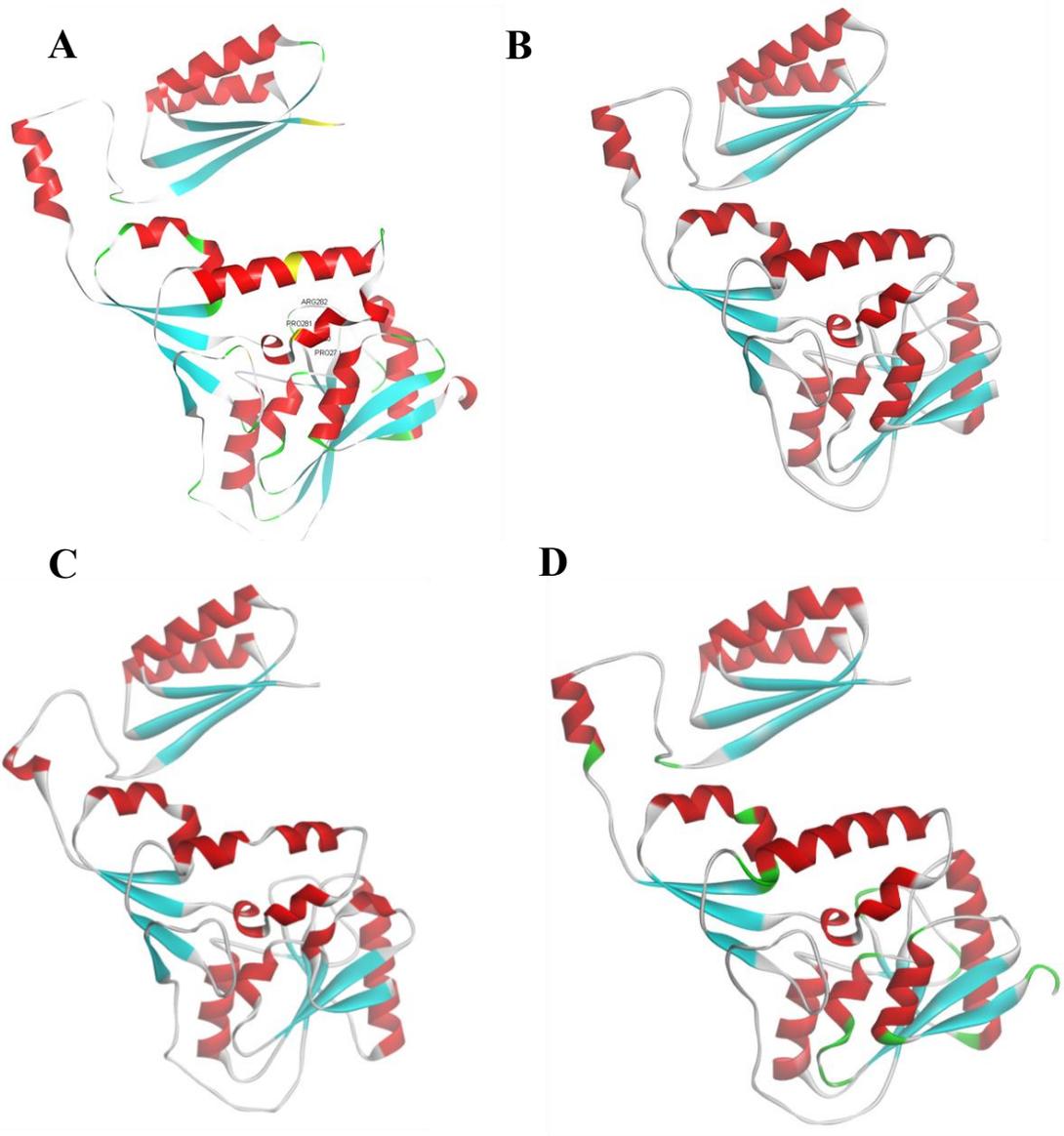
The protein data bank (PDB) contains four structures of Nst (2YK4, 2YK5, 2YK6, 2YK7) that have various ligands bound. The apo structure (2YK4) has the best resolution at 1.94 Å. 2YK4 was used as the template for building models of Lst, WfdL and Gtr16. The target proteins all show less than 30% sequence identity with NST, which is below the ideal threshold for homology modeling. However, in the absence of an alternative structurally characterised homologue and based upon genomic evidence, and experimental evidence for Lst, that these proteins function as PseTs it was decided that NST be used as the template for modeling. Models of this standard may be used to predict overall fold of the enzymes and may cautiously be used to predict the location of key conserved residues within the enzyme.

The GT-B fold of the enzymes can be seen in the NST structure and the putative PseT models (Figure 4.9). GT-B fold enzymes are typically metal-ion independent, NST and its homologues, with the exception of Lst, do indeed lack the candidate DXD motif required for metal-ion co-ordination in GT-A fold enzymes.<sup>171</sup> Sequence alignment shows a conservation of the PHPR motif, which has been characterized as a key motif in NST catalytic activity.<sup>32</sup> The conserved PHPR motif can be seen on what appears to be an equivalent loop on models for all four proteins. Within SiaTs the histidine residue of this conserved motif acts as an acid in the active site, to protonate the leaving group.<sup>32,38</sup> The second proline residue of this motif in NST interacts with the donor's cytidine ring. Finally, the arginine is proposed to form a hydrogen bond with the C-4 hydroxyl of Neu5Ac.<sup>32,185</sup> These residues interact with regions of CMP-Neu5Ac **1.5** that are common to both CMP-Neu5Ac and CMP-Pse5Ac7Ac **1.10**.



**Figure 4.8: Multiple sequence alignments of NST and putative PseTs Gtr16, Lst and WfdL, used to build homology models of putative PseTs. Green lines: highlight residue 49-130, domain-swapped region of NST homodimer. Red box: conserved PTHR motif.**

Structural data shows that NST forms a domain-swapped homodimer, which is believed to increase protein stability and be required for protein activity and specificity.<sup>32</sup> Based on sequence similarity and predicted structural homology of putative PseTs to NST there is likelihood that these putative inverting PseTs may also form dimers. Regions of conservation can be seen in alignments of the four sequences, between residues 49 and 130, which correspond to the domain-swapped region of NST homodimers (Figure 4.9).<sup>32</sup> SDS-PAGE analysis of NST showed the presence of two bands, with molecular weights corresponding to both the NST monomer and homodimer.<sup>41</sup> The homology of NST to Lst, Gtr16 and WfdL suggests that if present, dimerization caused by domain swapping may be visible on SDS-PAGE. However, a hydrophobic cavity is present in NST at the point of the domain-swap. Data suggests that a fatty acyl chain of the cell membrane may fill the hydrophobic cavity.<sup>32</sup> As the putative PseTs are not predicted to be a membrane-associated protein the fatty acyl filled cavity may not be present.<sup>104</sup>



**Figure 4.9: Structure of NST from crystal data and homology models of putative inverting PseTs. A: NST structure from PDB file 2YK4. B: WfdL homology model. C: Lst homology model. D: Gtr16 homology model.**

### 4.3.3 Initial Expression trials of Putative Pseudaminyltransferases

Plasmids encoding the five putative PseTs were constructed from pET vectors, which produce target proteins with a cleavable N-terminal hexahistidine tag (His-tag). The plasmids contain T7-lac promoters and expression of the gene insert is therefore induced by the addition of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Plasmids encoding His-tagged genes for each of the five putative PseTs to be studied were transformed into *E. coli* BL21 (DE3) to be used in expression trials, with the aim of obtaining soluble protein, which

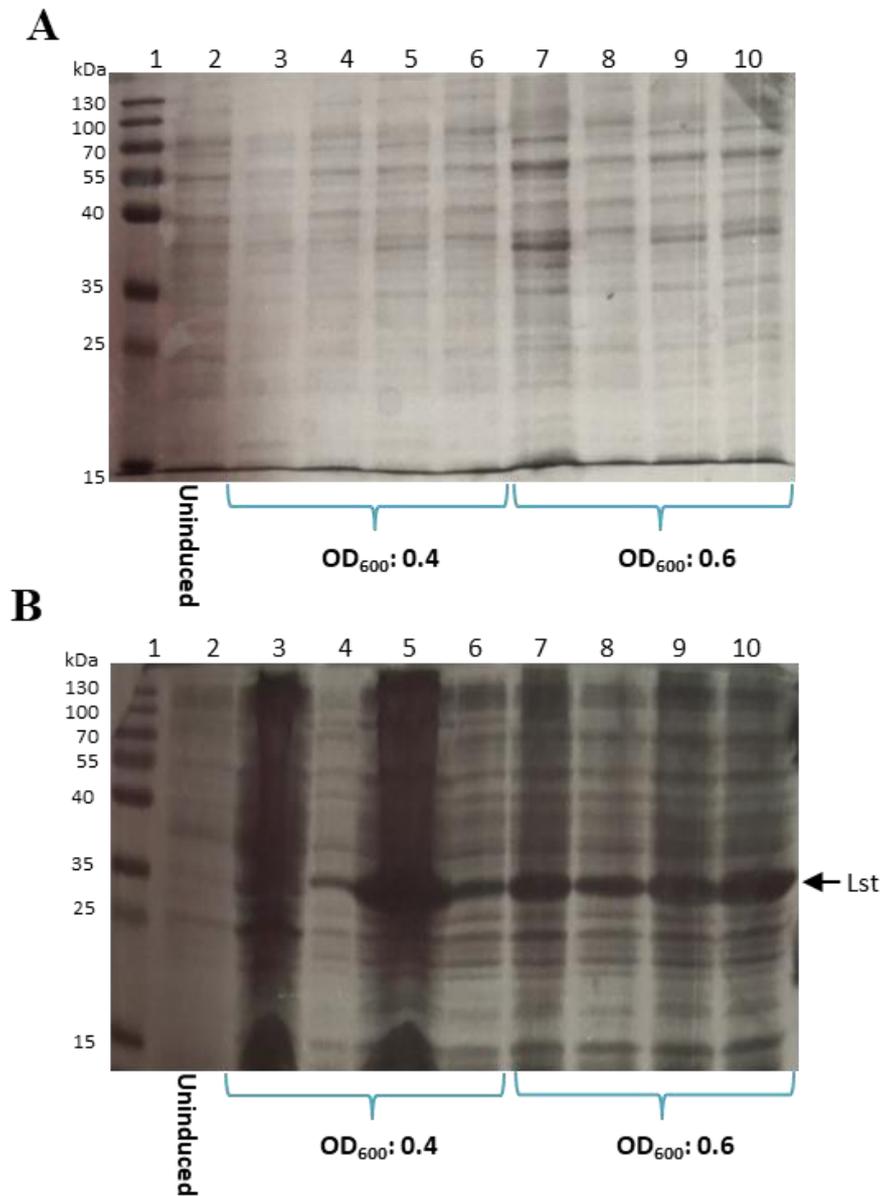
may be purified by affinity chromatography using the N-terminal His-tag, for use in characterisation studies.

Throughout these trials the following parameters were studied for their impact on the level of protein expression: temperature, OD<sub>600</sub> of cells when expression was induced, IPTG concentration used for induction and the time cells were cultured for post-induction. Samples of cells were lysed and analysed by SDS-PAGE to assess whether the putative PseT was present and soluble. SDS-PAGE analysis showed that all proteins were either not overexpressed or insoluble when produced under the conditions studied (Figure 4.10 shows Lst, which was insoluble when incubated at 25 °C (B) and Figure 4.11 shows KpsS1 which was insoluble when incubated at 30 °C (B)).

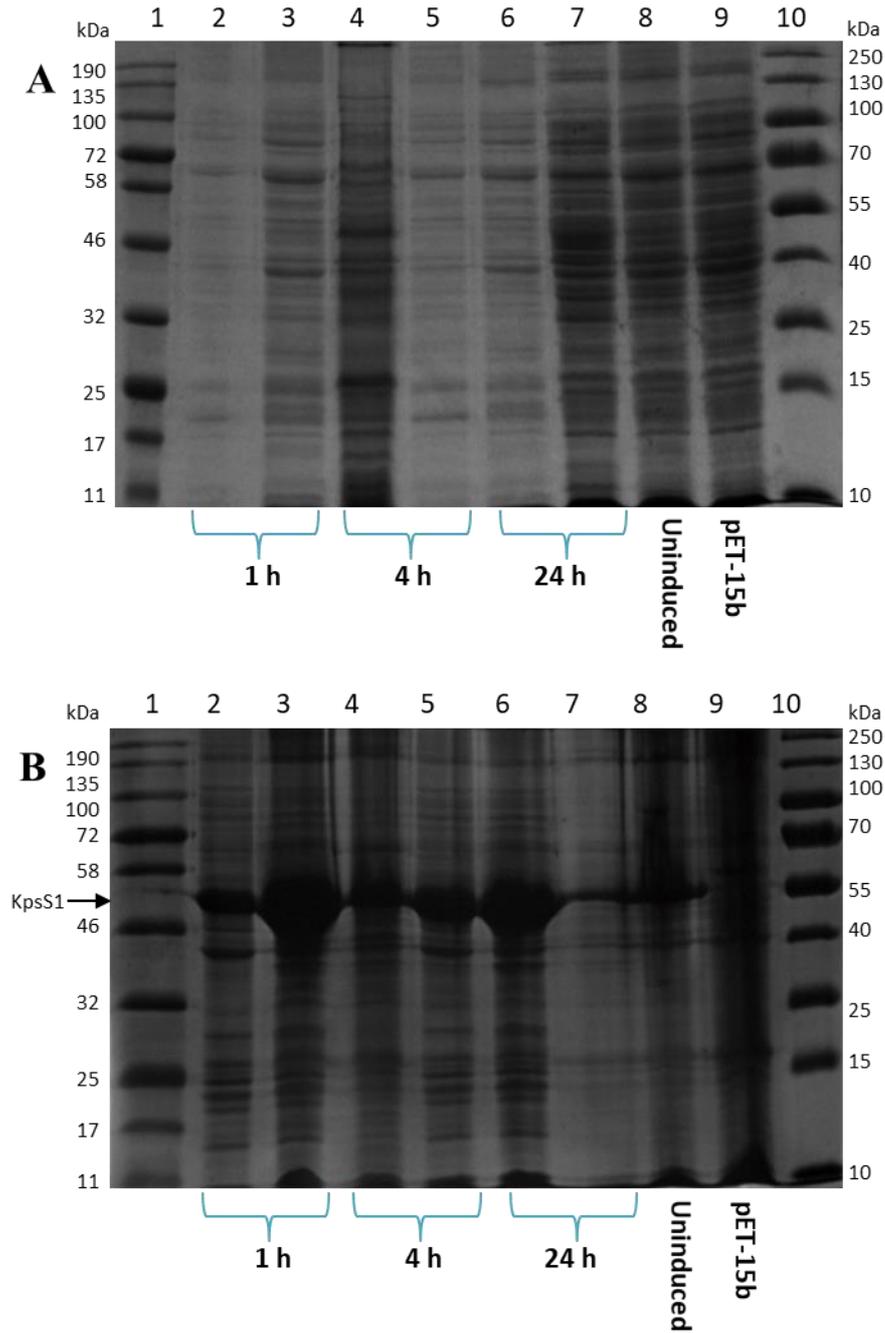
*E. coli* BL21 (DE3) encoding Lst were cultured for expression trials, however cells cultured at 20 and 30 °C post-induction did not show expression of Lst, after cell pellets were lysed in Bugbuster mix and analysed via SDS-PAGE. Lst was shown to be expressed in *E. coli* BL21 (DE3) cultured at 25 °C post-induction. However, the 38.3 kDa recombinant protein was insoluble when cells were lysed in Bugbuster protein extraction reagent (Figure 4.10, B, Lanes 4-10). No band corresponding to the approximate MW of Lst (38.3 kDa) can be seen in the soluble fractions.

Recombinant KpsS1 was not present in samples from expression trials which were performed at 20 or 25 °C (determined by the lack of a band at the approximate MW of KpsS1 in both soluble and insoluble fractions on SDS-PAGE). However, SDS-PAGE analysis revealed that KpsS1 was highly overexpressed in *E. coli* BL21 (DE3) cultured at 30 °C post-induction (Figure 4.11, B, Lanes 2-7. KpsS1 MW: 59.4 kDa). Some insoluble KpsS1 can be seen in the uninduced sample, suggesting basal or leaky expression occurred (Figure 4.11, B, Lane 8). There is no band visible at the approximate MW of KpsS1, either the insoluble or soluble samples of cells containing pET-15b vector, further indicating that the band seen at approximately 59 kDa in the other insoluble samples is KpsS1. However, no bands corresponding to KpsS1 can be seen in the soluble samples (Figure 4.11, A, Lanes 2-8). Similarly, Gtr16, Orf14 and WfdL were found to be overexpressed but insoluble, in the conditions screened (SDS-PAGE in Appendix).

Due to the availability of CMP-Pse5Ac7Ac to this study (described in Chapter 2), it was concluded that putative PseTs that utilise this donor should be studied further. On this basis, as Orf14 and WfdL are predicted to use alternative CMP-Pse derivatives as donors (CMP-Pse Pse5Hb7Fm and CMP-Pse5Ac7Hb respectively) these target proteins were not studied further. Given the mechanistic differences of inverting and retaining GTs, it was decided that both an inverting and retaining candidate protein should be studied further. Therefore, efforts were focused on the solubilisation of the putative inverting PseT, Lst and putative retaining PseT, KpsS1.



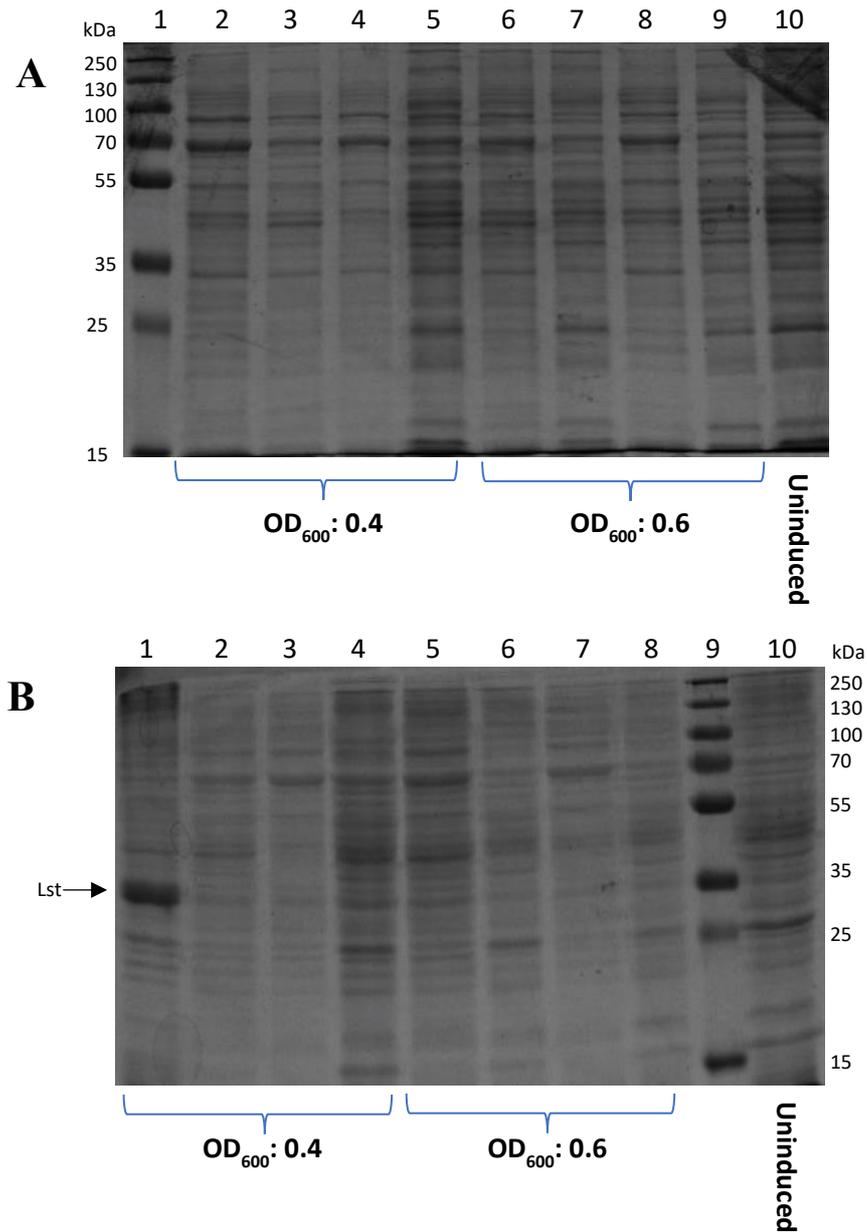
**Figure 4.10: 12% SDS-PAGE of expression trials for Lst (MW: 38.3 kDa) in *E. coli* BL21 (DE3) cultured at 25 °C. A: Cells lysed in Bugbuster mix, soluble fractions. B: Cells lysed in Bugbuster mix, insoluble fractions. Lanes loaded as follows: 1-PAGE-RULER MW marker 2- Uninduced sample; 3-OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 4 h; 4- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 24 h; 5- OD<sub>600</sub>: 0.4, 1 mM IPTG, 4 h; 6- OD<sub>600</sub>: 0.4, 1 mM IPTG, 24 h; 7- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 8- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 9- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h Uninduced sample; 10- : 0.6, 1 mM IPTG, 24 h.**



**Figure 4.11: 12% SDS-PAGE of expression trial for KpsS1 (MW: 59.4kDa) in *E. coli* BL21 (DE3) cultured at 30 °C. A: Cells lysed in Bugbuster mix, soluble fractions. B: Cells lysed in Bugbuster mix, insoluble fractions. Lanes loaded as follows: 1-Broad range MW marker 2- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 1 h; 3- OD<sub>600</sub>: 0.6, 1 mM IPTG, 1 h; 4- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 5- OD<sub>600</sub>: 0.6, 1 mM IPTG, 4h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 7- OD<sub>600</sub>: 0.6, 1 mM IPTG, 24 h; 8- Uninduced sample; 9- pET-15b vector (no KpsS1 insert) 10- PAGE Ruler MW marker.**

#### **4.3.4 Expression trials of Putative Pseudaminyltransferases, Lst and KpsS1 in *E. coli* Tuner (DE3)**

As expression trials using *E. coli* BL21 (DE3) had yielded highly overexpressed, insoluble protein, an alternative expression cell line was used in expression trials. *E. coli* Tuner™ (DE3) cells vary from BL21 (DE3) as they contain a *lacZY* gene knockout. Therefore, this expression strain was chosen with the aim of reducing basal expression of the putative PseTs, in the hope that lower level expression may promote solubility of putative PseTs. *E. coli* Tuner™ (DE3) encoding Lst or KpsS1, were grown at 20, 25 and 30 °C post-induction. SDS-PAGE analysis of expression trial samples revealed KpsS1 was not overexpressed in *E. coli* Tuner™ (DE3), under the conditions tested. SDS-PAGE analysis showed that for cells containing Lst plasmid, Lst was only potentially overexpressed in one condition and this protein was insoluble (Figure 4.12- B: Lane 1. Lst MW: 38.3 kDa.). Therefore, no soluble Lst was produced in *E. coli* Tuner™ (DE3) during these expression trials (shown by a lack of band in Figure 4.12- A: Lanes 1-8, corresponding to the approximate MW of Lst.) As this cell line failed to yield the proteins of interest, it was decided that work should focus on solubilisation of putative PseTs expressed in *E. coli* BL21 (DE3).

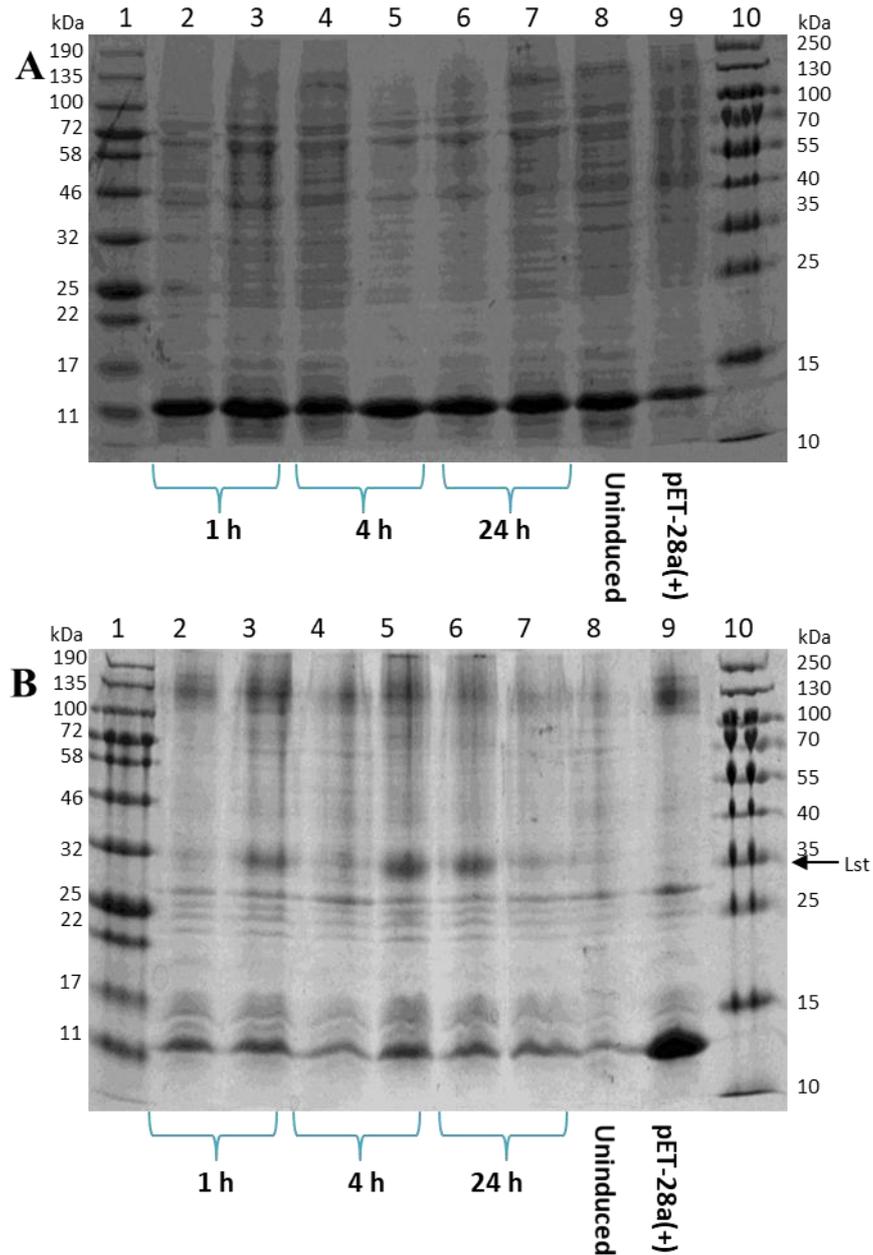


**Figure 4.12: 12% SDS-PAGE of expression trials for Lst (MW: 38.3 kDa) in *E. coli* Tuner™ cultured at 25 °C. A: Cells lysed in Bugbuster mix, soluble fractions. B: Cells lysed in Bugbuster mix, insoluble fractions. Lanes loaded as follows in A: 1- PAGE Ruler MW marker; 2- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 4 h; 3- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 24 h; 4- OD<sub>600</sub>: 0.4, 1 mM IPTG, 4 h; 5- OD<sub>600</sub>: 0.4, 1 mM IPTG, 24 h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 7- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 8- OD<sub>600</sub>: 0.6, 1 mM IPTG, 4 h; 9- OD<sub>600</sub>: 0.6, 1 mM IPTG, 24 h; 10- Uninduced sample. B: 1- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 4 h; 2- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 24 h; 3- OD<sub>600</sub>: 0.4, 1 mM IPTG, 4 h; 4- OD<sub>600</sub>: 0.4, 1 mM IPTG, 24 h; 5- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 7- OD<sub>600</sub>: 0.6, 1 mM IPTG, 4 h; 8- OD<sub>600</sub>: 0.6, 1 mM IPTG, 24 h; 9- PAGE Ruler MW marker; 10- Uninduced sample.**

### **4.3.5 Solubilisation trials of Putative Pseudaminyltransferase Lst**

Following the initial expression trials, it was concluded that further efforts should be made to soluble Lst from cells cultured at 25 °C post-induction. With the aim of solubilising Lst, several additives were used alongside Bugbuster during cell lysis. Additives included 1 M NaCl, 8 M urea, or 6 M guanidine HCl. The chaotropic agents, urea and guanidine HCl were both added to lysis buffer in concentrations that are frequently used to denature protein, in order to solubilise the protein. Following treatment with urea or guanidine HCl, solubilized proteins should then be refolded. However, the success of refolding to yield active protein is often limited.<sup>186,187</sup> NaCl was added to lysis buffer with the aim of overcoming any electrostatic interactions that were preventing Lst from solubilizing.

The guanidine HCl supplemented lysis buffer caused precipitation of proteins in the sample, therefore these samples were discarded. Neither NaCl nor urea was able to solubilise Lst (Figure 4.13 shows samples with urea. 4.13-A shows a lack of soluble Lst in all samples. Lst can be observed in the insoluble material, Figure 4.13 B, red box, Lanes 2-7). As these three additives were unsuccessful in solubilising Lst, it was concluded that detergents should be screened for their ability to yield soluble Lst.



**Figure 4.13: 12% SDS-PAGE of expression trials for Lst (MW: 38.3 kDa) in *E. coli* BL21 (DE3) cultured at 25 °C. Cells lysed in Bugbuster mix, plus 1 M NaCl. A: soluble fractions. B: insoluble fractions. Lanes loaded as follows: 1-Broad range MW marker 2- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 1 h; 3- OD<sub>600</sub>: 0.6, 1 mM IPTG, 1 h; 4- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 5- OD<sub>600</sub>: 0.6, 1 mM IPTG, 4h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 7- OD<sub>600</sub>: 0.6, 1 mM IPTG, 24 h; 8- Uninduced sample; 9- pET-28a vector (no Lst insert) 10- PAGE Ruler MW marker.**

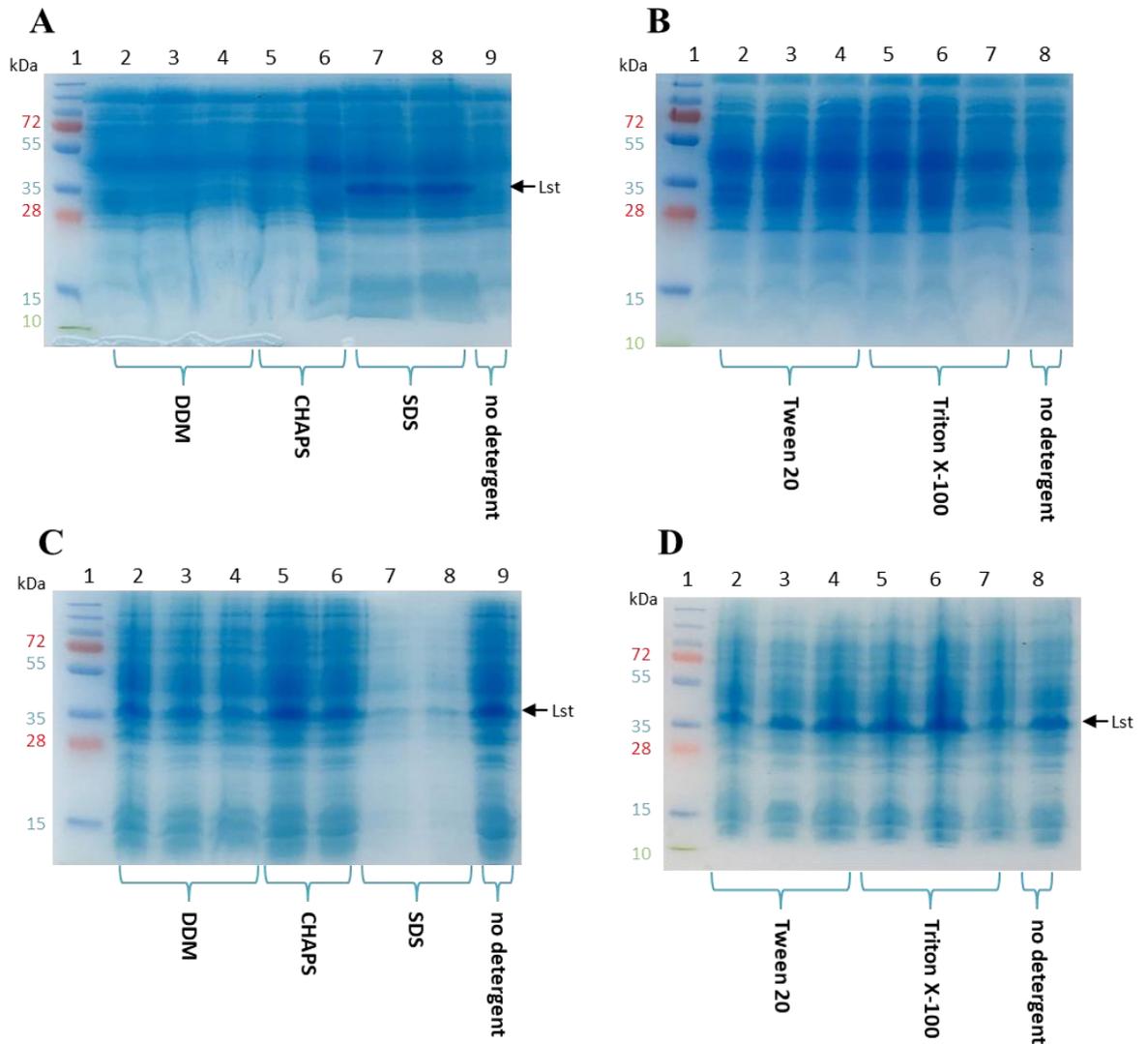
#### 4.3.6 Detergent Screen for Solubilisation of Putative Pseudaminyltransferases

##### Lst and KpsS1

*Lst*: As discussed previously, initial expression trials had found that Lst was insoluble when expressed *E. coli* BL21 (DE3) (Figure 4.10 B lanes 2-10). Based upon the results of

initial expression trials of Lst a number of detergents were screened for their ability to solubilize Lst. It was shown that for Lst homologue NST, Triton X-100 was required to stabilize the truncated protein during purification and structural characterization.<sup>32</sup> However, it is worth proceeding with caution as the addition of detergent such as Triton X-100, may interfere with protein function, impact structural studies and can be problematic to remove from samples.

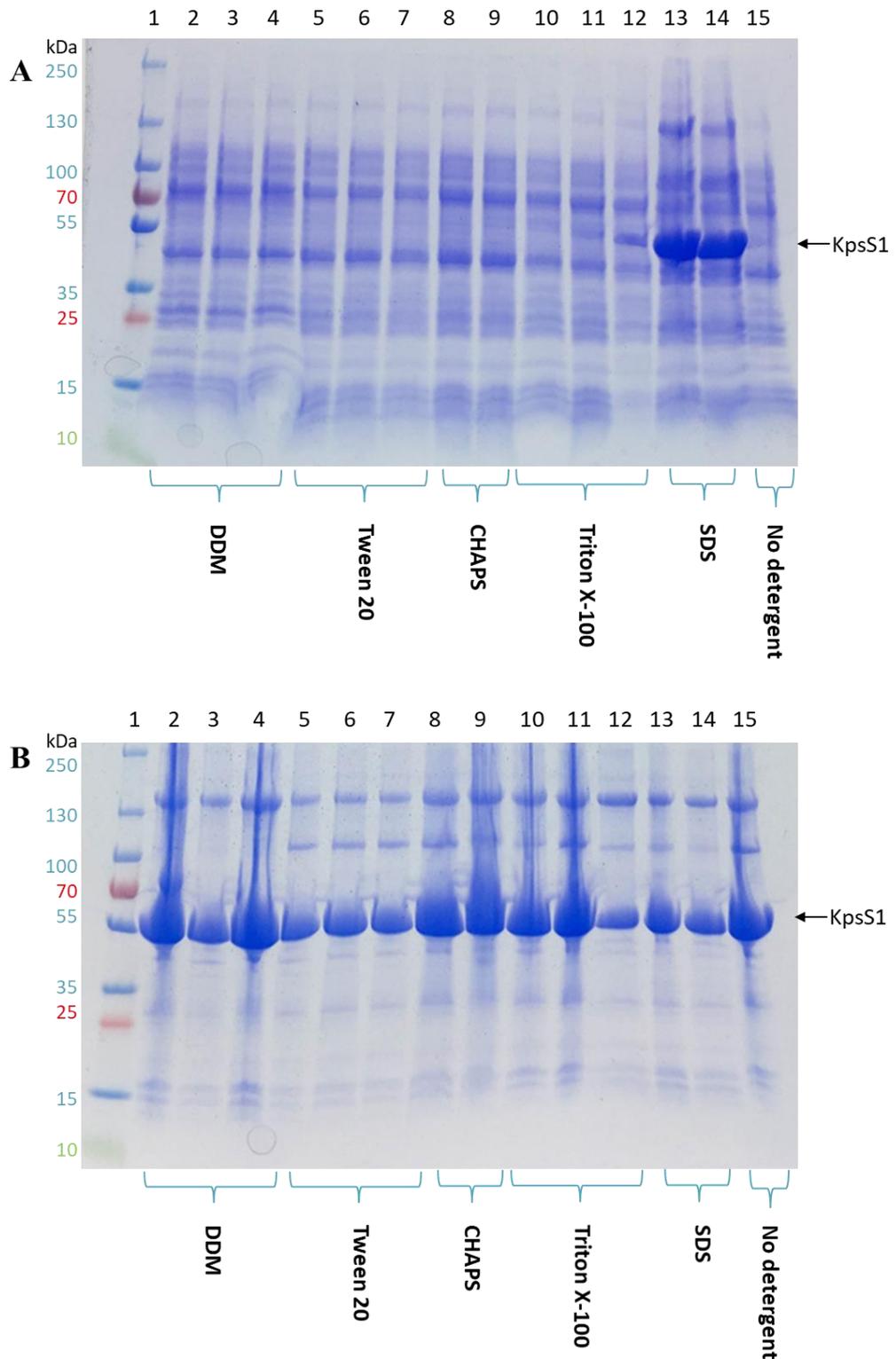
Each of the chosen detergents were added to lysis buffer in various concentrations, where necessary the detergent concentration was limited to that compatible with HisTrap chromatography to facilitate downstream purification of His<sub>6</sub>-tagged Lst. SDS-PAGE analysis showed that in the presence of SDS (Figure 4.14, A, lanes 7 and 8) a limited amount of Lst appeared to have solubilised. The use of SDS to solubilise protein that retains function is highly unfeasible as SDS denatures proteins and attempts at refolding Lst may not yield the functional protein. For all other detergents Lst appeared to remain in the insoluble fraction. Due to the inability to solubilise functional Lst from the strategies used, in combination with challenges that would likely be encountered in activity assays due to the undetermined identity of the predicted carbohydrate acceptor, no further studies with Lst were performed.



**Figure 4.14: Lst Detergent Screen: A and B: soluble material. C and D: insoluble material. Lanes loaded as follows: A, B, C, D: 1- PAGE Ruler MW marker; A and C: 2, 3, 4-DDM; 5, 6- CHAPS; 7, 8- SDS, 9- no detergent; B and D: 2, 3, 4- Tween 20, 5, 6, 7- Triton X-100; 8- no detergent; A: 10- Broad Range MW; B: 9- Broad Range MW marker; C: 10- Blank; D: 9-Blank. Concentrations of each detergent increase as lane number increases. (see Table 4.3 for concentrations of detergents.)**

*KpsS1*: As previously discussed, expression trials for the production of recombinant *KpsS1* had shown no expression in *E. coli* Tuner cells and had yielded insoluble *KpsS1* in *E. coli* BL21 (DE3). With the aim of solubilising *KpsS1* from *E. coli* BL21 (DE3) cell lysate, five detergents were trialled for their ability to solubilise the putative retaining PseT *KpsS1*, as described for Lst. SDS-PAGE showed that SDS (Figure 4.15, A, lanes 13 and 14) appeared to have solubilised a limited amount of *KpsS1*. In both cases a large proportion of *KpsS1* remained within the insoluble fraction (Figure 4.15, B, lanes 13 and 14). However, the use

of SDS to solubilise KpsS1 is unfavourable as this detergent will denature KpsS1 and refolded KpsS1 may not retain function. It appeared that some KpsS1 was solubilised by the addition of Triton X-100 to cell lysate (Figure 4.15, A, lanes 10-12, MW: 59.4 kDa) as SDS-PAGE analysis of soluble material showed the presence of a band appearing to correspond to KpsS1 (Figure 4.15, A, lanes 10-12) but the majority of this protein remained insoluble (Figure 4.15, A: lanes 2-15). 5% Triton X-100 yielded more soluble protein than the lower concentrations, however concentrations greater than 2% are not compatible with purification of His-tagged KpsS1, as they should not be applied to HisTrap columns. Additionally, as LC-MS was the preferred method to analyse activity screens of KpsS1, no further work was performed with KpsS1 and Triton X-100, as the presence detergent is not suitable for LC-MS analysis.<sup>188</sup>



**Figure 4.15: SDS-PAGE analysis of detergent screen with KpsS1. A: soluble material. B: insoluble material. Lanes loaded as follows: A and B: 2, 3, 4-DDM; 5, 6, 7- TWEEN 20; 9, 10- CHAPS, SDS, 10, 11, 12 Triton X-100 13, 14-SDS; 15- no detergent. Concentrations of each detergent increase as lane number increases. (See Table 4.3 for concentrations of detergents.)**

### 4.3.7 KpsS1 Fusion Plasmids

A common strategy for solubilising a recombinant insoluble protein of interest (POI) is to create a construct in which the POI is fused to a highly soluble protein. The highly soluble protein may aid the solubilisation of the POI, however the underlying mechanism for this enhanced solubility is unclear. It is likely that a fusion protein will act as a chaperone to the POI to promote proper folding, reducing the likelihood of POI aggregation and inclusion body formation.<sup>189</sup> As the addition of detergents to cell lysates containing KpsS1 had shown limited success in solubilising KpsS1 for use in functional studies, pET-based vectors each encoding a protein commonly used as fusion-proteins were used in the construction of KpsS1 fusion-protein plasmids. The use of MBP fusion to solubilise KpsS from *E. coli* provides precedent for the use of this strategy for the homologous KpsS1.<sup>30</sup>

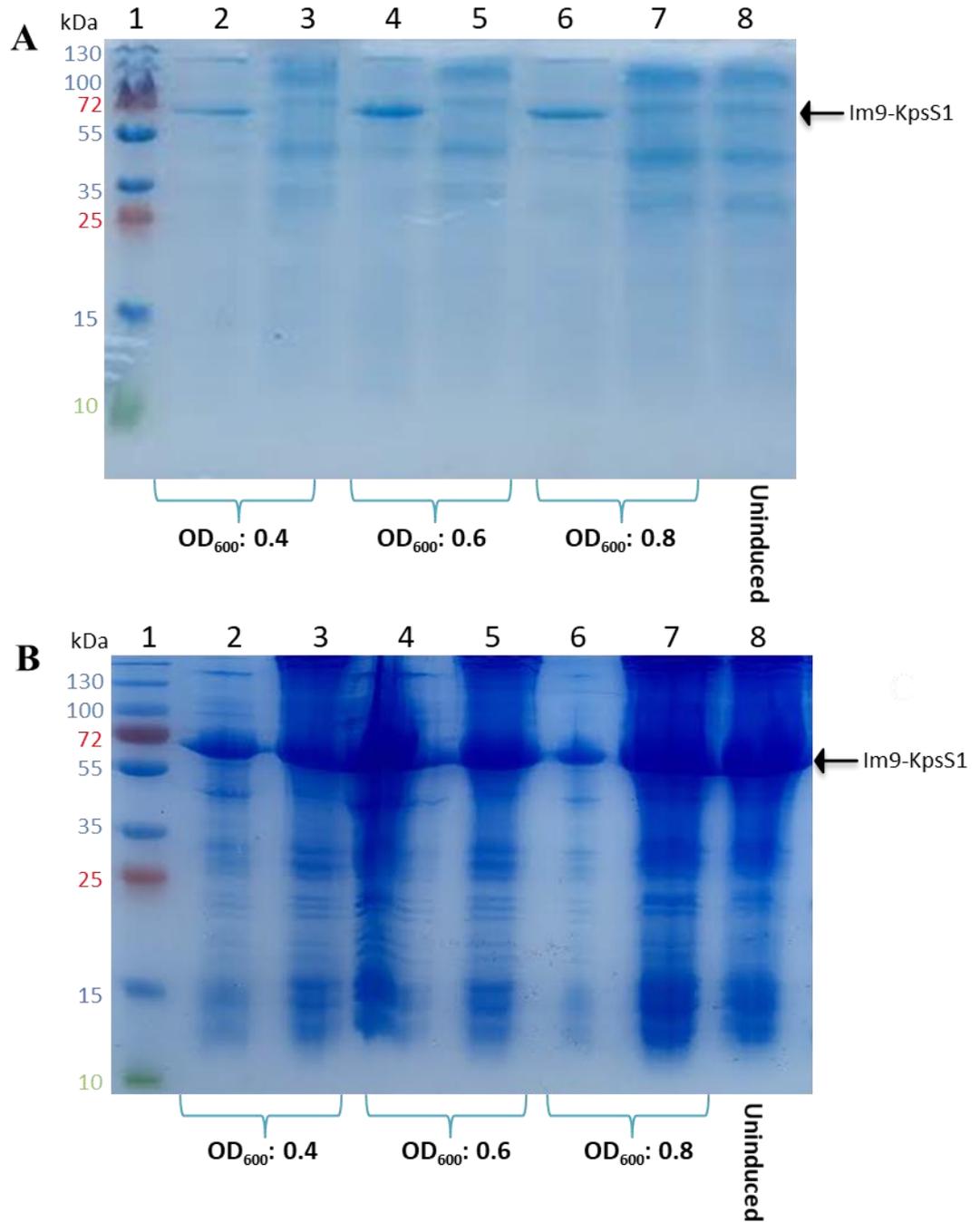
The MBP, GST and Im9 vectors were linearised by PCR, generating overhangs at both the 3' and 5' ends of the DNA. PCR was also used to amplify *kpsS1* from the pET-15b plasmid, where complementary overhangs were added to either end, to facilitate cloning (PCR products analysed by gel electrophoresis). Following the cloning reaction XL-1 Blue Supercompetent cells were transformed with product, and cells were screened for the presence of the desired plasmid. Colony PCR and sequencing confirmed that Im9-KpsS1 and GST-KpsS1 had been successfully transformed into *E. coli* XL-1 Blue Supercompetent cells, however no MBP-KpsS1 was obtained. The MBP-KpsS1 fusion reaction and transformation of product was repeated. However, no positive results were obtained, suggesting that the fusion reaction was unsuccessful. Im9-KpsS1 and GST-KpsS1 plasmid was transformed into *E. coli* BL21 (DE3) for use in expression trials.

### 4.3.8 KpsS1 Fusion Proteins Expression Trials

With the aim of producing soluble KpsS1-fusion proteins, the Im9-KpsS1 and Gst-KpsS1 plasmids were each transformed into *E. coli* BL21 (DE3) cells. The transformed cells were used in expression trials where growth temperature, OD<sub>600</sub> at induction, IPTG concentration and incubation period after induction were all varied.

Cells were lysed in Bugbuster mix, as described for initial expression trials, and subjected to SDS-PAGE analysis. Under many conditions KpsS1-fusion proteins were insoluble, however low levels of soluble Im9-KpsS1 (MW: 68.9 kDa) was recombinantly expressed in

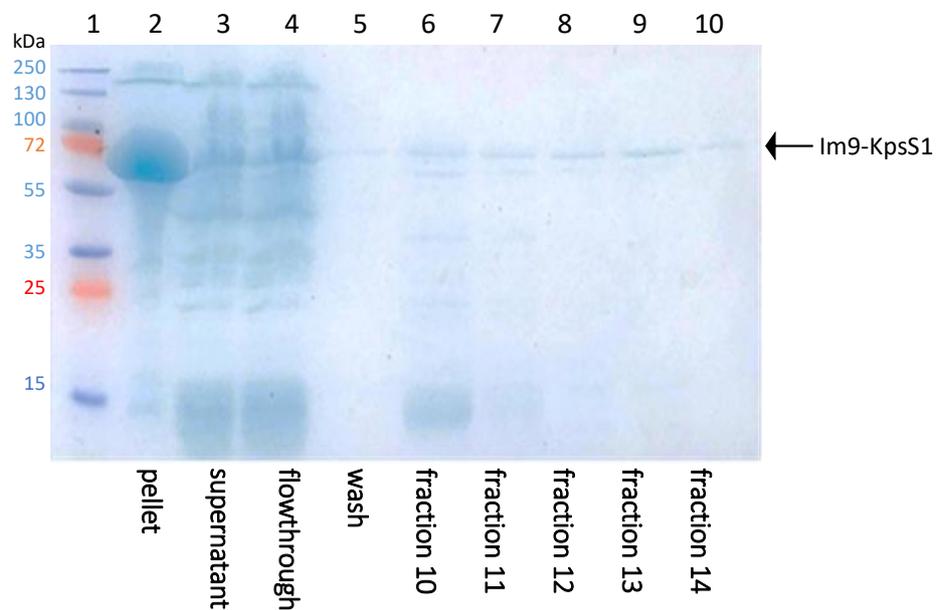
cells induced at an OD<sub>600</sub> value of 0.4, 0.6 or 0.8 with 0.1 mM IPTG, and incubated at 30 °C for 4 hrs post-induction (Figure 4.16, A, lanes 2, 4 and 6). All cultures analysed after overnight incubation show a lack of overexpressed Im9-KpsS1 (Figure 4.16, A, lanes 3, 5 and 7), suggesting that soluble protein may aggregate or may be degraded over time. Cells were cultured on a larger scale, using conditions that had appeared to yield soluble protein during the initial Im9-KpsS1 expression trial (OD<sub>600</sub> value of 0.6 when induced with 0.1 mM IPTG, incubated at 30 °C for 4 hrs post-induction) with the aim of purification for functional studies.



**Figure 4.16: SDS-PAGE analysis of Im9-KpsS1 Expression Trial.** Samples from *E. coli* BL21 (DE3) cells cultured at 30 °C, induced with 0.1 mM IPTG. A: soluble material. B: insoluble material. Lanes loaded as follows: 1- PAGE Ruler MW marker, 2- OD<sub>600</sub> 0.4, 4 h post-induction (PI), 3- OD<sub>600</sub> 0.4, 20 h PI, 4- OD<sub>600</sub> 0.6, 4 h PI, 5- OD<sub>600</sub> 0.6, 20 h PI, 6- OD<sub>600</sub> 0.8, 4 h PI, 7- OD<sub>600</sub> 0.8, 20 h PI, 8- Uninduced.

### 4.3.9 Im9-KpsS1 Purification

As expression trials had shown that the fusion of KpsS1 to Im9 did appear to yield limited amounts of soluble protein, *E. coli* BL21 (DE3) cells containing Im9-KpsS1 plasmid were cultured on a 2.4 L scale, with the aim of purifying Im9-KpsS1. SDS-PAGE analysis showed that His-tagged Im9-KpsS1 appeared to be purified from the supernatant of cell lysate using nickel-affinity chromatography (Figure 4.17, lanes 6-10, fractions 10-14, Im9-KpsS1 MW: 69.9 kDa). However, the majority of Im9-KpsS1 had remained insoluble and was found in the pellet after cells were lysed and centrifuged (Figure 4.17, lane 2). Fractions 10-14 were pooled and applied to a desalting column to exchange the protein into buffer that was suitable for use in activity assays.



**Figure 4.17: 10% SDS-PAGE analysis of Im9-KpsS1 purification. Lanes loaded as follows: 1: PAGE-RULER MW marker, 2: pellet, 3: supernatant, 4: flowthrough, 5: wash, 6: fraction 10, 7: fraction 11, 8: fraction 12, 9: fraction 13, 10: fraction 14.**

### 4.3.10 Im9-KpsS1 Protein Identification Mass Spectrometry

Following purification, to confirm the identity of the band that appeared to be Im9-KpsS1, a gel extraction was performed to obtain protein from the band (From SDS-PAGE, following desalting column). The protein was subjected to a trypsin digest and resulting peptide fragments were analysed mass spectrometry (MS), the data obtained was then analysed using a database to compare MS data to a database of peptide MS, in order to identify the protein. Results confirmed that Im9-KpsS1 had been purified, as peptides corresponding to

both Im9 and KpsS1 were identified in the protein ID MS analysis (Figure 4.18, blue text). Peptide was found which corresponded to residues 552-563 of this 587 amino acid protein, suggesting that full-length protein was being purified. Whilst SDS-PAGE results showed that the Im9-KpsS1 was not entirely pure, no further purification steps were attempted due to the low yield. This level of purity of Im9-KpsS1 was sufficient for use in functional activity screens.

```

1  MGSSHHHHHH SSMELEKHSIS DYTEAEFLQL VTTICNADTS SEEELVKLVT
51  HFEEMTEHPS GSDLIYYPKE GDDDSPSGIV NTVKQWRAAN GKSGFKQGLE
101 VLFQGPAMNF LILINSAPNY KYFFYELAKE IESRGHNIYF AIDSHRSKYL
151 EPLPELDNNQ NSFFFDSTYLE KNFDKNLSVS HNNNQEYWGD YFYSDYDRFL
201 THDFNLNKDK NYWLVNVKVSLS DSFFEDI IKD KQIDFVLYEN ISNSFAYAAY
251 LQCTKLGGKY IGLMGSRLPN HFEIQNSIVE EELKKLEILA QKPITQDEME
301 WFENYKKSIV DIQPDYMKQN GLDNVAISRI VKLNKFLKAL RLLTIGFKYK
351 HYYDYQFGNP FMVPIKAIRV NIKRYLNTKK SQKFYINNDE LEICSSKERF
401 YIYPIHFHPE SSTSVLAPEY TNEYSNIINI ANNLPFGTYL YVKDHKSAKG
451 VQSYEFYKKV SSLPNVRLVN FDVNIKRLIL KSLGVITVNS TAGYEALLLG
501 KPVYLLGRVF YENFNNVYNL KSFRDIRDIR DILDFQFLDV KKDFIAYKKY
551 VYKGVIFIDY GNRVNDKKRY FSELVDSIFL NINTDIE

```

**Figure 4.18: Im9-KpsS1 amino acid sequence. Protein-ID MS found peptides shown in blue to confirm identity of protein extracted from gel.**

#### 4.3.11 Im9-KpsS1 Activity Screen

As Im9-KpsS1 had been successfully solubilised and purified, to test whether Im9-KpsS1 functions as an active PseT, reactions were performed using the predicted donor CMP-Pse5Ac7Ac and a range of acceptor carbohydrates. Initial activity assays were performed using crude lysate of cells that were overexpressing Im9-KpsS1 and partially purified Im9-KpsS1. Reaction samples were subjected to LC-MS analysis, where the presence of starting material CMP-Pse5Ac7Ac and/or Pse-based products determined whether Im9-KpsS1 had activity towards these substrates, under the conditions tested (results summarised in Table 4.7 and Figure 4.19). Control reactions were also performed which lacked either Im9-KpsS1, CMP-Pse5Ac7Ac or no acceptor.

Given the structure of the *A. baumannii* CPS repeat unit **4.4** (Figure 4.4) where Pse5Ac7Ac is bound to  $\beta$ -D-Glcp; D-Glcp **4.6** was the first acceptor to be screened in this activity assay. LC-MS analysis of this sample did not show a peak corresponding to Pse5Ac7Ac-D-Glcp ( $[M-H]^-$ : 495) and a peak corresponding to CMP-Pse5Ac7Ac ( $[M-H]^-$ : 638.1) was present. A CMP-Pse5Ac7Ac peak was also observed in the no enzyme and no acceptor controls but was not present in the sample without donor. From these results it was concluded that enzymatic activity was not observed with Im9-KpsS1 and D-Glcp acceptor, suggesting that D-Glcp monosaccharide is not the native acceptor substrate for KpsS1.

Based upon this result and the structure of the KL2 CPS repeat unit structure (Figure 4.6), a non-reducing glucoside Me- $\beta$ -D-Glcp **4.7**, was the next substrate to be screened with Im9-KpsS1 as described for Glc. LC-MS analysis showed a lack of Im9-KpsS1 activity with Me- $\beta$ -Glcp, a molecular ion peak for CMP-Pse5Ac7Ac was observed at m/z 638.1, however, no peak corresponding to a Pse5Ac7Ac-Me- $\beta$ -Glc ( $[M-H]^-$ : 509.2) was present. Again, this result suggested that in CPS biosynthesis within *A. baumannii*, KpsS1 utilises a di-, tri- or tetrasaccharide acceptor. As the disaccharide  $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp is not commercially available, pNP- $\beta$ -D-Glcp **4.8** was used as a  $\beta$ -D-Glc-based disaccharide mimic. LC-MS analysis subsequently showed a peak corresponding to Pse5Ac7Ac-pNP- $\beta$ -D-Glcp **4.12** ( $[M-H]^-$ : 616.2) (Figure 4.19, B) and a lack of CMP-Pse5Ac7Ac **1.10**, observed with full conversion of CMP-Pse5Ac7Ac to disaccharide product **4.12** (summarised in Table 4.7).

As Im9-KpsS1 showed activity with a disaccharide mimic pNP- $\beta$ -D-Glcp **4.8**, the disaccharide  $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe **4.9** was produced in-house via a *de novo* synthesis. This compound was chosen as potential acceptor as **4.9** is structurally closer to that of the *A. baumannii* CPS repeat unit **4.8** than pNP- $\beta$ -D-Glcp. Whilst LC-MS analysis of the Im9-KpsS1 reaction suggested that 9 % turnover of **4.9** was observed, due to the presence of a peak corresponding to the molecular ion of Pse5Ac7Ac- $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe **4.13** ( $[M-H]^-$  m/z 671.3) (Figure 4.19, C), Pse5Ac7Ac ( $[M-H]^-$ : 332.8) peaks were observed and CMP-Pse5Ac7Ac remained in the samples.

The low level of turnover may be due to the presence of the Im9 tag on KpsS1, this may alter the folding of KpsS1 compared to that of the wild type enzyme. Alternatively, the low activity observed for KpsS1 with this substrate, maybe due to the effect of the C1 methyl substituent on Galp. The structure of the CPS repeat unit in *A. baumannii* **4.4** suggests that

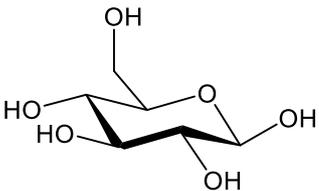
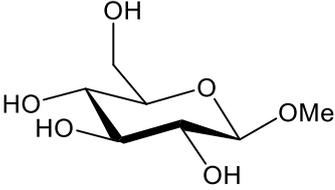
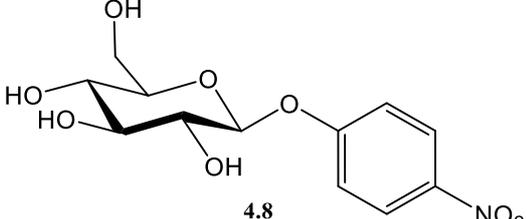
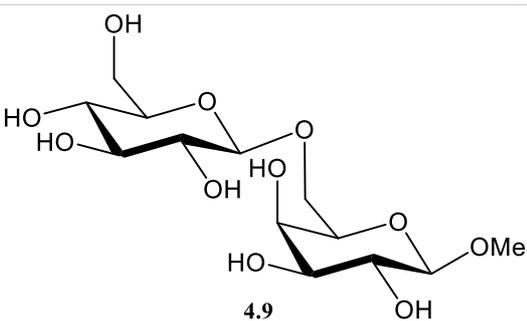
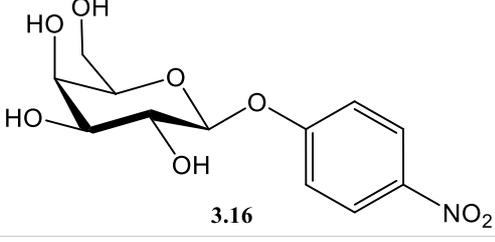
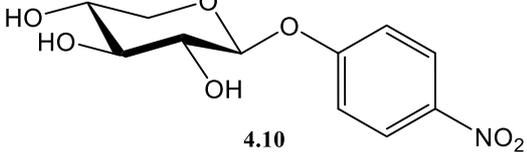
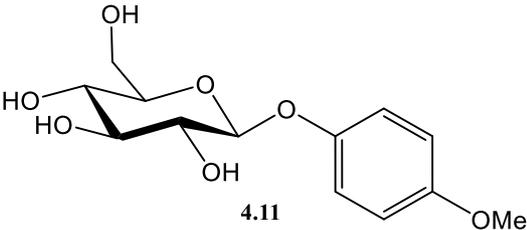
KpsS1 may require an acceptor closer to the natural tri- or tetrasaccharide moiety was the natural acceptor. The low level of activity of Im9-KpsS1 towards disaccharide **4.9** suggests that the third and fourth sugars of the acceptor may contribute significantly towards decreasing the  $K_m$  of binding. Kenyon *et al.* noted that GTs in *Acinetobacter baumannii* KL gene clusters are often distributed in inverse order to function.<sup>172,190</sup> KpsS1 is the first GT in the KL6 gene cluster, upstream from the remain putative GTs (Figure 4.3), this genetic evidence supports the hypothesis that the terminal Pse5Ac7Ac residue is the final carbohydrate to be installed into the tetrasaccharide repeat unit **4.4**.

To test substrate specificity towards the Glcp moiety of the acceptor, *p*NP- $\beta$ -D-Glcp was replaced as the acceptor by *p*NP- $\beta$ -D-Galp **3.16**, 21 % conversion of to Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp **4.14** was observed ([M-H]<sup>-</sup>: 616.2) via LC-MS however (Figure 4.19, D), CMP-Pse5Ac7Ac remained. This suggests that Im9-KpsS1 has selectivity for Glc moieties, over Gal moieties.

Structural characterisation of the CPS repeat unit **4.4** suggested that KpsS1 forms a 2,6 linkage between Pse5Ac7Ac and Glcp, therefore *p*NP- $\beta$ -D-Xylp **4.10** was screened due to a lack the of a sixth carbohydrate carbon atom.<sup>174</sup> The results of this screen provided further evidence in support of KpsS1 forming a 2,6 linkage, as turnover of **4.10** was not observed and CMP-Pse5Ac7Ac remained (summarised as no conversion observed, Table 4.7).

Finally, to examine the effect of the nitro group on the phenyl ring, *p*MeP- $\beta$ -D-Glcp **4.11**, was used as an acceptor, therefore the nitro group of *p*NP- $\beta$ -D-Glcp was substituted with a methyl group. A lack of CMP-Pse5Ac7Ac or Pse5Ac7Ac peaks and the presence of a peak corresponding to Pse5Ac7Ac-*p*MeP- $\beta$ -D-Galp molecular ions **4.15** ([M-H]<sup>-</sup>: 601.2) in LC-MS spectra indicated near-full conversion of CMP-Pse5Ac7Ac to glycoside product (Figure 4.19, E). No CMP-Pse5Ac7Ac remained, however, some Pse5Ac7Ac was present likely due to the inherent instability of CMP-Pse7Ac5Ac. This result was comparable to conversion with *p*NP- $\beta$ -D-Glcp as acceptor, suggesting that the presence of a substituted phenyl ring on the Glc-based acceptor aids turnover. Structural characterisation of KpsS1 or Im9-KpsS1 may provide further insight into why high levels of conversion are observed with this class of acceptor.

**Table 4.7: Acceptors screened for activity with Im9-KpsS1, CMP-Pse5Ac7Ac used as donor**

| Acceptor   | Conversion to Pse glycoside |
|--|-----------------------------|
|  <p>4.6</p>     | Not observed                |
|  <p>4.7</p>     | Not observed                |
|  <p>4.8</p>    | 100 %                       |
|  <p>4.9</p>   | 9 %                         |
|  <p>3.16</p> | 21 %                        |
|  <p>4.10</p> | Not observed                |
|  <p>4.11</p> | 99 %                        |

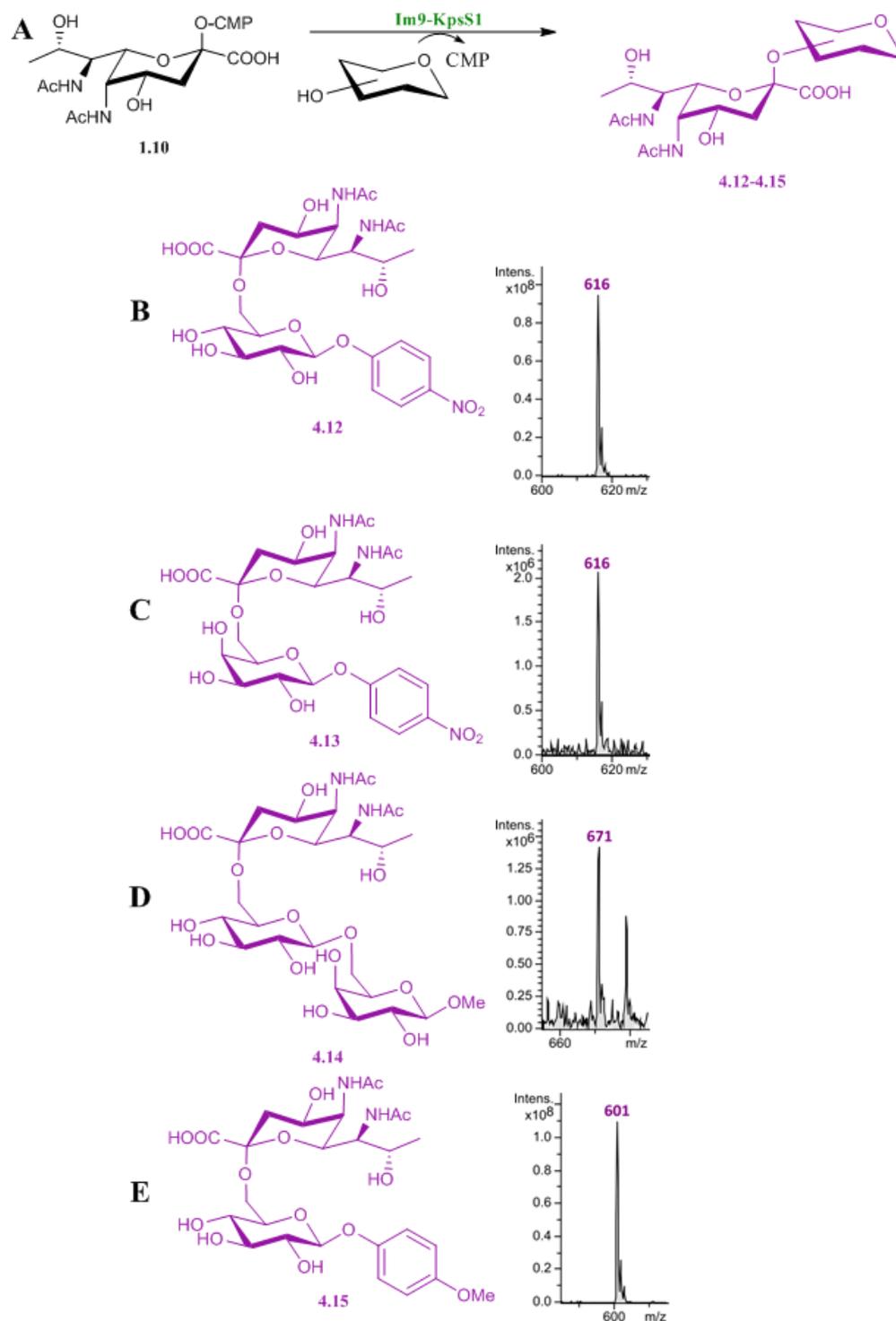
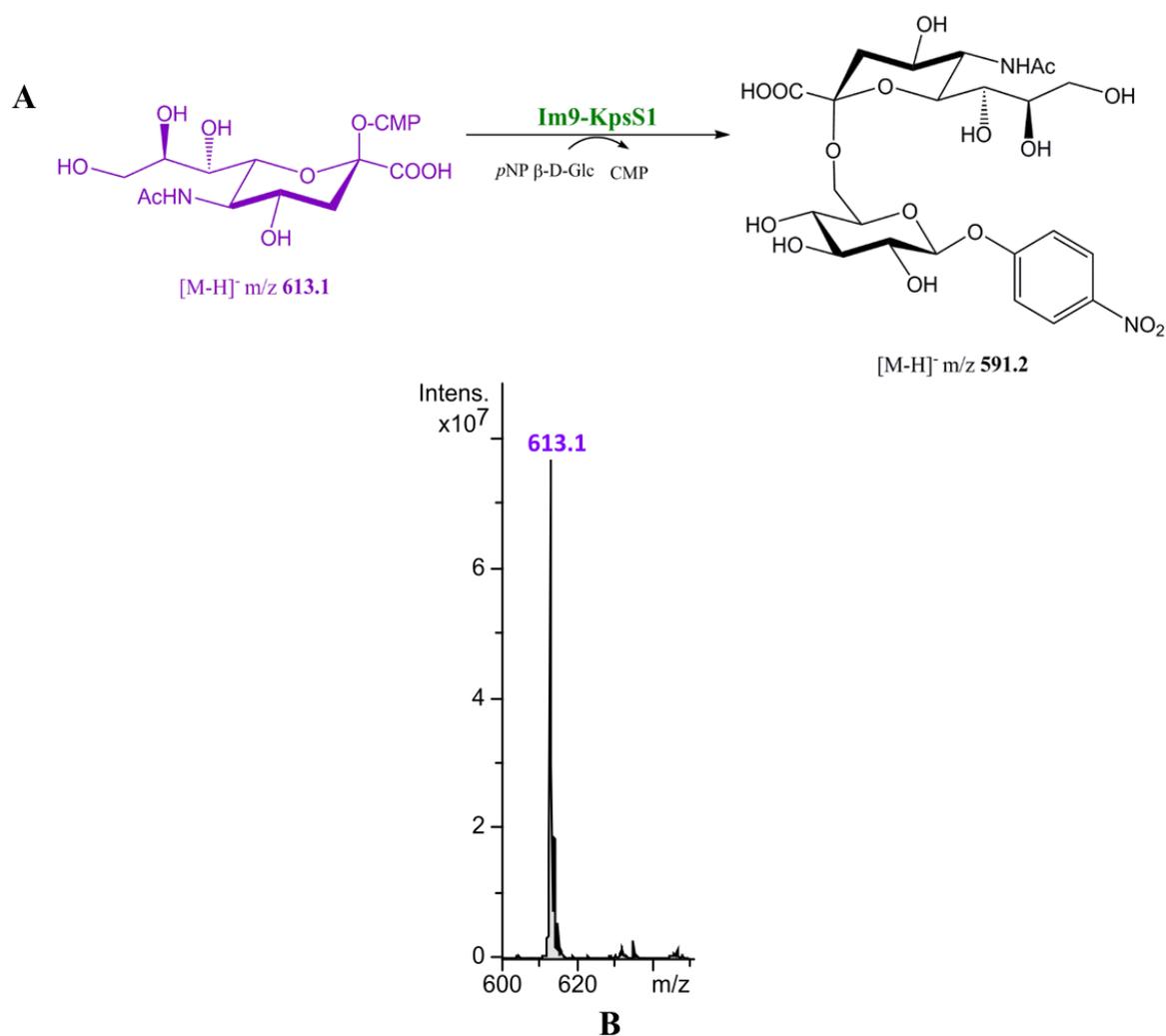


Figure 4.19: LC-MS data for positive results of Im9-KpsS1 acceptor screen. A: General reaction scheme, B:  $[M-H]^-$  peak for  $\beta$ -2,6-Pse5Ac7Ac-*p*NP- $\beta$ -D-Glcp 4.12 ( $m/z$  616), C:  $[M-H]^-$  peak for  $\beta$ -2,6-Pse5Ac7Ac- $\beta$ -D-Glcp-1,6- $\beta$ -D-GalpOMe 4.13 ( $m/z$  671), D:  $\beta$ -2,6-Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp 4.14 ( $m/z$  616), E:  $\beta$ -2,6-Pse5Ac7Ac-*p*OMeP- $\beta$ -D-Glcp 4.15 ( $m/z$  601).

In addition to screening acceptors, the selectivity of KpsS1 towards an alternative CMP-NuLO was studied. CMP-Neu5Ac **1.5** was used in place of CMP-Pse5Ac7Ac in the Im9-KpsS1 activity screen. In these reactions *p*NP- $\beta$ -D-Glcp was used as acceptor, due to the positive results obtained in the test with CMP-Pse5Ac7Ac. LC-MS analysis showed that CMP-Neu5Ac remained in the sample (Figure 4.20, [M-H]<sup>-</sup>: 613.1) and no peak corresponding to Neu5Ac-*p*NP- $\beta$ -D-Glcp **4.16** was present. Therefore, under the conditions tested, no turnover of CMP-Neu5Ac **1.5** by Im9-KpsS1 was observed. This lack of activity with CMP-Neu5Ac was further validated by a reaction that contained CMP-Pse5Ac7Ac as an internal positive control. In these samples Pse5Ac7Ac-*p*NP- $\beta$ -D-Glcp molecular ion peaks were present, Neu5Ac-*p*NP- $\beta$ -D-Glcp molecular ion peaks were absent and CMP-Neu5Ac remained (Appendix- Figure A.8).



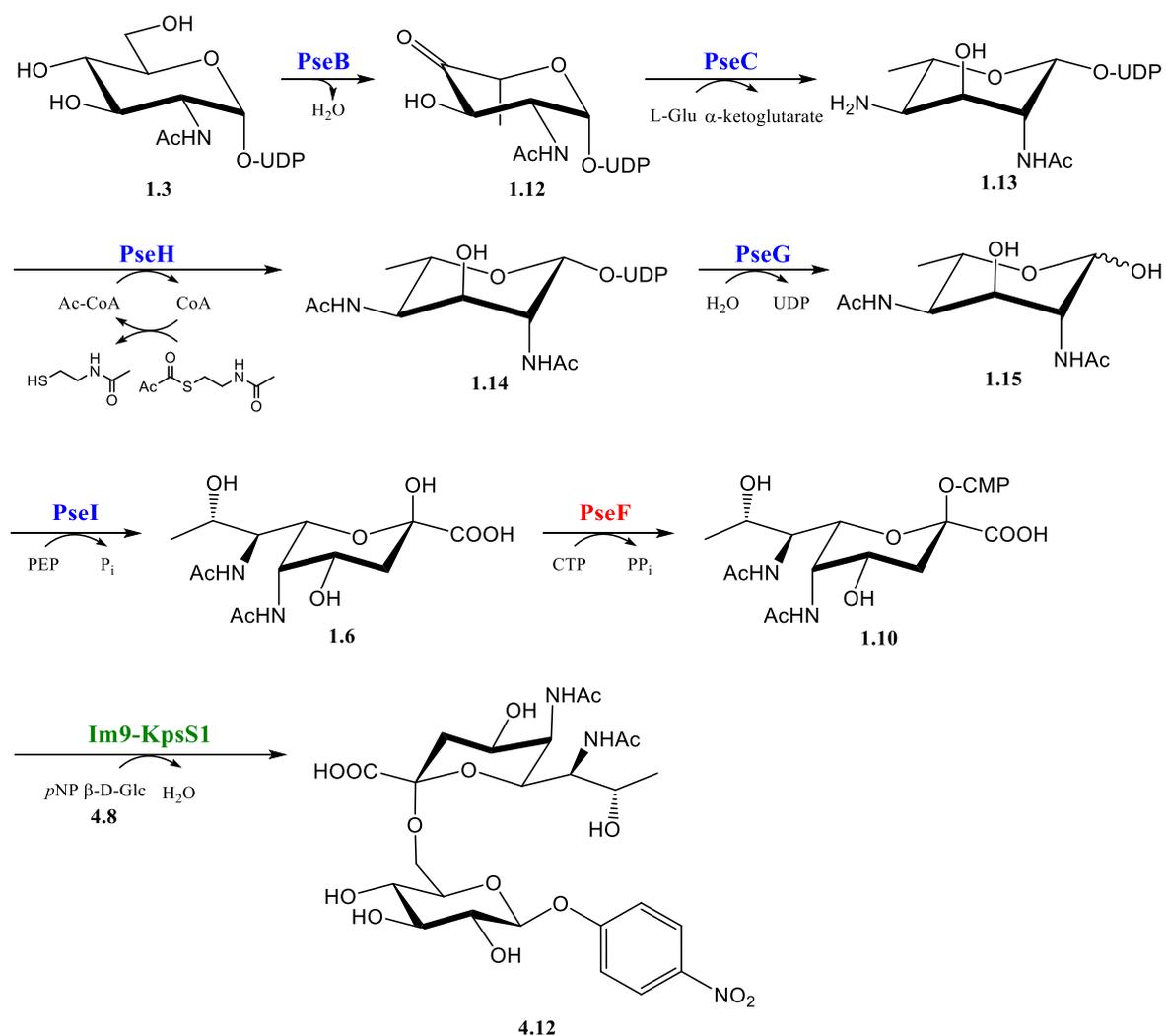
**Figure 4.20: Im9-KpsS1 activity screen with CMP-Neu5Ac.** A: Scheme of reaction tested between CMP-Neu5Ac and pNP-Glcp with Im9-KpsS1. B: LC-MS analysis shows CMP-Neu5Ac molecular ion peak, no peak corresponding to Neu5Ac-2,6-pNP- $\beta$ -D-Glcp 4.16 product was observed.

#### 4.3.12 Chemoenzymatic synthesis of glycoside containing Pse5Ac7Ac, via a one-pot, seven-enzyme reaction using Im9-KpsS1

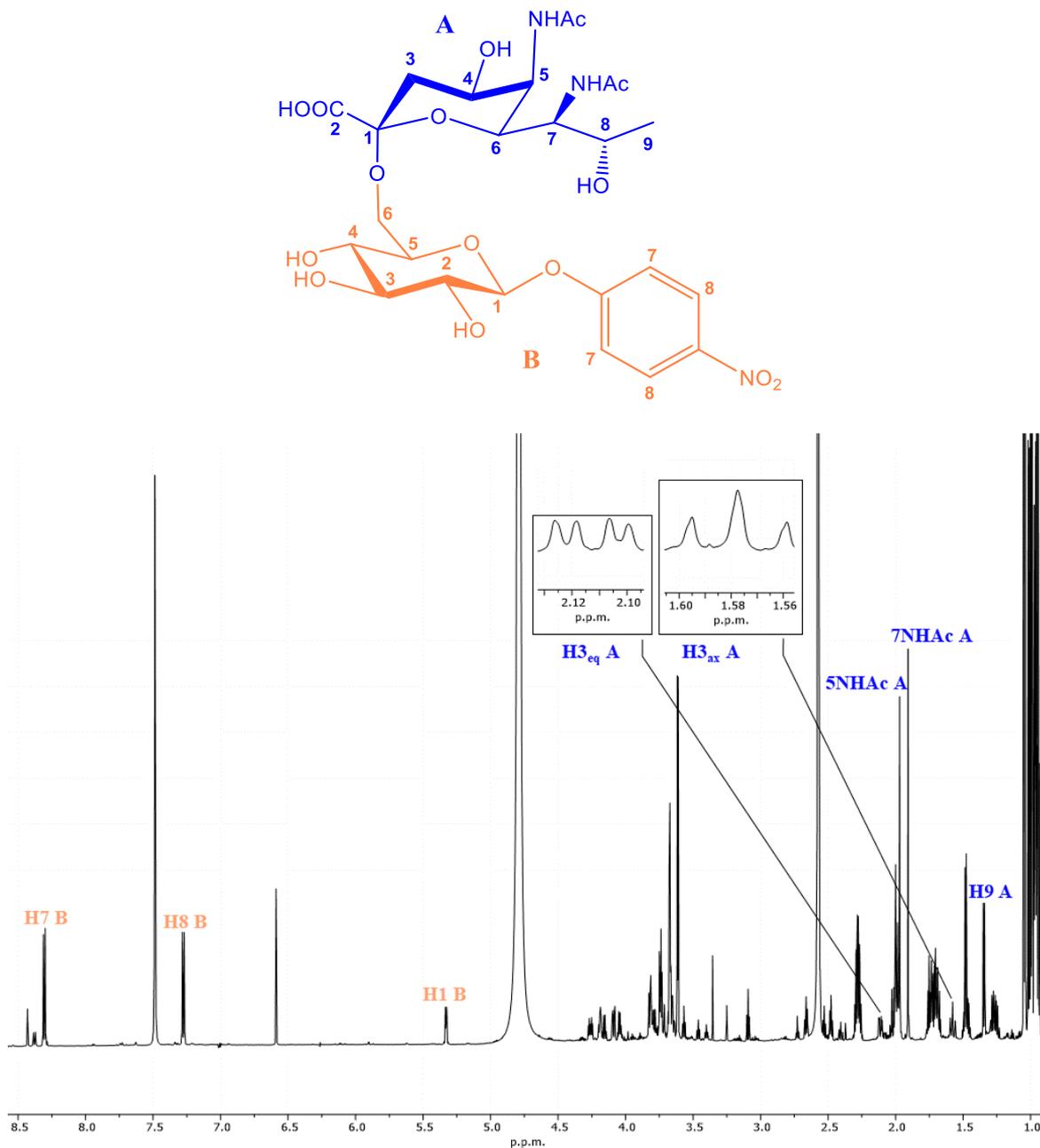
To validate the results of the initial activity screen and characterise the stereochemistry of the glycosidic bond formed by Im9-KpsS1, the reaction of CMP-Pse5Ac7Ac and pNP- $\beta$ -D-Glcp, catalysed by Im9-KpsS1 was performed on a larger scale (Scheme 4.1).  $\alpha$ -Pse5Ac7Ac-2,6-pNP- $\beta$ -D-Glcp **4.12** was produced in a seven enzyme one-pot reaction, starting with UDP-GlcpNAc **1.3**, in combination with the five *C. jejuni* enzymes (PseB, C, H, G and I) and reagents required to produce Pse5Ac7Ac. Once LC-MS analysis had confirmed that

Pse5Ac7Ac was present in the reaction mixture PseF, Im9-KpsS1 and the required reagents for the final enzymatic steps were added and were used to convert Pse5Ac7Ac to Pse5Ac7Ac-*p*NP- $\beta$ -D-Glcp, via CMP-Pse5Ac7Ac. In this reaction cleared cell lysate containing Im9-KpsS1 was used as it had been previously demonstrated that purification was low yielding and the Im9-KpsS1 obtained was not completely pure. This reaction was monitored via LC-MS and did not show full conversion to product, however over time enzymes precipitated therefore they were removed, and the reaction mixture was lyophilised (LC-MS data in Appendix- Figure A.9).

The product of the reaction was partially purified via gel-filtration and subsequent flash chromatography, using a process which has proved successful for the purification of  $\alpha$ -Pse5Ac7Ac-2,3-*p*NP- $\beta$ -D-Galp (results described in Chapter 3.3.3).<sup>106,128</sup> NMR analysis confirmed that crude  $\alpha$ -Pse5Ac7Ac-2,6-*p*NP- $\beta$ -D-Glcp had been produced. Key peaks have been highlighted in Figure 4.21, demonstrating the presence of the Pse5Ac7Ac, Glcp and *p*NP moieties of the product. The stereochemistry of Pse5Ac7Ac was assigned as  $\alpha$ , based on the 0.54 p.p.m. difference between  $\delta$  values for H-3<sub>ax</sub> and H-3<sub>eq</sub> peaks in the <sup>1</sup>H NMR spectrum (Figure 4.21) (reported as ~0.6 p.p.m. for the equatorial carboxy group and ~0.9 p.p.m for axial carboxy group (as shown in Chapter 3.3.3)).<sup>106,128</sup> Therefore, NMR characterisation of  $\alpha$ -Pse5Ac7Ac-2,6-*p*NP- $\beta$ -D-Glcp provides unequivocal evidence that *A. baumannii* ACUCI KpsS1 functions as a retaining PseT.



**Scheme 4.1: Chemoenzymatic synthesis of  $\alpha$ -Pse5Ac7Ac-2,6-pNP- $\beta$ -D-Glcp 4.12. Enzymes in blue (PseB, PseC, PseH, PseG and PseI) are from *Campylobacter jejuni*. PseF is from *A. caviae*. Im9-KpsS1 contains KpsS1 from *A. baumannii*.**



**Figure 4.21:** 700 MHz  $^1\text{H}$  NMR Spectrum of  $\alpha$ -Pse5Ac7Ac-2,6-*p*NP- $\beta$ -D-Glcp 4.12, in  $\text{D}_2\text{O}$ .  $^1\text{H}$  peaks for H9, H3<sub>ax</sub>, H3<sub>eq</sub>, 5NHAc, 7NHAc shown (blue labels). *p*NP- $\beta$ -D-Glcp H1, H7 and H8 shown (orange labels).

## 4.4 Discussion

The overall aim of this chapter was to characterize putative PseTs, which to our knowledge had previously not been studied *in vitro*. The successful solubilisation and purification of KpsS1, through the construction of a fusion protein, facilitated functional characterisation of this protein confirming its function as a retaining PseT. Results of the activity screen, where no turnover of Glcp acceptor and low turnover of  $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe were observed, suggest that KpsS1 likely acts as the final GT in the assembly of *A. baumannii* ACICU CPS tetrasaccharide repeat unit synthesis using a trisaccharide acceptor. This idea is supported by the order of putative GTs in the KL6 genetic island, where *kpsS1* is the fourth GT and additionally supported by CPS structural data which shows that Pse5Ac7Ac is a terminal carbohydrate.<sup>174</sup> As predicted by examining the CPS structure, Im9-KpsS1 showed enhanced selectivity towards Glcp in pNP- $\beta$ -D-glycosides, over Galp. Additionally, the lack of turnover of with CMP-Neu5Ac provides insight into the enzyme's selectivity for its glycosyl donor.

The results of this chapter highlight some of the many challenges faced when studying GTs. Given the abundance of genomic sequence data, genes may be putatively assigned GT functions based upon sequence homology with the limited empirically characterised examples, which are curated in the CAZY database (cazy.org). Structural characterisation of glycoconjugates can aid in putative assignment of GT function, as demonstrated for the proteins of interest in this chapter.<sup>144,159,168,174</sup> For the target putative PseTs this study, there is a distinct lack of functionally and structurally characterised homologues. As GTs are a relatively under-characterised class of proteins this is a reoccurring issue.<sup>24</sup> Structural data has been obtained for one putative PseT, *M. magenticum* Maf. However, this enzyme is predicted to pseudaminylate a protein acceptor, as opposed to the carbohydrate acceptors that are used by the target PseTs in this study.<sup>143,144,159,168,174</sup> This may account for the low sequence identity (18-22%) between Maf and the PseTs studied in this work. Therefore, whilst the Maf structure may be useful for modeling other putative PseTs, which act on protein substrates, it has not been used to model Lst, Gtr16, WfdL, KpsS1 or Orf14.

After identifying target proteins, the role of GTs in glycosylation cell surface glycoconjugates means they are frequently membrane bound or membrane associated

proteins and are therefore often insoluble when recombinantly expressed.<sup>30,32,191</sup> Sequence analysis of the five putative PseTs in this study suggested that the proteins are not membrane bound and as such there was not an obvious region of the proteins to truncate to drive solubility (with the exception of Lst which had nine N terminal residues truncated due to sequence similarity to membrane bound NST). However, their function in oligosaccharide repeat unit assembly, prior to export to the cell surface would suggest that these enzymes might be membrane associated.<sup>104,144,159,168,174</sup> The initial insolubility of all five proteins studied in this chapter would support this notion. Whilst there are a number of methods that can be used to solubilise membrane associated proteins, finding the optimal strategy is often a matter of trial and error.

A summary of methods trialled to solubilise the five putative PseTs can be found in Table 4.8. Results showed that the addition of NaCl, urea or Guanidine HCl to lysis buffer was not an effective strategy to solubilise Lst. Detergents were then screened for their ability to solubilise KpsS1 and Lst, with limited success. Detergents were not used further as they are not compatible with LC-MS which was used in analysis of Im9-KpsS1 reactions.<sup>188</sup> However detergents may be beneficial for further biochemical characterisation of putative PseTs and have frequently been used in studies of nonulosonic acid GTs.<sup>32,143</sup> For the Lst homolog NST, Triton X-100 was required to maintain activity, increase solubility during purification and is critical to the formation of ordered NST crystals.<sup>32</sup>

**Table 4.8: Summary of approaches taken to solubilise Putative PseTs. Black tick: technique did not yield soluble protein that was used in activity screen. Green tick: soluble protein obtained. Dash: technique was not used.**

| Putative PseT | BL21 (DE3) Expression Trial | Tuner Expression Trial | NaCl, Guanidine HCl, Urea | Detergent Screen | Fusion Protein |
|---------------|-----------------------------|------------------------|---------------------------|------------------|----------------|
| WfdL          | ✓                           | -                      | -                         | -                | -              |
| Gtr16         | ✓                           | -                      | -                         | -                | -              |
| Lst           | ✓                           | ✓                      | ✓                         | ✓                | -              |
| KpsS1         | ✓                           | ✓                      | -                         | ✓                | ✓              |
| Orf14         | ✓                           | -                      | -                         | -                | -              |

There was precedent in the literature for using fusion protein constructs to solubilise KpsS1, as an MBP-fusion had been used to solubilise many nonulosonic acid GTs, including the KpsS1 homolog *E. coli* KpsS.<sup>30,32,192</sup> Im9-KpsS1 fusion protein was produced and whilst

a large proportion of Im9-KpsS1 was found within the pellet of lysed cells, a sufficient amount of soluble protein was produced to facilitate functional characterisation. As purification of Im9-KpsS1 was low yielding, and a large amount of protein remained insoluble, further solubilisation of KpsS1 would be beneficial to future characterisation of this PseT. The use of detergents with Im9-KpsS1 may achieve this. Alternatively, there are a number of additional fusion partners that may be trialled, specifically the construction of MBP-KpsS1 should be reconsidered.<sup>189</sup>

As it was confirmed that Im9-KpsS1 had been produced, no further efforts were made to solubilise the other putative PseTs. For these remaining target proteins, fusion protein constructs may be a viable approach to accessing soluble enzymes. Alternatively, truncations may be made to enhance solubility. If present, transmembrane regions of the putative PseTs would have been the first residues to remove. It has been shown for initially insoluble GTs that truncations can be made to facilitate the production of soluble protein. For example the sequence of *Campylobacter jejuni* SiaTs Cst-I and Cst-II were aligned, to reveal that Cst-II lacked the C-terminal domain of Cst-I, therefore soluble, functional Cst-I was produced which lacked the C-terminal domain.<sup>79</sup> Truncations of the putative PseTs should be made cautiously, to avoid loss of enzymatic function.

Obtaining soluble protein is one of several challenges faced in studying GTs, however the undetermined identity of the acceptor sugar for Lst highlights a further barrier in functional characterisation of GTs, where both donor and acceptor sugar identity must be elucidated.<sup>24</sup> Before any soluble target protein was produced, it was determined that Lst should not be studied further as the unknown identity of the acceptor was likely to hinder functional studies. For many GTs including the remaining putative PseTs studied in this work the donor and acceptor moieties can be predicted from empirical evidence, where the resultant glycoconjugate structure has been characterised.<sup>144,159,168,174</sup> As demonstrated in this work, this knowledge can direct the choice of appropriate substrates to screen during preliminary functional studies. However, if the desired substrates are not commercially available functional studies may be impeded.

To perform the desired activity screen of Im9-KpsS1 both the donor CMP-Pse5Ac7Ac and a potential acceptor mimic **4.9**, were synthesised in house. However, only 9% conversion was seen with Im9-KpsS1 and **4.9**, further characterisation of the enzyme is required to provide

an informed explanation of this result. LC-MS analysis showed that Pse5Ac7Ac was present in this sample; this may be as a result of CMP-Pse5Ac7Ac hydrolysis. Ideally to assess this, the reaction could be performed with PseF present to convert any hydrolysed Pse5Ac7Ac to CMP-Pse5Ac7Ac, driving the glycosylation reaction (as shown in Chapter 3.3.3 with PmST1 reaction). Reactions could be analysed at several time points to monitor Pse5Ac7Ac-glycoside formation and potential hydrolysis. In addition to altering incubation time, it is possible that other reaction conditions, such as pH and temperature could be optimised to increase turnover of **4.9**.

In contrast to the Im9-KpsS1 activity screen result with **4.9**, full conversion or near-full conversion to Pse5Ac7Ac-glycoside was seen when *pNP-β-D-Glcp* **4.8** and *pOMeP-β-D-Glcp* **4.11** were used as acceptors. If aromatic residues are present in the KpsS1 active site, a possible explanation for the high level of turnover observed with *pNP-β-D-Glcp* and *pOMeP-β-D-Glcp* is that the aromatic moieties of these substrates are  $\pi$  stacking with aromatic residues in the active site. Aromatic residues are common in carbohydrate-protein interactions, where tryptophan is the most abundant residue within 5 Å of  $\beta$ -D-Glcp moieties in structurally characterised carbohydrate binding proteins.<sup>193</sup> Ultimately, an explanation of the different levels of conversion of the substrates screened relies on structural data for KpsS1.

Without structural data or a close homolog, a definitive GT fold cannot be assigned for KpsS1. A characteristic “DXD” motif can be found in many GT-A members, where a metal-ion co-ordinates with the residues side chains and two phosphate oxygen atoms of nucleotide donor.<sup>194</sup> Residues 88-90 of KpsS1 are DYD and therefore could be a candidate DXD motif. GT-B enzymes are also often metal-ion dependent, but a motif for coordination is yet to be identified.<sup>143</sup> Whilst NuLO GTs are not all metal dependent, divalent metal ions have been shown to enhance activity for many NuLO GTs.<sup>79,145</sup> WbbB and the homologous KpsS were both functionally assayed with 10 mM MgCl<sub>2</sub>, therefore this was included in all functional assays of KpsS1.<sup>30,31</sup> Additionally, Mg<sup>2+</sup> was required in the one-pot reaction for PseF activity. However, the metal-ion dependence of KpsS1 should be examined, in future mechanistic studies.

The exact mechanism used to pseudaminylate the acceptor is perhaps the most intriguing feature of KpsS1 that is yet to be determined. Debate surrounds mechanism of retaining

GTs. Structural data of WbbB<sub>2-401</sub>, the KpsS1 distant homolog, suggests that this retaining  $\beta$ -Kdo transferase uses an S<sub>N</sub>2-like mechanism. This conclusion was made based on the role of histidine residue from the HP motif, which coordinates the cytosine ring.<sup>30-32</sup> In WbbB this histidine is predicted to protonate the CMP leaving group.<sup>31</sup> This enzymatic acidic proton would not be required in the alternative S<sub>N</sub>i mechanism as it would be donated by the acceptor.<sup>37,48</sup> HP motifs are conserved in both inverting and retaining nonulosonic acid transferases, therefore the presence of these residues alone cannot be used to speculate whether KpsS1 uses an S<sub>N</sub>i or S<sub>N</sub>2-like mechanism.<sup>31,38</sup> Elucidation of the full KpsS1 mechanism likely requires protein structural data, mutagenesis and Pse-based probes.

## 4.5 Conclusions and Future Work

The work performed in this chapter has unequivocally confirmed that *A. baumannii* ACICU KpsS1 is a retaining PseT. To my knowledge, this is the first example of a PseT that has been shown to be active *in vitro* and one of a few examples of a retaining nonulosonic acid transferase. Additionally, as a result of the functional studies on Im9-KpsS1 the novel compound  $\alpha$ -Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp has been synthesised via a chemoenzymatic route.

Future work on KpsS1 should focus on further biochemical characterisation. The first step towards this would ideally require cleavage of a fusion tag. Structural studies of KpsS1 should be pursued, given that this enzyme will be the first in its family and that the mechanism of retaining GTs currently remains elusive. With access to large quantities of Pse5Ac7Ac and CMP-Pse5Ac7Ac within the Fascione laboratory, Pse-based probes could be designed for co-crystallisation studies, to help in determining the mechanism of this enzyme. Additionally, in the absence of structural data, the function of the HP motif could begin to be examined by mutagenesis.

For the other putative PseTs studied in this chapter, which are yet to be solubilised, fusion protein constructs could be produced with the aim of improving protein solubility. Soluble protein could then be studied in activity assays, using the general protocol followed for Im9-KpsS1. For the putative inverting PseT Gtr16, structural characterisation of the LPS O-antigen can be used to infer the identity of the putative acceptor and the predicted donor

CMP-Pse5Ac7Ac is available. Therefore, this may be the best of the remaining target proteins to focus future efforts on.

Work carried out within the Fascione laboratory on acetyl-CoA regeneration factors through the synthesis of N-acetyl-cysteamine (SNAc) provides a promising potential route to accessing CMP-Pse derivatives that vary at the C7 substituent. This methodology could be employed to synthesis N-hydroxybutanol-cysteamine (SNHb) for the chemoenzymatic synthesis of the predicted donor moiety for the putative PseTs WfdL, CMP-Pse5Ac7Hb. Additionally, if Orf14 were to be solubilised, functional studies may be performed with CMP-Pse5Ac7Fm **1.18**, previously synthesised in house. This CMP-Pse derivative is structurally closer to the predicted Orf14 donor CMP-Pse5Hb7Fm, than CMP-Pse5Ac7Ac, and so may be useful in studies of this enzyme.

Finally, CMP-Pse5Ac7Ac produced in this study was gifted to the Sulzenbacher laboratory to be used in crystal studies of *M. magneticum* Maf with the aim of obtaining a structure of Maf with CMP-Pse5Ac7Ac bound to the active site

## Chapter 5: Concluding Remarks and Future Prospectives

Non-eukaryotic, cell-surface carbohydrate pseudaminic acid (Pse5Ac7Ac) is known to contribute to the virulence of several multi-drug resistant bacterial pathogens.<sup>1</sup> Chemical synthesis of Pse5Ac7Ac and its derivatives has proved challenging.<sup>2-9</sup> Access to Pse5Ac7Ac and activated CMP-Pse5Ac7Ac has been a hindrance in studies into the biological significance of Pse5Ac7Ac, including Pse5Ac7Ac-processing enzymes, which may be novel therapeutic targets. This project aimed to characterise enzymes which process pseudaminic acid and to chemoenzymatically synthesise glycosides which contain pseudaminic acid.

This was achieved through three specific objectives. Firstly, nucleotide activated CMP-Pse5Ac7Ac **1.10**, was required. This was achieved through a chemoenzymatic synthesis described in Chapter 2. Access to CMP-Pse5Ac7Ac was crucial in facilitating further work of this study, described in Chapters 3 and 4, where it was the starting material for glycosyltransferase mediated reactions.

Six biosynthetic enzymes (*Campylobacter jejuni* PseB, PseC, PseH, PseG and PseI, plus *Aeromonas caviae* PseF) were recombinantly produced for the synthesis of **1.10**. This study has provided unequivocal evidence that *A. caviae* PseF functions as an  $\alpha$ -CMP-Pse5Ac7Ac synthetase. NMR characterisation confirmed the stereochemistry of this carbohydrate, revealing that  $\alpha$ -CMP-PseAc7Ac was produced, this data was crucial for subsequent conclusions of this study (Chapter 3 and 4).

Future work may revisit structural characterisation of PseF, where conditions used to crystallise the homologous NeuA may be screened.<sup>127</sup> Additionally, PseF mutants may be produced in order to probe the mechanism and substrate selectivity of this enzyme. The Pse5Ac7Ac and derivatives produced by methodology established within the Fascione laboratory, may be used in crystallisation of this enzyme.

Additionally, small-molecules were assayed as inhibitors of *C. jejuni* PseB and PseI plus *A. caviae* PseI, where some of the compounds tested showed a concentration dependent decrease in Pse5Ac7Ac **1.6** production. Future work could aim to obtain structural data for PseI to further aid inhibition studies and probe the substrate selectivity of this enzyme.

The work presented in this thesis, to my knowledge, provides the first examples of chemoenzymatically synthesised Pse5Ac7Ac containing glycosides, thus achieving the second objective of this thesis. As there are currently no published reports of *in vitro* activity assay of native pseudaminyltransferases, the use of well-characterised, highly stable, commercially available, promiscuous SiaTs, which utilise CMP-Pse5Ac7Ac as a donor, provides a promising route to synthesis of a range of novel and biologically relevant Pse-glycosides.

Crystal structure data of PmST with various ligands in the active site are available, including a structure with Lac acceptor and non-hydrolyzable donor substrate analogue, CMP-3-deoxy-3-fluoro-Neu5Ac (CMP-3F-Neu5Ac) both present in the activity site.<sup>38,164</sup> Future work may co-crystallise PmST with CMP-3F-Pse5Ac7Ac to gain insight into the molecular basis of donor substrate promiscuity that has been demonstrated in previous studies and expanded on in this work, with CMP-Pse5Ac7Ac.

Chapter 2 described a strategy developed in the Fascione laboratory for the chemoenzymatic synthesis of CMP-Pse5Ac7R derivatives. Future work may examine whether any of the SiaT which were shown to be active with CMP-Pse5Ac7Ac are able to utilise these CMP-Pse derivatives as donors, potentially expanding the range of Pse-glycosides which can be synthesised using the approach described in this study.

Finally, this project set out to characterise putative pseudaminyltransferases (PseTs) (Chapter 4). The work presented unequivocally confirmed that *A. baumannii* ACICU KpsS1 is a retaining PseT. To my knowledge, this is the first example of a PseT that has been shown to be active *in vitro* and one of a few examples of a retaining nonulosonic acid transferase. Additionally, as a result of the functional studies on Im9-KpsS1 the novel compound  $\alpha$ -Pse5Ac7Ac-*p*NP- $\beta$ -D-Galp was synthesised via a chemoenzymatic route, in a one-pot seven enzyme reaction.

With priority, future work on KpsS1 should focus on further biochemical characterisation. Considering that KpsS1 will be the first member of a novel GT family and that the mechanism of retaining GTs currently remains elusive, structural studies should be

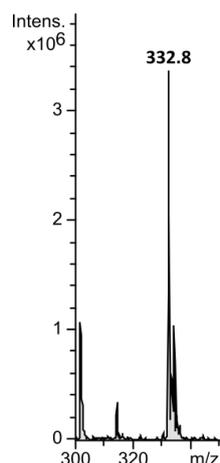
pursued. As large quantities of Pse5Ac7Ac and CMP-Pse5Ac7Ac may be produced using methodology described in Chapter 2, Pse-based probes could be designed for co-crystallisation studies, to help in determining the mechanism of this enzyme.

For the other putative PseTs studied in this project, which are yet to be solubilised, fusion protein constructs could be produced with the aim of improving protein solubility. Soluble protein could then be studied in activity assays, using the general protocol followed for Im9-KpsS1. For the putative inverting PseT Gtr16, structural characterisation of the LPS O-antigen can be used to infer the identity of the putative acceptor and the predicted donor CMP-Pse5Ac7Ac is available. Therefore, this may be the best of the remaining target proteins to focus future efforts on.

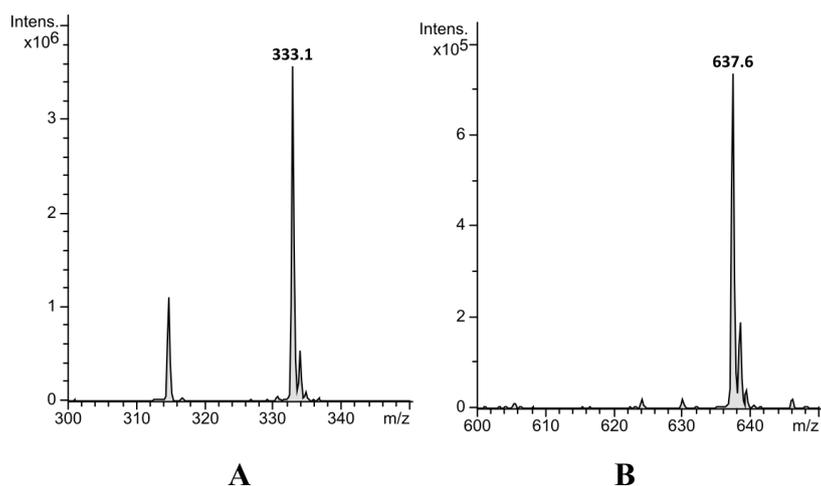
Work carried out within the Fascione laboratory on acetyl-CoA regeneration factors through the synthesis of N-acetyl-cysteamine (SNAc) provides a promising potential route to accessing CMP-Pse derivatives that vary at the C7 substituent. This methodology could be employed to synthesise N-hydroxybutanol-cysteamine (SNHb) for the chemoenzymatic synthesis of CMP-Pse5Ac7Hb, the predicted donor moiety for the putative PseT WfdL.

Finally, CMP-Pse5Ac7Ac produced in this study was gifted to the Sulzenbacher laboratory to be used in crystal studies of *M. magneticum* Maf with the aim of obtaining a structure of Maf with CMP-Pse5Ac7Ac bound to the active site.

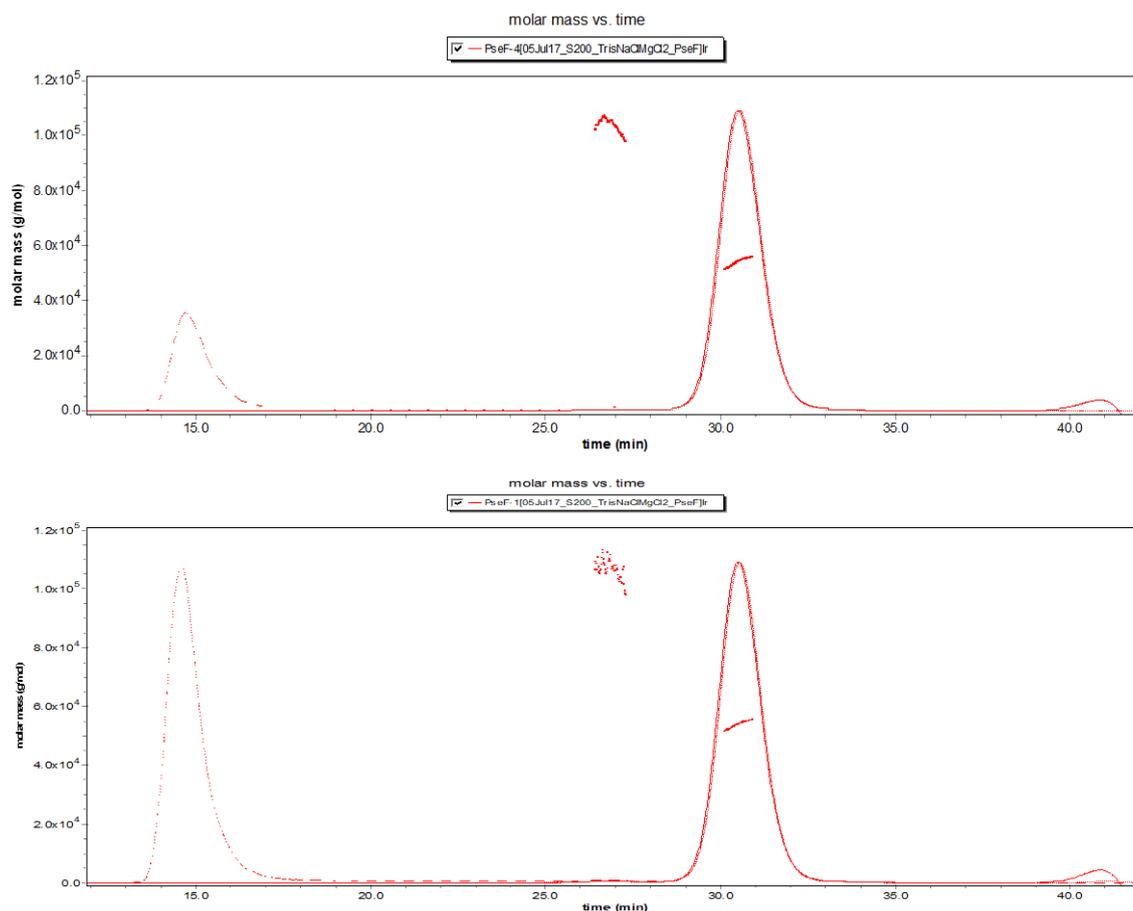
## Appendix



**Figure A.1:** MS data showing presence of Pse5Ac7Ac 1.6 [M-H]<sup>-</sup> peak when PseI and PEP were added to PseB-G reaction mixture to confirm that 6-deoxyAldiNac 1.15 had been produced in initial reaction



**Figure A.2:** -ESI LC-MS data for large scale synthesis of CMP-Pse5Ac7Ac described in 2.3.12. A: molecular ion peak for Pse5Ac7Ac 1.6 m/z 333.1 after enzymes PseB, PseC, PseH, PseG and PseI plus UDP-GlcNAc and co-factors. B: molecular ion peak for CMP-Pse5Ac7Ac 1.10 m/z 637.6 after PseF and CTP were added to the mixture in A to convert Pse5Ac7Ac into CMP-Pse5Ac7Ac.



**Figure A.3: SEC-MALLS data for PseF sample collected at 4 mg mL<sup>-1</sup> (top) and 1 mg mL<sup>-1</sup> (bottom). Large peak at around 30 minutes corresponds to a mass of ~54 kDa, consistent with dimer PseF.**

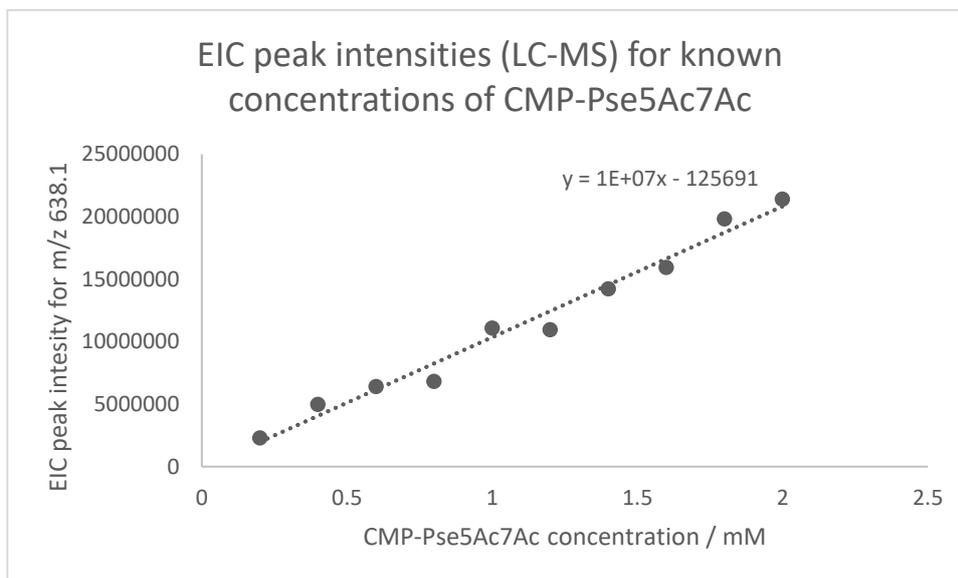
| Peak | from (min) | to (min) | MW (kDa) | μg |
|------|------------|----------|----------|----|
| 1    | 30.12      | 30.92    | 54.1     |    |
| 2    | 28.21      | 33.00    | 428      |    |
| 3    | 26.40      | 27.30    | 103.8    |    |
| 4    | 25.60      | 28.19    | 2.45     |    |

The total amount of protein seen is ~430 μg, quite consistent with the expected amount of 400 μg. The amount of "dimeric" material is less than 0.6% of the total.

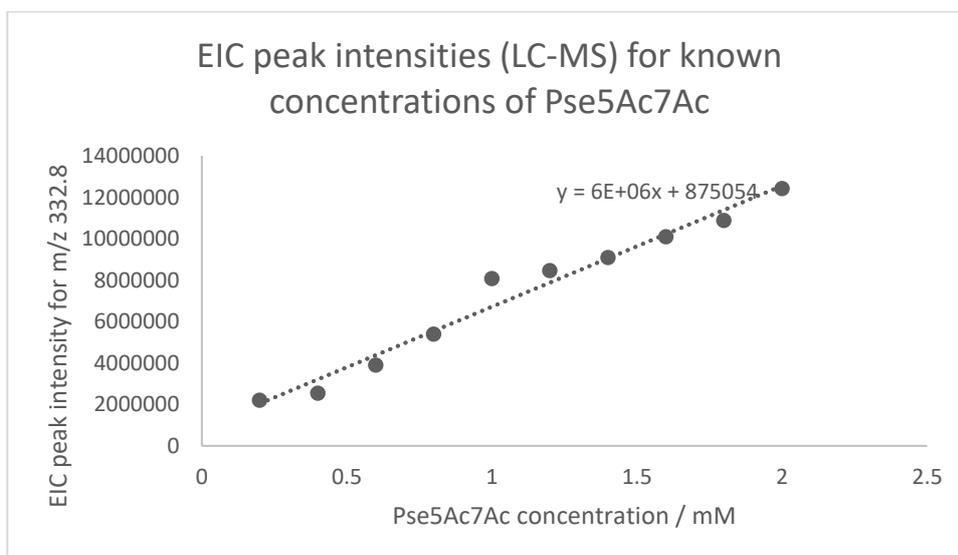
**Table A.1: 24 crystal well plate conditions for PseF crystallisation trials. Wells contained 10 mL hanging drops contained 1  $\mu$ L 10 mg mL<sup>-1</sup> PseF plus 1  $\mu$ L well liquor.**

|          | 1   | 2   | 3   | 4   | 5   | 6 |
|----------|---|---|---|---|---|---|
| <b>A</b> | 60 mM Bis Tris propane, 50 mM citric acid, pH 4.5, 12% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 4.5, 14% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 4.5, 16% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 4.5, 18% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 4.5, 20% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | / |
| <b>B</b> | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.0, 12% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.0, 14% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.0, 16% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.0, 18% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.0, 20% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | / |
| <b>C</b> | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 12% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 14% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 16% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 18% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 20% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | / |
| <b>D</b> | 60 mM Bis Tris propane, 50 mM citric acid, pH 6.0, 12% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 6.0, 14% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 6.0, 16% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 6.0, 18% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 60 mM Bis Tris propane, 50 mM citric acid, pH 5.5, 20% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | / |
|          | 1   | 2   | 3   | 4   | 5   | 6 |
| <b>A</b> | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 8% PEG 3350 (v/w)                    | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 10% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 12% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 14% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 16% PEG 3350 (v/w)                   | / |
| <b>B</b> | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 8% PEG 3350 (v/w), 0.3 % (v/v) C8E5  | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 10% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 12% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 14% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.0, 16% PEG 3350 (v/w), 0.3 % (v/v) C8E5 | / |
| <b>C</b> | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.7, 8% PEG 3350 (v/w)                    | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.7, 10% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.7, 12% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.7, 14% PEG 3350 (v/w)                   | 50 mM Bis Tris propane, 60 mM citric acid, pH 7.7, 16% PEG 3350 (v/w)                   | / |
| <b>D</b> | 50 mM Bis Tris propane, 60 mM citric acid,  | 50 mM Bis Tris propane, 60 mM citric acid,  | 50 mM Bis Tris propane, 60 mM citric acid,  | 50 mM Bis Tris propane, 60 mM citric acid,  | 50 mM Bis Tris propane, 60 mM citric acid,  | / |

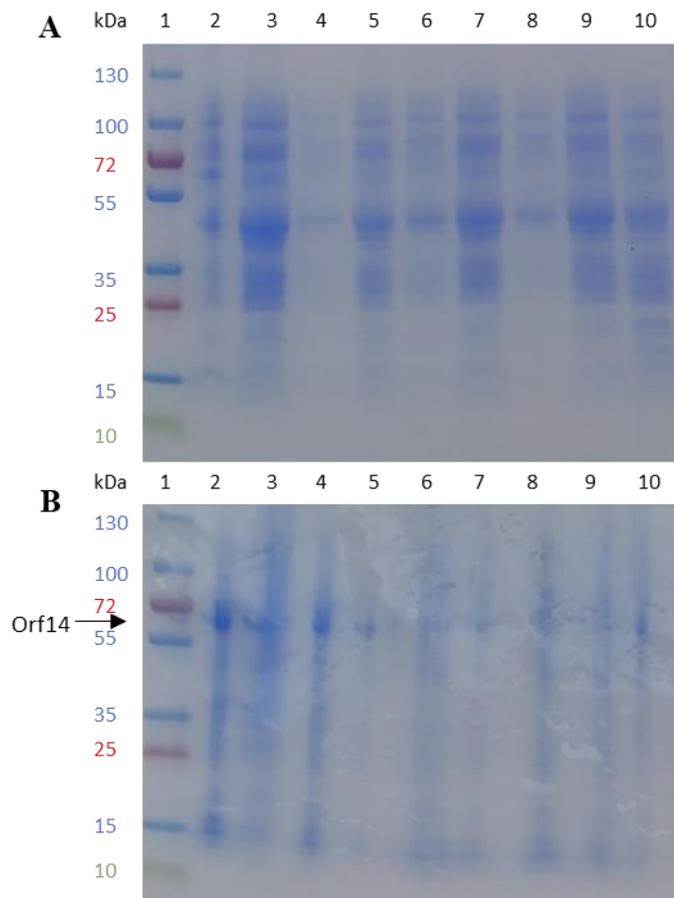
|                 |                 |                 |                 |                 |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| pH 7.7, 8% PEG  | pH 7.7, 10% PEG | pH 7.7, 12% PEG | pH 7.7, 14% PEG | pH 7.7, 16% PEG |
| 3350 (v/w), 0.3 |
| % (v/v) C8E5    |



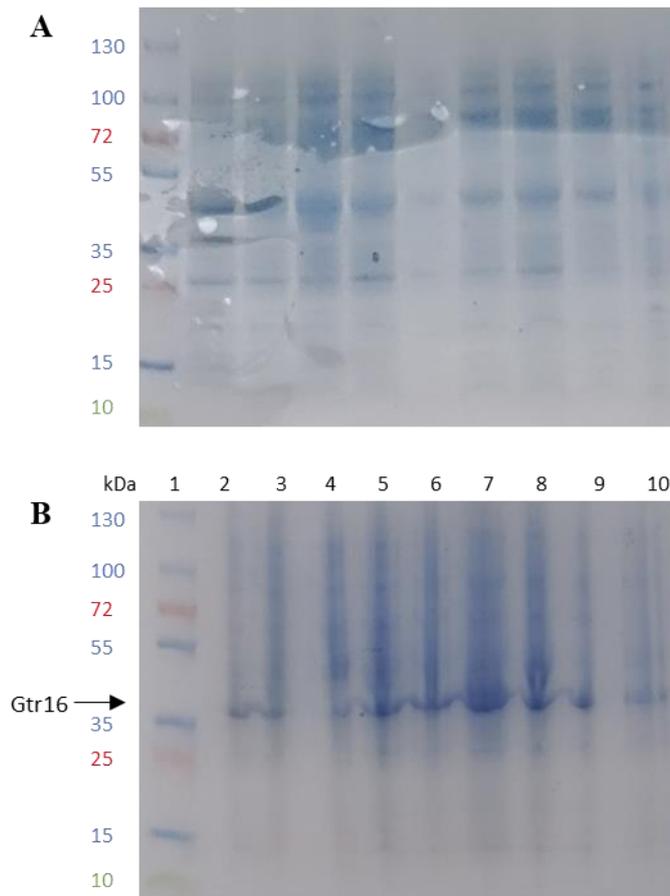
**Figure A.4: Extracted ion count peak intensities from LC-MS chromatograms of known concentrations of CMP-Pse5Ac. Data used in combination with EIC peak intensities of Pse5Ac7Ac to calculate conversions to products for SiaT reactions**



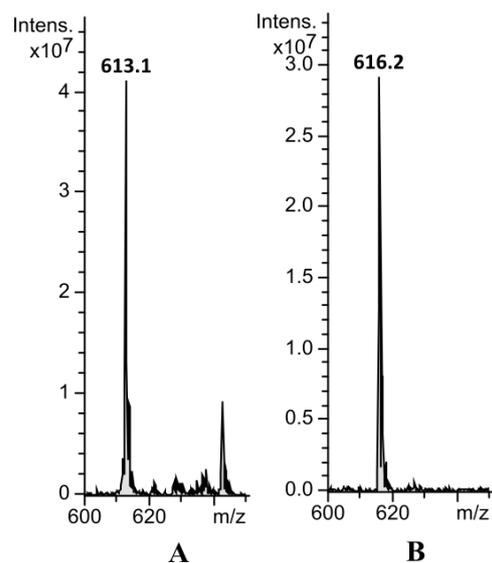
**Figure A.5: Extracted ion count peak intensities from LC-MS chromatograms of known concentrations of CMP-Pse5Ac. Data used in combination with EIC peak intensities of Pse5Ac7Ac to calculate conversions to products for SiaT reactions.**



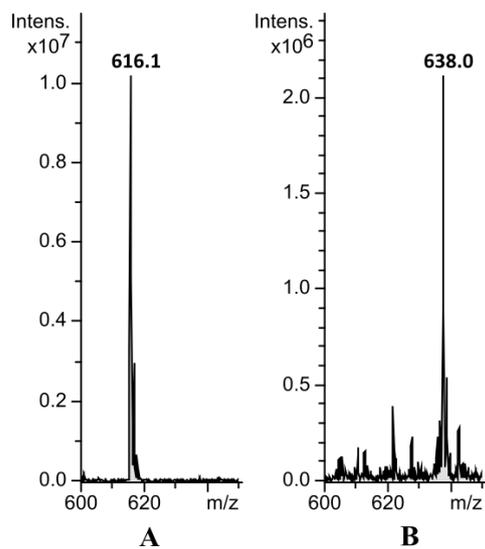
**Figure A.6: SDS-PAGE of expression trials for Gtr16 (MW: 59.5 kDa) in *E. coli* BL21 (DE3) cultured at 30 °C. A: Cells lysed in Bugbuster mix, soluble fractions. B: Cells lysed in Bugbuster mix, insoluble fractions. Lanes loaded as follows: 1-PAGE-RULER MW marker 2- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 4 h; 3- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 24 h; 4- OD<sub>600</sub>: 0.4, 1 mM IPTG, 4 h; 5- OD<sub>600</sub>: 0.4, 1 mM IPTG, 24 h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 7- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 8- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h Uninduced sample; 9- : 0.6, 1 mM IPTG, 24 h 10-Uninduced sample.**



**Figure A.7: SDS-PAGE of expression trials for Lst (MW: 40.3 kDa) in *E. coli* BL21 (DE3) cultured at 30 °C. A: Cells lysed in Bugbuster mix, soluble fractions. B: Cells lysed in Bugbuster mix, insoluble fractions. Lanes loaded as follows: 1-PAGE-RULER MW marker 2- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 4 h; 3- OD<sub>600</sub>: 0.4, 0.1 mM IPTG, 24 h; 4- OD<sub>600</sub>: 0.4, 1 mM IPTG, 4 h; 5- OD<sub>600</sub>: 0.4, 1 mM IPTG, 24 h; 6- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h; 7- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 24 h; 8- OD<sub>600</sub>: 0.6, 0.1 mM IPTG, 4 h Uninduced sample; 9- : 0.6, 1 mM IPTG, 24 h 10-Uninduced sample.**



**Figure A.8: LC-MS data for control reaction for activity screen of Im9-KpsS1 with CMP-Pse5Ac7Ac and CMP-Neu5Ac as potential donors with  $p$ NP- $\beta$ -D-Glc as acceptor. A: Molecular ion peak CMP-Neu5Ac (m/z 613.1). B: Molecular ion peak  $\alpha$ -2,6-Pse5Ac7Ac- $p$ NP- $\beta$ -D-Glc (m/z 616.2).**



**Figure A.9: LC-MS data for large scale synthesis of  $\alpha$ -2,6-Pse5Ac7Ac- $p$ NP- $\beta$ -D-Glc (m/z 616.2). Remaining CMP-Pse5Ac7Ac shown in A,  $\alpha$ -2,6-Pse5Ac7Ac- $p$ NP- $\beta$ -D-Glc molecular ion peak shown in B.**

## **Plasmid information, gene sequences and recombinant protein sequences**

Genes for putative pseudaminyltransferases, *KpsS1*, *Gtr16*, *WfdL* and *Orf14* were inserted into pET15b vectors between BamHI and NdeI restriction sites. *Lst* gene was in pET15b vector (RE sites? Truncated). AcPseF plasmid was kindly gifted by Jonathon Shaw.

### **PseB**

MGSSHHHHHSSGLVPRGSHMFNGKNILITGGTGSFGKTYTKVLENYKPNKIIYSRDELKQFEMSSI  
FNSNCMRYFIGDVRDKERLSVAMRDVDFVIHAAAMKHVPVAEYNPMECIKTNIHGAQNVIDACFEN  
GVKKCIALSTDKACNPVNLYGATKLASDKLFVAANNIAGNKQTRFSVTRYGNVVGSRGVSVPFFKKLIA  
QGSKELPITDRMTRFWISLEDGVKFLSNFERMHGGEIFIPKIPSMKITNLAHALAPNLSHKIIGIRAGE  
KLHEIMISSDSDLTYEFENYYAISPSIKLVDQESDFSINALGEKGQKVKDGFSSYSSDNNPQWASEKELL  
DIINHTEGF

### **PseC**

MGSSHHHHHSSGLVPRGSHMLTYSHQNIDQSDIDTLTKALKDEILTGGKKVNEFEEALCEYMGVKH  
ACVLNSATSALHLAYTALGVQEKIVLTTPLTFAATANAALMAGAKVEFIDIKNDGNIDEKKLEARLLKES  
ENIGAISVDFAGNSVEMDEISNLTKKYNIPLIDDASHALGALYKSEKVGKKADLSIFSHPVKPITTFEG  
GAVVSDNEELIDKIKLLRSHGIVKKRLWDSMVELGYNRLSDVACALGINQLKLDHNLEKREEIANF  
YDKEFEKNPYFSTIKIKDYKSSRHLYPILLPEFYCQKEELFESLLHAGIGVQVHYKPTYEFSFYKLLGEIK  
LQNADNFYKAELSIPCHQEMNLKDAKFKDTLFSILEKVKKGYCG

### **PseH**

ATGCTGATTAAGTGAAGAACTTCGCGGAAGTGAATAGCCAGGAAATTAAGTATCTTCAAATG  
GCGTAACCAACCGGACATTAGCCAATTCATGAAGACCAACACATCGACTTCGAGGAACACCTGC  
GTTTTATTTCGTAACCTGCACCAGGATAGCAACAAGAAATACTTCCTGGTGTTTCAGGACGAGCAA  
ATCATTGGTGTGATCGATTTTCGTTAACATTACCACCAAAGCTGCGAATTTGGCCTGTATGCGATC  
CCGGACCTGAAGGGTGTGGGCCAAGTTCTGATGAACGAGATCAAGAAATACGCGTTTCGAAATTC  
TGAAGGTGGACACCCTGAAAGCGTATGTTTTAAGGATAACCACAAGGCGCTGAAACTGTACCA  
GCAAAACCACTTTACCATTTATGATGAGGACAAGGACTTTTATTATGTGTGCCTGAAACAGAGCC  
ACTGCAAGGCGCTGCCGAGCTAA

MGSSHHHHHSSGLVPRGSHMIKLNFTELNSQEIELIFKWRNHPDINQFMKTKYIDFEEHLRFLKLL  
HQDSSKKYFLVFQDEQIIGVIDFVNITTKSCEFLYAKPNLKGVGQILMNEIIKYAFENLKVNTLKAYVFK  
DNRKALKLYQQNHFTIYDEDKDFYHICLKQSDCKALPS

### **PseG**

ATGAAGGTGCTGTTCCGTAGCGACAGCAGCAGCCAGATCGGTTTTGGCCACATTAAGCGTGACCT  
GGTGCTGGCGAAACAATACAGCGATGTTAGCTTTGCGTGCCTGCCGCTGGAGGGTAGCCTGATC  
GACGAAATTCGTACCCGTTTATGAACTGAGCAGCGAGAGCATCTACGAACTGATCAACCTGAT  
TAAAGAGGAGAAATTCGAGCTGCTGATCATTGATCACTATGGTATCAGCGTGGACGATGAGAAG  
CTGATTAAGTGAAGACCGGCGTTAAGATCCTGAGCTTTGACGATGAAATTAACCGCACCCTG  
CGACATCCTGCTGAACGTGAACGCGTACGCGAAGGCGAGCGATTATGAGGGTCTGGTGCCGTTT

AAATGCGAAGTTCGTTGCGGCTTTAGCTACGCGCTGATCCGTGAGGAATTCTATCAGGAAGCGAA  
GGAAAACCGTGAGAAGAAATACGACTTCTTTATTTGCATGGGTGGCACCGATATCAAAAACCTGA  
GCCTGCAGATTGCGAGCGAGCTGCCGAAGACCAAAATCATTAGCATCGCGACCAGCAGCAGCAA  
CCCGAACCTGAAGAACTGCAAAAGTTCGCGAACTGCACAACAACATCCGTCTGTTTATTGATC  
ACGAGAACATTGCGAAGCTGATGAACGAAAGCAACAACTGATCATTAGCGCGAGCAGCCTGGT  
GAACGAGGCGCTGCTGCTGAAGGCGAACTTTAAAGCGATCTGCTACGTTAAGAACCAAGAAAGC  
ACCGCGACCTGGCTGGCGAAGAAAGGCTATGAGGTTGAATACAAATATTA

MGSSHHHHHSSGLVPRGSHMLDKILYFKTLIRADSGSKIGHGHVRRDLILAKNFKDVSFACIDLPGSL  
TGEIPCPVFTLKSADINELVNLIKEHKFELLIIDHYGISAADekliEQTNVKILCFDDNYKEHFCDYLLNVN  
IYAQPQKYVNLVPANCELVFSPLVRSEFYDEAKIKREKKFDCFIALGGTDALNLTAKIASNLLAKNKKVA  
AITTSANANLANLQNLADSESNFSLFINSNEVARLMNESEILVISASSLVNEALVLGAKFKAVRVADNQ  
NEMAQWLAANGREIYEADEICLNL

### PseI

ATGCATATGCAAATTGGTAACTTTAACACCGACAAGAAGGTTTTTATCATTGCGGAACTGAGCGC  
GAATCATGCGGGTAGCCTGGAGATGGCGCTGAAGAGCATCAAAGCGGCGAAGAAAGCGGGTGC  
GGACGCGATCAAGATTCAGACCTACACCCCGGATAGCCTGACCCTGAACAGCGACAAAGAGGAC  
TTCATCATTAAGGTGGCCTGTGGGACAAGCGTAACTGTACGAACTGTATGAGAGCGCGAAAA  
CCCCGTATGAATGGCACAGCCAGATCTTCGAAACCGCGCAAAACGAGGGTATTCTGTGCTTCAGC  
AGCCCGTTTGCGAAGGAAGACGTGGAGTTCCTGAAACGTTTTGATCCGATCGCGTACAAGATTGC  
GAGCTTCGAAGCGAACGATGAGAACTTTGTGCGTCTGATTGCGAAAGAGAAGAAACCGACCATC  
GTTAGCACCGCATTGCGACCGAGGAAGAGCTGTTCAAGATCTGCGAAATTTTTAAGGAAGAGA  
AAAACCCGGACCTGGTGTTCCTGAAGTGCACCAGCACCTATCCGACCGCGATCGAGGATATGAAC  
CTGAAAGGTATTGTTAGCCTGAAGGAAAAATTTAACGTTGAGGTGGGTCTGAGCGACCACAGCTT  
CGGCTTTCTGGCGCCGGTATGGCGGTTGCGCTGGGTGCGCGTGTATCGAAAAGCACTTCATGC  
TGGACAAAAGCATTGAAAGCGAGGATAGCAAGTTTAGCCTGGACTTCGATGAATTTAAAGCGAT  
GGTGGATGCGGTTCTGCAAGCGGAGAGCGCGCTGGGTGATGGCAAGCTGGACCTGGATGAAAA  
GGTGCTGAAAAACCGTGTTCGCGCGTAGCCTGTACGCGAGCAAAGATATCAAGAAAGGCGAG  
ATGTTTAGCGAAGAGAACGTGAAGAGCGTTCGTCCGAGCTTCGGTCTGCACCCGAAATTTATCA  
AGAACTGCTGGGCAAGAAGGCGAGCAAGGACATCAAGTTCGGTGACGCGCTGAAGCAAGGCGA  
TTTCCAATAA

MGSSHHHHHSSGLVPRGSHMQIGNFNDDKVFIIAELSANHAGSLEMALKSIKAAKKAGADAIIQ  
TYTPDSLTLNSDKEDFIIKGLWDRKLYELYESAKTPYEWHSQIFETAQNEGILCFSSPFAKEDVEFLKR  
FDPIAYKIASFEANDENFVRLIAKEKKPTIVSTGIATEEELFKICEIFKEEKNPDLVFLKCTSTYPTAIEDMNL  
KGVLSLKEKFNVEVGLSDHSFGFLAPVMAVALGARVIEKHFMLDKSIESEDSKFSLDFDEFKAMVDAVR  
QAESALGDGKLDLDEKVLKNRVFARSLYASKDIKKGEMFSEENVKSVRPSFGLHPKFYQELLGKKASKD  
IKFGDALKQGDFQ

### CjPseF- Codon optimised

pET-15b NdeI-BamHI

CATATGCGTGCGATTGCGATTGTTCTGGCGCGTAGCAGCAGCAAACGTATCAAGAATAAGAATAT  
CATTGACTTTTTCAACAAGCCGATGCTGGCGTACCCGATCGAGGTGGCGCTGAACAGCAAGCTGT  
TCGAAAAAGTTTTTATTAGCAGCGACAGCATGGAGTACGTGAACCTGGCGAAGAAGTATGGTGC  
GAGCTTCTGAACCTGCGTCCGAAAATCCTGGCGGACGATCGTGCACCACCCTGGAAGTTATGG  
CGTACCACATGGAGGAACTGGAGCTGAAGGACGAAGATATTGCGTGCTGCCTGTATGGTGGCAG  
CGCGCTGCTGCAGGAGAAGCACCTGAAAAACGCGTTTGAACCCCTGAACAAAAACCAAAACACC  
GATTACGTTTTACCTGCAGCCCGTTTAGCGCGAGCCCGTATCGTAGCTTCAGCCTGGAGAACGG  
TGTGCAGATGGCGTTTAAGGAACACAGCAACACCCGTACCCAAGACCTGAAAACCCCTGTACCACG  
ATGCGGGTCTGCTGTATATGGGCAAGGCGCAGGCGTTCAAAGAGATGCGTCCGATCTTTAGCCA  
AAACAGCATTGCGCTGGAGCTGAGCCCGCTGGAAGTGCAGGACATCGCGCACTTCCGTCGTTTCC  
GTATTAGCCAGGCGCAGATTCAGCCGTTTGAAGAAGCGTATGCCGGTGTA (Codon optimised  
for *E. coli*)

#### CjPseF- Non-Codon optimised

CATATGAGAGCGATCGCTATTGTTTTAGCCAGAAGTCCAGTAAAAGGATCAAGAATAAAAATAT  
TATTGATTTTTCAATAAACCCATGCTCGCTTACCCTATTGAGGTGGCGCTAAATTCCAAGCTCTTT  
GAAAAGGTGTTTATCTCTAGCGATAGCATGGAGTATGTTAATCTGGCTAAAAATTATGGGGCGAG  
TTTTTTGAATTTGCGCCCTAAAATTTTAGCGGACGACAGAGCCACGACTTTAGAGGTGATGGCCTA  
TCACATGGAAGAATTAGAATTAAGATGAAGATATTGCGTGTTGTTTGTATGGCGCTTCAGCGC  
TTTTACAAGAAAAGCATTAAAAAACGCTTTTTGAACTTTAAACAAAAACCAAAATACGGATTATG  
TTTTACATGCTCTCCATTTAGCGCTTCGCCCTATCGTTCTTTTAGTCTTAAAACGGCGTTCAAAT  
GGCTTTAAAGAGCATTCAAACACGCGCACGCAAGATTTAAAAACGCTCTATCATGACGCGGGGT  
TGCTTTATATGGGGAAGGCTCAAGCCTTTAAAGAAATGCGGCCTATTTTTAGTCAAATTCTATCG  
CTTTAGAATTATCGCCCTTAGAAGTCCAAGATATTGCACACTTTAGAAGATTTAGAATTAGCCAAG  
CTCAAATACAGCCGTTTAAAAACGCATGCCAGTATA (Wild type sequence- non-codon  
optimised)

[MGSSHHHHHSSGLVPRGSHMRAIAIVLARSSSKRIKKNKIIDFFNKPLAYPIEVALNSKLFKVFIS  
DSMEYVNLAKNYGASFLNLRPKILADDRATLEV MAYHMEELKDEDIACCLYGASALLQEHLKNA  
FETLNKNQNTDYVFTCSPFASPYRSFSLENGVQMAFKEHSNTRTQDLKTLYHDAGLLYMGKAQAFK  
EMRPIFSQNSIALELSPLEVQDIAHFRRFRISQAQIQPFKRMPV\\*](#)

#### AcPseF

ATGAATATTGCCATCATCCCTGCCCGTGGTGGCAGTAAGCGTATTCCTAGGAAAAATATCAAACC  
ATTTATAGCAAGCCCATGATCGCATGGTCCATCTTAGCTGCTAAGAAGGCTGGTTGTTTTGAACG  
TATAATTGTTTCAACCGATGATGCTGAAATTGCTGCTGTTGCACTCGAATATGGTGTGAAAGTGC  
ATTTACTCGCCCGGCAGAGATTGCTAACGATTATGCCACTACTGGTGAGGTGATAAGCCATGCCA  
TTAATTGGTTGATAAATCAGCAAGGGCAAGTGCCGGAAAACGTATGCTGCCTCTATGCAACAGCA  
CCTTTTGTGAGCCTGATGATTTATGCCAGGGATTAGAATTGTTAACGTTCAACAAGGAATGCCAA  
TTTGTTCAGTGCTACTCGCTTTTCGTTCCGATTCAACGCGCTATCAAGCTTGATGAGTCAGGTT  
GGGTGAGTATGTTTCATCCGAGTATCAACTAACTCGTTCCCAAGATCTGGAGGAAGCCTACCAT  
GATGCGGGCAATTTTATTGGGAAAGGCTAATGCTTGGCTTAATAAATTACCTATATTTGCCGT

GCATACACAGGTAGTTCTATTACCCAGCCACAGGGTCAAGATATTGATACTCAGGATGATTGGC  
TGCGTGCTGAGAAGCTATTTACGCTAAGGTA

MGSSHHHHHSSGLVPRGSHMNIAIIPARGGSKRIPRKNIKPFHSPMIAWSILAACKAGCFERIIVST  
DDAEIAAAVALEYGA EVPFTRPAEIANDYATTGEVISHAINWLINQQGQVPENVCCLYATAPFVEPDDL  
CQGLELLTFNKECQFVFSATRFSPFIQRAIKLDESGWVSMFHPEYQLTRSQDLEEAYHDAGQFYWGK  
ANAWLNKLPFAVHTQVLLPSHRVQDIDTQDDWLRAEKLFTLR\*

### Gtr16

Vector: pET-15b

Restriction Enzymes: NdeI, BamHI

Sequence of Insert:

ATGTACGTTTATGGTGTGAACTGGTGGTGTCTCAGCGTAACCTGCTGATCTGCTTCACCCCGCTG  
CAAATCCTGATTGCGAGCAAGATCCTGGTTGAGAAAGACTTTGATACCCTGCTGATTAGCTACGT  
GGACAACGATAAGTACCGTTTCTACTTCGATAAGATCAGCGCGATTAGCCGTAAGAGCTGGTTCT  
TTAAAATCAACAGCACCAACAAGTTTAGCCGTATGATGGACATGATCAAGCTGAAGAAAATCATC  
CGTGAATTCGATCCGCACTACAACATCGTTTATTTTTCGAGCCTGGACAACCGCTTCCTGCACCTG  
GTGGTTAGCAACATCAGCTTTAACAGCATTGAGACCTTCGACGATGGTAGCGCGAACATTAACAA  
GGATAGCACCTATTTCAAGGGCGAACGTAAGAGCAGCTTTCAGCTGCTGTTTAGCGCGCTGCTGG  
GTATCAAGTTTAAACAAAAGCATCATTCTGGACAAGATCTACAAACTATAGCATTTTTCGAGGGCT  
ACAGCAACATTGTTCCGAACGTGGAATACATCAAGATCTTCGAGAGCGAAAACCTGGTTCCGCCG  
AACAAAGGTGATCAAAATTTTCTGGGGTCAGCCGTTTGGAGGAAATGGGCTTCATCGATAAAGAGG  
AGCTGTACCTGTTTCTGCGTAAAATCGGTATTGACTACTATTTCCCGCACCCGCGTGAGAAGCACG  
ACAAGGATTTCTACTTCGAAATCGTGCAAAGCAAAGTATCTTCGAGGAATTTATTCTGGAGCTGC  
TGACCGATGGTAACCTGATTGAAGTTTACACCCTGCTGAGCACCCGCGGGCCTGAACGTGAGCATG  
CTGGACTATGTGACCGTTAAGGTGATCCGTAGCAAAGATCTGTACATGCGTTATAGCAGCCTGTA  
CAAGGTTTTCCAAGACATGAAAGTGGAGTTCATCGACTTTGATGGCATTAGCCTGTAA

MGSSHHHHHSSGLVPRGSHMYVYGVKLVVSQRNLLICFTPLQILIASKILVEKDFDILLISYVDNDKYR  
FYFDKISAI SRKSWFFKINSTNKF SRMMDMIK LKKIREFDPHYNIVYFASLDNAFLHLVVSNISFNSETF  
DDGSANINKDSTYFKGERKSSFQLLFSALLGIFKFNKSIILDKIYKHYSIFEGYSNIVPNVEYIKIFESENLVPP  
NKVIKIFLGQPFEEMGFIDKEELYFLRKGIDYFPHPREKHKDFYFEIVQSKLIFEFILELLTDGNLIEV  
YTLLSTAGLNVSM LDYVTVK VIRSKDLYMRYSSLYKVFQDMKVEFIDFDGSL

### KpsS1

Vector: pET-15b

Restriction Enzymes: NdeI, BamHI

Sequence of Insert:

ATGAACTTTCTGATCCTGATTAACAGCGCGCCGAACTACAAGTATTTCTTTTATGAACTGGCGAAA  
GAAATCGAGAGCCGTGGCCACAACATTTATTTTCGCGATCGATAGCCACCGTAGCAAGTACCTGGA

ACCGCTGCCGGAGCTGGACAACAACCAGAACAGCTTCTTTTTTCGATAGCTACCTGGAAAAGAACT  
TTGACAAAAACCTGAGCGTTAGCCACAACAACAACCAAGAGTATTGGGGTGATTACTTCTATAGC  
GACTACGATCGTTTTCTGACCCACGATTTCAACCTGAACAAGGACAAAACTACTGGCTGAACGT  
GAAAGTTAGCCTGGACAGCTTTTTCGAAGACATCATTAAAGGATAAACAGATCGACTTTGTGCTGT  
ATGAGAACATTAGCAACAGCTTCGCGTACGCGGCGTATCTGCAATGCACCAAACCTGGGCAAGAA  
ATACATCGGTCTGATGGGCAGCCGTCTGCCGAACCACTTTGAAATTCAGAACAGCATCGTTGAGG  
AAGAGCTGAAGAACTGGAGATCCTGGCGCAGAAGCCGATTACCCAAGATGAAATGGAGTGTT  
CGAAAACATAAGAAAAGCATCGTGGATATTCAGCCGACTACATGAAACAAAACGGTCTGGAC  
AACGTGGCGATCAGCCGTATTGTTAAGCTGAACAAATTTCTGAAGGCGCTGCGTCTGCTGACCAT  
CGGCTTCAAGTACAAGCACTACTACGATTACCAGTTCGGTAACCCGTTTCATGGTGCCGATCAAGG  
CGATTCGTGTTAACATTAACGTTATCTGAACACCAAGAAAAGCCAAAAGTTTTACATCAACAACG  
ACGAACTGGAGATTTGCAGCAGCAAAGAACGTTTTTACATCTATCCGATTCACTTCCATCCGGAGA  
GCAGCACCAGCGTTCTGGCGCCGAATACACCAACGAGTATAGCAACATCATTAAACATCGCGAAC  
AACCTGCCGTTTCGGCACCTACCTGTATGTGAAAGATCACAAGAGCGCGAAAGGTGTTCCAGAGCTA  
CGAGTTTTATAAGAAAGTGAGCAGCCTGCCGAACGTGCGTCTGGTTAACTTCGACGTGAACATTA  
AGCGTCTGATCCTGAAAAGCCTGGGCGTGATTACCGTTAACAGCACCGCGGGTTACGAAGCGCT  
GCTGCTGGGCAAGCCGGTGTACCTGCTGGGTCGTGTGTTCTACGAGAACTTCAACAACGTTTACA  
ACCTGAAAAGCTTCCGTGACATCCGTGATATTCGTGACATCCTGGATTTTCAATTCCTGGATGTGA  
AGAAAGACTTTATCGCGTACAAGAAATACGTGTATAAAGGCGTTATCTTCATTGATTATGGTAACC  
GTGTTAACGACAAGAAACGTTACTTTAGCGAACTGGTGGATAGCATTTTCTGAACATCAACACC  
GACATTGAGTAA

MGSSHHHHHSSGLVPRGSHMNFLILINSAPNYKYFFYELAKEIESRGNHNYFAIDSHRSKYLEPLPELD  
NNQNSFFFDSYLEKNFDKNLSVSHNNNQEYWGDYFYSDYDRFLTHDFNLNKDKNYWLNVKVSLSDF  
FEDIKDKQIDFVLYENISNSFAYAAYLQCTKLGKKYIGLMGSRPNHFEIQNSIVEEELKKLEILAQKPITQ  
DEMEWFENYKKSIVDIQPDYMKQNGLDNVAISRIVKLNKFLKALRLLTIGFKYKHYYDYQFGNPFMVP  
IKAIRVNIKRYLNTKKSQKFYINNDELEICSSKERFYIPIHFHPESSTSVLAPEYTNESNIINIANNLPFGT  
YLYVKDHKSAKGVQSYEFYKKVSSLPNVRLVNFVNIKRLILKSLGVITVNSTAGYEALLGKPVYLLGRV  
FYENFN NVYNLKSFRDIRDIRDILDFQLDVKKDFIAYKKYVYKGVIFIDYGNRVNDKKRYFSELVDSIFLN  
INTDIE

Lst

Vector: pET-28a

Restriction Enzymes: NdeI, BamHI

Sequence of Insert:

ATGAATTTAGTTGTTTGTAAATACACCATTTCAAGCTTTCGAGATTATCAATTTGGTTGCCAAAGGA  
ATAATAACAACTTCGATTTTTTTTTATTTTTGTAAAAATAAACTGCACAAGTAGAATATTACTTTG  
AGCAAGTTAAGAAGCACGCTATTTTCATCACAACCTCTATATTAGTGACAAGCGTTATCCTTATCATA  
TATTAGACATCAGGCGTCTATTCAAAGGTAGGTATTACCAATCTGTCTACAGTGCCTCTGTTGATA  
GTGTATTCACACACAGTATTTTATCGTTTATAAAATTTGACAATTTTTACTCTTTTGATGATGGGTC  
TGCTAATTTAAATAGAACTCTACATATTACATTGAGCAGAGAGGTTATTTTAAACGGATATTGTT

CAGATTGACAGGTTGTAAATATGACTTGAGAAAAACGAAATCAATAATTCGTTCTCATTATACCGT  
ATATAAGAATAAAAAAATATAGTATCCAATACAATATATATTGATTTTGAGTTTCCTGCTCAATG  
TTCTGATGGCCGAAAGGAACTAACGACAAAACTCAGTCGCCAATGTATTGTTAGGTACAGTTT  
ATGATGAGGTGTTTTCGGATAATGAGGTAATAATTAATAAATTTGTCATGCTTTTTTGGCGATAAAG  
ATTTCTATTATATCCCACACCCAAGAGATACACATAACTTCTTTATAAATGGAATGCGTATCGATG  
GGCCTGAGATTGCTGAAAGCAAAATTCTATCTTTGTTAAAAAGATATGAAACAATAAATTTATATG  
GTTTTGGAAGTTCAGTTCAAATAAATATGAGCTCACATCGTGGAGTTATCAATTATATATTTGATA  
TAGGAAGTCATGAATATGAATATGACCATGATTATGATTATGGTTATGAAATACTGAAGATATAG

Recombinant Protein Sequence:

MGSSHHHHHSSGLVPRGSHMNLVVCNTPFQALQIINLVAKGIITNFDFYFCKNKTAQVEYFEQV  
KKHAISSQLYISDKRYPYHILDIRRLFKGRYYQSVYSASVDSVFTHSILSFIKFDNFYSFDDGSANLNRNST  
YYIEQRGYFKRILFRLTGCKYDLRKTCSIIRSHYTVYKNKNIVSNTIYIDFEFPAQCSDGRKETNDKNSVA  
NVLLGTVYDEVFSDNEVIKKLSCFFADKDFYIPHRDTHNFFINGMRIDGPEIAESKILSLKRYETINLY  
GFGSSVQINMSSHRGVINYIFDIGSHEYEDHDYDYGYEILKI

Orf14

Vector: pET-15b

Restriction Enzymes: NdeI, BamHI

Sequence of Insert:

ATGAACACCTACTTCGACATCCCGCACCGTCTGGTGGGTAAAGCGCTGTATGAGAGCTACTATGA  
TCACTTTGGTCAGATGGACATCCTGAGCGATGGCAGCCTGTACCTGATTTATCGTCGTGCGACCG  
AGCACGTTGGTGGCAGCGACGGTCGTGTGGTTTTAGCAAAGTGGAAAGGTGGCATCTGGAGCGC  
GCCGACCATTGTGGCGCAGGCGGGTGGCCAAGACTTTCGTGATGTTGCGGGTGGCACCATGCCG  
AGCGGTCGTATTGTGGCGGCGAGCACCGTTTACGAGACCGGTGAAGTTAAGGTGTATGTTAGCG  
ACGATAGCGGCGTGACCTGGGTTTACAAATTCACCCTGGCGCGTGGTGGCGCGGATTACAACCTC  
GCGCACGGCAAGAGCTTTCAAGTGGGCGCGGTTACGTTATCCCGCTGTATGCGGCGACCGGTG  
TGAACACGAGCTGAAATGGCTGGAAAGCAGCGATGGTGGCGAGACCTGGGGTGAAGGCAGCA  
CCATTTATAGCGGCAACACCCCGTACAACGAGACCAGCTATCTGCCGGTGGGTGATGGCGTTATC  
CTGGCGGTGGCGCGTGTGGTAGCGGTGCGGGTGGCGCGCTGCGTCAATTCATTAGCCTGGACG  
ATGGTGGCACCTGGACCGACCGGTAACGTGACCGCGCAAACGGCGACAGCACCGATATCCT  
GGTTGCGCCGAGCCTGAGCTACATTTATAGCGAAGGTGGCACCCCGCACGTGGTTCTGCTGTACA  
CCAACCGTACCACCACTTTTGTACTATCGTACCATCCTGCTGGCGAAAGCGGTTGCGGGTAGCA  
GCGGTTGGACCGAGCGTGTCCGGTTTACAGCGCGCCGGCGGCGAGCGGTTATACCAGCCAGGT  
GGTTCTGGGTGGCCGTCGTATTCTGGGTAACCTGTTCCGTGAAACCAGCAGCACCACCAGCGGCG  
CGTACCAATTTGAGGTGTATCTGGGTGGCGTTCCGGACTTCGAAAGCGATTGGTTTACGCTGAGC  
AGCAACAGCCTGTACACCTGAGCCATGGTCTGCAGCGTAGCCCGCGTCGTGTGGTTGTGGAGTT  
CGCGCGTAGCAGCAGCCCGAGCACCTGGAACATCGTTATGCCGAGCTATTTTAACGACGGTGGCC  
ACAAAGGTAGCGGCGCGCAAGTGGAAAGTTGGTAGCCTGAACATTCGTCTGGGTACCGGTGCGGC  
GGTGTGGGGTACCGGCTACTTCGGTGGCATTGATAACAGCGCGACCACCCGTTTTGCGACCGGCT  
ACTATCGTGTTCTGTCGTGGATTTAA

Recombinant Protein Sequence:

MGSSHHHHHSSGLVPRGSHMYVYGVKLVVSQRNLLICFTPLQLIASKILVEKDFDILLISYVDNDKYR  
FYFDKISAIRKSWFFKINSTNKF SRMMDMIKLLKIIREFDPHYNIVYFASLDNAFLHLVVSNI SFNSIETF  
DDGSANINKDSTYFKGERKSSFQLLFSALLGIKFNKSIILDKIYKHYSIFEGYSNIVPNVEYIKIFESENLVPP  
NKVIKIFLGQPFEEMGFIDKEELYFLFRKIGIDYFPHPREKHDKDFYFEIVQSKLIFEFILELLTDGNLIEV  
YTLLSTAGLNVSM LDYVTVK VIRSKDLYMRYSSLYKVFQDMKVEFIDFDGISL

WfdL

Vector: pET-15b

Restriction Enzymes: NdeI, BamHI

Sequence of Insert:

ATGCCGAGCCTGTTTATTTGCGTGACCCCGCTGCAGATGCTGATCGCGGAGAAGATCATTGACAA  
AACCCGTCCGGTGAACATTGAAATCATTGTTCTGGCGTACCAAAGAACGATAAGTACATGCACT  
ACATCAAACGTCTGGAGAAGAAATGCACCAACTTCACCGTGCTGGCGGTTACCCGAAGAACAAA  
TTTGTGACCGTTATTGCGTTCGCGAAGCTGCACACCATCCTGAACAAGAACATGAGCAAAACCTA  
TAGCGAAGTTTACCTGAGCAGCATTGACAACAAGTACGTGCAGCTGATCGTTAGCAAACCTGAACT  
ATGCGCGTCTGTACACCTTTGACGATGGTACCGCGAACATCATTAAAAGCAGCGCGTACTATCAG  
GAAGAGAAGAAGACCCTGAAGACCAACATCCTGCGTTGGATCTTCGGTATTAACAAAGGCCTGC  
AAGAGATCAAGAGCGAAATTTGCAAACACTATACCATCTACCCGAGCGTGAGCAACATTGTTAGC  
AACACCGAGCTGATCGAAATGTTTACCCAGTGCAAGAAACGTAACAAGACAAGAAAGTGGTTC  
GTGTGTTTATCGGTCAACCGTTCGATGAGCTGGGCATCCCGCTGAGCCTGATCGAGGAATTCTTTT  
TCAAGTACAAGATGGATTACTATTACCCGCACCCGCGTGAGAAGATCATCAACAACAAGTTCACC  
TACATCCACAGCCACCTGATTTTCGAGGAATACATCATTGAAGGTCTGCAGCACCAAGACATCGT  
GTATAAGATTTACGGCGCGGTTTGACCCAGCATTCTGAACCTGGCGAGCAGCAAAAACGATATCG  
AGATTTGCAGCATCTATACCGACGAACTGCGTACCAAGTACAGCGATTATTACGCGCTGGCGGAA  
AAGATGAACATCACCTGCTGAAACTGACCTAA

Recombinant Protein Sequence:

MGSSHHHHHSSGLVPRGSHMPSLFICVTP LQMLIAEKIIDKTRPVNIEIIVLAYQKNDKYMHYIKRLEK  
KCTNFTVLAVTPKNKFVTVIAFAKLHTILNKNMSKTYSEVYLSSIDNKYVQLIVSKLNYARLYTFDDGTA  
NIIKSSAYYQEEKTLKTNILRWIFGINKGLQEIKSEICKHYTIYPSVSNIVSNTELIEMFTQCKKRKQDKKV  
VRVFIGQPFDLGIPLSLIEFFFFKYKMDYYPHPREKIINNKFTYIHSHLIFEYIIEGLQHQDIVYKIYGAV  
CTSILNLASSKNDIEICSIYDELRTKYSYDYYALAEKMNITLLKLT



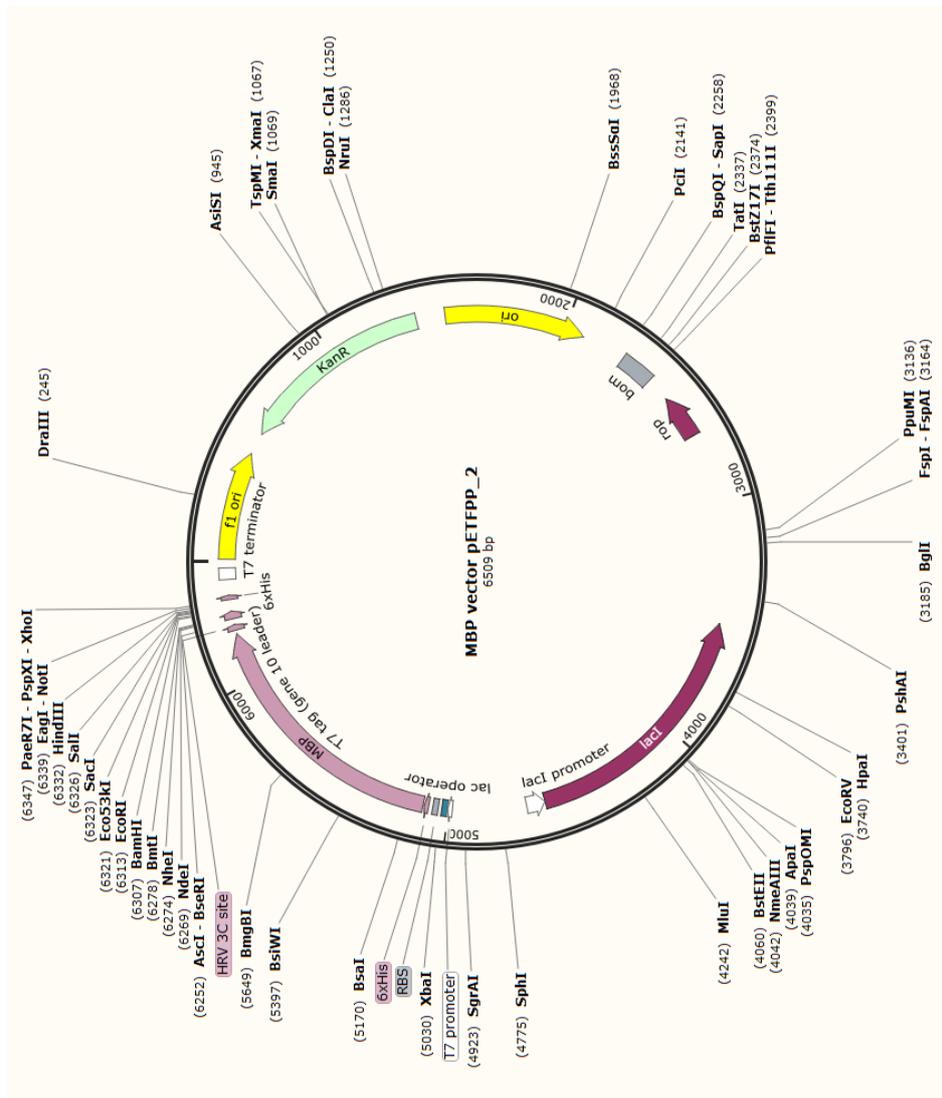


Figure A.11: MBP vector map



TACTGGCTGAACGTGAAAGTTAGCCTGGACAGCTTTTTCGAAGACATCATTAAGGATAAACAGAT  
CGACTTTGTGCTGTATGAGAACATTAGCAACAGCTTCGCGTACGCGGCGTATCTGCAATGCACCA  
AACTGGGCAAGAAATACATCGGTCTGATGGGCAGCCGTCTGCCGAACCACTTTGAAATTCAGAAC  
AGCATCGTTGAGGAAGAGCTGAAGAACTGGAGATCCTGGCGCAGAAGCCGATTACCCAAGATG  
AAATGGAGTGGTTCGAAAACATAAGAAAAGCATCGTGGATATTCAGCCGGACTACATGAAACA  
AAACGGTCTGGACAACGTGGCGATCAGCCGTATTGTTAAGCTGAACAAATTTCTGAAGGCGCTGC  
GTCTGCTGACCATCGGCTTCAAGTACAAGCACTACTACGATTACCAGTTCGGTAACCCGTTTCATGG  
TGCCGATCAAGGCGATTCTGTGTTAACATTAACGTTATCTGAACACCAAGAAAAGCCAAAAGTTT  
TACATCAACAACGACGAACTGGAGATTTGCAGCAGCAAAGAACGTTTTTACATCTATCCGATTCAC  
TTCCATCCGGAGAGCAGCACCAGCGTTCTGGCGCCGGAATACACCAACGAGTATAGCAACATCAT  
TAACATCGCGAACAACTGCCGTTCCGGCACCTACCTGTATGTGAAAGATCACAAGAGCGCGAAAG  
GTGTTACAGAGCTACGAGTTTTATAAGAAAGTGAGCAGCCTGCCGAACGTGCGTCTGGTTAACTTC  
GACGTGAACATTAAGCGTCTGATCCTGAAAAGCCTGGGCGTGATTACCGTTAACAGCACC GCGG  
GTTACGAAGCGCTGCTGCTGGGCAAGCCGGTGTACCTGCTGGGTCGTGTGTTCTACGAGAACTTC  
AACACGTTTACAACCTGAAAAGCTTCCGTGACATCCGTGATATTCGTGACATCCTGGATTTTCAA  
TTCCTGGATGTGAAGAAAGACTTTATCGCGTACAAGAAATACGTGTATAAAGGCGTTATCTTCATT  
GATTATGGTAACCGTGTTAACGACAAGAAACGTTACTTTAGCGAACTGGTGGATAGCATTTCCT  
GAACATCAACACCGACATTGAGTAACGCGCCTTCTCCTCA

Recombinant Protein Sequence:

MGSSHHHHHSSMELKHSISDYTEAEFLQLVTTICNADTSSEEELVKLVTHFEEMTEHPSGSDLIYYPK  
EGDDDSPSGIVNTVKQWRAANGKSGFKQGLEVLFGQPAMNFLILINSAPNYKYFFYELAKEIESRGHN  
IFYAIDSHRSKYLEPLPELDNNQNSFFFDSTYLEKNFDKNLSVSHNNNQEYWGDFYSYDRFLTHDFNL  
NKDKNYWLNVKVSLDSFFEDIKDKQIDFVLYENISNSFAYAAYLQCTKLGKKYIGLMGSRLPNHFEIQN  
SIVEEELKLEILAQKPITQDEMEWFENYKKSIVDIQPDYMKQNGLDNVAISRIVKLNKFLKALRLLTIGF  
KYKHYYDYQFGNPFMVPIKAIRVNIKRYLNTKKSQKFYINNDELEICSSKERFYIPIHFHPSSTSVLAPE  
YTNEYSNIIANIANNLPGTYLYVKDHKSAGVQSYEFYKKVSSLPNVRLVNFVDVNIKRLILKSLGVITVNST  
AGYEALLLGKPVYLLGRVYENFNNVYNLKSFRDIRDIRDILDQFLDVKKDFIAYKKYVYKGVIFIDYGN  
RVNDKKRYFSELVDSIFLNINTDIE\*



GTCTGCTGACCATCGGCTTCAAGTACAAGCACTACTACGATTACCAGTTCGGTAACCCGTTTCATGG  
TGCCGATCAAGGCGATTTCGTGTTAACATTAAACGTTATCTGAACACCAAGAAAAGCCAAAAGTTT  
TACATCAACAACGACGAACTGGAGATTTGCAGCAGCAAAGAACGTTTTTACATCTATCCGATTAC  
TTCCATCCGGAGAGCAGCACCAGCGTTCTGGCGCCGGAATACACCAACGAGTATAGCAACATCAT  
TAACATCGCGAACAACTGCCGTTCCGGCACCTACCTGTATGTGAAAGATCACAAGAGCGCGAAAG  
GTGTTCAAGAGCTACGAGTTTTATAAGAAAAGTGAGCAGCCTGCCGAACGTGCGTCTGGTAACTTC  
GACGTGAACATTAAGCGTCTGATCCTGAAAAGCCTGGGCGTGATTACCGTTAACAGCACCGCGG  
GTTACGAAGCGCTGCTGCTGGGCAAGCCGGTGTACCTGCTGGGTCGTGTGTTCTACGAGAACTTC  
AACACGTTTTACAACCTGAAAAGCTTCCGTGACATCCGTGATATTCGTGACATCCTGGATTTTCAA  
TTCCTGGATGTGAAGAAAGACTTTATCGCGTACAAGAAATACGTGTATAAAGGCGTTATCTTCATT  
GATTATGGTAACCGTGTTAACGACAAGAAACGTTACTTTAGCGAACTGGTGGATAGCATTTCCT  
GAACATCAACACCGACATTGAGTAACGCGCCTTCTCCTCA

MGSSHHHHHSSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLP  
YYIDGDVCLTQSMAIIRYIADKHNMLGGCPKERAISMLEGAVLDIRYGVSRIAYSKDFETLKVDLFLSKL  
PEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPIQIDKYL  
KSSKYIAWPLQGWQATFGGGDHPKGLEVLFGQPAMNLFILINSAPNYKYFFYELAKEIESRGNHNIYFA  
IDSHRSKYLEPLPELDNNQNSFFFDSYLEKNFDKNLSVSHNNNQEYWGDFYSYDRFLTHDFNLNKD  
KNYWLNVKVSLSDFEDIIKDKQIDFVLYENISNSFAYAAYLQCTKLGKKYIGLMGSRLPNHFEIQNSIVE  
EELKLEILAQKPITQDEMEWFENYKKSIVDIQPDYMKQNGLDNVAISRIVKLNKFLKALRLLTIGFKYK  
HYYDYQFGNPFMVPIKAIRVNIKRYLNTKKSQKFYINNDELEICSSKERFYIPIHFHPESSSVLAPEYTN  
EYSNIINIANNLPGTYLYVKDHKSAKGVQSYEFYKVSLLPNVRLVNFVNIKRLILKSLGVITVNSTAG  
YEALLGKPVYLLGRVYENFNNVYNLKSFRDIRDIRDILDFQLDVKKDFIAYKKYVYKGVIFIDYGNRV  
NDKKRYFSELVDSIFLNINTDIE



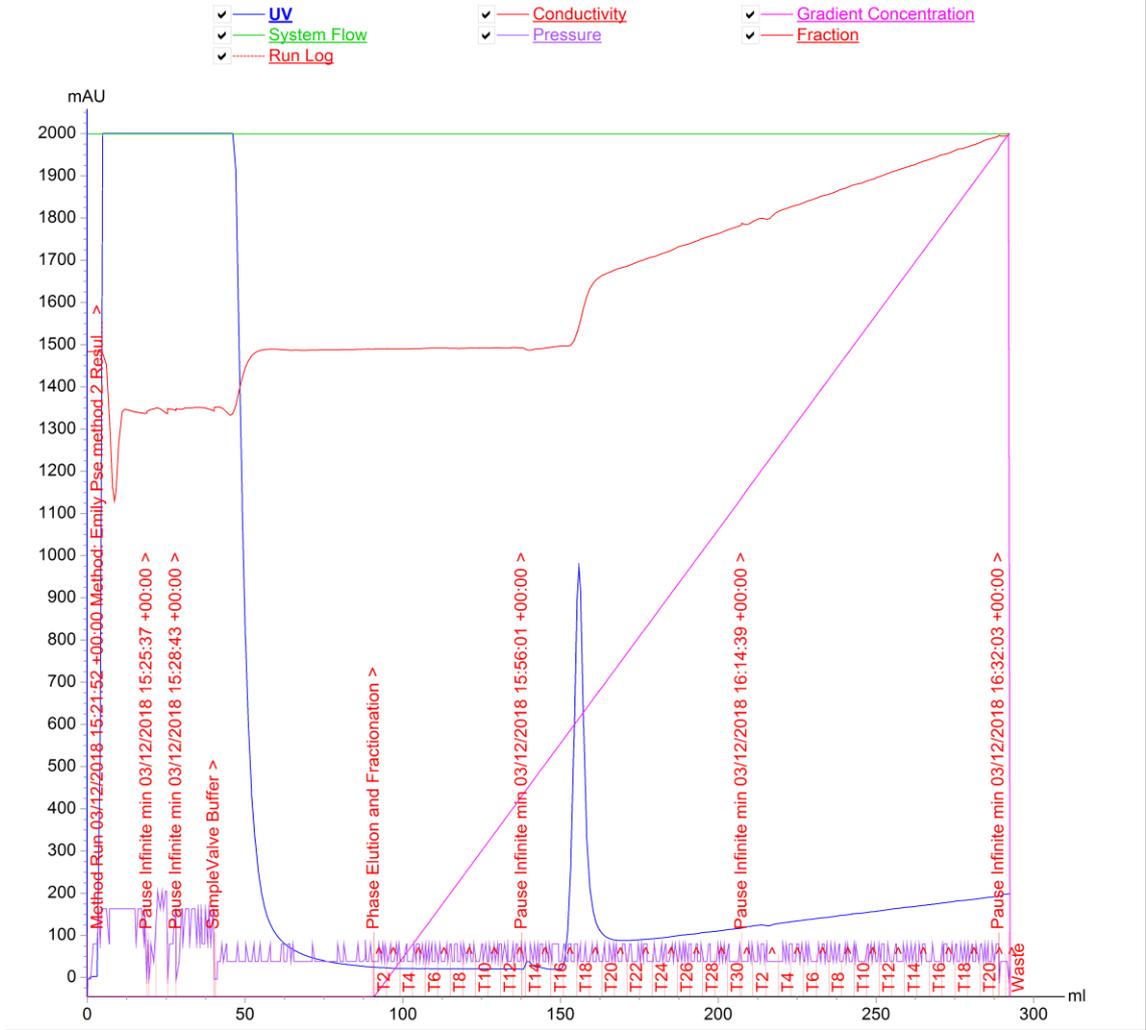


Figure A.14: Chromatogram for PseI Purification using Ni-affinity column

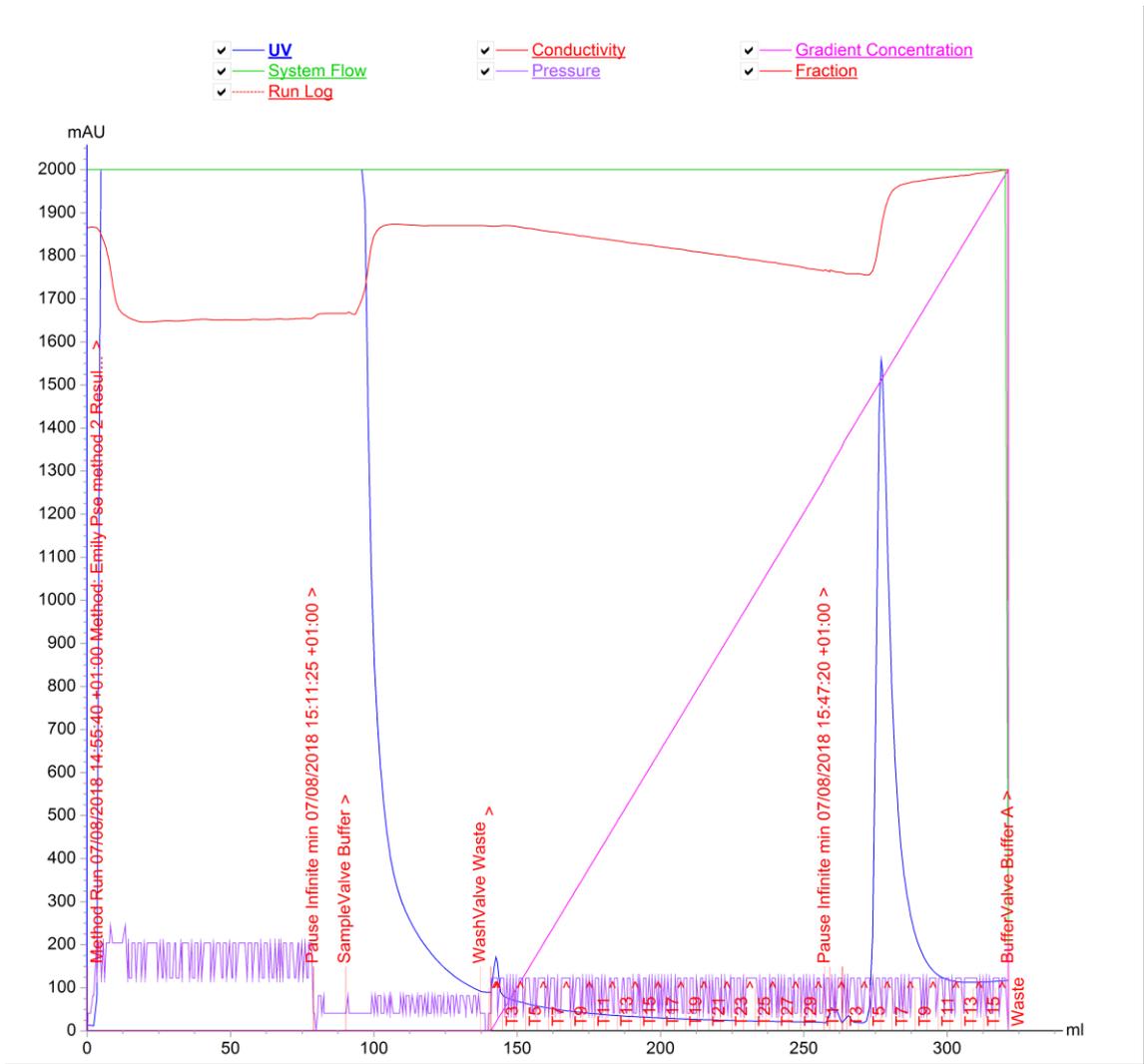


Figure A.15: Chromatogram for PseC Purification using Ni-affinity column

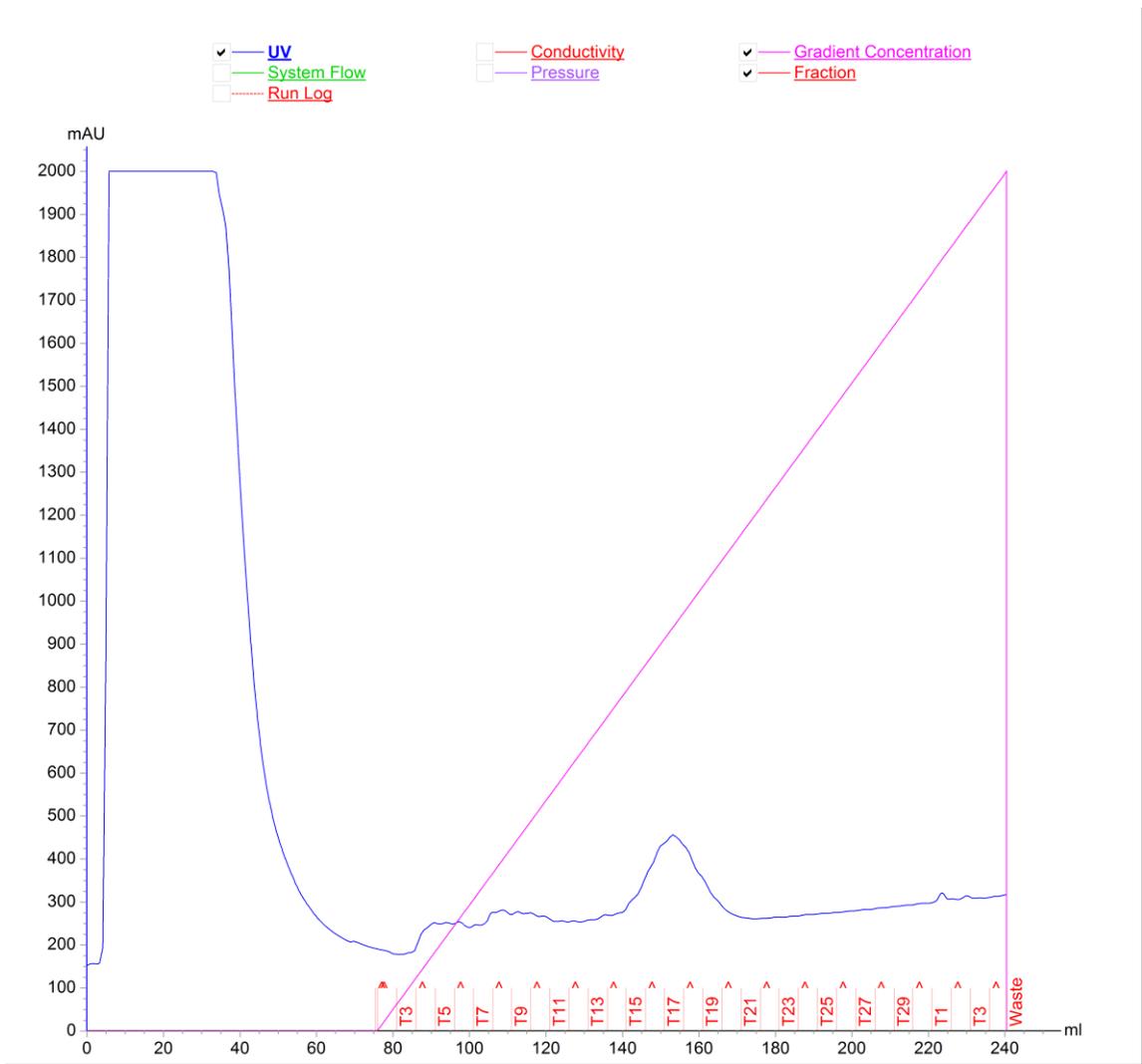


Figure A.16: AKTA Trace for PseG Purification using Ni-affinity column

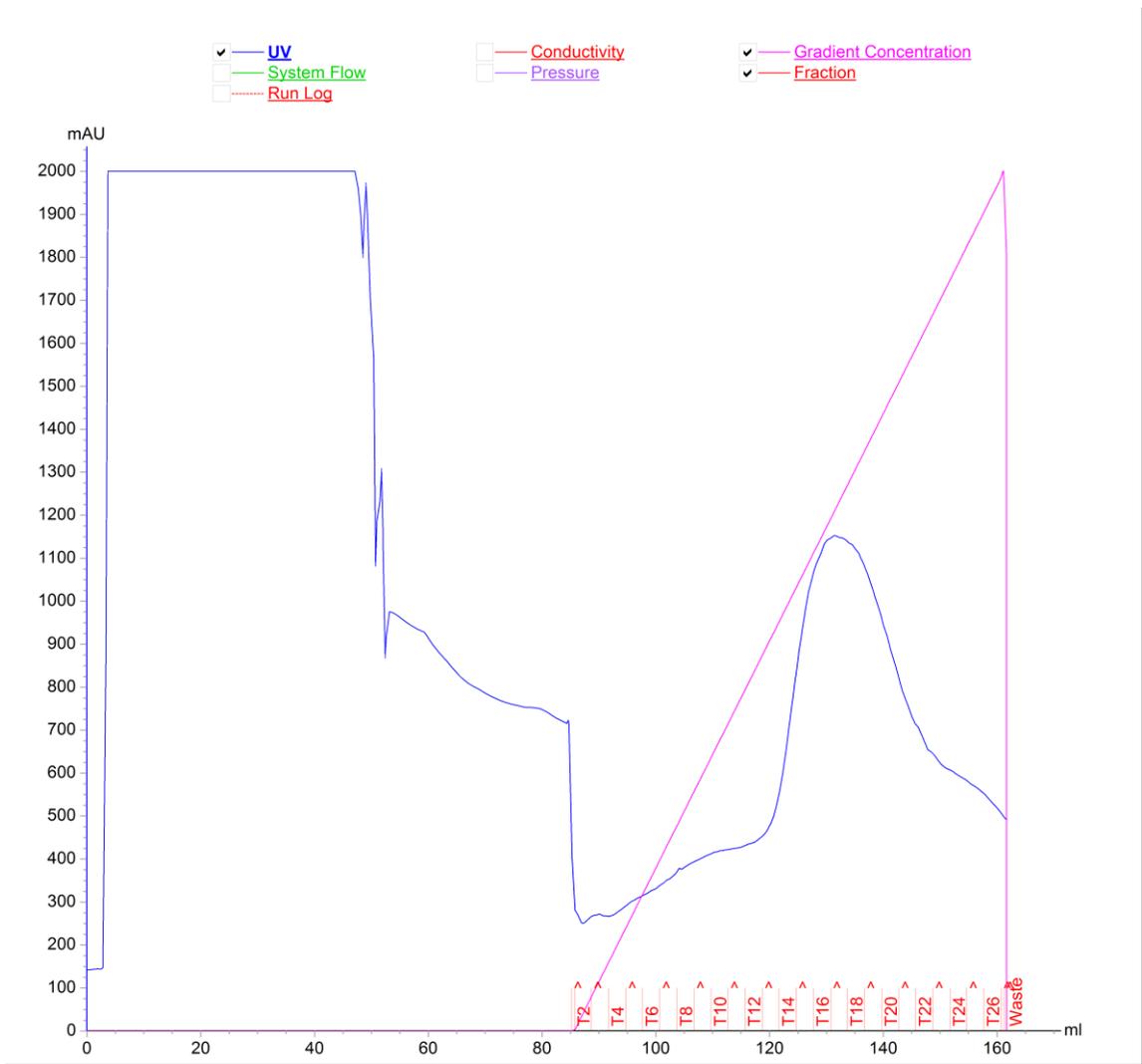
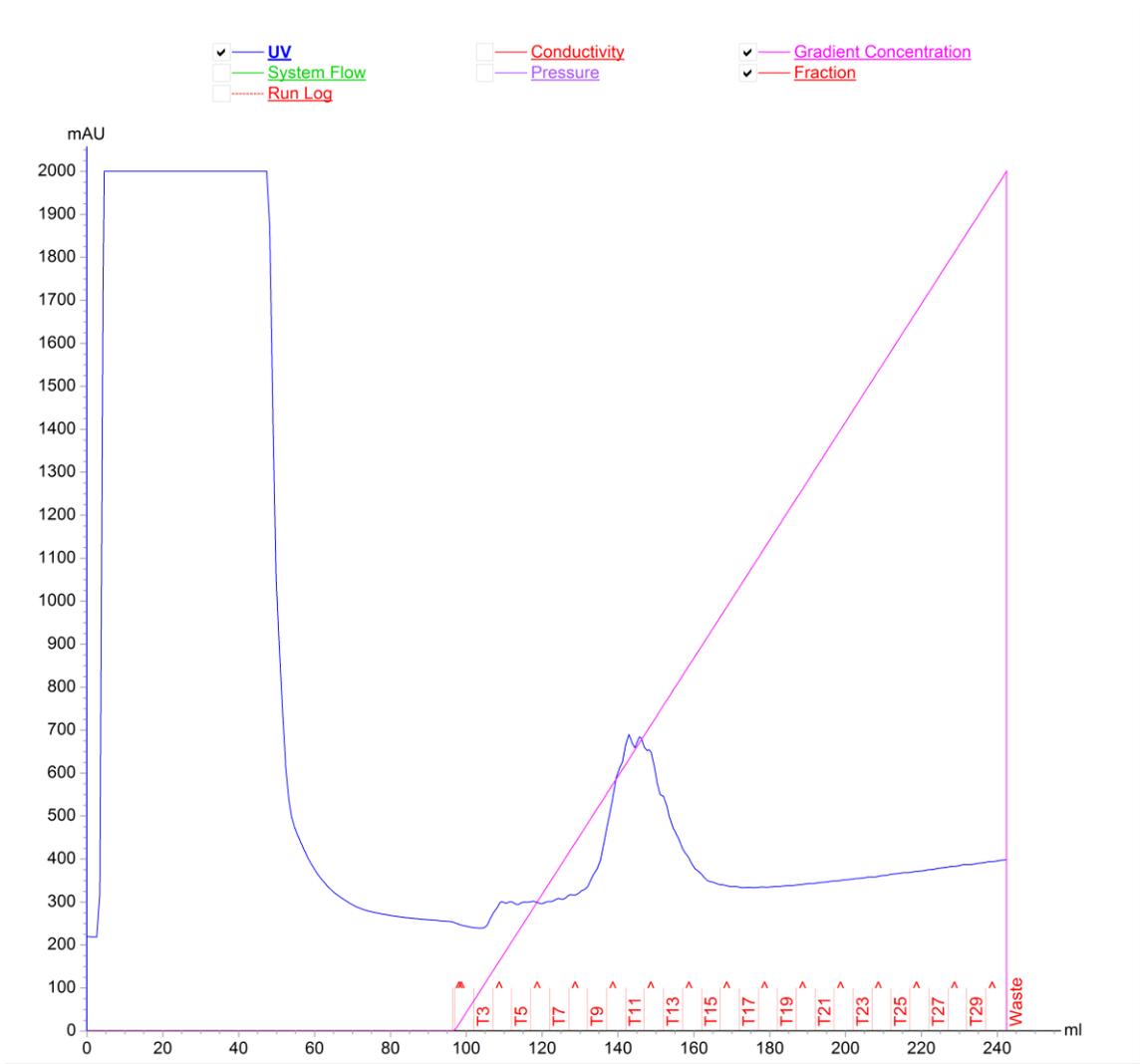
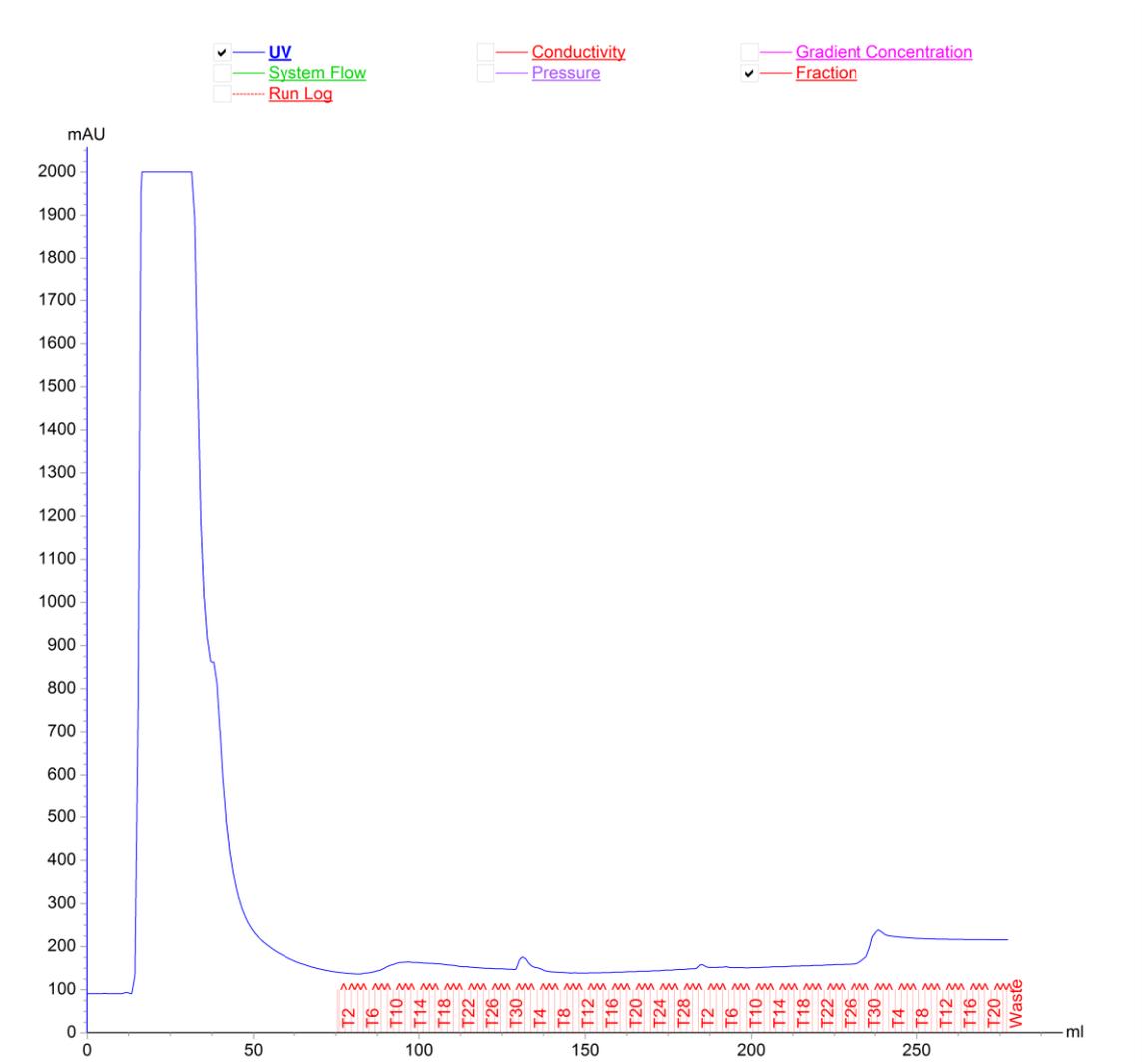


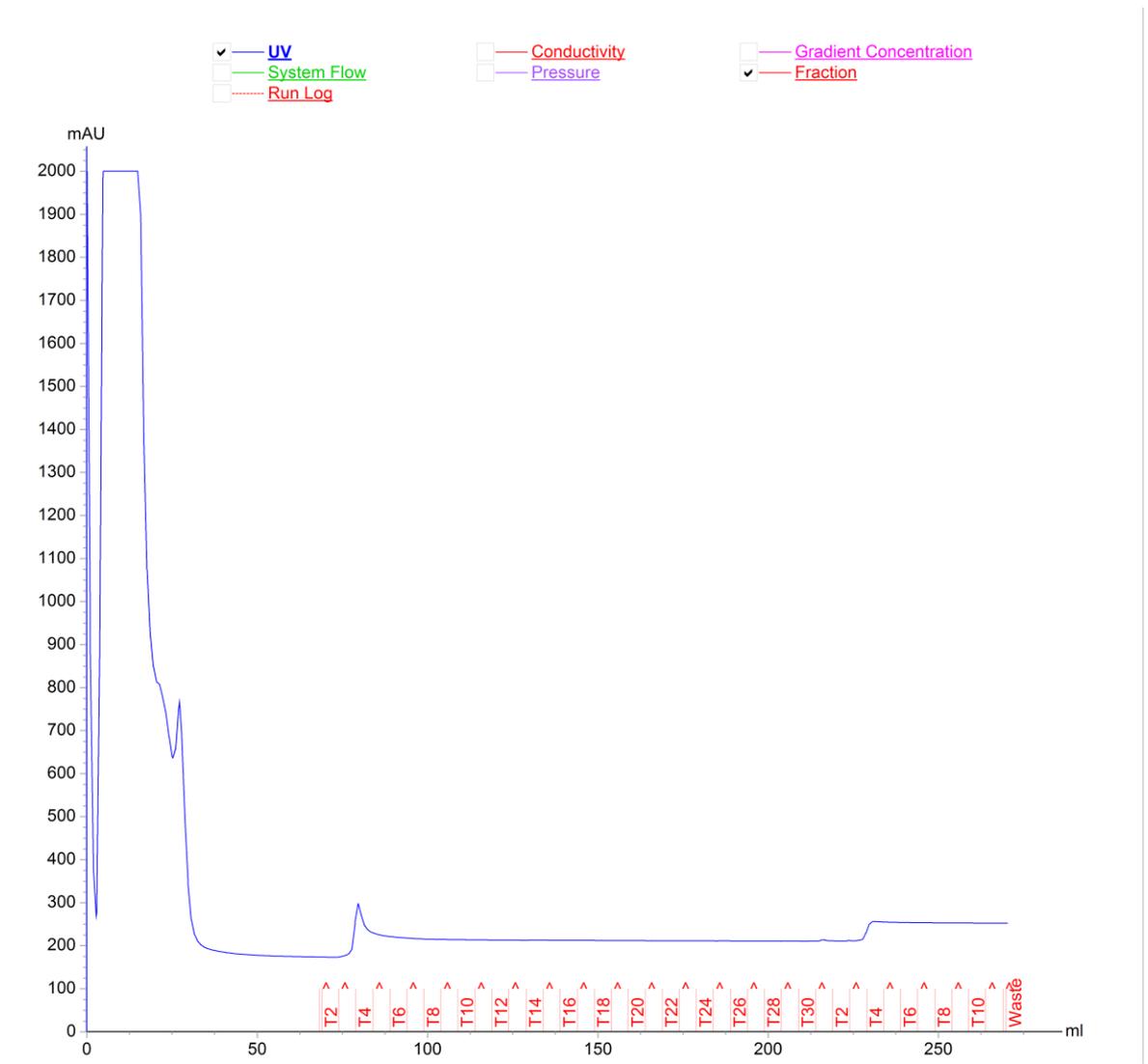
Figure A.17: AKTA Trace for PseH Purification using Ni-affinity column



**Figure A.18: Chromatogram for PseI Purification using Ni-affinity column**



**Figure A.19: AKTA Trace for CjPseF (codon optimised for *E. coli*) Purification using Ni-affinity column. SDS-PAGE in Figure 2.12 contains material from this purification attempt.**



**Figure A.20: AKTA Trace for CjPseF (non-codon optimised for *E. coli*) Purification using Ni-affinity column. SDS-PAGE in Figure 2.12 contains material from this purification attempt.**

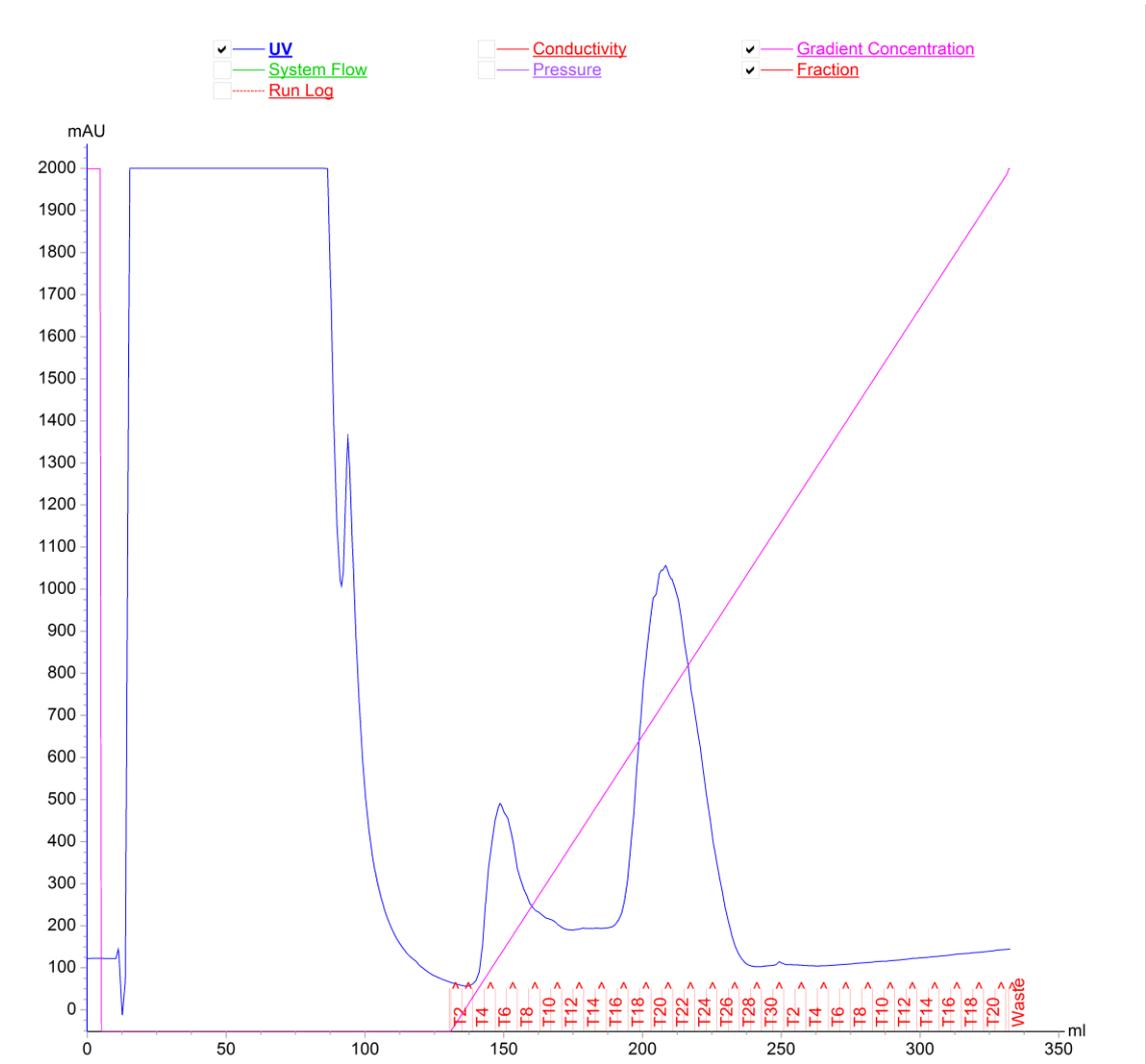


Figure A.21: AKTA Trace for AcPseF Purification using Ni-affinity column

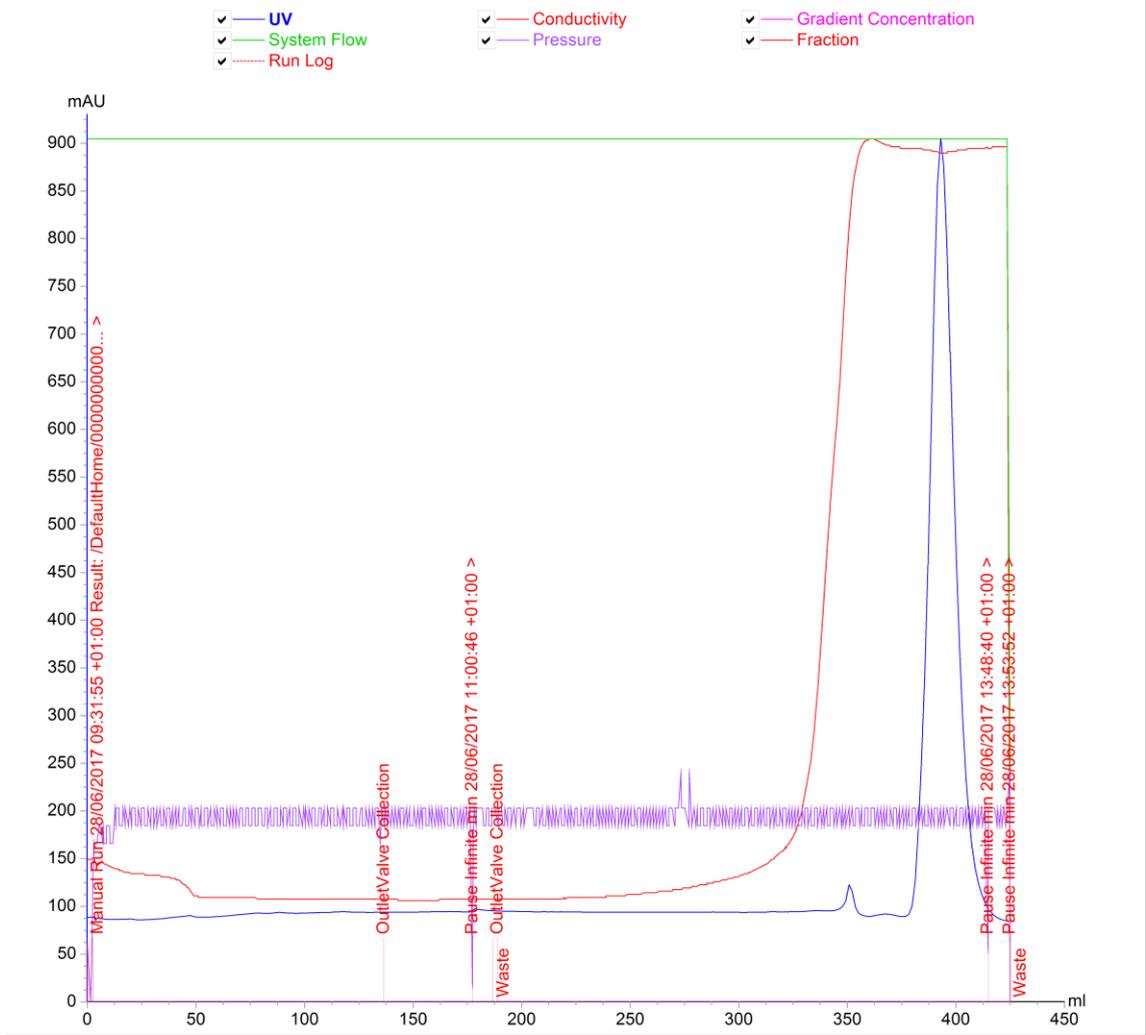


Figure A.22: AKTA Trace for PseF Purification using SEC

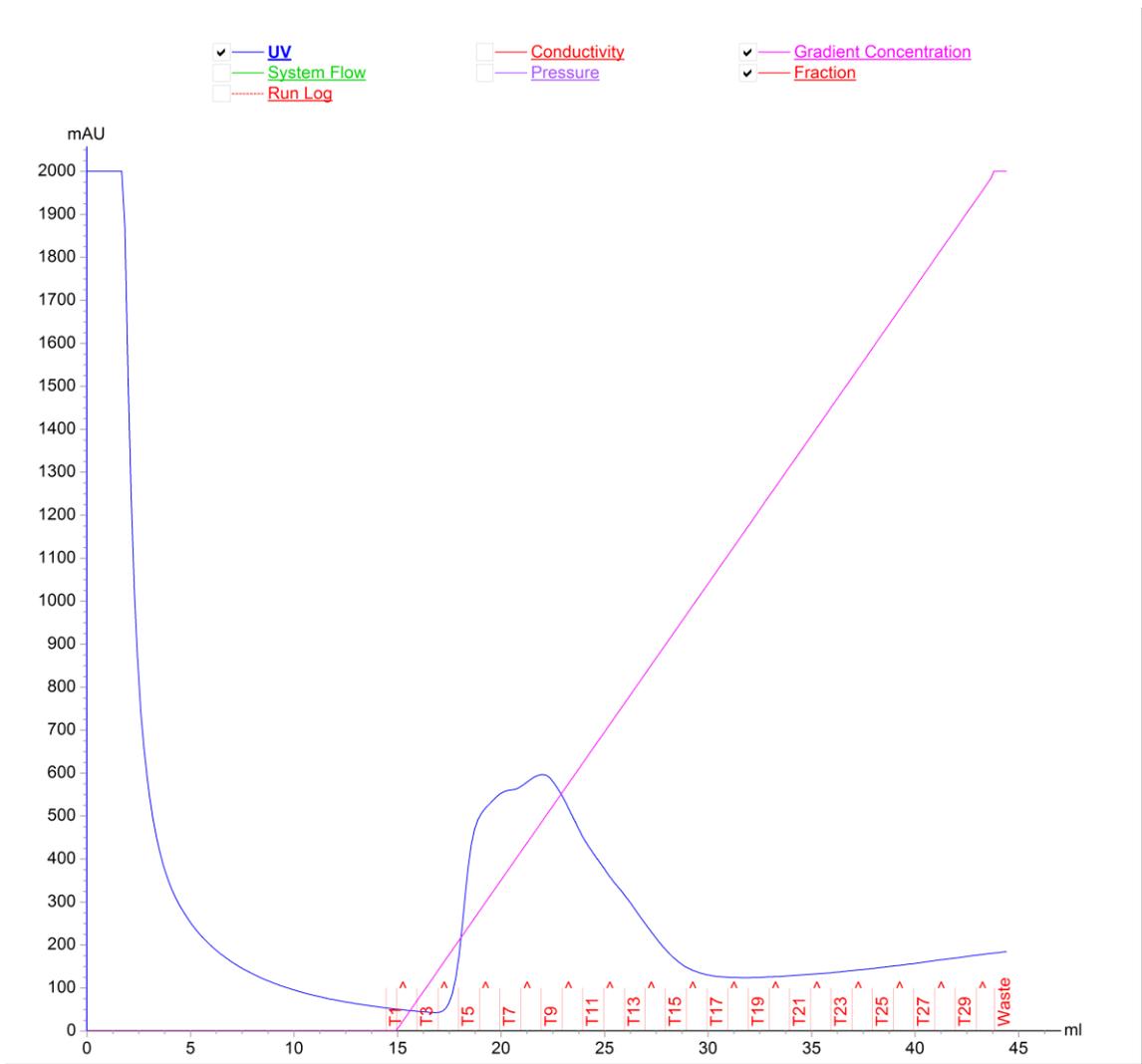
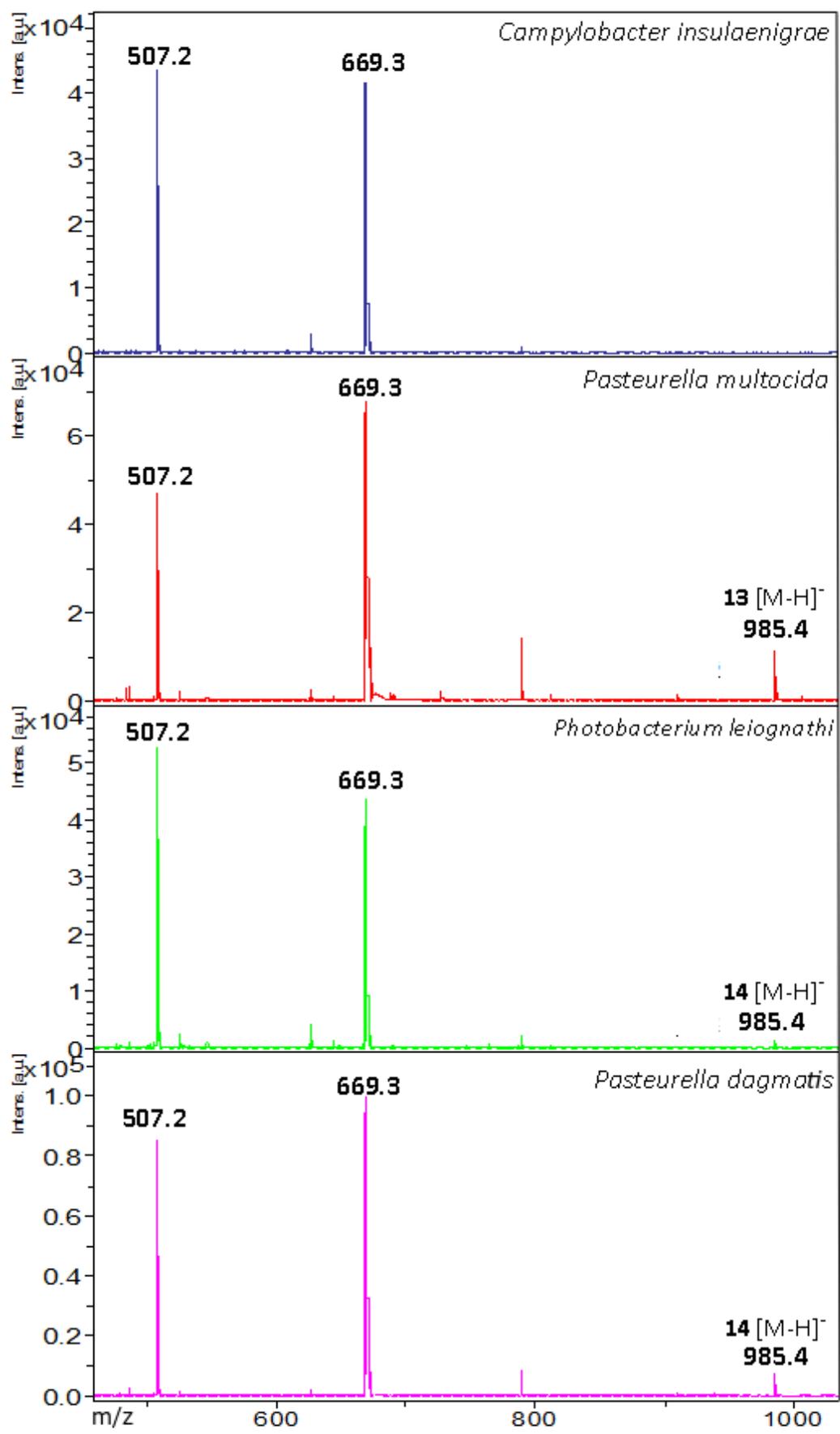
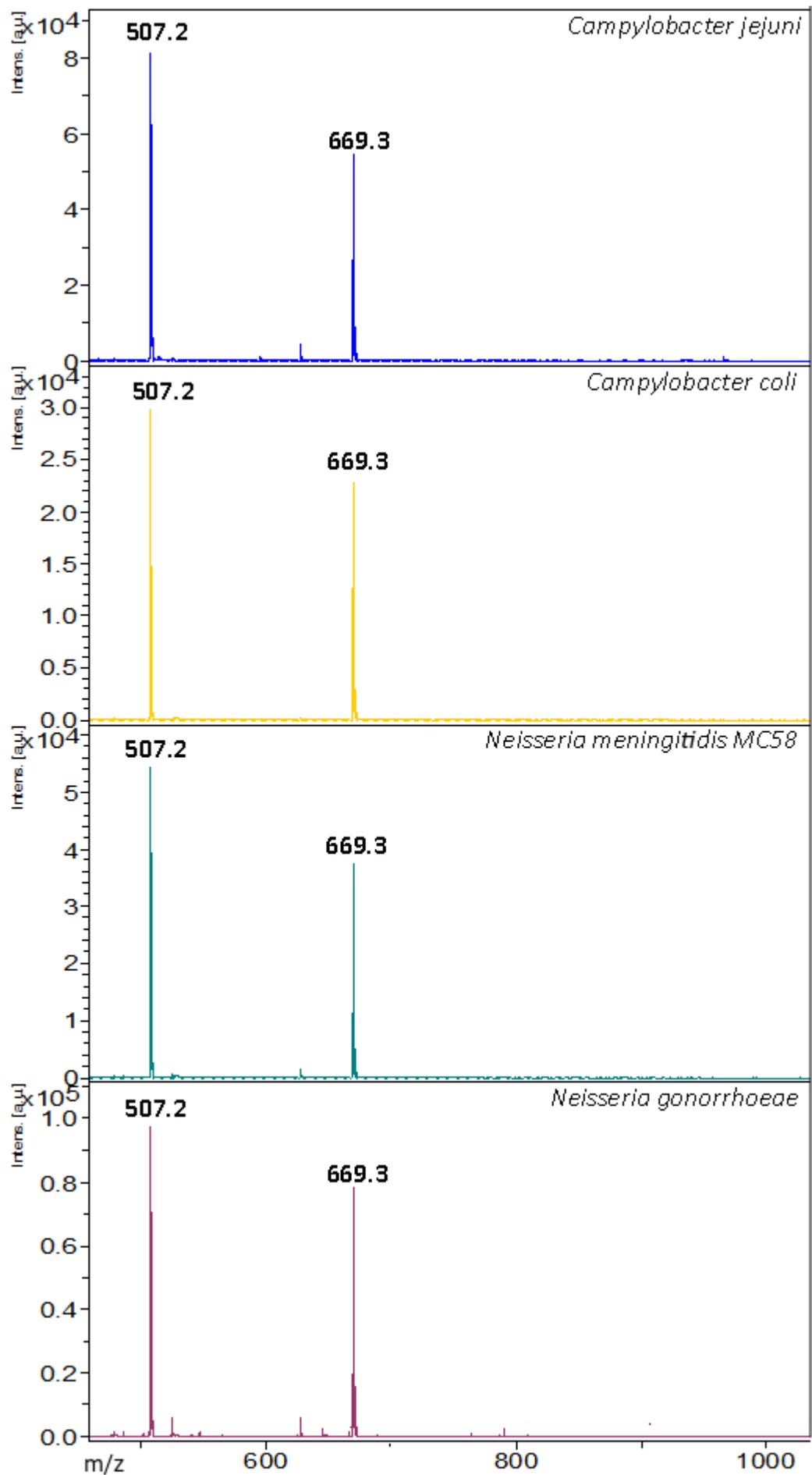
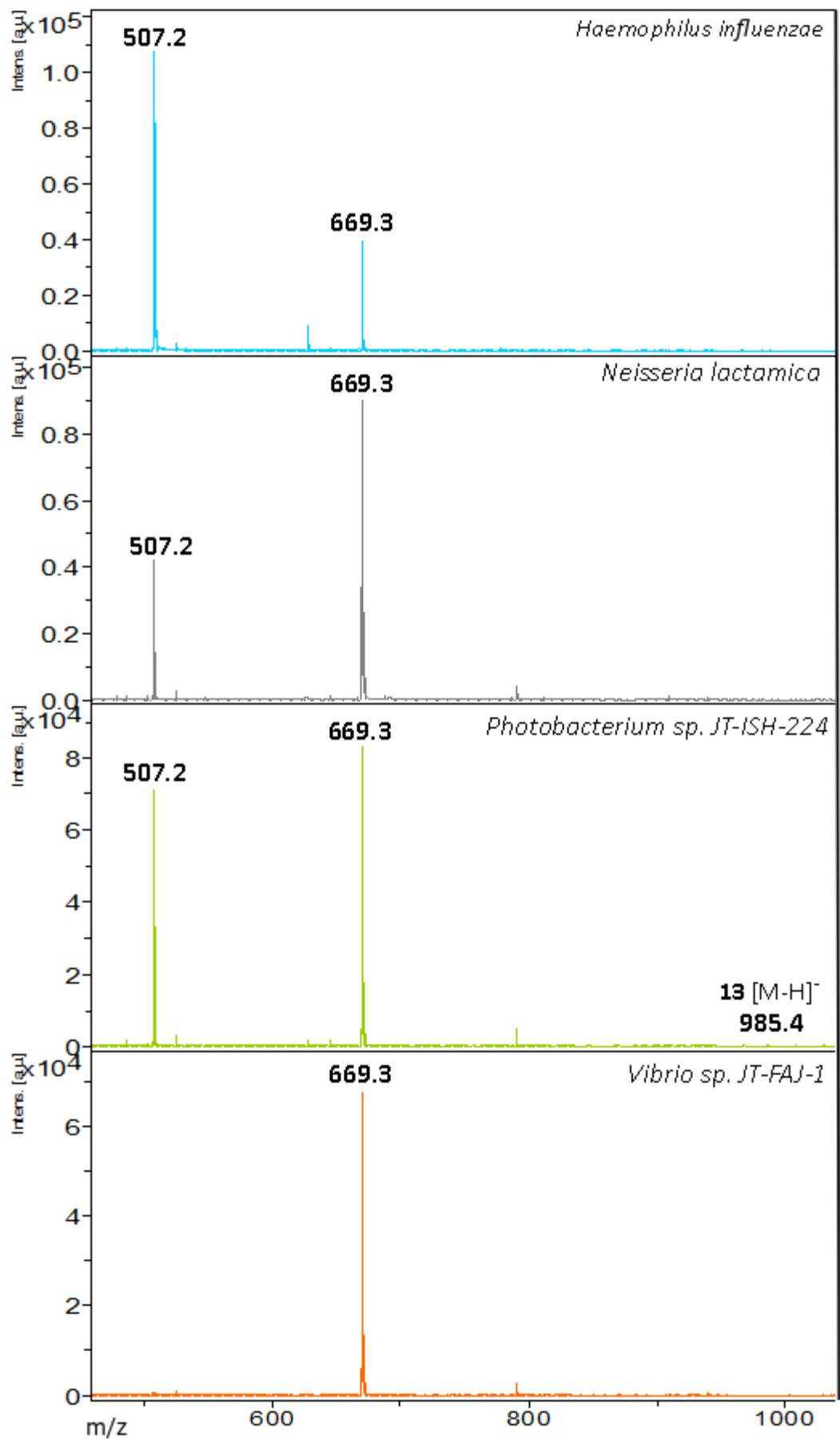


Figure A.23: AKTA Trace for Im9-KpsS1 Purification using Ni-affinity column







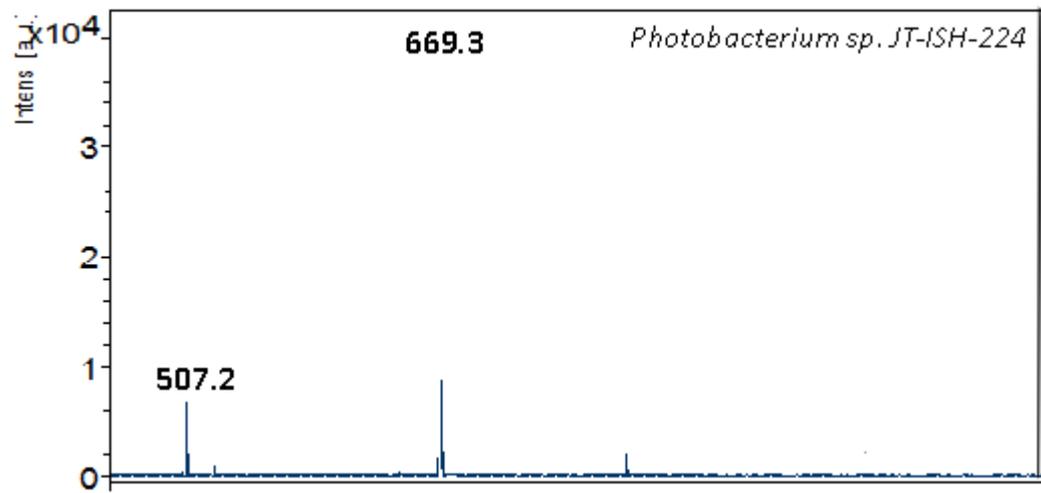
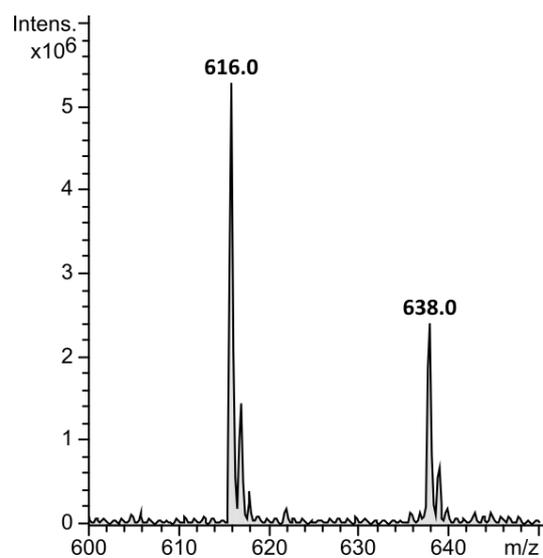


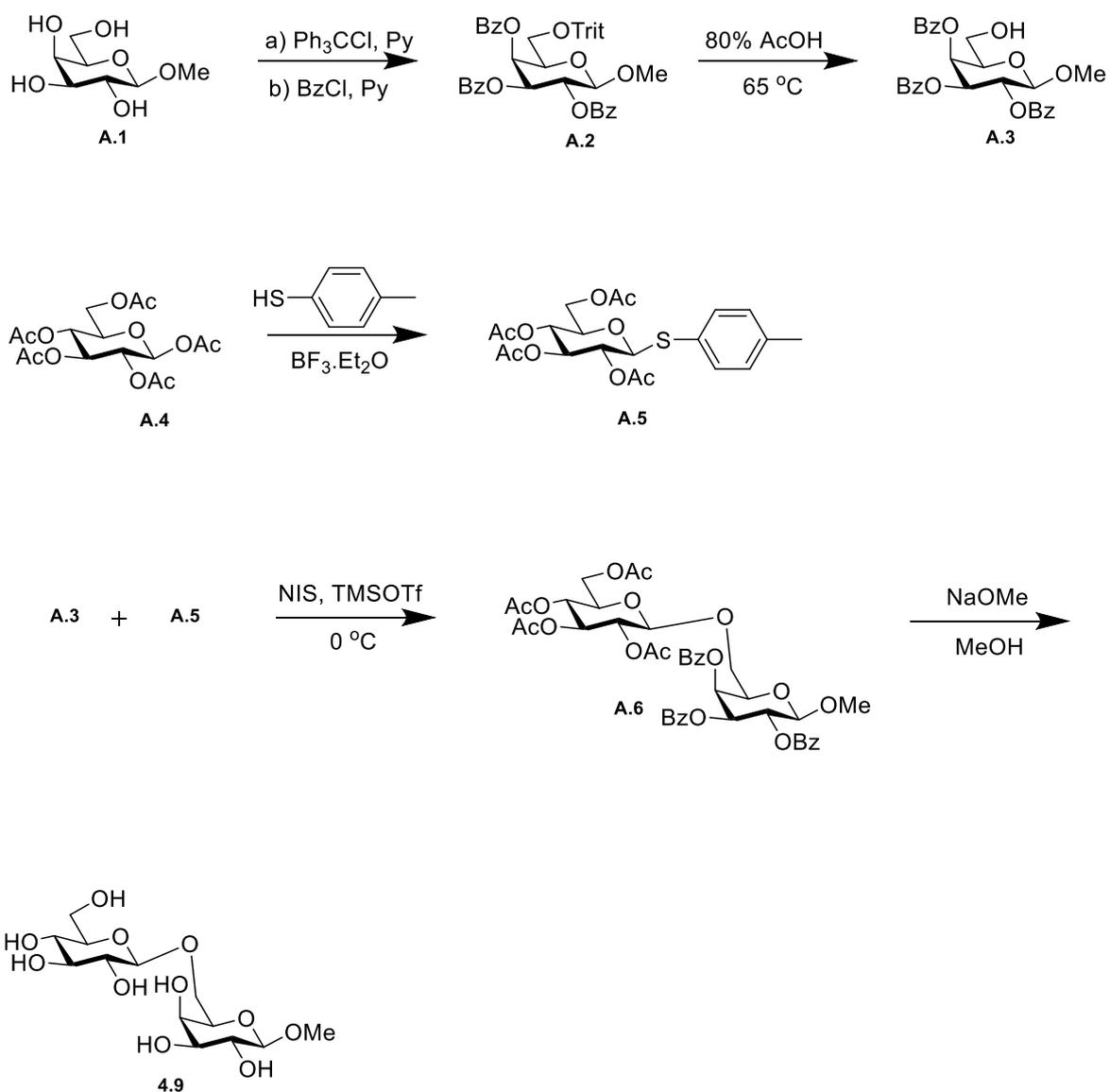
Figure A.24: MALDI-TOF-MS Spectra for Initial Sialyltransferase Activity Screen with CMP-Pse5Ac7Ac and ITag-LacNAc, as described in 3.2.1 and 3.2.3

Table A.2: Secretion signal analysis for putative PseTs

| Protein | Signal peptide (Sec/SPI) likelihood | TAT signal peptide (Tat/SPI) likelihood | Lipoprotein signal peptide (Sec/SPII) likelihood |
|---------|-------------------------------------|---|--|
| KpsS1   | 0.0098                              | 0.0003                                  | 0.0008   |
| Orf14   | 0.0571                              | 0.0453                                  | 0.0183   |
| Lst     | 0.0254                              | 0.0012                                  | 0.003  |
| Gtr16   | 0.0361                              | 0.0007                                  | 0.0449   |
| WfdL    | 0.1531                              | 0.001                                   | 0.0476   |



**Figure A.25: LC-MS data of synthesis of Pse5Ac7Ac 4.X, from CMP-Pse5Ac7Ac 1.10 in a one-pot reaction. Molecular ion peaks [M-H]<sup>-</sup> for 1.10 (m/z 638.0) and 4.X (m/z 616.0).**



**Scheme A.1: Synthesis of 4.9**

## General methods

### Solvents and Starting Materials

Solvents used for flash chromatography purposes were GPR-grade. All commercially-available reagents were used as received and were supplied by Sigma-Aldrich, Fisher Scientific, VWR International, Carbosynth, Sussex Research and TCI.

### LB recipe

NaCl 1 % (w/v), Tryptone 1 % (w/v), Yeast 0.5 % (w/v) in dH<sub>2</sub>O and autoclave.

### LB agar recipe

NaCl 1 % (w/v), Tryptone 1 % (w/v), Yeast 0.5 % (w/v), Agar 1.5 % (w/v) in dH<sub>2</sub>O and autoclave.

### Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE);

5 x SDS reducing sample buffer recipe: SDS 10 % (w/v), Glycerol 20 % (w/v), Bromophenol blue 0.05 % (w/v), Tris-HCl 200 mM pH 6.8,  $\beta$ -mercaptoethanol 10 mM, in dH<sub>2</sub>O.

SDS resolving gel buffer: SDS 0.4 % (w/v), Tris-HCl 1.5 M pH 8.8, in dH<sub>2</sub>O.

SDS stacking gel buffer: SDS 0.4 % (w/v), Tris-HCl 0.5 M pH 6.8, in dH<sub>2</sub>O.

SDS running buffer 4X: Glycine 160 mM, Tris 0.1 M

Staining solution: 0.1 % Coomassie Brilliant Blue R-250, 50 % MeOH, 10 % glacial acetic acid

Destaining solution: 50 % MeOH, 10 % glacial acetic acid

### Protein Purification

All protein purification, with the exception of PseF size-exclusion chromatography (SEC) was performed on a AKTA Start, GE Healthcare Life Technologies. PseF SEC was performed on a AKTA Prime, GE Healthcare Life Technologies.

Analytical TLC was performed on silica gel 60-F<sup>254</sup> with detection by fluorescence and/or charring following immersion in a solution of MeOH, 5% H<sub>2</sub>SO<sub>4</sub>.

### Mass Spectrometry (MS)

Small-molecule HRMS data were obtained at room temperature on a Bruker Daltonics micrOTOF.

### Liquid Chromatography Mass Spectrometry (LC-MS)

LC-MS data presented in 2.3.2, 2.2.3 (PseB inhibition only) and 2.3.6 was carried out using negative ion mode electrospray ionisation (-ESI) on a Bruker HCTultra ETD II system (Bruker Daltonics) MS in The University of York Centre of Excellence in Mass Spectrometry (CoEMS). A Waters C18 column was fitted to a high performance Dionex UltiMate<sup>®</sup> 3000 LC system (ThermoScientific) equipped with an UltiMate<sup>®</sup> 3000 Diode Array Detector. Chromeleon<sup>®</sup> 6.80 SR12 software (ThermoScientific) combined with esquireControl version 6.2, Build 62.24 software (Bruker Daltonics), and Bruker compass Hystar 3.2-SR2, Hystar version 3.2, Build 44 software (Bruker Daltonics). General procedure: Solvent A - Water, Formic Acid 0.1 % (v/v); Solvent B -

Acetonitrile, Formic Acid 0.1 % (v/v). A flow-rate of 300  $\mu\text{L min}^{-1}$ , followed pre-equilibrium of the column in solvent A 95 % for 30 seconds followed by application of a linear gradient; solvent B 70 % to 30 % over 6 minutes. The column was washed in 95 % solvent B for 1 minute before solvent A 95 % was applied for 1 minute to re-equilibrate the column.

All other LC-MS data was carried out using negative ion mode electrospray ionisation (-ESI) on a Bruker HCTultra ETD II system (Bruker Daltonics) MS in The University of York Centre of Excellence in Mass Spectrometry (CoEMS). A Waters C18 column was fitted to an ultra-performance Acquity Class I LC system. General procedure: Solvent A - Water, Formic Acid 0.1 % (v/v); Solvent B - Acetonitrile, Formic Acid 0.1 % (v/v). A flow-rate of 600  $\mu\text{L min}^{-1}$ , followed pre-equilibrium of the column in solvent A 100 % for followed by application of a linear gradient; solvent B 0 % to 40 % over 5 minutes. The column was washed in 95 % solvent B for 1 minute before solvent A 95 % was applied for 4 minute to re-equilibrate the column.

All data was processed using Bruker Daltonik DataAnalysis 4.1 software.

#### Nuclear Magnetic Resonance Spectroscopy

For CMP-Pse5Ac7Ac **1.10**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra NMR data was collected on a Jeol ECS400 NMR Spectrometer or the Bruker AVIIIHD500 FT-NMR Spectrometer made available at The University of York Centre for Magnetic Resonance, using an internal deuterium lock at room temperature. Signals were assigned using additional DEPT135, COSY, HSQC experiments. Chemical shifts are reported in ppm according to the following references:  $\text{D}_2\text{O}$ :  $\delta$ : 4.80

For  $\beta$ -2,3-Pse5Ac7Ac-*p*NP- $\beta$ -D-Gal **3.17** and  $\alpha$ -2,6-Pse5Ac7Ac-*p*NP- $\beta$ -D-Glc **1.12**  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded at 700 MHz and 126 MHz respectively on a Bruker Avance Neo 700 MHz spectrometer operating at 658.78 MHz using an internal deuterium lock at room temperature. Signals were assigned using additional DEPT135, COSY, HSQC experiments. Chemical shifts are reported in ppm according to the following references:  $\text{D}_2\text{O}$ :  $\delta$ : 4.80

The following abbreviations were used to describe signal multiplicities or appearances: s, singlet; d, doublet; t, triplet; dt, doublet of triplets; q, quartet; qd, quartet of doublets; quint, quintet; dq, doublet of quintets; m, multiplet.

## Abbreviations

- $\beta$ -D-Glcp-2,6- $\beta$ -D-Galp-OMe**  $\beta$ -D-glucopyranosyl-2,6-methyl- $\beta$ -D-galactopyranoside
- 6-deoxy-AltdiNAc** 2,4-diacetamido-2,4,6-trideoxy-L-altropyranose
- 6-deoxy-AltNAc4NAcN<sub>3</sub>** 2-acetamido-4-azidoacetamido-2,4,6-trideoxy- $\beta$ -L-altropyranose
- 6-F-Galp** 6-deoxy-6-fluoro-D-Galactose
- Abs** Absorbance
- Acetyl-CoA** Acetyl-Coenzyme A
- AcOH** Acetic acid
- Ada** Adamantanyl
- CAZy** Carbohydrate-Active EnZymes
- CDG** Congenital Disorders of Glycosylation
- CMP-Aci5Ac7Ac** di-N-acetylacinetaminic acid
- CMP-Kdo** Cytidine-5'-monophosphate-2-keto-3-deoxy-*manno*-octonic acid
- CMP-Neu5Ac** Cytidine-5'-monophosphate N-acetylneuraminic acid
- CMP-3F-Neu5Ac** CMP-3-deoxy-3-fluoro-N-acetylneuraminic acid
- CMP-Pse5Hb7Fm** CMP-5-N-[(R)-3-hydroxybutanoyl]-7-N-formyl-3,5,7,9-tetradeoxy-L-glycero-L-manno-2-nonulosonic acid
- CMP-Pse5Ac7Ac** Cytidine-5'-monophosphate-5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno-2-nonulosonic acid
- CoA** Coenzyme-A
- CPS** Capsular polysaccharide
- CTP** Cytidine-5'-triphosphate
- C.V.** Column volumes
- D<sub>2</sub>O** Deuterium oxide
- DCM** Dichloromethane
- diNAcBac** 2,4-diacetamido-2,4,6-trideoxy-D-glucose
- DNA** Deoxyribose nucleic acid
- EIC** Extracted ion chromatograph

**-ESI Negative Ion-Mode** Electrospray Ionisation

**Equiv** Equivalents

**EPS** extracellular polysaccharide

**FucNAc** *N*-acetyl fucosamine

**Galp**  $\beta$ -D-Galactopyranose (Galactose, Galp)

**GalNAcp** 2-Acetamido-2-deoxy- $\beta$ -D-galactopyranose (GalNAc)

**Glc**  $\beta$ -D-Galactopyranose (Glucose, Glcp)

**GH** Glycosylhydrolase

**GlcNAc** *N*-acetylglucosamine

**GRAVY** the GRand AVerage of hydrophathY

**GST** glutathione-S-transferase colicin

**GT** Glycosyltransferase

**HOBt** Hydroxybenzotriazole

**HPLC** High Performance Liquid Chromatography

**Hrs** Hours

**HR-MS** High Resolution-Mass Spectrometry

**IC<sub>50</sub>** 50% inhibitory concentration

**Im9** E9 immunity protein

**IPTG** Isopropyl  $\beta$ -D-1-thiogalactopyranoside

**kb** Kilobases

**kDa** KiloDalton

**Kdo** 3-deoxy-d-manno-oct-2-ulosonic acid

**KDN** 2-keto-3-deoxy-  $\beta$ -D-glycero- $\beta$ -D-galacto-2-nonulosonic acid

**Lac**  $\beta$ -D-galactopyranosyl-1,4-D-glucose (lactose)

**LacNAc** *N*-Acetyl-D-lactosamine

**LacNAc-ITag** imidazolium-tagged *N*-Acetyl-D-lactosamine

**LB** Lysogeny Broth

**LB<sub>amp</sub>** LB supplemented with 100  $\mu$ g mL<sup>-1</sup> ampicillin

**LB<sub>kan</sub>** LB supplemented with 100  $\mu\text{g mL}^{-1}$  kanamycin

**LC-MS** Liquid Chromatography-Mass Spectrometry

**Leg5Ac7Ac** N, N'-diacetyl-legionaminic acid

**L-Glu** L-Glutamate

**LPS** Lipopolysaccharide

**LOS** Lipooligosaccharide

**Maf1** Motility associated factor-1

**ManNAc** N-acetyl-D-mannosamine

**MBP** maltose binding protein

**MeOH** Methanol

**Me- $\beta$ -D-Glcp** methyl- $\beta$ -D-galactopyranoside

**Mins** Minutes

**$\mu\text{L}$**  Microlitre

**$\mu\text{M}$**  Micromolar

**mM** Millimolar

**MWCO** molecular weight cut off

**MS** Mass spectrometry

**MurNAc** N-acetylmuramic acid

**N<sup>10</sup>-fTHF** N<sup>10</sup>-formyltetrahydrofolate

**Neu5Ac** 5-acetamido-3,5-D-glycero-D-galacto-2-nonulosonic acid, N-acetyl-neuraminic acid

**NMR** Nuclear Magnetic Resonance

**NuIO** Nonulosonic Acid

**PCR** Polymerase chain reaction

**PEG** Poly(ethylene glycol)

**PEP** Phosphoenolpyruvate

**PdST** *Pasteurella dagmatis*  $\alpha$ -2,3-SiaT

**PJT-ISH-224ST** *Photobacterium* sp. JT-ISH-224  $\alpha$ -2,3-SiaT

**PIST** *Photobacterium leiognathi*  $\alpha$ -2,6-SiaT

**PmST** *Pasteurella multocida*  $\alpha$ -2,3-/ $\alpha$ -2,6-SiaT

**PLP** Pyridoxal 5'-phosphate

**pNP- $\beta$ -D-Galp** 4-nitrophenyl- $\beta$ -D-galactopyranoside

**pNP- $\beta$ -D-Glcp** 4-nitrophenyl- $\beta$ -D-glucopyranoside

**pMeP- $\beta$ -D-Galp** 4-methoxyphenyl  $\beta$ -D-galactopyranoside

**pNP- $\beta$ -D-Xylp** 4-nitrophenyl- $\beta$ -D-xylopyranoside

**Pse** Pseudaminic Acid (generic term)

**Pse5Ac7Ac** 5,7-diacetamido-3,5,7,9-tetra-deoxy-L-glycero-L-manno-2-nonulosonic acid

**Pse5Ac7Hb** 5-N-acetyl- 7-(3-hydroxybutanoyl)- 3,5,7,9-tetra-deoxy-L-glycero-L-manno-2-nonulosonic acid

**Pse5Ac7R** 5-acetamido-7-R-3,5,7,9-tetra-deoxy-L-glycero-L-manno-2-nonulosonic acid

**Pse5Hb7Fm** 5-N-[(R)-3-hydroxybutanoyl]-7-N-formyl-3,5,7,9-tetra-deoxy-L-glycero-L-manno-2-nonulosonic acid

**R.p.m.** Revolutions per minute

**SDS-PAGE** Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

**SEC** Size-Exclusion Chromatography

**SNAc** N-acetyl-S-acetylcysteamine

**SNFm** N-acetyl-S-formylcysteamine

**TLC** Thin layer chromatography

**Tris** Tris(hydroxymethyl)aminomethane

**UDP** Uridine Diphosphate

**UDP-6-deoxy-AltNAc4NAcCl** UDP-2-acetamido-4-chloroacetamido-2,4,6-trideoxy- $\beta$ -L-Alt

**UDP-diNAcBac** Uridine Diphosphate N, N'-diacetylbacillosamine

**UDP-GlcNAc** Uridine 5'-diphosphate-N-acetyl-D-glucosamine

**UPLC** Ultraperformance Liquid Chromatography

**UV** Ultraviolet

**Xyl** xylose



## References

1. Zunk M, Kiefel MJ. The occurrence and biological significance of the  $\alpha$ -keto-sugars pseudaminic acid and legionaminic acid within pathogenic bacteria. *RSC Adv.* 2014;4(7):3413-3421. doi:10.1039/c3ra44924f
2. Tsvetkov YE, Shashkov AS, Knirel YA, Zaehring U. Synthesis and NMR Spectroscopy of Nine Stereoisomeric 5,7-Diacetamido-3,5,7,9-tetra-deoxy-non-2-ulosonic Acids. *Carbohydr Res.* 2001;335(4):221-243. doi:10.1016/S0008-6215(01)00235-X
3. Lee YJ, Kubota A, Ishiwata A, Ito Y. Synthesis of pseudaminic acid, a unique nonulopyranoside derived from pathogenic bacteria through 6-deoxy-AltdiNAc. *Tetrahedron Lett.* 2011;52(3):418-421. doi:10.1016/j.tetlet.2010.11.078
4. Zunk M, Williams J, Carter J, Kiefel MJ. A new approach towards the synthesis of pseudaminic acid analogues. *Org Biomol Chem.* 2014;12(18):2918-2925. doi:10.1039/c3ob42491j
5. Williams JT, Corcilius L, Kiefel MJ, Payne RJ. Total Synthesis of Native 5, 7-Diacetyl-pseudaminic Acid from N-Acetylneuraminic Acid. *J Org Chem.* 2016;81(6):2607–2611. doi:10.1021/acs.joc.5b02754
6. Liu H, Zhang Y, Wei R, Andolina G, Li X. Total Synthesis of Pseudomonas aeruginosa 1244 Pilin Glycan via de Novo Synthesis of Pseudaminic Acid. *J Am Chem Soc.* 2017;139(38):13420-13428. doi:10.1021/jacs.7b06055
7. Dhakal B, Crich D. Synthesis and Stereocontrolled Equatorially Selective Glycosylation Reactions of a Pseudaminic Acid Donor: Importance of the Side-Chain Conformation and Regioselective Reduction of Azide Protecting Groups. *J Am Chem Soc.* 2018;140(44):15008-15015. doi:10.1021/jacs.8b09654
8. Andolina G, Wei R, Liu H, et al. Metabolic Labeling of Pseudaminic Acid-Containing Glycans on Bacterial Surfaces. *ACS Chem Biol.* 2018;13(10):3030-3037. doi:10.1021/acscchembio.8b00822
9. Wei R, Liu H, Tang AH, Payne RJ, Li X. A Solution to Chemical Pseudaminylation via a Bimodal Glycosyl Donor for Highly Stereocontrolled  $\alpha$ - And  $\beta$ -Glycosylation. *Org Lett.* 2019;21(10):3584-3588. doi:10.1021/acs.orglett.9b00990
10. Biological Roles of Glycans - Essentials of Glycobiology - NCBI Bookshelf.
11. SCHMIDT RR. ChemInform Abstract: New Methods for the Synthesis of Glycosides and Oligosaccharides - Are there Alternatives to the Koenigs-Knorr-Method? *Angew Chemie - Int Ed.* 1986;17(27):212-235. doi:10.1002/chin.198627354
12. R. A. Laine. A Calculation of All Possible Oligosaccharide Isomers Both Branched And Linear Yields  $1.05 \times 10^{12}$  Structures For A Reducing Hexasaccharide. *Glycobiology.* 1994;4(6):759-767.
13. Hurlley S, Service R, Szuromi P. Cinderella's coach is ready. *Science (80- ).* 2001;291(5512):2337. doi:10.1126/science.291.5512.2337
14. Marth JD. A unified vision of the building blocks of life. *Nat Cell Biol.* 2008;10(9):1015-1016. doi:10.1038/ncb0908-1015
15. Varki A. Evolutionary forces shaping the Golgi glycosylation machinery: Why cell

- surface glycans are universal to living cells. *Cold Spring Harb Perspect Biol.* 2011;3(6):1-14. doi:10.1101/cshperspect.a005462
16. Varki A. Biological roles of oligosaccharides : all of the theories are correct. *Glycobiology.* 1993;3(2):97-130. doi:10.1093/glycob/3.2.97
  17. Varki A. Biological roles of glycans. *Glycobiology.* 2017;27(1):3-49. doi:10.1093/glycob/cww086
  18. Szymanski CM, Ruijin Y, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol.* 1999;32(5):1022-1030. doi:10.1046/j.1365-2958.1999.01415.x
  19. Imperiali B. Bacterial carbohydrate diversity — a Brave New World. *Curr Opin Chem Biol.* 2019;53:1-8. doi:10.1016/j.cbpa.2019.04.026
  20. Stephenson HN, Mills DC, Jones H, et al. Pseudaminic acid on *Campylobacter jejuni* flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. *J Infect Dis.* 2014;210(9):1487-1498. doi:10.1093/infdis/jiu287
  21. Lewis AL, Desa N, Hansen EE, et al. Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proc Natl Acad Sci U S A.* 2009;106(32):13552-13557. doi:10.1073/pnas.0902431106 [pii]
  22. Lombard J. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biol Direct.* 2016;11(1). doi:10.1186/s13062-016-0137-2
  23. Sinnott ML. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chem Rev.* 1990;90(7):1171-1202. doi:10.1021/cr00105a006
  24. Gloster TM. Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol.* 2014;28:131-141. doi:10.1016/j.sbi.2014.08.012
  25. Ardèvol A, Rovira C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J Am Chem Soc.* 2015;137(24):7528-7547. doi:10.1021/jacs.5b01156
  26. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 2008;77:521-555. doi:10.1146/annurev.biochem.76.061005.092322
  27. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2013;42(D1):490-495. doi:10.1093/nar/gkt1178
  28. Schmid J, Heider D, Wendel NJ, Sperl N, Sieber V. Bacterial glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front Microbiol.* 2016;7(FEB):1-7. doi:10.3389/fmicb.2016.00182
  29. Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. Structures and mechanisms of glycosyltransferases. *Glycobiology.* 2006;16(2):29-37. doi:10.1093/glycob/cwj016
  30. Willis LM, Whitfield C. KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. *Proc*

- Natl Acad Sci.* 2013;110(51):20753-20758. doi:10.1073/pnas.1312637110
31. Ovchinnikova OG, Mallette E, Koizumi A, Lowary TL, Kimber MS, Whitfield C. Bacterial  $\beta$ -Kdo glycosyltransferases represent a new glycosyltransferase family (GT99). *Proc Natl Acad Sci U S A.* 2016;113(22):E3120-9. doi:10.1073/pnas.1603146113
  32. Lin LYC, Rakic B, Chiu CPC, et al. Structure and mechanism of the lipooligosaccharide sialyltransferase from *Neisseria meningitidis*. *J Biol Chem.* 2011;286(43):37237-37248. doi:10.1074/jbc.M111.249920
  33. Ovchinnikova OG, Doyle L, Huang BS, Kimber MS, Lowary TL, Whitfield C. Biochemical characterization of bifunctional 3-deoxy- $\beta$ -D-manno-oct-2-ulosonic acid ( $\beta$ -Kdo) transferase KpsC from *Escherichia coli* involved in capsule biosynthesis. *J Biol Chem.* 2016;291(41):21519-21530. doi:10.1074/jbc.M116.751115
  34. Brown C, Leijon F, Bulone V. Radiometric and spectrophotometric in vitro assays of glycosyltransferases involved in plant cell wall carbohydrate biosynthesis. *Nat Protoc.* 2012;7(9):1634-1650. doi:10.1038/nprot.2012.089
  35. Glycan-Recognizing Probes as Tools - Essentials of Glycobiology - NCBI Bookshelf.
  36. Schmölder K, Ribitsch D, Czabany T, et al. Characterization of a multifunctional  $\alpha$ -2,3-sialyltransferase from *Pasteurella dagmatis*. 2013;23(11):1293-1304. doi:10.1093/glycob/cwt066
  37. Yu H, Takeuchi M, LeBarron J, et al. Notch-modifying xylosyltransferase structures support an SNI-like retaining mechanism. *Nat Chem Biol.* 2015;(September). doi:10.1038/nchembio.1927
  38. Chen X, Fisher AJ. Crystal Structures of *Pasteurella multocida* Sialyltransferase Complexes with Acceptor and Donor Analogues Reveal Substrate Binding Sites and Catalytic. 2007:6288-6298. doi:10.1021/bi700346w
  39. Takakura Y, Tsukamoto H, Yamamoto T. Molecular Cloning , Expression and Properties of an  $\alpha$  /  $\beta$  -Galactoside  $\alpha$  2 , 3-Sialyltransferase from *Vibrio* sp . JT-FAJ-16. 2018;412(May):403-412. doi:10.1093/jb/mvm147
  40. Tsukamoto H, Takakura Y, Mine T, Yamamoto T. *Photobacterium* sp . JT-ISH-224 Produces Two Sialyltransferases ,  $\alpha$  -/ $\beta$ -Galactoside  $\alpha$  2,3-Sialyltransferase and  $\beta$ -Galactoside. 2008;197:187-197. doi:10.1093/jb/mvm208
  41. Gilbert M, Cunningham A, Watson DC, Martin A, Richards JC, Wakarchuk WW. Characterization of a recombinant *Neisseria meningitidis*  $\alpha$ -2 ,3-sialyltransferase and its acceptor specificity. 1997;194:187-194. doi:10.1111/j.1432-1033.1997.t01-1-00187.x
  42. Watson DC, Wakarchuk WW, Gervais C, et al. Preparation of legionaminic acid analogs of sialo-glycoconjugates by means of mammalian sialyltransferases. 2015:729-734. doi:10.1007/s10719-015-9624-4
  43. Yu H, Chokhawala H, Karpel R, et al. A Multifunctional *Pasteurella multocida* Sialyltransferase : A Powerful Tool for the Synthesis of Sialoside Libraries. 2005;127:17618-17619. doi:10.1021/ja0561690
  44. Koshland DE. STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC

- REACTIONS. 1953;(February):416-436. doi:10.1111/j.1469-185X.1953.tb01386.x
45. Ardèvol A, Iglesias-fernández J. The reaction mechanism of retaining glycosyltransferases. 2016. doi:10.1042/BST20150177
  46. Persson K, Ly HD, Dieckelmann M, Wakarchuk WW, Withers SG, Strynadka NCJ. Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat Struct Biol.* 2001;8(2):166-175. doi:10.1038/84168
  47. Schuman B, Evans S V., Fyles TM. Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *PLoS One.* 2013;8(8). doi:10.1371/journal.pone.0071077
  48. Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG. Mechanistic evidence for a front-side, S<sub>N</sub>i-type reaction in a retaining glycosyltransferase. *Nat Chem Biol.* 2011;7(9):631-638. doi:10.1038/nchembio.628
  49. Withers SG, Aebersold R. Approaches to labeling and identification of active site residues in glycosidases. *Protein Sci.* 1995;4:361-372. doi:10.1002/pro.5560040302
  50. Jamaluddin H, Tumbale P, Withers SG, Acharya KR, Brew K. Conformational Changes Induced by Binding UDP-2F-galactose to  $\alpha$ -1,3 Galactosyltransferase-Implications for Catalysis. *J Mol Biol.* 2007;369:1270-2181. doi:10.1016/j.jmb.2007.04.012
  51. Lee SS, Hong SY, Errey JC, Izumi A, Davies GJ, Davis BG. Mechanistic evidence for a front-side, S<sub>N</sub>i-type reaction in a retaining glycosyltransferase. *Nat Chem Biol.* 2011;7(9):631-638. doi:10.1038/nchembio.628
  52. Lairson LL, Chiu CPC, Ly HD, et al. Intermediate trapping on a mutant retaining  $\alpha$ -galactosyltransferase identifies an unexpected aspartate residue. *J Biol Chem.* 2004;279(27):28339-28344. doi:10.1074/jbc.M400451200
  53. McArthur JB, Chen X. Glycosyltransferase engineering for carbohydrate synthesis. 2018;44(1):129-142. doi:10.1042/BST20150200.Glycosyltransferase
  54. Schmolzer K, Czabany T, Luley-goedl C, et al. Complete switch from  $\alpha$ -2,3- to  $\alpha$ -2,6-regioselectivity in *Pasteurella dagmatis* b-D-galactoside sialyltransferase by active-site redesign. *Chem Commun.* 2015;51:3083-3086. doi:10.1039/C4CC09772F
  55. Angata T, Varki A. Chemical diversity in the sialic acids and related  $\alpha$ -keto acids: An evolutionary perspective. *Chem Rev.* 2002;102(2):439-469. doi:10.1021/cr000407m
  56. Samuel BS, Hansen EE, Manchester JK, et al. Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A.* 2007;104(25):10643-10648. doi:10.1073/pnas.0704189104
  57. Li Y, Chen X. Sialic acid metabolism and sialyltransferases: natural functions and applications. 2012;94(4):887-905. doi:10.1007/s00253-012-4040-1.Sialic
  58. Tanner ME. The enzymes of sialic acid biosynthesis. *Bioorg Chem.* 2005;33(3):216-228. doi:10.1016/j.bioorg.2005.01.005
  59. Varki A. Uniquely human evolution of sialic acid genetics and biology. *Proc Natl*

- Acad Sci.* 2010;107(Supplement\_2):8939-8946. doi:10.1073/pnas.0914634107
60. Vimr ER. Unified theory of bacterial sialometabolism: how and why bacteria metabolize host sialic acids. *ISRN Microbiol.* 2013;2013:816713. doi:10.1155/2013/816713
  61. Schwarzkopf M, Knobloch K-P, Rohde E, et al. Sialylation is essential for early development in mice. *Proc Natl Acad Sci U S A.* 2002;99(8):5267-5270. doi:10.1073/pnas.072066199
  62. Traving C, Schauer R. Structure, function and metabolism of sialic acids. *Cell Mol Life Sci.* 1998;54(12):1330-1349. doi:10.1007/s000180050258
  63. Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM. Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev.* 2004;68(1):132-153. doi:10.1128/MMBR.68.1.132
  64. Varki A. Diversity in the sialic acids. *Glycobiology.* 1992;2(2):169. doi:10.1093/glycob/2.2.169
  65. Severi E, Randle G, Kivlin P, et al. Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol Microbiol.* 2005;58(4):1173-1185. doi:10.1111/j.1365-2958.2005.04901.x
  66. Allen S, Zaleski A, Johnston JW, Gibson BW, Apicella MA. Novel sialic acid transporter of *Haemophilus influenzae*. *Infect Immun.* 2005;73(9):5291-5300. doi:10.1128/IAI.73.9.5291-5300.2005
  67. Steenbergen SM, Lee YC, Vann WF, Vionnet J, Wright LF, Vimr ER. Separate pathways for O acetylation of polymeric and monomeric sialic acids and identification of sialyl O-acetyl esterase in *Escherichia coli* K1. *J Bacteriol.* 2006;188(17):6195-6206. doi:10.1128/JB.00466-06
  68. Thomas GH. Sialic acid acquisition in bacteria-one substrate, many transporters. *Biochem Soc Trans.* 2016;44(3):760-765. doi:10.1042/bst20160056
  69. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens. *Microbiology.* 2007;153(9):2817-2822. doi:10.1099/mic.0.2007/009480-0
  70. Vimr E, Lichtensteiger C. To sialylate, or not to sialylate: that is the question. *Trends Microbiol.* 2002;10(6):254-257. <http://www.ncbi.nlm.nih.gov/pubmed/12088651>.
  71. Ng MK, Ferreyra AJ, Higginbottom KS, et al. Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature.* 2013;504(4):996-999.
  72. Yu H, Chen X. Carbohydrate post-glycosylational modifications. *Org Biomol Chem.* 2007;5(6):865-872. doi:10.1039/b700034k
  73. Harvey HA, Swords WE, Apicella MA. The mimicry of human glycolipids and glycosphingolipids by the lipooligosaccharides of pathogenic *Neisseria* and *Haemophilus*. *J Autoimmun.* 2001;16(3):257-262. doi:10.1006/jaut.2000.0477
  74. Nizet V, Uchiyama S, Lewis AL, Varki A, Carlin AF, Chang Y-C. Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood.* 2009;113(14):3333-3336.

doi:10.1182/blood-2008-11-187302

75. Fong JJ, Tsai C-M, Saha S, Nizet V, Varki A, Bui JD. Siglec-7 engagement by GBS  $\beta$ -protein suppresses pyroptotic cell death of natural killer cells. *Proc Natl Acad Sci*. 2018;115(41):10410-10415. doi:10.1073/pnas.1804108115
76. Schauer R. Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol*. 2009;19(5):507-514. doi:10.1016/j.sbi.2009.06.003
77. Vimr E, Lichtensteiger C, Steenbergen S. Sialic acid metabolism's dual function in *Haemophilus influenzae*. *Mol Microbiol*. 2000;36(5):1113-1123. doi:10.1046/j.1365-2958.2000.01925.x
78. Kakuta Y, Okino N, Kajiwara H, et al. Crystal structure of *Vibrionaceae* Photobacterium sp. JT-ISH-224  $\alpha$  2, 6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition. 2018;18(1):66-73. doi:10.1093/glycob/cwm119
79. Chiu CPC, Watts AG, Lairson LL, et al. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. *Nat Struct Mol Biol*. 2004;11(2):163-170. doi:10.1038/nsmb720
80. Yamamoto T, Hamada Y, Ichikawa M, et al. A  $\beta$ -galactoside  $\alpha$  2, 6-sialyltransferase produced by a marine bacterium, *Photobacterium leiognathi* JT-SHIZ-145, is active at pH 8. 2018;17(11):1167-1174. doi:10.1093/glycob/cwm086
81. Knirel, Y. A., Vinogradov, E. V., L'vov, V. L., Kocharova, N. A., Shashkov AS, Dmitriev, B. A., Kochetkov NK. Sialic acids of a new type from the lipopolysaccharides of *Pseudomonasaeruginosa* and *Shigella boydii*. *Elsevier Sci Publ BV*. 1984;133:6-8. doi:10.1016/0008-6215(84)85213-1
82. Kandiba L, Eichler J. Analysis of putative nonulosonic acid biosynthesis pathways in Archaea reveals a complex evolutionary history. *FEMS Microbiol Lett*. 2013;345(2):110-120. doi:10.1111/1574-6968.12193
83. Miller, Terry L; Wolin, M. J., Conway de Macario, Everly, Macario JL. Isolation of *Methanobrevibacter smithii* from Human Feces. *Appl Environ Microbiol*. 1982;43(1):227-232.
84. Castric P, Cassels FJ, Carlson RW. Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. *J Biol Chem*. 2001;276(28):26479-26485. doi:10.1074/jbc.M102685200
85. Thibault P, Logan SM, Kelly JF, et al. Identification of the Carbohydrate Moieties and Glycosylation Motifs in *Campylobacter jejuni* Flagellin. *J Biol Chem*. 2001;276(37):34862-34870. doi:10.1074/jbc.M104529200
86. Kiss E, Kereszt a, Barta F, et al. The rkp-3 gene region of *Sinorhizobium meliloti* Rm41 contains strain-specific genes that determine K antigen structure. *Mol Plant Microbe Interact*. 2001;14(12):1395-1403. doi:10.1094/MPMI.2001.14.12.1395
87. DiGiandomenico A, Matewish MJ, Bisailon A, Stehle JR, Lam JS, Castric P. Glycosylation substrate specificity of *Pseudomonas aeruginosa* 1244 pilin: glycan substrate specificity. *Mol Microbiol*. 2002;46(2):519-530. doi:10.1074/jbc.M510975200
88. Russo TA, Luke NR, Beanan JM, et al. The K1 capsular polysaccharide of *Acinetobacter baumannii* strain 307-0294 is a major virulence factor. *Infect Immun*.

2010;78(9):3993-4000. doi:10.1128/IAI.00366-10

89. Parker JL, Day-Williams MJ, Tomas JM, Stafford GP, Shaw JG. Identification of a putative glycosyltransferase responsible for the transfer of pseudaminic acid onto the polar flagellin of *Aeromonas caviae* Sch3N. *Microbiologyopen*. 2012;1(2):149-160. doi:10.1002/mbo3.19
90. Tomás JM. The Main *Aeromonas* Pathogenic Factors. *ISRN Microbiol*. 2012;2012:1-22. doi:10.5402/2012/256261
91. Schirm M, Soo EC, Aubry AJ, Austin J, Thibault P, Logan SM. Structural, genetic and functional characterization of the flagellin glycosylation process in. 2003;48(6). doi:10.1046/j.1365-2958.2003.03527.x
92. Guerry P, Alm RA, Power ME, Logan SM, Trust TJ. Role of two flagellin genes in *Campylobacter* motility. *J Bacteriol*. 1991;173(15):4757-4764. doi:10.1128/jb.173.15.4757-4764.1991
93. Thibault P, Logan SM, Kelly JF, et al. Identification of the Carbohydrate Moieties and Glycosylation Motifs in *Campylobacter jejuni* Flagellin. *J Biol Chem*. 2001;276(37):34862-34870. doi:10.1074/jbc.M104529200
94. Ewing CP, Andreishcheva E, Guerry P. Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. *J Bacteriol*. 2009;191(22):7086-7093. doi:10.1128/JB.00378-09
95. Goon S, Kelly JF, Logan SM, Ewing CP, Guerry P. Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Mol Microbiol*. 2003;50(2):659-671. doi:10.1046/j.1365-2958.2003.03725.x
96. Guerry P. *Campylobacter* flagella: not just for motility. *Trends Microbiol*. 2007;15(10):456-461. doi:10.1016/j.tim.2007.09.006
97. Ottemann KM, Lowenthal AC. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect Immun*. 2002;70(4):1984-1990. doi:10.1128/IAI.70.4.1984-1990.2002
98. Knirel YA, Sergei SD, Perepelov A V. Higher aldulosonic acids: components of bacterial glycans. *Mendeleev Commun*. 2011;21(3):173-182. doi:10.1016/j.mencom.2011.
99. Schoenhofen IC, McNally DJ, Brisson JR, Logan SM. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: Synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. *Glycobiology*. 2006;16(9). doi:10.1093/glycob/cwl010
100. Margaret I, Crespo-Rivas JC, Acosta-Jurado S, et al. *Sinorhizobium fredii* HH103 rkp-3 genes are required for K-antigen polysaccharide biosynthesis, affect lipopolysaccharide structure and are essential for infection of legumes forming determinate nodules. *Mol Plant-Microbe Interact*. 2012;25(6):825-838. doi:10.1094/MPMI-10-11-0262
101. Datta De D, Roychoudhury S. To be or not to be: The host genetic factor and beyond in *Helicobacter pylori* mediated gastro-duodenal diseases. *World J Gastroenterol*. 2015;21(10):2883-2895. doi:10.3748/wjg.v21.i10.2883
102. Hofreuter D. Defining the metabolic requirements for the growth and colonization capacity of *Campylobacter jejuni*. *Front Cell Infect Microbiol*.

2014;4(September):137. doi:10.3389/fcimb.2014.00137

103. McNally DJ, Schoenhofen IC, Houliston RS, et al. CMP-Pseudaminic Acid is a Natural Potent Inhibitor of PseB, the First Enzyme of the Pseudaminic Acid Pathway in *Campylobacter jejuni* and *Helicobacter pylori*. *ChemMedChem*. 2008;3(1):55-59. doi:10.1002/cmdc.200700170
104. Tabei SMB, Hitchen PG, Day-Williams MJ, et al. An *Aeromonas caviae* Genomic island is required for both O-antigen lipopolysaccharide biosynthesis and flagellin glycosylation. *J Bacteriol*. 2009;191(8):2851-2863. doi:10.1128/JB.01406-08
105. Li Z, Hwang S, Ericson J, Bowler K, Bar-Peled M. Pen and Pal are nucleotide-sugar dehydratases that convert UDP-GlcNAc to UDP-6-deoxy-D-GlcNAc-5,6-ene and then to UDP-4-keto-6-deoxy-L-AltNAc for CMP-pseudaminic acid synthesis in *Bacillus thuringiensis*. *J Biol Chem*. 2015;290(2):691-704. doi:10.1074/jbc.M114.612747
106. Gil-serrano AM, Rodriguez-Carvajal MA, Tejero-Mateo P, et al. Structural determination of a 5-acetamido-3,5,7,9-tetra-deoxy-7-(3-hydroxybutyramido)-L-glycero-L-manno-nonulosonic acid-containing homopolysaccharide isolated from *Sinorhizobium fredii* HH103. *Biochem J*. 1999;342:527-535.
107. Reimer JM, Harb I, Ovchinnikova OG, Jiang J, Whitfield C, Schmeing TM. Structural Insight into a Novel Formyltransferase and Evolution to a Nonribosomal Peptide Synthetase Tailoring Domain. *ACS Chem Biol*. 2018;13(11):3161-3172. doi:10.1021/acscchembio.8b00739
108. McNally DJ, Hui JPM, Aubry AJ, et al. Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *J Biol Chem*. 2006;281(27):18489-18498. doi:10.1074/jbc.M603777200
109. Rangarajan ES, Proteau A, Cui Q, et al. Structural and Functional Analysis of *Campylobacter jejuni* PseG. 2009;284(31):20989-21000. doi:10.1074/jbc.M109.012351
110. Ménard R, Schoenhofen IC, Tao L, et al. Small-molecule inhibitors of the pseudaminic acid biosynthetic pathway: targeting motility as a key bacterial virulence factor. *Antimicrob Agents Chemother*. 2014;58(12):7430-7440. doi:10.1128/AAC.03858-14
111. Shirai R, Ogura H. Improved syntheses of two 3-deoxyald-2-ulosonic acids (KDN, KDO) by condensation of oxalacetic acid with aldoses followed by Ni<sup>2+</sup> catalyzed decarboxylation<sup>1</sup>). *Tetrahedron Lett*. 1989;30(17):2263-2264. doi:10.1016/S0040-4039(00)99665-5
112. Sharma M, Bernacki RJ, Paul B, Walter K. Fluorinated carbohydrates as potential plasma membrane modifiers. Synthesis of 40 and 60 fluoro derivatives of 2-acetamido-2-deoxy-d-hexopyranosides. *Carbohydr Res*. 1990;198:205-221.
113. Zunk M, Williams J, Carter J, Kiefel MJ. A new approach towards the synthesis of pseudaminic acid analogues. *Org Biomol Chem*. 2014;12(18):2918-2925. doi:10.1039/c3ob42491j
114. Marra A, Sinaÿ P. Stereoselective synthesis of 2-thioglycosides of N-acetylneuraminic acid. *Carbohydr Res*. 1989;187(1):35-42. doi:10.1016/0008-6215(89)80054-0
115. Crich D, Li W.  $\alpha$ -selective sialylations at -78°C in nitrile solvents with a 1-

- adamantanyl thiosialoside. *J Org Chem.* 2007;72(20):7794-7797.  
doi:10.1021/jo7012912
116. Liu F, Aubry AJ, Schoenhofen IC, Logan SM, Tanner ME. The engineering of bacteria. Bearing azido-pseudaminic acid-modified flagella. *ChemBioChem.* 2009;10(8):1317-1320. doi:10.1002/cbic.200900018
  117. Canals R, Ramirez S, Vilches S, et al. Polar flagellum biogenesis in *Aeromonas hydrophila*. *J Bacteriol.* 2006;188(2):542-555. doi:10.1128/JB.188.2.542-555.2006
  118. Xu G, Ryan C, Kiefel MJ, Wilson JC, Taylor GL. Structural Studies on the *Pseudomonas aeruginosa* Sialidase-Like Enzyme PA2794 Suggest Substrate and Mechanistic Variations. *J Mol Biol.* 2009;386(3):828-840.  
doi:10.1016/j.jmb.2008.12.084
  119. Soong G, Muir A, Gomez MI, et al. Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. *J Clin Invest.* 2006;116(8):2297-2305. doi:10.1172/JCI27920
  120. Knirel YA, Vinogradov E V., Shashkov AS, et al. Somatic Antigens of *Pseudomonas aeruginosa*. 1986;651:129-138. doi:10.1111/j.1432-1033.1987.tb10913.x
  121. Comer JE, Marshall MA, Blanch VJ, Deal CD, Castric P. Identification of the *Pseudomonas aeruginosa* 1244 pilin glycosylation site. *Infect Immun.* 2002;70(6):2837-2845. doi:10.1128/iai.70.6.2837-2845.2002
  122. McNally DJ, Schoenhofen IC, Houlston RS, et al. CMP-pseudaminic acid is a natural potent inhibitor of PseB, the first enzyme of the pseudaminic acid pathway in *Campylobacter jejuni* and *Helicobacter pylori*. *ChemMedChem.* 2008;3:55-59.  
doi:10.1002/cmdc.200700170
  123. Anonsen JH, Vik Å, Egge-Jacobsen W, Koomey M. An extended spectrum of target proteins and modification sites in the general O-linked protein glycosylation system in *Neisseria gonorrhoeae*. *J Proteome Res.* 2012;11(12):5781-5793.  
doi:10.1021/pr300584x
  124. Scott NE, Parker BL, Connolly AM, et al. Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of *Campylobacter*. *Mol Cell Proteomics.* 2010;10(2):M000031-MCP201. doi:10.1074/mcp.m000031-mcp201
  125. Shashkov AS, Liu B, Kenyon JJ, et al. Structures of the K35 and K15 capsular polysaccharides of *Acinetobacter baumannii* LUH5535 and LUH5554 containing amino and diamino uronic acids. *Carbohydr Res.* 2017;448:28-34.  
doi:10.1016/j.carres.2017.05.017
  126. Heyes DJ, Levy C, Lafite P, et al. Structure-based mechanism of CMP-2-keto-3-deoxymanno-octulonic acid synthetase: Convergent evolution of a sugar-activating enzyme with DNA/RNA polymerases. *J Biol Chem.* 2009;284(51):35514-35523.  
doi:10.1074/jbc.M109.056630
  127. Horsfall LE, Nelson A, Berry A. Identification and characterization of important residues in the catalytic mechanism of CMP-Neu5Ac synthetase from *Neisseria meningitidis*. *FEBS J.* 2010;277(13):2779-2790. doi:10.1111/j.1742-

4658.2010.07696.x

128. Knirel Y a, Rietschel ET, Marre R, Zähringer U. The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. *Eur J Biochem.* 1994;221:239-245.
129. Song WS, Nam MS, Namgung B, Yoon S II. Structural analysis of PseH, the *Campylobacter jejuni* N-acetyltransferase involved in bacterial O-linked glycosylation. *Biochem Biophys Res Commun.* 2015;458(4):843-848. doi:10.1016/j.bbrc.2015.02.041
130. Ud-Din AI, Liu YC, Roujeinikova A. Crystal structure of *Helicobacter pylori* pseudaminic acid biosynthesis N-acetyltransferase PseH: implications for substrate specificity and catalysis. *PLoS One.* 2015;10(3):e0115634.
131. Ishiyama N, Creuzenet C, Miller WL, et al. Structural studies of FlaA1 from *Helicobacter pylori* reveal the mechanism for inverting 4,6-dehydratase activity. *J Biol Chem.* 2006;281(34):24489-24495. doi:10.1074/jbc.M602393200
132. Umme S, Wahid H. Structural and functional characterization of the *Helicobacter pylori* cytidine 5' -monophosphate- pseudaminic acid synthase PseF : molecular insight into substrate recognition and catalysis mechanism. 2017:79-88.
133. Kyte J, Doolittle RF, Diego S, Jolla L. Kyte1982.Pdf. 1982:105-132.
134. Wei Y, Thompson J, Floudas CA. CONCORD: A consensus method for protein secondary structure prediction via mixed integer linear optimization. *Proc R Soc A Math Phys Eng Sci.* 2012;468(2139):831-850. doi:10.1098/rspa.2011.0514
135. Yu NY, Wagner JR, Laird MR, et al. PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics.* 2010;26(13):1608-1615. doi:10.1093/bioinformatics/btq249
136. Mosimann SC, Gilbert M, Dombrowski D, To R, Wakarchuk W, Strynadka NCJ. Structure of a Sialic Acid-activating Synthetase, CMP-acylneuraminate Synthetase in the Presence and Absence of CDP. *J Biol Chem.* 2001;276(11):8190-8196. doi:10.1074/jbc.M007235200
137. Mosimann SC, Gilbert M, Dombrowski D, To R, Wakarchuk W, Strynadka NCJ. Structure of a Sialic Acid-activating Synthetase , CMP- acylneuraminate Synthetase in the Presence and Absence of CDP \*. 2001;276(11):8190-8196. doi:10.1074/jbc.M007235200
138. Ambrose MG, Freese SJ, Reinhold MS, Warner TG, Vann WF. <sup>13</sup>C NMR investigation of the anomeric specificity of CMP-N-acetylneuraminic acid synthetase from *Escherichia coli*. *Biochemistry.* 1992;31(3):775-780. <http://www.ncbi.nlm.nih.gov/pubmed/1731934>.
139. Mu AK, Tiralongo J, Krapp S, et al. Structure and function of vertebrate CMP – sialic acid synthetases. *Glycobiology.* 2004;14(10):43-51. doi:10.1093/glycob/cwh113
140. Tsvetkov YE, Shashkov AS, Knirel YA, Za U. Synthesis and NMR spectroscopy of nine stereoisomeric. 2001;335:221-243.
141. Boons G, Demchenko A V. Recent Advances in O -Sialylation. *Chem Rev.* 2000;100:4539-4565. doi:10.1021/cr990313g

142. Boons G-J. Recent developments in chemical oligosaccharide synthesis. *Contemp Org Synth.* 2004;3(3):173. doi:10.1039/co9960300173
143. Mansuelle P, Yamakawa N, Qian X, Vincentelli R, Bourne Y. Glycosylate and move! The glycosyltransferase Maf is involved in bacterial flagella formation. 2018;20:228-240. doi:10.1111/1462-2920.13975
144. Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K6 capsular polysaccharide from *Acinetobacter baumannii* isolate RBH4. *Carbohydr Res.* 2015;409:30-35. doi:10.1016/j.carres.2015.03.016
145. Yu H, Chokhawala H, Karpel R, et al. A Multifunctional *Pasteurella multocida* Sialyltransferase : A Powerful Tool for the Synthesis of Sialoside Libraries. 2005:17618-17619. doi:10.1021/ja0561690
146. Mine T, Katayama S, Kajiwara H, et al. An  $\alpha$  2,6-sialyltransferase cloned from *Photobacterium leiognathi* strain JT-SHIZ-119 shows both sialyltransferase and neuraminidase activity. *Glycobiology.* 2010;20(2):158-165. doi:10.1093/glycob/cwp157
147. Kakuta Y, Okino N, Kajiwara H, et al. Crystal structure of vibrionaceae *Photobacterium* sp. JT-ISH-224  $\alpha$ 2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: Catalytic mechanism and substrate recognition. *Glycobiology.* 2008;18(1):66-73. doi:10.1093/glycob/cwm119
148. Galan MC, Tran AT, Bernard C. Ionic-liquid-based catch and release mass spectroscopy tags for enzyme monitoring. *Chem Commun.* 2010;46(47):8968-8970. doi:10.1039/c0cc04224b
149. Tsukamoto H, Takakura Y, Mine T, Yamamoto T. *Photobacterium* sp . JT-ISH-224 Produces Two Sialyltransferases , a - / b -Galactoside a 2 , 3-Sialyltransferase and b -Galactoside. 2008;197:187-197. doi:10.1093/jb/mvm208
150. Mine T, Katayama S, Kajiwara H, et al. An  $\alpha$  2 , 6-sialyltransferase cloned from *Photobacterium leiognathi* strain JT-SHIZ-119 shows both sialyltransferase and neuraminidase activity. 2018;20(2):158-165. doi:10.1093/glycob/cwp157
151. Schmo K, Czabany T, Luley-goedl C, et al. sialyltransferase by active-site redesign †. 2015;14:3083-3086. doi:10.1039/C4CC09772F
152. Sittel I, Galan MC. Chemo-enzymatic synthesis of imidazolium-tagged sialyllactosamine probes. *Bioorganic Med Chem Lett.* 2015;25(19):4329-4332. doi:10.1016/j.bmcl.2015.07.049
153. Xiao X, Liu S, Liu X, Jiang S. Application of ionic liquid in analytical chemistry. *Fenxi Huaxue.* 2005;33(4):569-574. doi:10.1016/j.trac.2004.09.005
154. Toukach P V. Bacterial carbohydrate structure database 3: Principles and realization. *J Chem Inf Model.* 2011;51(1):159-170. doi:10.1021/ci100150d
155. Tulsakaya EM, Streshinskaya GM, Shashkov AS, et al. Novel teichulosonic acid from cell walls of some representatives of the genus *Kribbella*. *Carbohydr Res.* 2011;346(13):2045-2051. doi:10.1016/j.carres.2011.06.003
156. Lee I-M, Tu I-F, Yang F-L, et al. Structural basis for fragmenting the exopolysaccharide of *Acinetobacter baumannii* by bacteriophage  $\Phi$ AB6 tailspike protein. *Sci Rep.* 2017;7(September 2016):42711. doi:10.1038/srep42711

157. Perepelov A V, Wang M, Guo X, et al. Structure and genetics of the O-antigen of *Enterobacter cloacae* G3054 containing di-N-acetylpsseudaminic acid. *Carbohydr Res.* 2015;407:59-62.
158. Han R, Perepelov A V., Wang Y, et al. Structural and genetic characterization of the O-antigen of *Enterobacter cloacae* C5529 related to the O-antigen of *E. cloacae* G3054. *Carbohydr Res.* 2017;443-444:49-52. doi:10.1016/j.carres.2017.02.006
159. Knirel Y a, Bystrova O V, Kocharova N a, Zähringer U, Pier GB. Conserved and variable structural features in the lipopolysaccharide of *Pseudomonas aeruginosa*. *J Endotoxin Res.* 2006;12(6):324-336. doi:10.1179/096805106X118906
160. Kenyon JJ, Hall RM. Variation in the complex carbohydrate biosynthesis loci of *Acinetobacter baumannii* genomes. *PLoS One.* 2013;8(4):e62160. doi:10.1371/journal.pone.0062160
161. Arbatsky NP, Shneider MM, Shashkov AS, et al. Structure of the N-acetylpsseudaminic acid-containing capsular polysaccharide of *Acinetobacter baumannii* NIPH67. *Russ Chem Bull.* 2016;65(2):588-591. doi:10.1007/s11172-016-1342-y
162. Benito-Alifonso D, Tremell S, Sadler JC, Berry M, Galan MC. Imidazolium-tagged glycan probes for non-covalent labeling of live cells. *Chem Commun.* 2016;52:4906-4909. doi:10.1039/C5CC10040B
163. Kim DU, Yoo JH, Ryu K, Cho HS. Crystallization and preliminary X-ray crystallographic analysis of the  $\alpha$ -2,6-sialyltransferase PM0188 from *Pasteurella multocida*. *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2006;62(2):142-144. doi:10.1107/S1744309106000844
164. Kim D, Yoo J, Lee YJ, Kim KS, Cho H. Structural analysis of sialyltransferase PM0188 from *Pasteurella multocida* complexed with donor analogue and acceptor sugar.
165. Parker JL, Lowry RC, Couto N a S, Wright PC, Stafford GP, Shaw JG. Maf-dependent bacterial flagellin glycosylation occurs before chaperone binding and flagellar T3SS export. *Mol Microbiol.* 2014;92(2):258-272. doi:10.1111/mmi.12549
166. Sulzenbacher G, Roig-Zamboni V, Lebrun R, et al. Glycosylate and move! The glycosyltransferase Maf is involved in bacterial flagella formation. *Environ Microbiol.* 2018;20(1):228-240. doi:10.1111/1462-2920.13975
167. Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K2 capsule associated with the KL2 gene cluster of *Acinetobacter baumannii*. *Glycobiology.* 2014;24(6):554-563. doi:10.1093/glycob/cwu024
168. Liu B, Knirel Y a, Feng L, et al. Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev.* 2008;32(4):627-653. doi:10.1111/j.1574-6976.2008.00114.x
169. Canals R, Vilches S, Wilhelms M, Shaw JG, Merino S, Tomás JM. Non-structural flagella genes affecting both polar and lateral flagella-mediated motility in *Aeromonas hydrophila*. *Microbiology.* 2007;153(Pt 4):1165-1175. doi:10.1099/mic.0.2006/000687-0
170. Hyma KE, Lacher DW, Nelson AM, Bumbaugh AC, Janda JM, Strockbine NA. Evolutionary Genetics of a New Pathogenic *Escherichia* Species : *Escherichia albertii* and Related *Shigella boydii* Strains. 2005;187(2):619-628. doi:10.1128/JB.187.2.619-628.2005

171. Bourne Y, Henrissat B. Glycoside hydrolases and glycosyltransferases: Families and functional modules. *Curr Opin Struct Biol.* 2001;11(5):593-600. doi:10.1016/S0959-440X(00)00253-0
172. Kenyon JJ, Arbatsky NP, Sweeney EL, et al. Production of the K16 capsular polysaccharide by *Acinetobacter baumannii* ST25 isolate D4 involves a novel glycosyltransferase encoded in the KL16 gene cluster. *Int J Biol Macromol.* 2019;128:101-106. doi:10.1016/j.ijbiomac.2019.01.080
173. Kasimova AA, Shneider MM, Arbatsky NP, et al. Structure and Gene Cluster of the K93 Capsular Polysaccharide of *Acinetobacter baumannii* B11911 Containing 5 N Acetyl 7 N [( R ) 3 hydroxybutanoyl ] pseudaminic Acid. *Biochem.* 2017;82(4):483-489.
174. Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K2 capsule associated with the KL2 gene cluster of *Acinetobacter baumannii*. *Glycobiology.* 2014;24(6):554-563. doi:10.1093/glycob/cwu024
175. Senchenkova SN, Shashkov AS, Shneider MM, et al. Structure of the capsular polysaccharide of *Acinetobacter baumannii* ACICU containing di-N-acetyl pseudaminic acid. *Carbohydr Res.* 2014;391:89-92. doi:10.1016/j.carres.2014.04.002
176. Senchenkova SN, Popova A V, Shashkov AS, et al. Structure of a new pseudaminic acid-containing capsular polysaccharide of *Acinetobacter baumannii* LUH5550 having the KL42 capsule biosynthesis locus. *Carbohydr Res.* 2015;407:154-157. doi:10.1016/j.carres.2015.02.006
177. Kenyon JJ, Marzaioli AM, Hall RM, De Castro C. Structure of the K12 capsule containing 5,7-di-N-acetylacinetaminic acid from *Acinetobacter baumannii* isolate D36. *Glycobiology.* 2015;25(8):881-887. doi:10.1093/glycob/cwv028
178. Lee IM, Yang FL, Chen TL, et al. Pseudaminic Acid on Exopolysaccharide of *Acinetobacter baumannii* Plays a Critical Role in Phage-Assisted Preparation of Glycoconjugate Vaccine with High Antigenicity. *J Am Chem Soc.* 2018;140(28):8639-8643. doi:10.1021/jacs.8b04078
179. Pier GB. *Pseudomonas aeruginosa* lipopolysaccharide: A major virulence factor, initiator of inflammation and target for effective immunity. *Int J Med Microbiol.* 2007;297(5):277-295. doi:10.1016/j.ijmm.2007.03.012
180. Castric P, Cassels FJ, Carlson RW. Structural Characterization of the *Pseudomonas aeruginosa* 1244 Pilin Glycan. *J Biol Chem.* 2001;276(28):26479-26485. doi:10.1074/jbc.M102685200
181. Smedley JG, Jewell E, Roguskie J, et al. Influence of pilin glycosylation on *Pseudomonas aeruginosa* 1244 pilus function. *Infect Immun.* 2005;73(12):7922-7931. doi:10.1128/IAI.73.12.7922-7931.2005
182. Qutyan M, Neely AN, Cross AS, et al. Immunization with a *Pseudomonas aeruginosa* 1244 Pilin Provides O-Antigen-Specific Protection. *Clin Vaccine Immunol.* 2008;15(4):590-597. doi:10.1128/cvi.00476-07
183. Meccas J, Green ER. Bacterial Secretion Systems – An overview. *Microbiol Spectr.* 2016;4(1):1-32. doi:10.1128/microbiolspec.VMBF-0012-2015.Bacterial
184. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol.* 2003;328(2):307-317.

doi:10.1016/S0022-2836(03)00307-3

185. Ni L, Chokhawala HA, Cao H, et al. Crystal structures of *Pasteurella multocida* sialyltransferase complexes with acceptor and donor analogues reveal substrate binding sites and catalytic mechanism. *Biochemistry*. 2007;46(21):6288-6298. doi:10.1021/bi700346w
186. Bennion BJ, Daggett V. The molecular basis for the chemical denaturation of proteins by urea. *Proc Natl Acad Sci U S A*. 2003;100(9):5142-5147. doi:10.1073/pnas.0930122100
187. Maxwell KL, Bona D, Liu C, Arrowsmith CH, Edwards AM. Refolding out of guanidine hydrochloride is an effective approach for high-throughput structural studies of small proteins. *Prot Sci*. 2003;12:2073-2080. doi:10.1110/ps.0393503.der
188. Yee-Guide Y, Edward N, Ruth A, E. Richard S. Removal of detergents from protein digests for mass spectrometry anal. *Anal Biochem*. 2008;382(2):135-137. doi:10.1016/j.cgh.2008.07.016.Cytokeratin
189. Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: The novel Fh8 system. *Front Microbiol*. 2014;5(FEB):1-20. doi:10.3389/fmicb.2014.00063
190. Kenyon JJ, Kasimova AA, Shneider MM, et al. The KL24 gene cluster and a genomic island encoding a Wzy polymerase contribute genes needed for synthesis of the K24 capsular polysaccharide by the multiply antibiotic resistant *Acinetobacter baumannii* isolate RCH51. *Microbiol (United Kingdom)*. 2017;163(3):355-363. doi:10.1099/mic.0.000430
191. Welner DH, Shin D, Tomaleri GP, et al. Plant cell wall glycosyltransferases: Highthroughput recombinant expression screening and general requirements for these challenging enzymes. *PLoS One*. 2017;12(6):1-20. doi:10.1371/journal.pone.0177591
192. Gilbert M, Karwaski M-F, Bernatchez S, et al. The Genetic Bases for the Variation in the Lipo-oligosaccharide of the Mucosal Pathogen, *Campylobacter jejuni*. *J Biol Chem*. 2002;277(1):327-337. doi:10.1074/jbc.m108452200
193. Spiwok V. CH/ $\pi$  interactions in carbohydrate recognition. *Molecules*. 2017;22(7):1-11. doi:10.3390/molecules22071038
194. Wiggins CAR, Munro S. Activity of the yeast MNN1-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. *Proc Natl Acad Sci*. 2002;95(14):7945-7950. doi:10.1073/pnas.95.14.7945