# Using chemical glycobiology to further our understanding of the enzymatic processing of poly-*N*-acetylglucosamine (PNAG) in bacteria

**Joshua Benjamin Fair**

**PhD**

**University of York**

**Chemistry**

**November 2025**

## Abstract

Poly-*N*-acetylglucosamine (PNAG) is a polysaccharide that is found in the biofilms of several pathogenic species. As such it has been the subject of several studies since its discovery. However, synthesising long polysaccharides *in vitro* has proven difficult due to the challenges of carbohydrate chemistry, and as such access to pure PNAG of defined length and modification has not yet been achieved. This has led to there still being some gaps in the literature surrounding PNAG. The glycosyltransferase enzymes responsible for PNAG polymerisation *in vivo* are complex membrane proteins, and therefore unsuitable for *in vitro* synthesis. It has been shown that glycoside hydrolases can be mutated to perform the opposite reaction, creating what is known as a glycosynthase[1]. Therefore, we proposed to screen mutants of a suitable PNAGase, such as DspB or a member of new class of potential *staphylococci* PNAGases, for glycosynthase activity using a set of successfully synthesised carbohydrate substrates alongside commercially available substrate. This strategy would have allowed us to use synthetic monosaccharide donors for enzymatic polymerisation of PNAG oligosaccharides which could be used for further study. However, these efforts failed due to the selected mutants of DspB not acting as glycosynthases, and the putative *staphylococci* PNAGases proving difficult to express without extensive method development that was outside the scope of this project.

We also aimed to develop a method to chemically modify PNAG *in vitro*. PNAG can be isolated *in vivo*, and can by modified by the selective incorporation of an azide at the reducing terminus of the PNAG polysaccharide[2], deacetylation, or desuccinylation. This modifiation of PNAG opens the door for many avenues of further research, such as bioconjugation to something antigenic potentially creating a vaccine candidate[3]. Some initial evidence shows this modification might have been successful but further experiments are required to conclusively prove this.

# Contents

## Acknowledgements

Firstly, thank you to my supervisors, Dr. Martin Fascione and Prof. Gavin Thomas. Thank you for the opportunity to achieve my dreams, thank you for the support and guidance you've given me throughout the years, and thank you for not giving up on me even after I'd given up on myself.

I wish I could list every person who has been a part of this incredible journey over the last 4 years, but I fear we'd be here forever. From the Ladies Who Lunch, to the Architects of the Monstument, the Organics Friday Football Team, to the ChemBio Rushes B server, you've all served to make York feel truly like home. Thank you for everything you've done for me throughout the years, from putting up with my awful jokes, to feigning sympathy for one of my several self-inflicted injuries. I will forever be grateful for your friendship and will treasure the memories we made together. Hopefully there will be many more to come.

To ChemBio, thank you all for making me feel welcome here. Thank you for helping me settle in and answering any and all questions I may have, even the really stupid ones. Specifically, thank you to Tessa and your help and advice with all the biological experiments I was inexperienced with when I started. Thank you to Nick and Tasha for all their help with the synthetic chemistry questions I had, and for brainstorming potential new synthetic routes with me. And thank you to Julia and Helen for their excellent management of the lab, and for teaching me how to use the equipment that would become a vital part of my research.

Thank you to the Monster Beverage Corporation for delighting the senses with your delicious beverages. The energy I gained from your drinks fuelled most of the experiments detailed in this thesis.

To the Thomas Group, thank you for welcoming the chemist into your lab and never making me feel like I didn't belong. The insight I gained from listening to your projects and presentations, and from the feedback you gave mine was invaluable to my research. To Reyme, thank you for helping me around the lab, and for guiding me through some of the more complex biological experiments.

To my family, thank you for always being there for me throughout this whole journey. Thank you for giving me a place to retreat to when things were getting too much, for always being on the other end of the phone if I needed it, and thank you giving me the foundation which this whole thing has been built off of. Your love and support got me through this. I love you all, and always will do.

## Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references. The work presented was carried out by me with the exception of:

Chapter 1.1:

- Structural model of the IcaADBC complex with PNAG and succinyl-CoA produced by Phil Stansfield, University of Warwick

Chapter 3.2:

- ITAG **59** was provided by the Galen Group, University of Bristol

Chapter 3.3:

- Phylogenetic analysis of putative PNAGases performed by Andrew Higginson

Chapter 3.4:

- Purification of PNAG done in collaboration with Reyme Herman, Connor Munroe and Beth Kinniment-Williams, University of York
- TSDS-PAGE analysis of PNAG Click reactions performed by Dr Joe Nabarro, University of York

# Abbreviations

| | | | |
|---|---|---|---|
| $(COCl)_2$ | Oxayl Chloride | DBU | 1,8-diazobicycloundec-7-ene |
| Ac | Acetyl | DCC | N,N'-Dicyclohexylcarbodiimide |
| $Ac_2O$ | Acetic anhydride | DCM | Dichloromethane |
| AcOH | Acetic acid | DDM | n-dodecyl-β-D-maltoside |
| ADMP | 2-azido-1,3-dimethylimidazolinium hexafluorophosphate | $dH_2O$ | Distilled water |
| AGA | Automated glycan assembly | DMAP | 4-dimethylaminopyridine |
| AgOTf | Silver triflate | DMAPA | Dimethylaminopropylamine |
| Alloc | Allyloxycarbonyl | DMC | 2-chloro-1,3-dimethylimidazolium chloride |
| Asn | Asparagine | DMF | Dimethylforamide |
| Asp | Aspartic acid | DMSO | Dimethylsulfoxide |
| BdhI | *Bifidobacterium bifidum* GH20 enzyme | dPNAG | Partially deacetylated PNAG |
| $BF_3.OEt_2$ | Boron trifluoride etherate | DspB | Dispersin B |
| $Bi(OTF)_3$ | Bismuth (III) triflate | EIC | Extracted ion chromatogram |
| Boc | *t*-butyloxycarbonyl | ELISA | Enzyme linked immunosorbent assay |
| BSA | Bovine serum albinum | EPS | Extracellular polymeric substance |
| $Bu_4NBr$ | Tetrabutylammonium bromide | eq. | Equivalents |
| Bz | Benzoyl | ESI | Electrospray ionisation |
| BzCl | Benzoyl Chloride | EtOAc | Ethyl acetate |
| C.V. | Column volume | Fmoc | Fluorenylmethyloxycarbonyl |
| CAZy | Carbohydrate Active Enzymes | FTIR | Fourier transformed infra red |
| CDMBI | 1,3-dimethyl-1H-benzimidazol-3-ium chloride | GH | Glycoside hydrolase |
| CHAPS | 3-(Dimethy[3-(3α,7α,12α-trihydroxy-5β-cholan-24-amido)propyl]azaniumyl) propane-1-sulfonate | Glc | Glucose |
| $Cl_3CCN$ | Trichloroacetonitrile | GlcNAc | *N*-acetylglucosamine |
| CMC | Critical micelle concentration | Glu | Glutamaic acid |
| CoA | Coenzyme A | GNAT | Gcn5-related N-acetyltransferases |
| COSY | Correlation syn-spectroscopy | $H_3PO_4$ | Phosphoric acid |
| CuAAC | Copper (I) catalysed azide-alkyne cycloaddition | HCl | Hydrogen chloride/Hydrochloric acid |
| $CuSO_4$ | Copper sulphate | Hex | Hexane |
| CycloOct-GFP | Cyclooctyne green fluorescent protein | HF | Hydrogen fluoride/Hydrofluroic acid |
| Cys | Cysteine | HILIC LC-MS | Hydrophilic interaction chromatography for liquid chromatography mass spectrometry |
| $D_2O$ | Dueterated water | His | Histidine |
| DAST | Diethylaminosulfur trifluoride | HRMS | High resolution mass spectrometry |

| | | | |
|---|---|---|---|
| HRP | Horse raddish peroxidase | PAGE | Polyacrylamide gel electrophoresis |
| HSQC | Heteronuclear single quantum coherence | PBS | Phosphate buffer saline |
| Ica | Intercullar adhesion | $Pd(PPh_3)$ | Tetrakis(triphenylphosphine) palladium(0) |
| IPTG | Isopropyl-β-D-1-thioglacatopyranoside | PDB | Protein database |
| ITAG | Imidazolium tag | Peg | Polyethylene glycol |
| KOH | Potassium hydroxide | PG | Polyglucosamine (Polysaccharide) |
| LB | Lysogeny broth | Pga | Polyglucosamine (Synthase Protein and Genes) |
| LC-MS | Liquid chromatography mass spectrometry | $PhSiH_3$ | Phenylsilane |
| MALDI | Matrix-assisted laser desorption/ionisation tandem mass spectroscopy | PIA | Polysaccharide Intercellular Adhesion |
| Man | Mannose | pLDDT | Predicted Local Distance, Difference Test |
| MeCN | Acetonitrile | *p*NA | *p*-nitroaniline |
| MeI | Methyl iodide | PNAG | Poly-*N*-acetylglucosamine |
| MeOH | Methanol | pNP | *p*-nitrophenol |
| $MgSO_4$ | Magnesium sulphate | pNP-GlcNAc | *p*-nitrophenyl-GlcNAc |
| MS | Mass spectrometry | Pyr. | Pyridine |
| MTBE | Methyl *t*-butyl ether | QCM-D | Quartz Crystal Microbalance with Dissipation monitoring |
| MurNAc | *N*-acetylmuramic Acid | R.T. | Room temperature |
| $N_2$ | Nitrogen | SaHex | *Staphylococcus aureus* Hexoamidase |
| NaCl | Sodium chloride | $Sc(OTF)_3$ | Scandium (III) triflate |
| $NaHCO_3$ | Sodium hydrogen carbonate | SDS | Sodium dodecyl sulfate |
| $NaN_3$ | Sodium azide | SEC | Size exclusion chromatography |
| NaOH | Sodium hydroxide | Ser | Serine |
| NaOMe | Sodium methoxide | ShHex | *Staphylococcus hominis* Hexoamidase |
| NCBI | National centre for biotechnology information | SlHex | *Staphylococcus lugdunensis* Hexoamidase |
| NGP | Neighbouring group participation | SPAAC | Strain promoted azide-alkyne cycloaddition |
| $NH_2NH_2^{\cdot}\bullet H_2O$ | Hydrazine monohydrate | STol | Silyl toluene |
| $NH_4OH$ | Ammonium hydroxide | TBAF | Tetra-N-butylammonium |
| NHS | *N*-hydroxysuccinimide | TBDPS | *t*-butyldiphenylsilyl |
| NMR | Nuclear Magnetic Resonance | TBDPS | *t*-butyldiphenylsilyl chloride |
| $OD_{600}$ | Optical denisty (600 nm) | TBDPSCl | t-butyldiphenylchlorosilane |
| ON | Overnight | TCA | Trichloroacetic acid |
| OPAL | Organocatalyst-mediated protein aldol ligation | TEA | Triethylamine |

| | |
|---|---|
| TFA | Trifluoroacetic acid |
| TFAA | Trifluoroacetic anhydride |
| TfHex | *Tessaracococcus flavus* Hexoamidase |
| THF | Tetrahydrofuran |
| THPTA | Tris(benzyltriazolylmethyl) amine |
| Thr | Threonine |
| TLC | Thin layer chromatography |
| TMSOTf | Trimethylsilyl trifluoromethanesulfonate |
| Tol | Toluene |
| TolSH | *p*-toluenesulfoniuc acid |
| Troc | 2,2,2-tricchloroethoxycarbonyl |
| TSB | Tryptic soy broth |
| TSDS-PAGE | Tricine SDS-PAGE |
| TT | Tetanus toxoid |
| UDP | Uridine diphosphate |
| UV/Vis | Ultraviolet-visible light spectroscopy |
| WGA | Wheat germ aggultinin |
| WT | Wild type |
| YSBL | York Structural Biology Lab |

# List of Figures

# 1. Introduction:

## 1.1. Part 1: *N*-acetylglucosamine, and the biosynthesis of poly-*N*-acetylglucosamine

### 1.1.1. Carbohydrates – nature's LEGO blocks

Carbohydrates are an important building block within biological systems. They can covalently link together to create what are known as oligosaccharides and polysaccharides, depending on the chain length, which are widely used in biological systems due to their diverse range of properties. Through a process known as glycosylation they can also be covalently attached to various biomolecules, such as proteins and lipids. This can serve several functions such as aiding with protein folding[4], protein stability[5], and immune evasion[6]. This wide range of properties can be attributed to the chemical structure of carbohydrates. In 1891, Emil Fischer, in his Nobel Prize winning work, defined a method to describe the structures of a new class of biomolecules that was emerging. This new system for representing carbohydrate molecules, now referred to as Fischer Projections (Figure 1.1.1A), began to reveal the potential for structural diversity presented by carbohydrate molecules. Fischer described 16 unique structures in 8 enantiomeric pairs, which he labelled as D/L- defined by the side of the chain the hydroxy group on the penultimate carbon sits on, with D- being on the right, and L- the left[7]. This, however, was only the tip of the iceberg of carbohydrate structural diversity, as it now known that carbohydrates can also form what are known as -ose rings (Figure 1.1.1B).

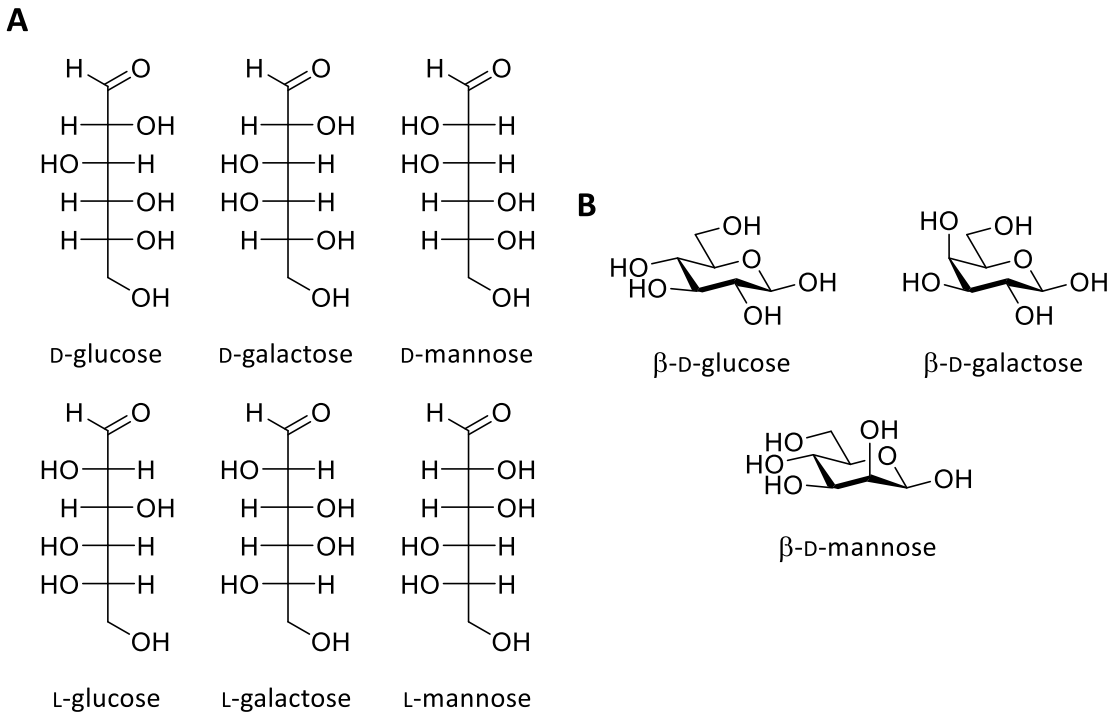



Figure 1.1.1: A) Fischer projections of D-glucose, D-galactose, and D-mannose, and their enantiomeric pairs in D-glucose, L-galactose, and L-mannose. B) The hexose, or pyranose, ring forms of D-glucose, D-galactose, and D-mannose in their β-anomers.

Carbohydrates have the empirical formula $C_n(H_2O)_m$, although different types of sugars regularly deviate from this. Carbohydrates can exist in an acetal or hemiacetal form (ketal and hemiketal for 5 membered ring sugars). In their acetal form, that being where the anomeric hydroxy group is glycosylated in some form, they are referred to as non-reducing sugars. In their hemiacetal form, that being where the anomeric hydroxy group is unglycosylated, they are referred to as reducing sugars. The anomeric centre is defined as the carbon that bears the acetal or hemiacetal when it in its cyclic form. While non-reducing sugars are locked into a ring structure, reducing sugars exist in an equilibrium between their ring forms and their ring opened form (Figure 1.1.2). Upon ring closing, two different sized rings can be formed, the pyranose form (6 membered ring) and the furanose form (5 membered ring), and further still, each ring form can have two anomers, the α-anomer and the β-anomer. This is determined by the position of the anomeric hydroxy group in relation to the substituent on the carbon on the other side of the ring oxygen, where α means they're on different faces of the ring, and β means they're on the same face. This means that each sugar molecule can form 5 different structural isomers.

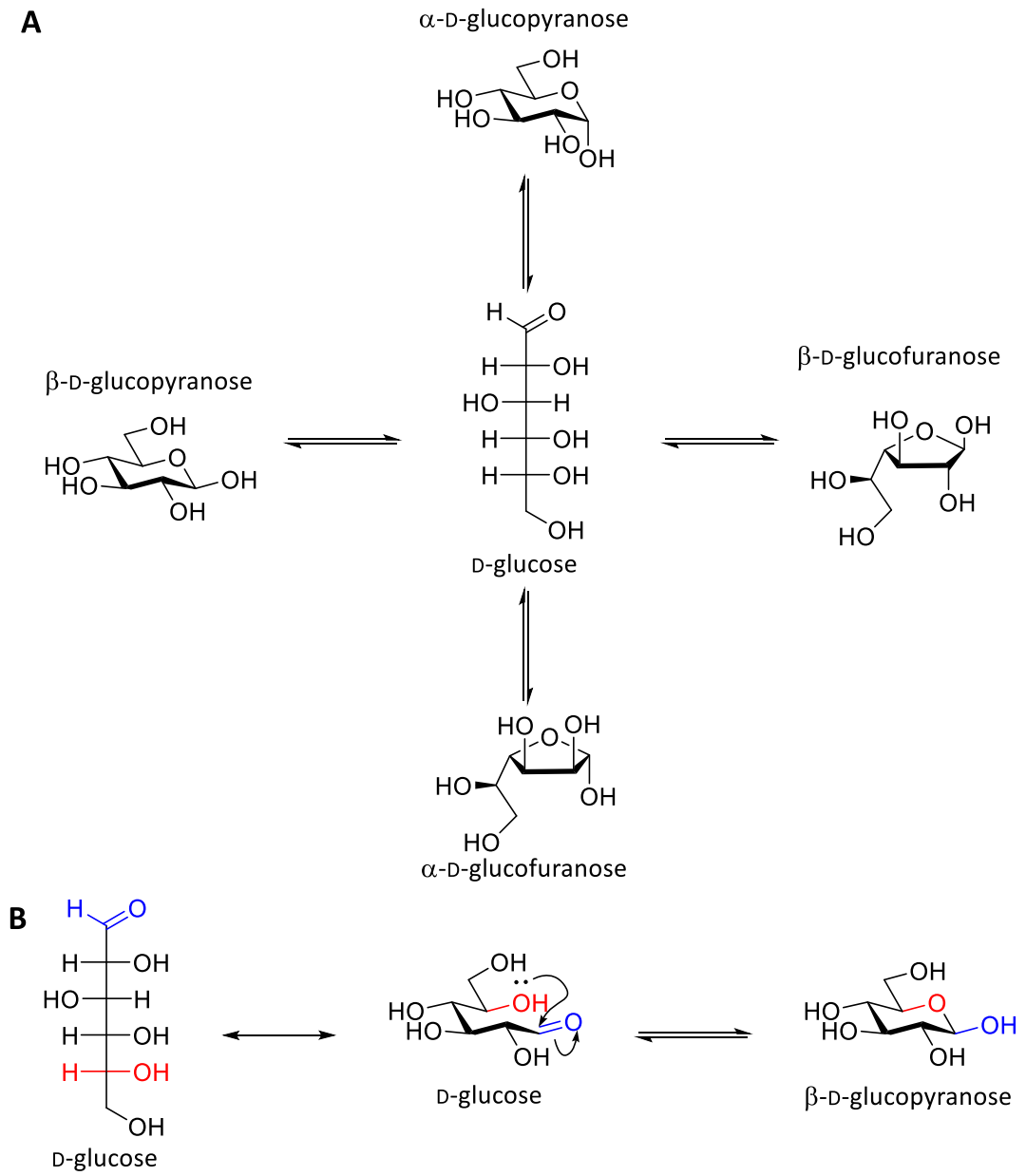



Figure 1.1.2: A) The five structural isomers of D-glucose that can be formed. B) Ring closing mechanism of D-glucose showing the anomeric carbon and oxygen (blue) and the oxygen that forms part of the ring (red).

To complicate matters even further, each sugar molecule has an enantiomer, which can also form the aforementioned 4 structural isomers. In total then, each reducing monosaccharide sugar, i.e. glucose or galactose, can be found in 10 different isomers, and each isomer has different properties, and confers different properties to any polysaccharide or bioconjugate it is incorporated into. Adding different functional groups to the hydroxy groups then also allows for additional properties to be conferred to the sugar or conjugated biomolecule. Calculations made in 1994 show there are up to $1.025 \times 10^{12}$ possible structures of a reducing hexasaccharide if you account for branching polysaccharides as well as linear[8]. This level of complexity has meant that research into carbohydrates and their roles in biological systems has often been neglected, and what research

was conducted was bogged down by the challenges of working with carbohydrates. This has changed with the turn of the millennium and research into carbohydrates is beginning to really shine a light onto what this diverse set of molecules is capable of[9]. One sugar of particular interest is an amide derivative of glucose that is found ubiquitously across the full spectrum of life.

### 1.1.2. *N*-acetylglucosamine – a ubiquitous sugar

β-D-2-(acetylamino)-2-deoxy-glucopyranose, otherwise known as *N*-acetyl-D-glucosamine (Figure 1.1.3A), or simply GlcNAc, is a reducing sugar, that is found in multiple forms across the natural world. Most commonly GlcNAc is found within long polysaccharide chains. Chitin is the second most abundant biopolymer in the biosphere, second only to cellulose, and it plays a major role in the mycelia of fungi[10,11], as well as being utilised extensively by arthropods[12,13] molluscs[13], and even in some fish species[14]. First determined in 1929 by Albert Hofmann[15], Chitin is formed of β-1-4-linked GlcNAc monomers (Figure 1.1.3B), in a structure that is reminiscent of that of cellulose. This leads to a strong, chemically resistant polymer that works very well as a barrier, leading to its use in cell walls, and within exoskeletons and shells. GlcNAc can also been found within the repeating units of both peptidoglycan and hyaluronic acid (Figure 1.1.3C-D). Peptidoglycan, sometimes referred to as murein, forms an important component of the cell wall of bacteria. First identified by researchers attempting to determine the method of action of penicillin[16,17], subsequent research has shown that peptidoglycan is important to bacterial growth[18]. Inhibition of any part of the biosynthesis of peptidoglycan, for example the inhibition of DD-transpeptidases by penicillin[19], causes cell death by cytolysis[20]. Peptidoglycan is formed of β-1-4-linked GlcNAc and *N*-acetylmuramic acid (MurNAc), with a tetrapeptide bound to the lactic acid residue of the MurNAc (Figure 1.1.3C). Hyaluronic acid is found across a wide variety of organs and tissues in mammals[21], and it performs several important functions, ranging from wound healing[22], lubricating joints[23], and forming part of connective tissues and cartilage[24]. First described in 1934 by Meyer[25], hyaluronic acid is formed of alternating β-1-3 and β-1-4 linked D-glucuronic acid and GlcNAc (Figure 1.1.3D), and has found use in extensive use in surgery due to its wound healing properties.

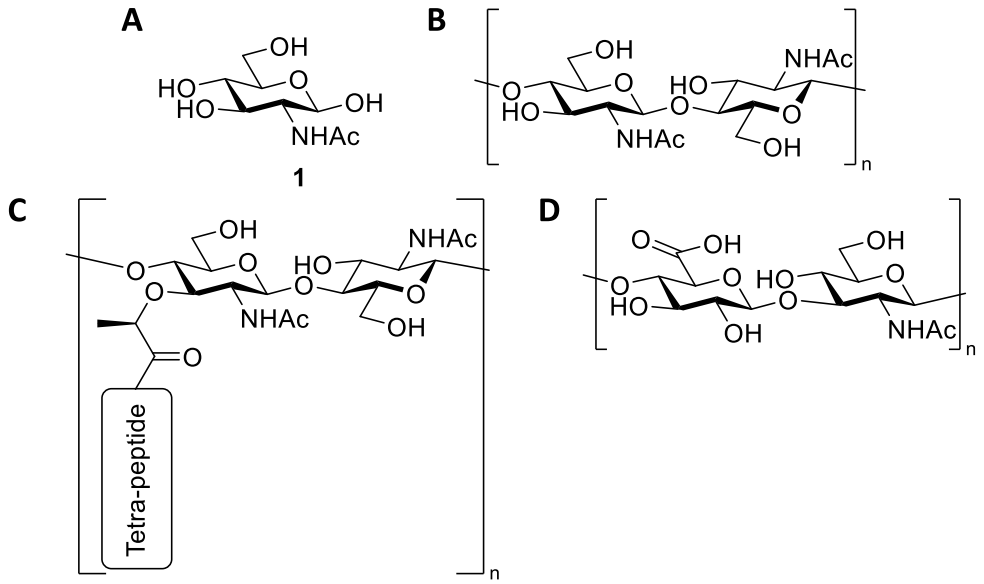



Figure 1.1.3: A) Chemical structure of *N*-acetyl-D-glucosamine (GlcNAc **1**). B) Chemical structure of the repeating unit of the biopolymer chitin formed of two β-1-4-linked GlcNAc molecules. C) General chemical structure of the repeating unit of the biopolymer peptidoglycan formed of β-1-4-linked GlcNAc and MurNAc. The structure of the tetrapeptide is species specific[18]. D) Chemical structure of the repeating unit of hyaluronic acid formed of alternating β-1-3 and β-1-4 linked D-glucuronic acid and GlcNAc.

As well as major functions in polymers, GlcNAc monomers and oligosaccharides also find use within biological systems. *O*-GlcNAcylation[26,27] and *N*-GlcNAcylation[28] are the two processes through which GlcNAc is attached to proteins. *N*-GlcNAcylation takes place at the amide nitrogen of an asparagine residue (Asn), and the involves a precursor oligosaccharide containing 2 consecutive GlcNAc molecules followed by 3 D-mannose (Man) molecules being attached via the reducing GlcNAc molecule (Figure 1.1.4). This oligosaccharide, known as the core glycan structure, forms the basis of all *N*-glycan modifications of proteins[28,29], and thus GlcNAc is essential in mediating protein stability and function[30-32].

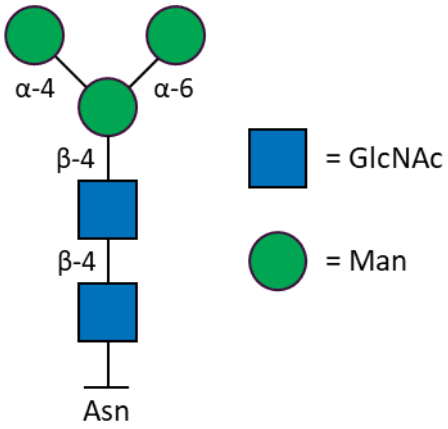



Figure 1.1.4: The core glycan structure of *N*-glycans. The reducing end of the oligosaccharide is bound to the amide nitrogen of the asparagine residue of the protein. The core structure consists of 2 β-1-4 linked GlcNAc residues following by 3 Man residues which are α-1-4 and α-1-6 linked.

*O*-GlcNAcylation however, is the glycosylation of monosaccharide GlcNAc onto the hydroxy group of a serine (Ser) or threonine (Thr) residue. This is a dynamic process with GlcNAc being attached and removed from proteins at various points. This dynamic attachment and removal of GlcNAc has been shown to be important to modulating various signalling and regulatory pathways[33-35], and disrupting the natural levels of *O*-GlcNAcylation has been implicated in several different diseases such as diabetes, and Alzheimer's[36]. This makes *O*-GlcNAcylation analogous to phosphorylation[37], and indeed the two are heavily linked. They often share the same modification site have been known to modulate each other[38].

This non-exhaustive exploration of GlcNAc's role and function within biological systems demonstrates the importance of the ubiquitous sugar. However, there is another biopolymer that GlcNAc can form that has important implications for the antibiotic resistance crisis.

### 1.1.3. Poly-N-acetylglucosamine – GlcNAc's evil cousin

Several microorganisms, including pathogenic species such as *Staphylococcus aureus*[39], *P. aeruginosa*[40], *Neisseria gonorrhoea*[41], and *Escherichia coli*[42], produce a matrix of extracellular polymeric substances (EPS), such as polysaccharides and proteins, known as a biofilm[43]. The biofilm performs several functions for the organism, including promoting adhesion on abiotic surfaces[42,44-48] and intercellular adhesion[45,46], providing resistance to antibiotic and antimicrobial compounds[44,48-50], and giving protection from phagocytes[47,49]. When combined, these factors all contribute to an improved virulence of the pathogen[51]. Biofilm presenting strains of a pathogen are known to cause more severe infections than strains of the same species that do not present biofilms, especially for patients with implants, and in dentistry where large abiotic surfaces can be

found. They also make the infection more difficult to fully eradicate[52]. This coupled with the antibiotic resistance conferred by biofilms has meant that EPS have drawn a lot of clinical attention.

In 1982 scanning electron microscopy was initially used by Christensen *et al*[53] to reveal that the biofilms formed during Staphylococcal infections had a major polysaccharide component, however the exact chemical structure wasn't identified until 1996 by Mack *et al*[54]. They were investigating the polysaccharide in *S. epidermidis* using NMR, amongst other chemical analytical methods. They were able to determine the structure to be a linear, non-branching polysaccharide that they called polysaccharide intercellular adhesion (PIA) but it is now generally referred to as poly-*N*-acetylglucosamine, or PNAG (Figure 1.1.5). It is a β-1-6-linked polysaccharide consisting of purely GlcNAc monomers that is excreted by bacteria creating the structural basis for the biofilms.

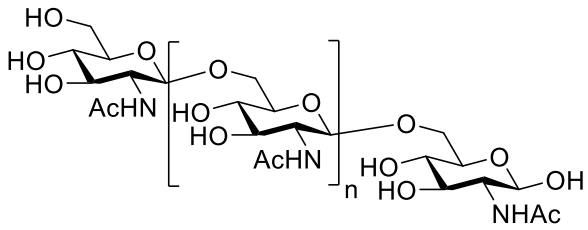



Figure 1.1.5: Chemical structure of PNAG, where n = 100-120[55].

Despite having a core homopolymeric structure, preparations could be separated into two distinct molecular species, determined by the charge displayed by each species. Polysaccharide I was positively charged and contained a higher percentage of *N*-deacetylation (≈15% of monomers deacetylated), although this appeared to be randomly spaced throughout the structure. Polysaccharide II was anionicly charged, with evidence of succinates and phosphates being present, however, these could not be conclusively shown to be part of the polysaccharide structure at the time.

The structural differences between PNAG taken from various different species is relatively minor with the biggest difference being the amount of *N*-deacetylation found along the chain length (ranging from 5-15% of the monomers[42,56]). The other notable difference is whether or not the PNAG has undergone O-succinylation of the hydroxy group at either position 3 or 4. Initially misidentified as *N*-succinylation, it has since been demonstrated in both *S. epidermidis*[57], and *S. aureus*[58] PNAG that it is indeed *O*-succinylation, however, at the time of writing, it is unclear as to mechanism with which these species perform this modification[59] (Figure 1.1.6).

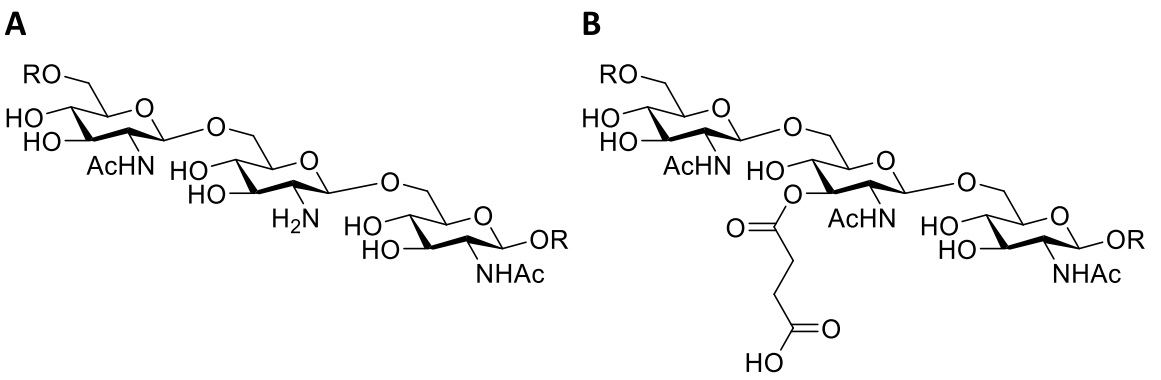



Figure 1.1.6: A) *N*-deacetylated PNAG monomer in the polymer chain. B) *O*-succinylated PNAG monomer in the polymer chain.

Since its discovery in 1982, PNAG has been discovered to form a part of many species' biofilms. This includes, as mentioned above, *S. aureus*, but also *A. baumannii*[60], both of which were included in the WHO Bacterial Priority Pathogen List 2024[61], along with *E. coli*[42], *Y. pestis*[62], *A. actinomycetemcomitans*[48], *B. pertussis*[63], and many others. Whilst the core structure of PNAG is the same in Gram -ve bacteria, the *O*-succinylation appears to be limited to only Gram +ve bacteria. There is also some evidence to suggest this list can be expanded to include two more pathogens listed by the WHO as critical and high priority respectively, *M. tuberculosis* and *N. gonorrhoeae*, alongside up to 150 other species. However, whilst these two species, amongst others, did give a positive result when tested with a PNAG binding antibody, they, along with a number of the other species tested, do not have an operon that has been identified as encoding for proteins capable of producing the polysaccharide, and this should be taken into account[43].

This list of pathogenic bacteria, whilst not exhaustive and likely to be expanded too, is already evidence enough that studying PNAG and developing counter measures is of vital importance if we wish to improve the outcomes of patients presenting with biofilm infections. PNAG, having been discovered nearly 30 years prior to the time of writing, has been the subject of many research papers, and efforts were initially focussed on determining its biosynthetic pathway.

### 1.1.4. The Biosynthetic Pathway of PNAG and its biomodification

The *icaABC* locus was the first biosynthetic pathway of PNAG to be discovered in 1996 in *S. epidermis* by Heilmann *et al*[46], who, via knockout mutations of genes *icaA*, *icaB*, and *icaC* showed that biofilm formation was disrupted and intercellular adhesion was completely lost. IcaA was hypothesised to be the *N*-acetylglucosaminyltransferase protein of the locus due to its sequence similarity to other *N*-acetylglucosaminyltransferases, NodC from *R. meliloti*[64], and HasA from *S. pyogenes*[65]. This was later confirmed by Gerke *et al*[66], alongside the discovery of the protein IcaD, which expanded the locus to *icaADBC*. They showed that while a strain of *S. carnosus* pTXicaA was

capable of producing GlcNAc oligosaccharides, however, for 'full' *N*-acetylglucosaminyltransferase activity, IcaD was required to be expressed alongside IcaA. This wasn't the whole story though, as unless IcaC was also expressed alongside IcaAD, the PNAG produced would only reach a maximum length of 20 monomers of GlcNAc. All three proteins contain several predicted transmembrane helices (4 in IcaA, 2 in IcaD, and 6 in IcaC)[46,66], which points to a three protein complex that is involved in the biosynthesis of PNAG, although the exact function that IcaD and IcaC play in biosynthesis was not determined. At that time no homologues were known for IcaB, however Heilmann *et al* made two predictions of possible functions it could perform, one of which was that it might act as an *N*-deacetylase. Vuong *et al* later proved this hypothesis via SEC/ESI-MS analysis of a Δ*icaB* strain of *S. epidermidis*[56], which revealed a lack of the signature fragmentation pattern of partially deacetylated PNAG, which was present in the wild type (WT). Despite the exact function of IcaD and IcaC remaining unclear, these results have led to an overall hypothesis of how the *IcaADBC* operon functions becoming accepted. IcaA is the *N*-acetylglucosaminyltransferase enzyme that forms a complex with IcaD to become active and synthesise the PNAG. IcaC then forms a complex with IcaAD to allow for the translocation of the synthesised PNAG across the membrane, where IcaB, the only non-membrane bound protein of the locus, performs partial *N*-deacetylation.

The locus in *E. coli* is known as *pga*, and it contains the four genes encoding for the proteins *pgaABCD*[42]. Wang *et al* showed by deletions of all four genes, that the biofilm production was disrupted, and intracellular adhesion was lost, as with *S. epidermis* when *icaADBC* was knocked out, as well as loss of attachment to abiotic surfaces. Sequence analysis revealed that the two loci do share proteins that perform the same key functions, however the naming system for *pgaABCD* does not match that of *icaADBC*. PgaA had the sequence of a large outer membrane protein suggesting some form of membrane translocation function, making it unique from any protein in the *ica* locus as Staphylococcal species lack an outer membrane. Later studies have shown that PgaA and PgaB form a complex together[67], with a negatively charged transmembrane β-barrel that works to transport the positively charged PNAG produced by PgaB across the membrane. PgaB is an *N*-deacetylase with a glycoside hydrolase domain[68-70] that requires PgaA to be bound to it to have maximum *N*-deacetylase activity[67,71]. PgaB is orthologous to IcaB, as this is also an *N*-deacetylase. PgaC is the *N*-acetylglucosaminyltransferase protein[69], orthologous to IcaA, which, like IcaA, needs to form a complex with PgaD in order to exhibit full transferase activity.

The study of *pgaABCD* has led to full determination of the function of each protein, however, this has also led to speculation about the accepted hypothesis of the function of *icaADBC*. More specifically the role of IcaC in PNAG presenting *staphylococci* species. Atkin *et al* performed an up-to-date sequence analysis of IcaC in 2014[59]. This revealed that in the intervening years

between Heilmann *et al* publishing their paper in 1996 and Atkin *et al* publishing theirs in 2014, a large family of acyltransferases (Acyl_transf_3 family (PF01757), InterPro family IPR002656[72]) had been characterised, and their sequence analysis showed homology with this family. Whilst *icaADBC* orthologues can be found among gram +ve organisms, the operons rarely contain a protein with a homologous sequence to IcaC. The orthologous operon in *E. coli*, *pgaABCD*, is also capable of transporting PNAG across the inner membrane without an IcaC homologue. This has led to the suggestion that IcaC could function as an *O*-succinyltransferase[59] responsible for the production of PNAG with anionic character as described by Mack *et al*[54]. Evidence for *O*-succinylation of PNAG being found in species that contain IcaC, and not in *E. coli* and others without it, leading to the belief that IcaC is involved in the incorporation of the succinyl groups that may be found in *staphylococci* PNAG[54]. It is speculated that IcaC would utilise a succinyl-CoA substrate to transfer the succinyl group the PNAG chain, and is membrane bound to allow for the addition of the succinyl group as the PNAG is translocated across the membrane by IcaAD (Figure 1.1.7). Since this hypothesis was made, no evidence for or against this hypothesis has been published, at the time of writing. As described above, PNAG is a key virulence factor for a number of dangerous pathogens, and as such, fully characterising its biosynthetic pathway is a key challenge on the path to effective treatment of these diseases. In this work, one of the aims we set out to achieve was to experimentally determine whether IcaC is in fact an *O*-succinyltransferase.

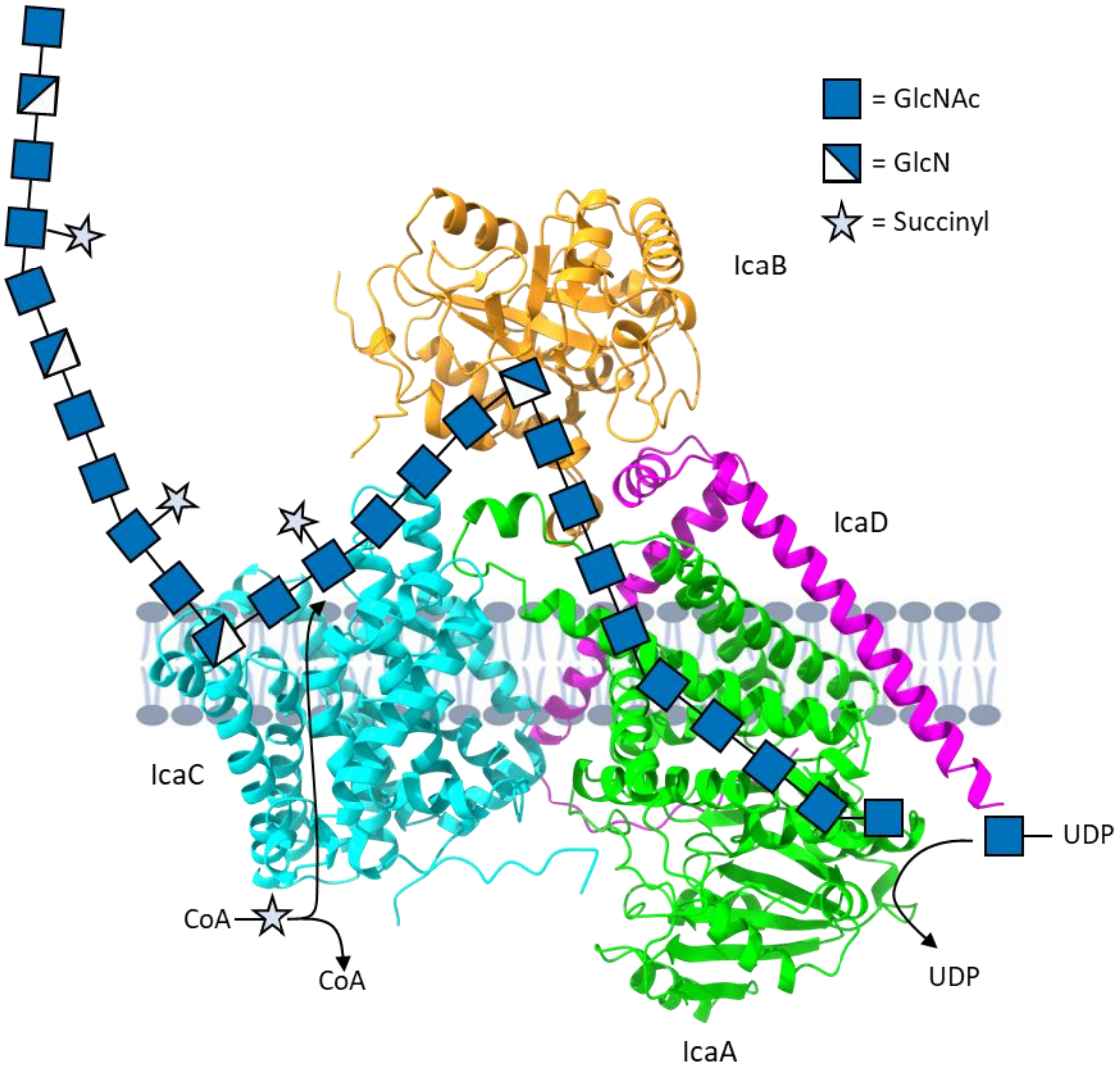
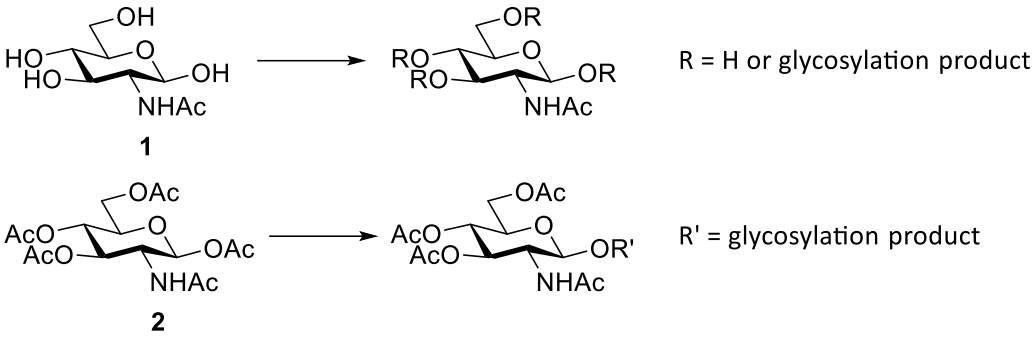



Figure 1.1.7: The hypothesised biosynthetic pathway of PNAG in *staphylococci*. The model for the IcaADBC complex was generated by Phil Stansfield (University of Warwick) using AlphaFold3[73,74] and the image was produced using Chimera[75].

### 1.1.5. The Synthetic Chemistry approach to making PNAG

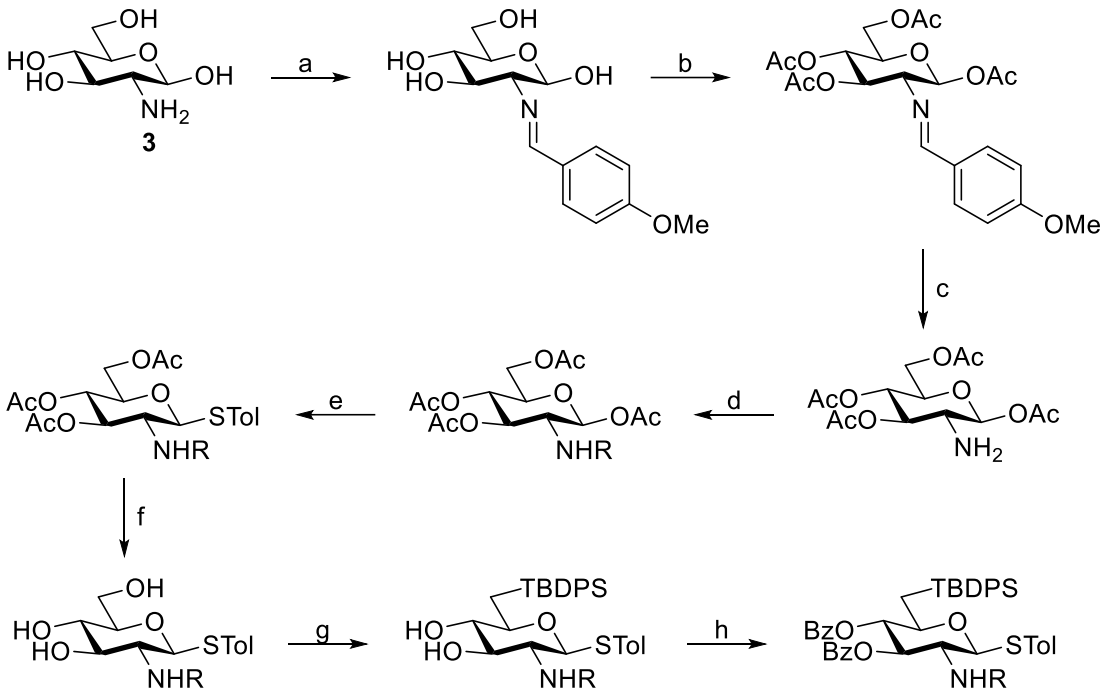
Another area that has received a lot of attention from researchers since the discovery of PNAG is the development of a synthetic route to make PNAG. The chemical synthesis of any long chain polysaccharide poses unique challenges due to the structural complexity of carbohydrates. Each monosaccharide has several hydroxy groups in its structure, and therefore it has several possible locations where reactions can take place. This, coupled with the fact that any reaction that a monosaccharide hydroxy group undergoes can either maintain or invert the stereochemistry at that location means that the potential for unwanted side reactions with each additional monosaccharide that is added to the polysaccharide chain becomes exponentially greater if not performed correctly. The classic solution to this problem is to utilise protecting group chemistry. For example, if you were to attempt to perform a glycosylation at the anomeric position of a deprotected GlcNAc molecule there is the potential for that bond to be formed at any of the 4 other

substituent groups. However, if you protected each other hydroxy group with an acetyl group, for example, the only position which that reaction can occur is the anomeric position (Scheme 1.1.1).




Scheme 1.1.1: A) General scheme of a glycosylation of an unprotected GlcNAc **1** monosaccharide showing the potential side products. B) A general scheme of a glycosylation of a GlcNAc monosaccharide **2** that has been protected so only the anomeric centre can react.

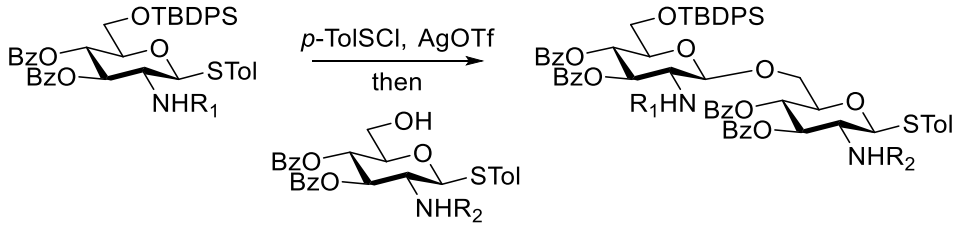
Carbohydrate protecting group chemistry has developed to the point it is possible to perform individual protection of each hydroxy group in a sugar ring[76] and this has been successfully applied in the synthesis of PNAG oligosaccharides[3,77-84], partially deacetylated PNAG (dPNAG) oligosaccharides[3,85-88], and β-1-6-polyglucosamine (PG)[3,84,89,90]. Most recently, Tan *et al*[61] used this method was able to generate a library of dPNAG pentasaccharides, alongside PNAG pentasaccharide and PG pentasaccharides. Starting from glucosamine **3** (GlcN), a precursor molecule to GlcNAc, the amine can be selectively protected using anisaldehyde as it is more nucleophilic than the hydroxy groups. Next, the hydroxy groups can be acetylated. The *p*OMe-phenyl group can then be removed leaving the free amine, which can be reacted with the desired protecting group without reaction occurring on any of the hydroxy groups[91,92]. The anomeric hydroxy group can then be selectively reacted to form a thioglycoside due to the anomeric effect increasing its reactivity versus the other hydroxy groups. The remaining hydroxy groups can then be deacetylated, and the C-6 hydroxy group can then be converted into a silyl ether via reaction with *t*-butyl(chloro)diphenylsilane chloride (TBDPSCl). This works due to the bulky TBDPS group preferentially reacting with the least sterically hindered hydroxy group. Finally, the C-3 and C-4 hydroxy groups can be protecting using benzoyl groups (Scheme 1.1.2).




R = Boc, Alloc, Troc, Fmoc, or Ac

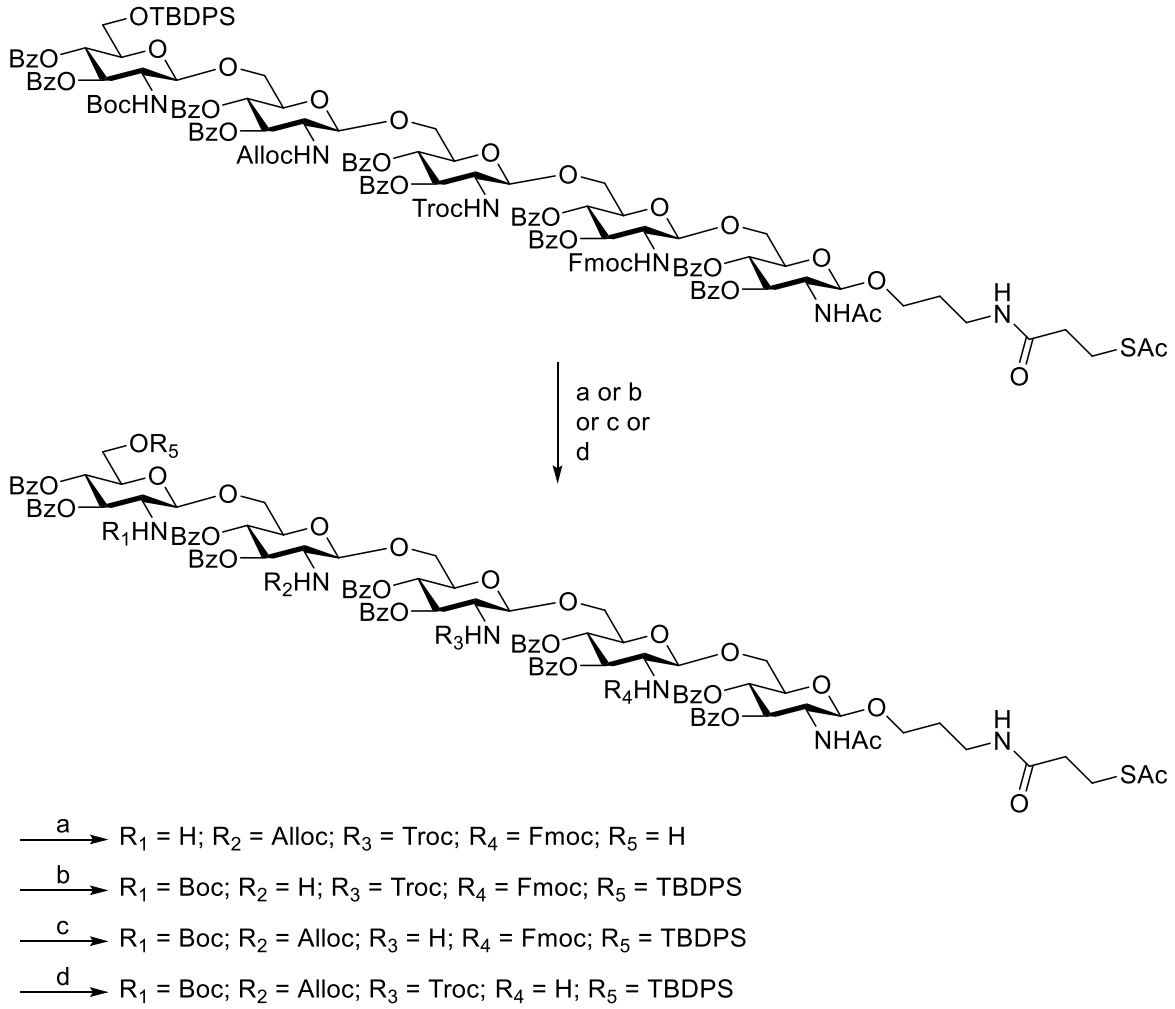
Scheme 1.1.2:. a) *p*-$OMeC_6H_4CHO$, 1M NaOH. b) $Ac_2O$, Pyr, DMAP. c) 5N HCl, acetone.[91] d) R-Cl, base (for Ac this step is skipped). e) TolSH, $BF_3$•$Et_2O$.. f) MeONa in MeOH.[92] g) TBDPSCl, imidazole. h) BzCl, Pyr.[3]

This was repeated to generate 5 separate molecules where R could be Boc, Alloc, Troc, Fmoc, or Ac respectively. Next, the TBDPS group can be removed for 4 of the 5 molecules, which allows for the *p*TolSCl/AgOTF promotor system developed by Huang *et al*[93] to be utilised to perform the 1,6 glycosylation (Scheme 1.1.3).



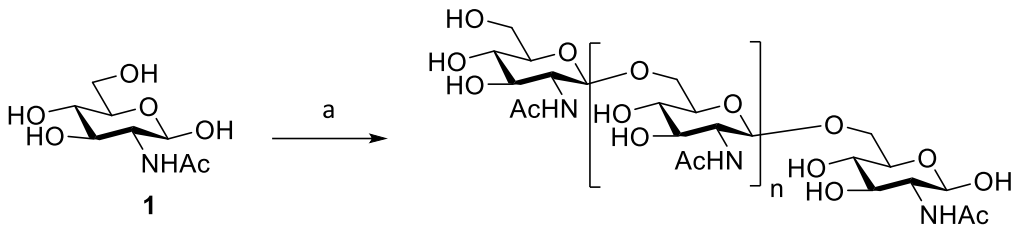

Scheme 1.1.3: Glycosylation utilising the *p*TolSCl/AgOTF promotor system to create a disaccharide with two different protecting groups.

This was repeated using each different protecting group, generating a protected PNAG pentasaccharide with a different protecting group on each amine with the final monomer having a handle for bioconjugation at the anomeric position. Each amine could be selectively deprotected leaving a deacetylated amine. Each amine can then be acetylated, or either left deacetylated allowing for the generation of 32 different PNAG pentasaccharides (Scheme 1.1.4).




Scheme 1.1.4: a) 90% TFA in $H_2O$. b) $Pd(PPh_3)_4$, $PhSiH_3$. c) Zn, AcOH. d) 20% piperidine in DMF. Each deprotected amine can be acetylated using $Ac_2O$ in pyridine, or left deacetylated, then a global deprotection can be performed using 20% hydrazine hydrate in MeOH.

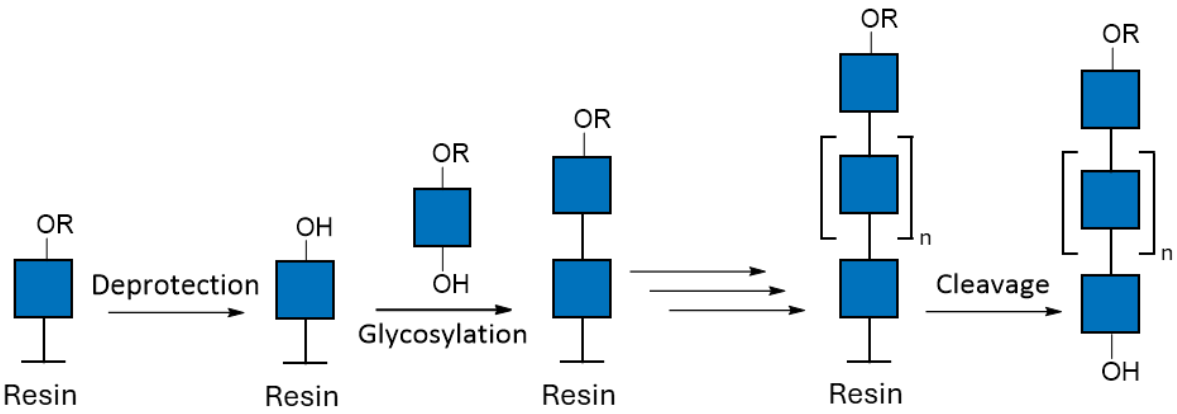
Whilst these strategies work well, they can be time consuming, often taking multiple weeks to complete due to the large number of steps involved. They also can be fairly low yielding, again due to the large number of steps leading to diminishing returns. Another method that has been used successfully in the synthesis of PNAG oligosaccahrides is called the acid reversion reaction[94,95]. This method involves reacting GlcNAc **1** monosaccharides together with hydrogen fluoride (HF) in pyridine (Pyr) (Scheme 1.1.5). This method has one reaction step and can be completed in as few as 15 hours[94], however reaction times up to 5 days can lead to better ratios of longer products[95].




Scheme 1.1.5: Acid reversion reaction of GlcNAc **1** to produce PNAG oligosaccharides where n≤6. a) 70% HF in Pyr, up to 5d.

This initially may seem appealing, with a quicker reaction time and comparable yields to the protecting group strategies, however, the purification can be challenging due to the mix of different oligosaccharide lengths. Homogeneity can be achieved using size exclusion chromatography (SEC), however this is also a time consuming process. The maximum size of the oligosaccharides produced is also lower than the protecting group strategies as well, being capped at 6. Finally, there are safety concerns surrounding the use of HF. HF can cause lethal burns, and as well as fluoride poisoning, if not handled with the proper care[96]. It also requires extensive quenching before any purification can be attempted. For these reasons, researchers have tended to avoid this method in recent years.

Improvements in the field of chemical peptide and nucleotide synthesis have led to the development of machines capable of reliable, automated, solid phase synthesis to rapidly produce the desired biomolecule[97,98]. This is a field in which polysaccharide synthesis has fallen behind. It wasn't until 2012 when Kröck *et al* published their method for automated oligosaccharide or polysaccharide synthesis, herein referred to as Automated Glycan Assembly (AGA), using a machine they call the Glyconeer that the field had anything comparable[99]. The Seeberger group had been working towards this with their work on automated synthesis of oligosaccharides in the early 2000s[82,100], and following their creation in 2012, another other group in the Demchenko group have created similar systems that allow for AGA[101], and at least one group has been able to utilise the Glyconeer to perform AGA[102]. AGA works by anchoring a selectively protected mono/oligosaccharide to some sort of resin, then washing the resin with the correct reagents to deprotect the hydroxy group you wish to glycosylate. The resin is then washed again to remove the reagents, but the leave the sugar attached. This can then be reacted with another mono/oligosaccharide, and a glycosylation reaction can occur on the resin. This can be performed iteratively, and when the sugar reaches the desired length, the resin can be washed with a reagent to cleave the sugar, which can be collected and purified (Scheme 1.1.6).




Scheme 1.1.6: General scheme for the process of AGA.

The automated nature of this method gives it a certain appeal, and certainly has been demonstrated to be effective, with it recently being used to create a 151-mer polysaccharide[103]. However, the technology is still prohibitively expensive, and whilst some have been able to repeatedly publish data using AGA, it is mainly confined to the two aforementioned groups, and it has only been able to produce PNAG oligosaccharides of a comparable size to regular synthetic routes[99].

From this we can see that there is yet to be developed a robust method for the production of PNAG polysaccharides comparable to biologically produced PNAG. These synthetic methods have provided substrates for the characterisation of PNAG degrading enzymes [85-88], as well as showing promise in the development of vaccine technology[3].

### 1.1.6. Immunological response towards PNAG

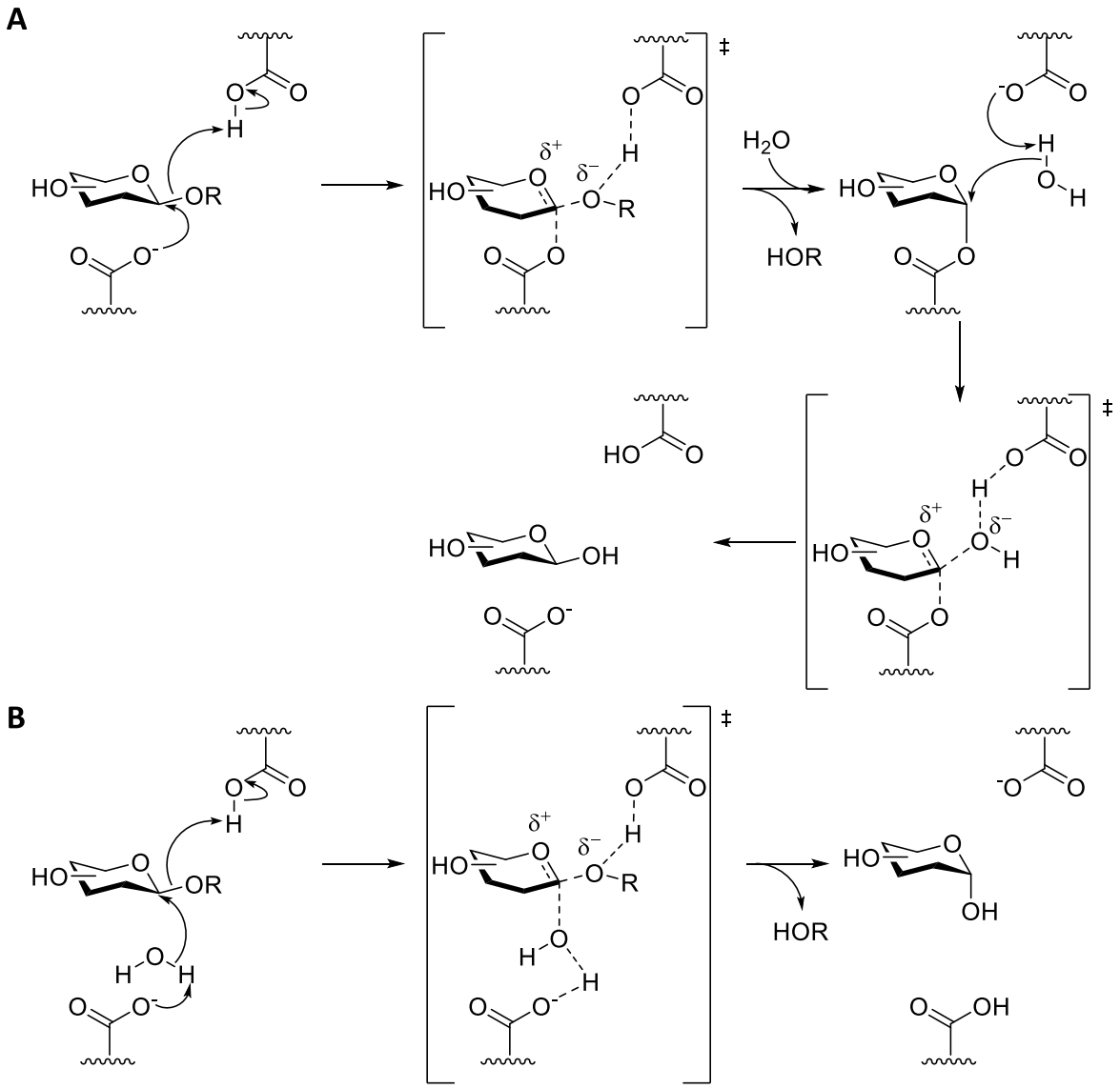
It has long been known that the human immune system has natural antibodies to PNAG, but these do not induce an effective immune response to PNAG presenting pathogenic species [104-106]. In fact, in order to be able to detect any immune system activity, conjugation to tetanus toxoid (TT) was required[78,84]. This was a known property of PNAG immunology as McKenny *et al* had previously shown conjugation to something antigenic was essential for immune system recognition[107]. The initial studies utilising these TT conjugated PNAG oligosaccharides were in determining the viability of PNAG as a vaccine antigen, and they used both PNAG and PG oligosaccharides to investigate this possibility. They found that the antibodies produced by the immune system in response to fully acetylated PNAG actually disrupted killing activity[55]. Furthermore, they showed that PG oligosaccharides were required to produce opsonic killing activity demonstrating the need for deacetylation in any potential vaccine candidate. However, a different study conducted by Pozzi *et al*[108] showed that killing activity of the antibodies produced by PG oligosaccharides caused interference in the opsonic killing activity of other antibodies to

*S*. *aureus* antigens, thus suggesting that another approach was required. These findings informed many attempts at producing a PNAG vaccine candidate, but as noted by Tan *et al*[3], it remained unclear whether or not using fully acetylated or deacetylated PNAG oligosaccharides rather than partially deacetylated PNAG oligosaccharides. With this in mind, Tan *et al* generated a library of different partially acetylated dPNAG pentasaccharides, which were then conjugated to antigenic proteins and used to induce immune to *S. aureus* infection in mice. This study demonstrated that for the most effective vaccine, a portion of the GlcNAc monomers have to be deacetylated. These findings also suggest that a vaccine utilising PNAG oligosaccharides is not only possible, but would be effective as a treatment against PNAG presenting pathogens. However, at the time of writing, no vaccine candidate has been developed for human trials, so this area of research is still ongoing. It should also be noted that no study has investigated the effect of *O*-succinylation on the immune response, which, as with the levels of *N*-deacetylation, could play a key role in the immunology of PNAG.

## 1.2. **Part 2: PNAG Degradation**

### 1.2.1. Glycoside Hydrolases

Glycoside hydrolases (GH) are carbohydrate active enzymes (CAZy) that catalyse the breakdown of a large variety of glycans via the cleavage of glycosidic bonds[109]. Enzymes within this group, known as the GH group, are split into families based on sequence similarity[110]. These enzymes can be furthered classified into retaining or inverting enzymes, where retaining enzymes retain the stereochemistry of the glycosidic bond in the new anomeric centre, where inverting enzymes invert the stereochemistry (e.g. $\alpha \rightarrow \beta$). The GH enzymes all utilise a catalytic pair of acidic residues that act as both the nucleophilic catalyst, and as the proton donor. The major difference between the two types of GH enzymes comes from the distance between the two catalytic residues, with the inverting enzymes having a greater distance between the two to allow for the water molecule to attack from below the ring[111,112]. It was eventually revealed that the mechanisms also proceed through an oxocarbenium ion[113] (Scheme 1.2.1).




Scheme 1.2.1: A) Generalised scheme for the mechanism of a retaining β-glycoside hydrolase enzyme. B) Generalised scheme for the mechanism of an inverting β-glycoside hydrolase enzyme. For the α mechanism for both the retaining and inversion mechanisms, the two residues swap function with the top residue becoming deprotonated and acting as the nucleophile, and the bottom residue acting as the proton source.

These enzymes can also be classified as *exo* or *endo* which refers to whether or not the enzyme cleaves the first glycosidic bond of a polysaccharide chain, usually at the non-reducing terminus, or if the enzyme will cleave the polysaccharide in the middle of the chain (Figure 1.2.1).

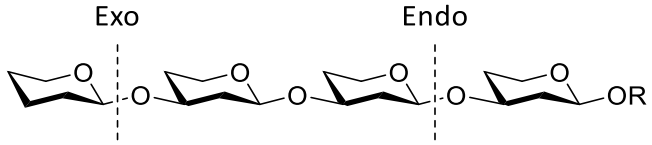


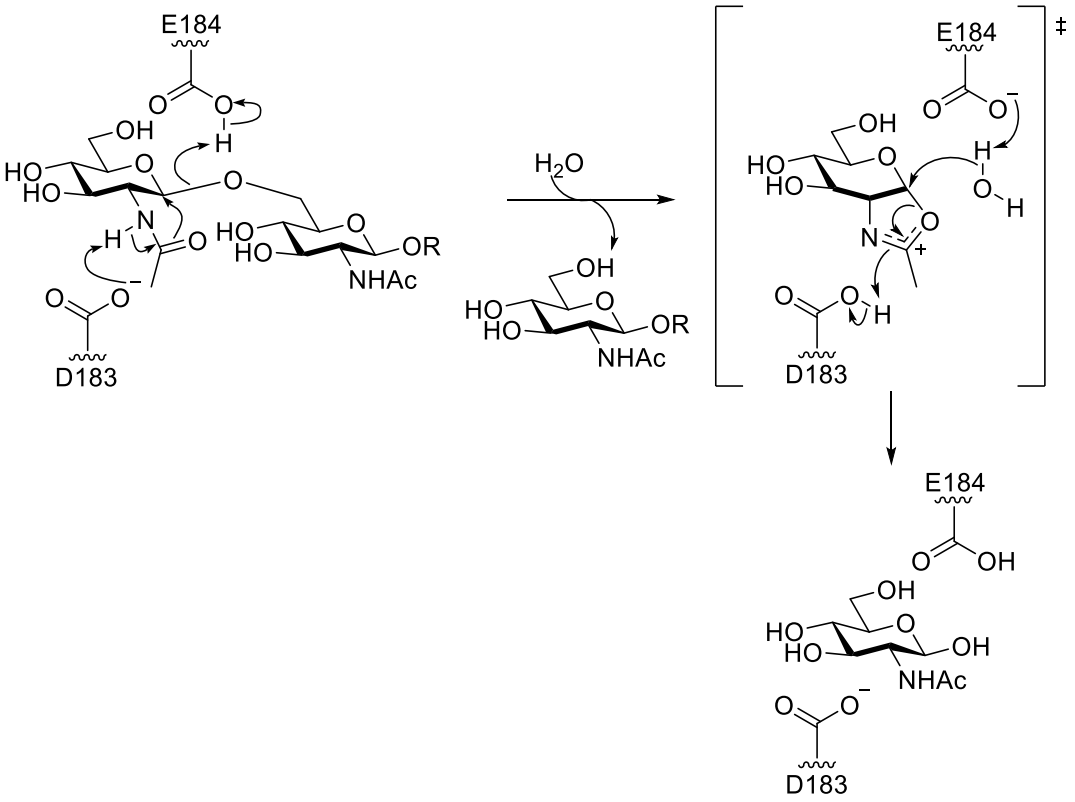

Figure 1.2.1: The locations of the glycosidic bond cleavage of exo acting glycoside hydrolases and endo glycoside hydrolases

Certain families can have slightly different mechanisms, such as a neighbouring group participation (NGP) mechanism, where a functional group present on the sugar ring participates in the cleavage of the glycosidic bond. This sort of mechanism is the most common enzymes that break down glycosidic bonds where there is an *N*-acetyl sugar, such as GlcNAc, at the non-reducing terminus of a glycosidic bond, however it has been shown to occur with hydroxy groups as well[114]. One family that perform this mechanism is the GH20 family, and this family contains a protein of particular interest to this study.

### 1.2.2. The GH20 family and Dispersin B

The GH20 family, which is sometimes referred to as the β-*N*-acetylhexosamidase enzymes, characteristically cleave glycosidic bonds between a GlcNAc residue and another sugar moiety. This is typically a β-1,4-linkage, although individual members of the family have been shown to catalyse the hydrolysis of different GlcNAc glycosidic linkages as well[115]. In particular, a member of this family isolated from *Aggregatibacter actinomycetemcomitans*, Dispersin B (DspB), first reported in 2003[116], has been found to cleave the GlcNAc-GlcNAc β-1,6-linkage found in PNAG polysaccharides[117], which until very recently was the only known member of the GH20 family to be able to do so[118], alongside the only known non-GH20 enzyme with PNAGase activity, the aforementioned PgaB[70]. Interestingly though, this enzyme can act as both an *endo* and *exo* glycoside hydrolase with an initial preference for *endo* activity[79,85-87,119,120], in contrast to PgaB which can only act in an *endo* manner[70]. Initially, clinical interest in this protein was primarily focussed on its use in biofilm disruption and detachment[117] to help clean medical devices and implants, and although this approach has potential as an *ex-vivo* anti-microbial strategy, it is likely limited for use targeting pathogens in the lungs or other organs *in-vivo*.

The mechanism of GH20 enzymes, including DspB, has been well studied, and the hydrolysis reaction proceeds via a retaining exo mechanism involving neighbouring group participation from the *N*-acetyl group at position-2 forming an oxazolium ion intermediate[121-124]. The DspB specific mechanism proceeds with E184 acting as the acid/base catalyst of the reaction, whilst D183 acts to activate the nucleophilic *N*-acetyl group and stabilise the resulting oxazolium ion formed transition state[119] (Scheme 1.2.2).




Scheme 1.2.2: The β-retaining neighbouring group participation (NGP) mechanism of DspB.

As part of their study that allowed them to propose the DspB mechanism, Manuel *et al* also conducted an extensive study of the residues found in and around the active site [119]. Through mutagenesis assays, they identified several residues that were essential for DspB activity. Amongst these were the two catalytic residues, D183 and E184, and another glutamic acid residue, E332. These three residues form part of a larger anionic groove on the surface of the protein first demonstrated by Breslawec *et al*[85]. Having previously demonstrated that DspB has faster rates of hydrolysis on dPNAG oligosaccharides[87], they noticed that the electrostatic surface charge of DspB shows a number of negatively charged residues forming a groove leading into the catalytic pocket and out of it on the other side (Figure 1.2.2A). This groove contains four potential binding sites, including the active site, which are given the numerical designations of -2, -1 (active site), +1, and +2, with each binding site corresponds to a GlcNAc monomer along the PNAG chain ( Figure 1.2.2B).

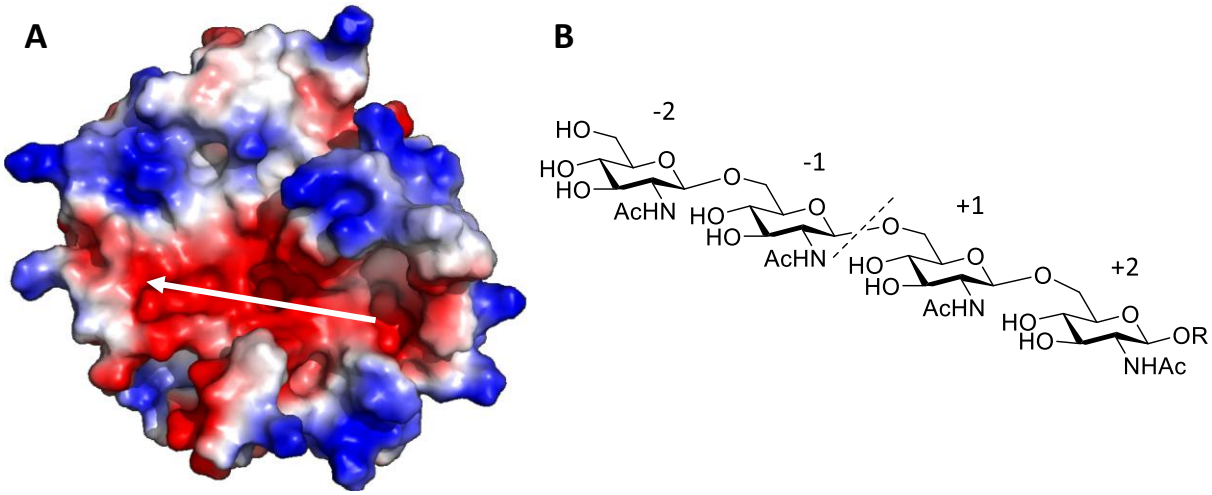



Figure 1.2.2: A) Surface electrostatics showing the anionic charge (red) that lines the groove in DspB's surface formed by the PNAG binding sites (white arrow). Structure of DspB obtained from PDB entry 1HYT[116]. Image produced using PyMOL Molecular Graphics System[125]. B) The 4 GlcNAc monomers bound by each potential binding site in the anionic groove on DspB's surface. The dashed line represents the location of the cleavage performed by the enzyme.

They hypothesised that this groove was conferring the substrate preference for dPNAG as the deacetylated regions of dPNAG carry a positive charge and as such would bind better into the potential binding sites. To test this, they mutated various residues along this anionic binding groove. For this initial study they selected three residues that lined the groove, D147, D242, and E248, and mutated each residue to remove the anionic character. Then, using selectively deacetylated dPNAG trisaccharides with each mutant, they were able to determine which residues were important in recognising deacetylated GlcNAc. Fully acetylated GlcNAc trisaccharide was used as a control, and surprisingly, when their E248Q was incubated with the glycan, the rate of hydrolysis observed was approximately 5 times greater than the WT. When mutated to an alanine, activity returned to WT levels, suggesting that the E248 is important for substrate recognition of deacetylated GlcNAc, and that the mutation at this position may afford a more efficient biocatalyst when using fully acetylated GlcNAc substrate over partially deacetylated substrates.

D147 was revealed to form the +2 site, as hydrolysis activity of a GlcNAc trisaccharide with an $NH_3^+$ group in the +2 site was decreased when this residue was mutated. D147 is flanked by two other negatively charged residues in D146 and E148, so it is likely these residues are involved in binding in this area as well, as mutation of D147 did not see as dramatic of a reduction in hydrolysis activity as mutation of other residues. E245 and E248 mutants both also saw a reduction in hydrolysis activity of a trisaccharide with an $NH_3^+$ in the +1 site, suggesting these residues form part of this binding site[85]. In a further study by Breslawec *et al*, a further three residues were selected to be mutated, D242, E332, and D333[86]. D242 showed similar results to D147 when a D242N mutant was incubated with a trisaccharide with an $NH_3^+$ group in the -2 position. This would suggest D242

binds the substrate in this region. Further to this, D242 mutants showed a significant drop off in *endo* glycoside hydrolase activity compared to wild type. This would suggest that D242's recognition of deacetylated PNAG monomers is what confers DspB's preference for *endo* activity. Mutants of residues E332 and D333 were also mutated as part of this study, and this resulted in complete loss of activity, which corresponds with Manuel *et al*'s investigation E332 which showed this residue was essential for DspB activity[119], and equivalents residues in *Streptomyces plicatus* hexosaminidase[126] and *Serratia marcescens* chitobiase[127] to E332 form hydrogen bonds with the C4 hydroxy group of the GlcNAc monomer in the active site, suggesting these residues form part of the -1 site.

Taken together these experiments paint a picture of the constituent residues of each binding site (Figure 1.2.3A), and the possible mechanisms of action of DspB. DspB can either move along a PNAG chain cleaving in an *exo* manner (Figure 1.2.3B), or it can bind onto the PNAG chain so that a deacetylated monomer is in the -2 binding site, where it then cleaves the PNAG in an *endo* manner (Figure 1.2.3C).

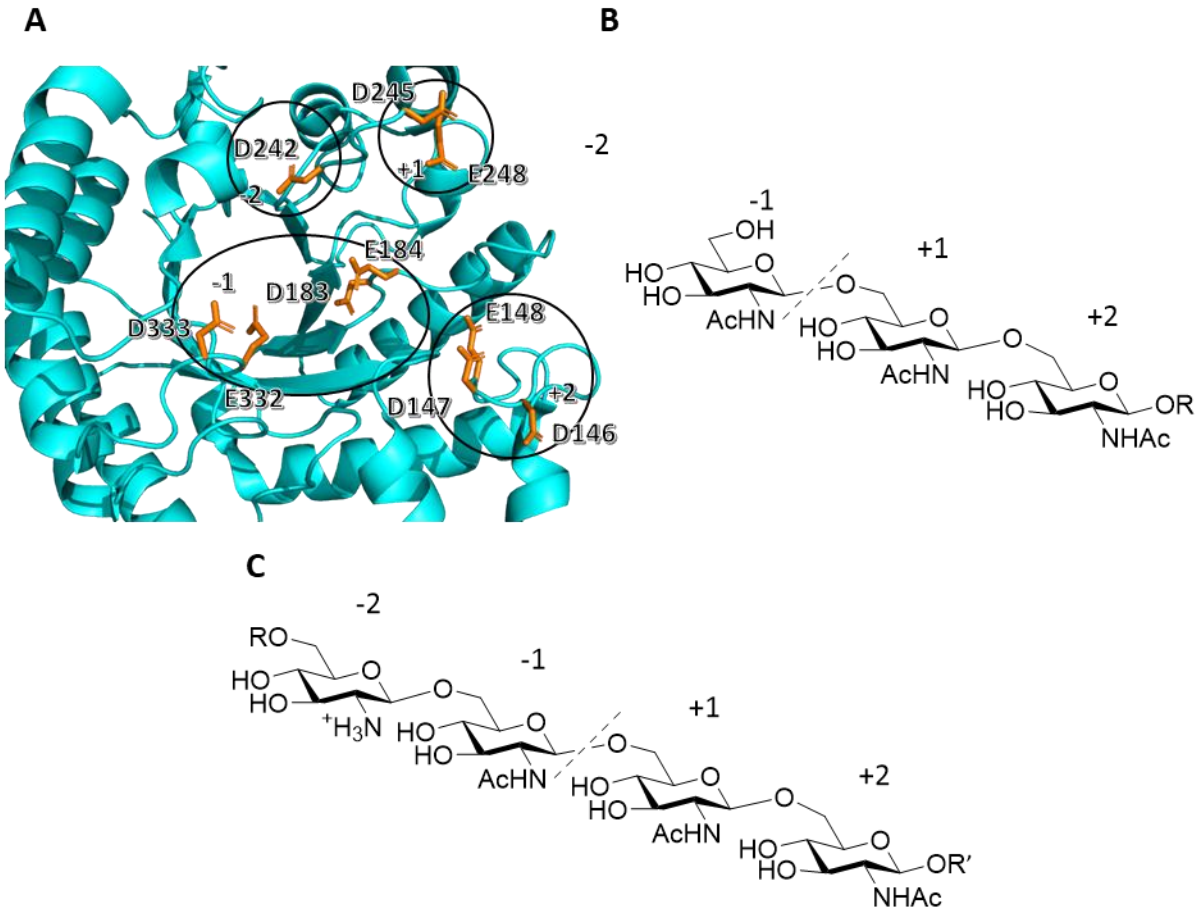



Figure 1.2.3: A) Residues that form the anionic binding groove and the binding sites of DspB, including the active site. Structure of DspB obtained from PDB entry 1HYT[116]. Image produced using PyMOL Molecular Graphics System[125]. B) Position of the GlcNAc monomers of the PNAG polymer in the binding sites during the exo mechanism of DspB. The -2 binding site remains unoccupied as DspB moves down the PNAG chain cleaving one GlcNAc monomer at a time. C) Position of the GlcNAc monomers of the PNAG polymer in the binding sites during the endo mechanism of DspB. The deacetylated monomer binds preferentially into the -2 binding site leading to multiple GlcNAc monomers being cleaved at once.

### 1.2.3. PgaB – the other fully characterised PNAGase

DspB has been extensively studied due to the current rarity of PNAGases in the literature, the same is also true of the other known fully characterised PNAGase, PgaB, which until 2025 was the only other known PNAGase[118]. As discussed previously, PgaB is encoded within the *pgaABCD* locus responsible for the biosynthesis of PNAG in *E. coli*. PgaB's function was initially determined to be as a *N*-deacetylase protein in 2008 by Itoh *et al*[69], and in this paper, they identified a second C-terminus domain that had sequence identity with a β-galactosidase *T. thermophilus*. This led them to hypothesise that this domain was responsible for PNAG binding, as deleting this region caused a loss of *N*-deacetylation. Little *et al*[68] further explored this hypothesis in 2012, and their analysis of the structure of the C-terminus showed there was a similarity to DspB, however, unlike DspB, and

other GH enzymes with similar structures, they were unable to show activity using PNAG oligosaccharides or *para*-nitrophenyl GlcNAc **4** (*p*NP-GlcNAc **4**), a commonly used substrate for displaying GH activity (Figure 1.2.4).

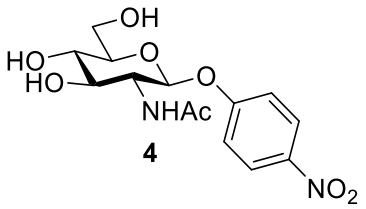



Figure 1.2.4: Structure of *p*NP-GlcNAc **4**. The *para*-nitrophenyl group both acts as a mimic of a second sugar ring, and as a chromaphore when cleaved from the sugar allowing for monitoring via UV/Vis Spectrometry.

This led to them also concluding that the C-terminus was just important for binding to the PNAG and aiding in the *N*-deacetylase activity, as they also saw a complete loss of *N*-deacetylase activity upon deletion of this region of the protein. It wasn't until 2018 that Little *et al*[70] were able to show GH activity in PgaB. Having proven the hypothesis that PgaB's C-terminus was involved in PNAG binding in 2014[128], they set out to demonstrate GH activity in both the PgaB of *E. coli* and of *B. bronchiseptica* by utilising dPNAG isolated from biological sources such as *S. aureus*, and tested both the full protein with the *N*-deacetylase domain, and a truncated version with just the GH domain. These experiments showed that these enzymes were able to hydrolyse dPNAG, producing the trisaccharide GlcN-GlcNAc-GlcNAc, showing this protein operated with *endo* activity. Furthermore, they showed that the GH domain of PgaB was capable of hydrolysing dPNAG produced *in situ* by PgaCD, but this only occurred when the *N*-deacetylase domain was present, demonstrating the intrinsic link between these two functions.

This discovery created a new GH family, GH153, and expanded the known number of PNAGases at the time to two. However, it would probably be more accurate to describe PgaB as a dPNAGase given that it requires deacetylation to function. This, and the fact that PgaB is a periplasmic protein, separate it from DspB, which is excreted by the cell.

### 1.2.4. Putative PNAGases in *Staphylococci* species

Despite the extensive study of the *icaADBC* locus in Staphylococcal species, no enzyme with similar activity to DspB has been characterised in any Staphylococcal species to date[129]. The fact that biofilm producing *staphylococci* species have been shown to produce solely PNAG as the only extracellular polymeric substance (EPS) present in their biofilm[130] makes this an even more apparent hole in the current understanding of PNAG within biological systems. Frank *et al* identified an ORF encoding a protein with high sequence identity with DspB in *S. lugdunensis* isolates IDRL-

2414 and IDRL-2664[131] which they hypothesised is a novel glycoside hydrolase of the GH20 family. Upon further investigation by Andrew Higginson, a member of the Thomas Lab (University of York)[132], using this sequence as the query for a BLASTp search in the UniProtKB and NCBI databases, a further 16 *staphylococci* proteins were identified, alongside two other proteins from *T. goriensis* and *M. lamae* respectively. A number of homologous proteins to DspB (>50% sequence identity) were also identified via a BLASTp search. These novel *staphylococci* proteins have high enough sequence identity with DspB (≈ 28%) to suggest that they are in fact PNAGases, however, upon further inspection, the protein sequence is lacking some key features of DspB that potentially point to different substrate specificity.

A groove in DspB's surface leading into the active site that contains several anionic residues identified by Breslawec *et al* (Figure 1.2.2A) has been previously shown to confer a substrate preference to DspB for dPNAG, particularly when a deacetylated monomer sits in the +2 position[85,87]. D147, alongside the neighbouring residues D146 and E148, was revealed to form the +2 site (Figure 1.2.3). This residue is relatively well conserved among the 19 sequences with 95% having a D/E residue within 2 positions of the equivalent residue, however the full DDE motif is not very well conserved (Figure 1.2.5).

The residues that form the other non-active site binding sites, D242, E245, and E248 form part of an α-helix that lines one side of the binding groove. When we look at the sequence of the putative PNAGases, we see that they all lack the residues that make up this α-helix, and therefore lack the residues that form two important binding sites. This combined with the lack of the full DDE might suggest that these proteins would show less of a preference towards dPNAG. In fact, Breslawec *et al* found through their mutational studies of DspB that an E248Q mutant showed an increased preference for fully acetylated PNAG[85], and the *staphylococci* PNAGases all have a glutamine residue within two residues of the equivalent residue to E248. This perhaps suggests these PNAGases preferentially hydrolyse fully acetylated PNAG (Figure 1.2.5).

These putative *staphylococci* PNAGases can be further categorised by whether the gene for the putative PNAGase is found upstream of *icaA* or if it is convergent with *icaC*. These two clades were named after a highly conserved Q (convergent) or conserved K (upstream) in place of the E248 residue of DspB (Figure 1.2.5). The two separate clades have a few differences in their sequences, however it is unclear if this will have any impact on function or substrate activity. They are both, however, more similar to each other than to DspB suggesting similar roles despite this difference. Why these *staphylococci* proteins would have evolved to have a substrate specificity for fully

acetylated PNAG is unclear. There are three hypotheses within the research group at the time of writing.

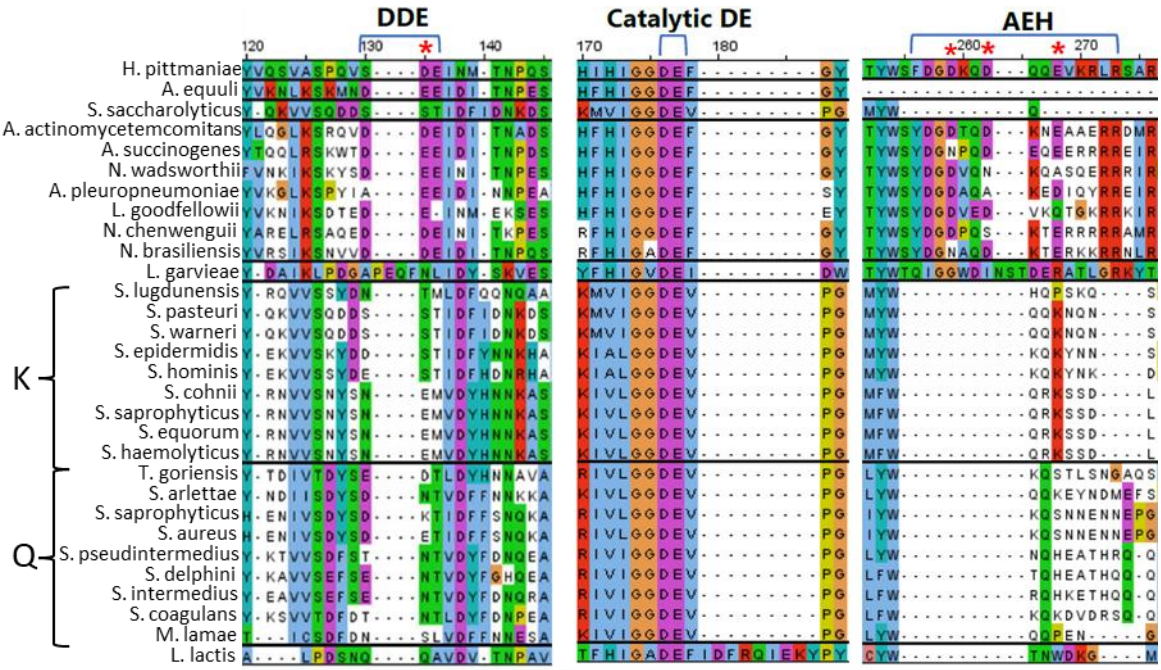



Figure 1.2.5: Sequence alignment of the 19 putative PNAGases, with DspB and the other DspB homologues. Figure created by Andrew Higginson[132], reused with permission. DspB's sequence is marked with a black asterix. Residues D147, D242, D245, E248 are marked with a red asterisk. The DDE motif found around D147 in DspB, DspB's active site, and the location of the α-helix (AEH) are also marked. The K-clade and Q-clade are also marked.

Firstly, it is possible these proteins are an evolutionary remnant from before Staphylococcal species began producing deacetylated PNAG. This hypothesis is supported by the fact that phylogenetic analysis of the 19 putative PNAGases reveals that a number of these proteins evolved before DspB. This would therefore suggest that initially most species produced fully acetylated PNAG, but some evolved to deacetylate their PNAG in order to disrupt other species ability to breakdown their biofilms. Then *A. actinomycetemcomitans*, amongst others, evolved in response to begin producing DspB, a protein better able to hydrolyse dPNAG. This could also explain why Staphylococcal species have begun *O*-succinylating their PNAG[57-59], as this would place a negative charge in the anionic binding groove, disrupting the ability of the protein to hydrolyse the PNAG.

The second hypothesis states that this protein evolved because of *staphylococci* species beginning to *O*-succinylate their PNAG. The potential presence of a +2 binding site would support this theory as this could point a mechanism for allowing both the recognition of dPNAG, whilst also stopping complete disruption of substrate recognition by the presence of an *O*-succinyl group. As Staphylococcal species have been shown to have both an N-deacetylase and a potential *O*-succinyltransferase within the *icaADBC* operon, which is responsible for the biosynthesis of PNAG

in *staphylococci* species[59]. This theory is also supported by the fact that *A. actinomycetemcomitans* is a dental pathogen, whilst the two strains of *S. lugdunensis* in which the initial putative PNAGase was discovered, IDRL-2414 and IDRL-2664, were isolated from a patient with endocarditis[133], and from a prosthetic joint infection[134] respectively. This perhaps suggests that the gene encoding for these PNAGases evolved separately from the selection pressure of competing microorganisms producing a DspB like protein.

Finally, the third hypothesis states that these proteins have evolved as a way of regulating PNAG production. The ORF that encodes for the putative *S. lugdunensis* PNAGase (SlHex) was found upstream from the *icaADBC* operon in place of IcaR, and no further homologue could be found within the genome[131]. Upon further investigation into the genes encoding for all 19 putative PNAGases, it was discovered that ORF encoding for the PNAGase is closely associated with the *icaADBC* operon in all 19 species[132]. IcaR is a negative regulator of the *icaADBC* operon[135-137], meaning its absence in *S. lugdunensis* perhaps suggests another protein is fulfilling the same role in some capacity, as it is unlikely the operon is fully unregulated. It is therefore possible that the putative PNAGases fulfil this role, possibly by cleaving PNAG strands formed by a malfunctioning *N*-deacetylase. Alternatively, the two proteins could potentially form a heterodimer with a part of the operon, perhaps effecting a structural change upon doing so. This would mimic the dual function of PgaB, and perhaps the putative glycoside hydrolase is required for full *N*-deacetylase activity in this instance as well.

One of the aims in this thesis was to test these hypotheses, and to investigate whether or not these proteins are in fact PNAGases. Three of the putative PNAGases were selected for investigation, the aforementioned SlHex from *S. lugdunensis*, SaHex from *S. aureus*, and ShHex from *S. hominis*. These were chosen as SlHex was the original query sequence, and SaHex and ShHex are members of the Q clade and the K clade respectively.

## 1.3. Part 3: Glycosynthases and a new path towards PNAG synthesis

### 1.3.1. Glycosynthases – history and theory

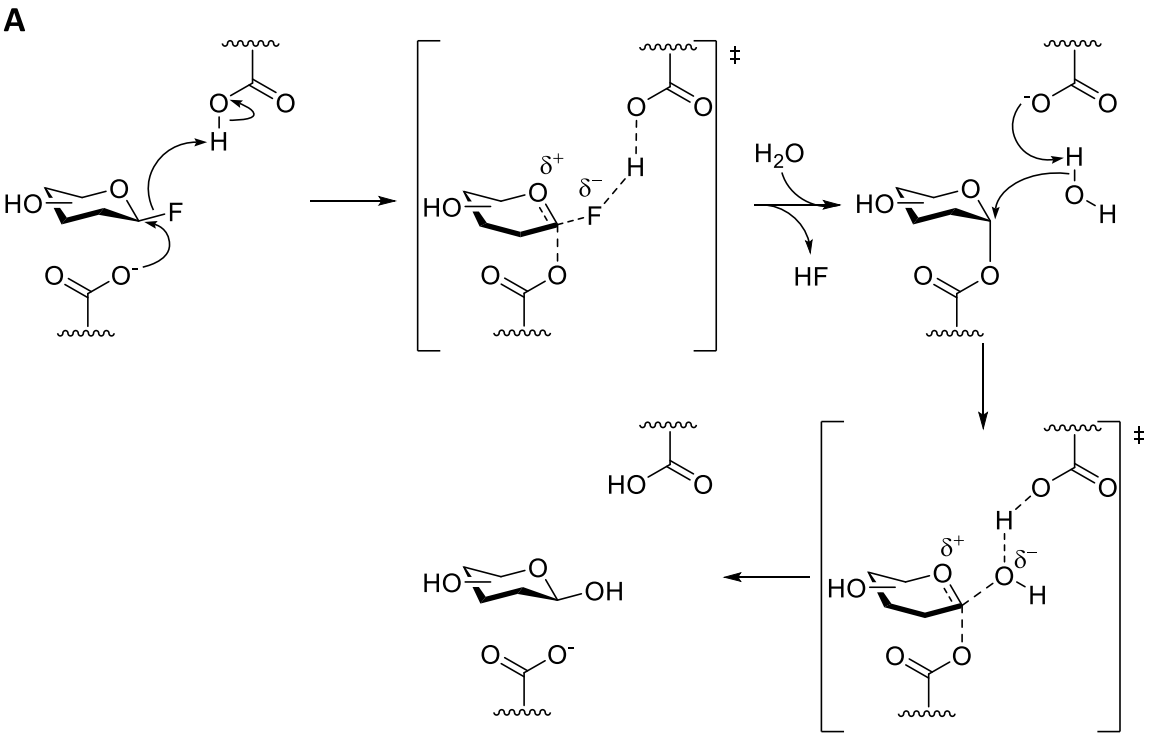
The various mechanisms of GH type enzymes are well understood and extensively studied. It was through these studies it was first noticed that if the nucleophilic residue in the active site was mutated to a non-nucleophilic residue, the protein still folded properly, yet had no GH activity[138]. This is important, as the protein being properly folded meant the enzyme was still capable of binding oligosaccharides, and therefore monosaccharide molecules. However, it wasn't until an interesting quirk of the reaction of enzymes utilising glycosyl fluorides was realised that a whole new field of study became apparent that could be exploited to generate oligosaccharides.

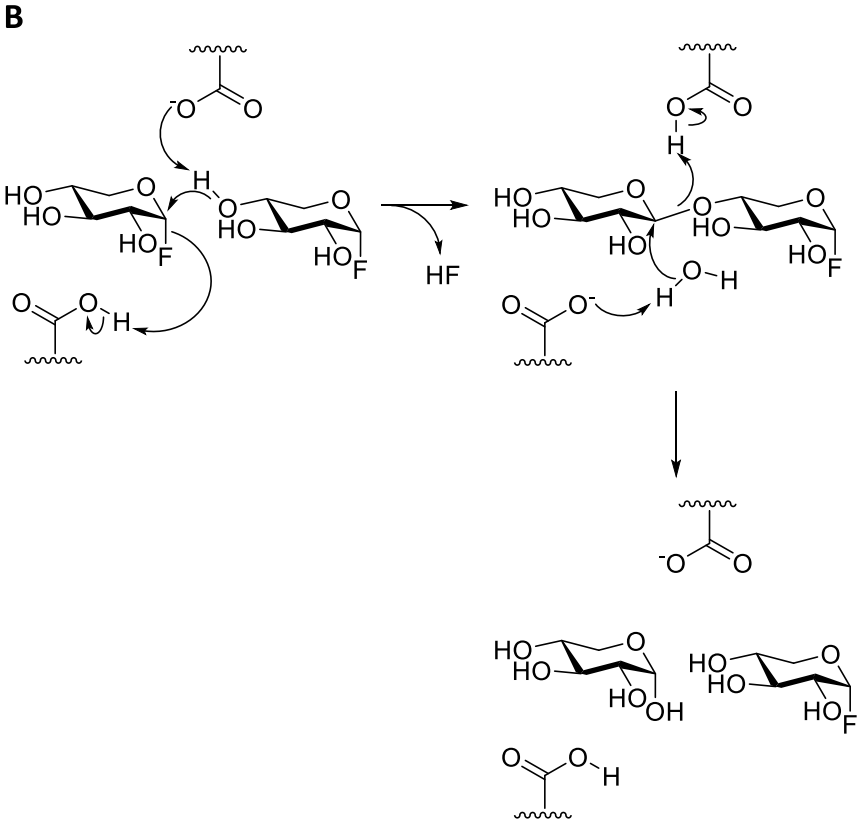
It has long been known that GH-enzymes can hydrolyse glycosyl fluorides, with this being discovered in 1961 by v. Hofstein[139]. When presented with a glycosyl fluoride, the enzyme will proceed as normal as if breaking a glycosidic bond, however, in the released product will be HF alongside a monosaccharide, rather than the oligosaccharide released during the normal mechanism (Scheme 1.3.1A). It was quickly realised these substrates have a very high $\left(\frac{k_{cat}}{K_M}\right)$ compared to other substrates. This, coupled with the fact that they don't require acid catalysis[140], and that they will readily react with nucleophiles in aqueous solutions[141], has led to their extensive use in the study of GH enzymes, both to determine their stereochemistry[142,143], and to determine the individual roles of amino acid residues via mutagenesis studies[144].

It was during one of these studies it was also discovered that GH enzymes can hydrolyse glycosyl fluorides with the incorrect stereochemistry[145] (i.e. a α-glycosyl fluoride being hydrolysed by an β-GH enzyme). What was most interesting about this though, was that enzymes proceeded via what is now known as the Hehre resynthesis-hydrolysis mechanism. The enzymes binds two of the glycosyl fluoride monosaccharides into the active site, and performs a transglycosylation reaction. This causes HF to be released and a disaccharide molecule with the correct stereochemistry to be generated *in situ* of the active site. This disaccharide molecule being a substrate of the enzyme is quickly hydrolysed, inverting the stereochemistry again, giving the normal hydrolysis product which now has the same stereochemistry as the original glycosyl fluoride (Scheme 1.3.1B).

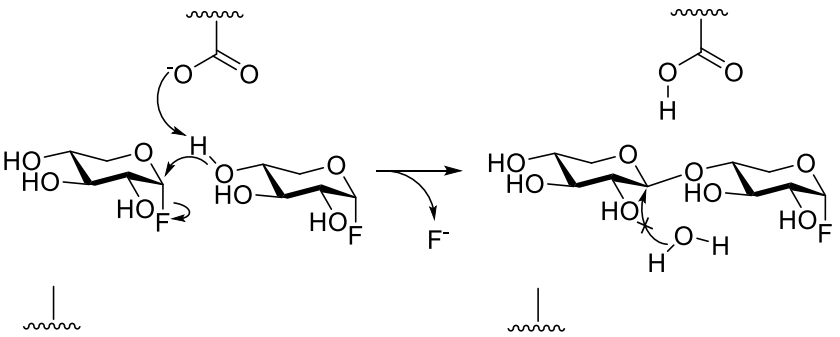
This transglycosylation behaviour caught the attention of Mackenzie *et al*[1]. They realised that the transglycosylation only requires the residue that normally acts as the proton donor. As the group had previously shown that the nucleophilic residue can be mutated and not disrupt the protein folding, they hypothesised that if you mutate this residue to an alanine, and then give the

resulting mutated enzyme a glycosyl fluoride substrate, the transglycosylation reaction would proceed, but the subsequent hydrolysis reaction wouldn't be able to (Scheme 1.3.1C). This would mean the resulting enzymes were capable of producing oligosaccharides. Mackenzie *et al* tested this with a D358A mutant of a β-GH from *Agrobacterium* that catalysed the hydrolysis of glucose and galactose oligosaccharides. When incubated with the glycosyl fluoride substrate and an aryl glycoside acceptor, the enzyme was successfully able to produce disaccharides and trisaccharides using both glucose and galactose as substrate, proving the hypothesis correct, and creating a new class of, synthetic, enzyme, which they called Glycosynthases[1].


A
HO
O
F
H
δ+
δ−
‡
H2O
HF
OH


B
HO
O
F
H
HF
OH

**C**



Scheme 1.3.1: A) Generalised scheme for the mechanism of a retaining β-glycoside hydrolase enzyme hydrolysing a glycosyl fluoride. B) Hehre resynthesis-hydrolysis mechanism[145] showcasing an α-inverting enzyme utilising the wrong anomeric conformation of the glycosyl fluoride substrate whilst still producing the correct anomeric conformation in the product. C) The mechanism of a glycosynthase proposed by Mackenzie *et al*[1].

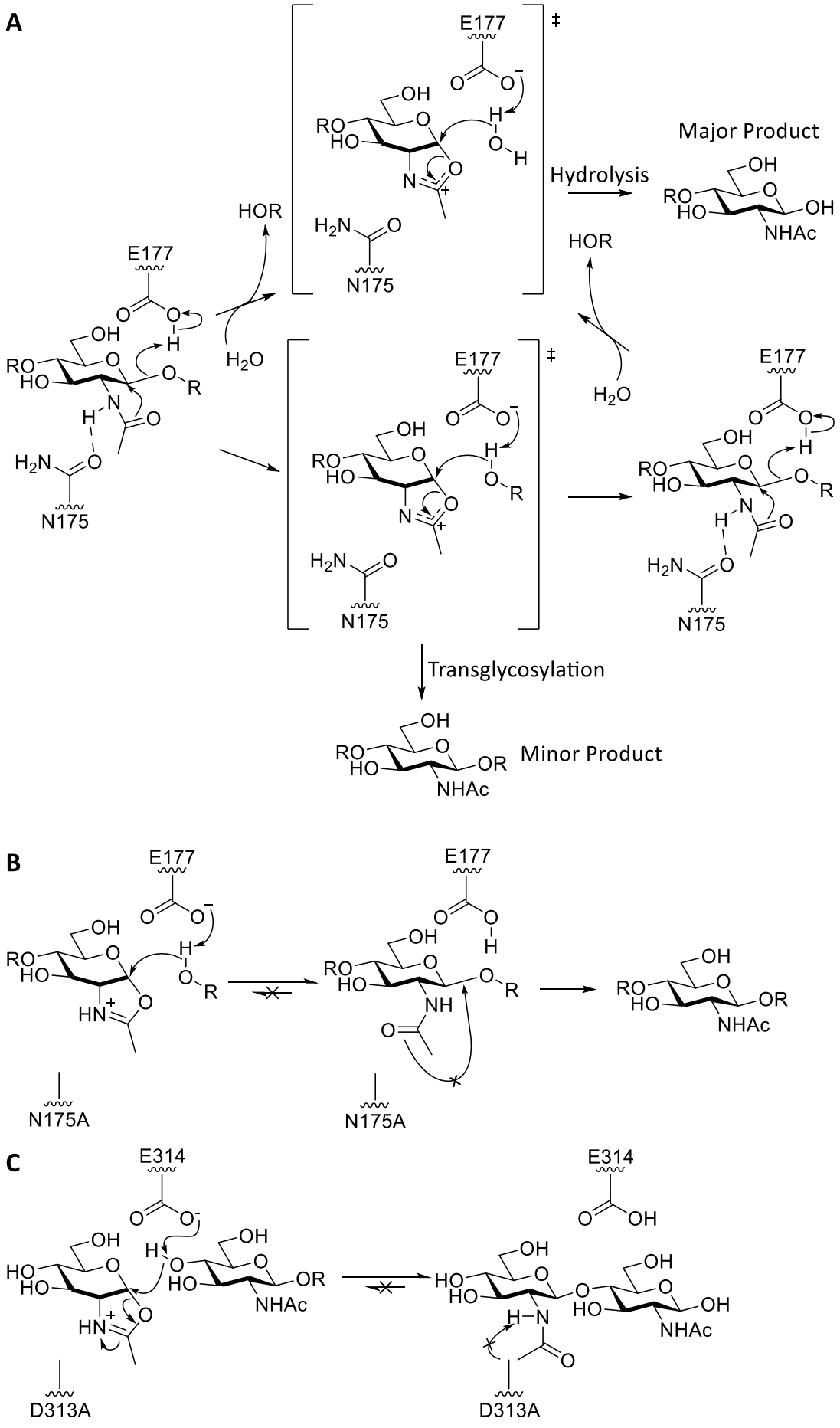
The methodology behind the production of glycosynthases since their discovery has increased the scope of the type of oligosaccharides they can produce. There now exist glycosynthases capable of producing galactoses, mannoses, xyloses, and fucoses[146], amongst others. Glycosynthases allow access to the production of oligosaccharides where the corresponding glycosyltransferase enzyme might otherwise not, for example if the glycosyltransferase enzyme is membrane bound versus the glycosynthase which is not associated to any membrane. Glycosyltransferase enzymes also often require an expensive substrate that mimics the biological precursor molecule, such as UDP-Glc[147], whilst glycosynthases can operate with inexpensive glycosyl derivatives. However, the downside to glycosynthases is that they require a corresponding GH enzyme to exist, and it must be capable of being turned into a glycosynthase, which is not always possible.

### 1.3.2. Can DspB be made into a glycosynthase?

In this review, various synthetic methods for producing PNAG have been discussed, as well as the biosynthetic pathway which produces it *in vivo*. However, it has yet to be shown whether the glycosynthase method can be used to produce PNAG oligosaccharides. Of the two known PNAGases, DspB would be the more suitable enzyme for this task, being non-membrane associated, and only containing the one catalytic domain. However, DspB acts via the NGP mechanism, whilst all of the previous examples of glycosynthases have discussed turning a typical GH hydrolase, i.e. one that operates via the oxocarbenium intermediate, into a glycosynthase. NGP GH enzymes cannot be converted into glycosynthases in the same way that other GH enzymes can. Glycosynthases that utilise glycosyl fluorides as a substrate require the removal of the nucleophilic residue from the protein to function. However, in NGP mechanisms, the nucleophile is the *N*-acetyl

group (Scheme 1.2.2) and therefore cannot be removed, and glycosyl fluorides cannot be utilised in the same way. The development of a NGP glycosynthase was therefore slowed and progress was not made until 2008.

The first glycosynthase of a NGP GH was created by Umekawa *et al* when they produced an N175A mutant of EndoM[148]. EndoM is a GH85 enzyme that hydrolyses *N,N*-diacetylchitobiose moieties found in *N*-glycans. It is known to have low levels of transglycosylation activity, but the major product is the hydrolysis product (Scheme 1.3.2A). The mutation used for this enzyme differs from the classic method of creating a glycosynthase as this residue is not part of the active site, but did still remove hydrolysis activity by stopping the enzyme from being able to properly orientate the *N*-acetyl group for oxazolium formation. The other major difference was the choice of substrate. To circumnavigate the issue of the *N*-acetyl group acting as the nucleophile, Umekawa *et al* proposed that a transition state mimic, a sugar with an oxazoline ring formed between the anomeric centre and the *N*-acetyl group, can be used instead. They hypothesised that this substrate could bind into the active site and react with another oligosaccharide to undergo transglycosylation by breaking the oxazoline ring open, regenerating the *N*-acetyl group. As the mutated enzyme cannot produce the oxazolium intermediate, the new glycosidic bond cannot be hydrolysed, and the enzyme acts a glycosynthase (Scheme 1.3.2B). When presented with an oxazolium mimicking substrate, the EndoM N175A mutant did show glycosynthase activity successfully proving NGP GH enzymes can be converted into glycosynthases. Later in 2012 it was then shown that NGP GH enzymes can also be converted into glycosynthases through the same mutation as the traditional glycosynthase method as long as the substrate remains as the oxazoline[149] (Scheme 1.3.2C). Since then there have been multiple examples of glycosynthases of NGP enzymes that utilise both types of mutations to generate the glycosynthase, however the substrate scope has been expanded to include *para*-nitrophenyl sugars as well[150-153].


A
E177
N175
HOR
H2O
Hydrolysis
Major Product
Transglycosylation
Minor Product
B
N175A
C
E314
D313A

Scheme 1.3.2: A) The wild type mechanism of EndoM showing both the transglycosylation and hydrolysis mechanistic pathways. N175 is pictured hydrogen bonding to the proton of the *N*-acetyl group to correctly orientate it, and allow it to act as a nucleophile for hydrolysis. B) The glycosynthase mechanism of the mutated N175A EndoM. The *N*-acetyl group is unable to be correctly orientated to allow for hydrolysis so only the transglycosylation reaction can proceed. C) Example of the mutated GH20 enzyme SpHex acting as a glycosynthase, as presented by Tegl *et al.*[154]. The mutated D313A residue can no longer activate the *N*-acetyl group to act as a nucleophile and so the hydrolysis reaction cannot proceed, however the E314 residue can deprotonate the acceptor sugar, causing a transglycosylation reaction to occur.

Whilst no studies have been published on the conversion of DspB specifically into a glycosynthase, other studies on different GH20 enzymes provide evidence that it is indeed possible for this GH family. Santana *et al* showed that the GH20 enzyme SpHex could be converted into a glycosynthase via the traditional mutation of the active site to remove the activating residue[153]. Another study by Slámová *et al.*[155] investigated the GH20 enzyme TfHex from the fungus *Tessaracoccus flavus*, which is active on GlcNAc-β,1,4-GlcNAc, to see if they could create a glycosynthase from this enzyme. They selected residue Y470 for mutation as it is involved in stabilisation the oxazolium intermediate[156], and therefore may affect hydrolysis activity, potentially promoting glycosynthase activity. They investigated Y470F, Y470H, and Y470N mutants and noted that all three had glycosynthase activity when presented with *p*-nitrophenyl-GlcNAc (pNP-GlcNAc), with the Y470N mutant so effective, the resulting oligosaccharide product appeared to precipitate out of solution due to its length. Interestingly, the mutants showed no activity towards an oxazoline substrate. The residue Y470 in TfHex is conserved as residue Y278 in DspB (Figure 1.3.1).

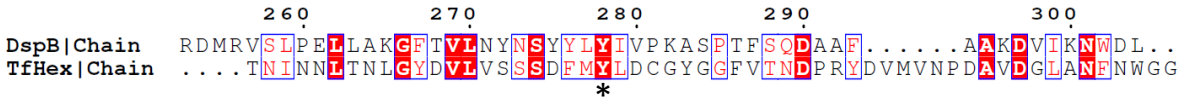



Figure 1.3.1: Sequence alignment of DspB and TfHex showing DspB residue Y278 and the equivalent Tyr residue in TfHex, Y470 (marked with a *).

Manuel *et al.*[119] had previously performed a comprehensive study into the active site residues of DspB to determine which were important for hydrolysis activity. Y278 was among those investigated, and a Y278A mutant was found to have drastically reduced (<10% of the wild type (WT) DspB) hydrolysis activity. This is consistent with TfHex, and therefore suggests that a mutant similar to that targeted in TfHex may also promote glycosynthase activity in DspB.

A further two studies investigating GH20 enzymes from *Bifidobacterium bifidum*, known as BdhI and LnbB, found both could be converted into a glycosynthase via mutation of an active site oxazolium stabilising aspartic acid residue[151,152] (equivalent to D183 in DspB). However, both of these studies showed that the more effective mutation for these particular enzymes was in fact an D→E mutation, rather than an D→A. It also revealed that whilst an oxazoline substrate was superior

to *p*NP-GlcNAc, *p*NP-GlcNAc would be turned over by these enzymes. These results taken together suggest a range of mutations could potentially be used to create a glycosynthase of DspB that was capable of producing PNAG. In this thesis, the mutants that were selected were D183A, Y278N, and Y278F. An E248Q was also selected for study as this had been previously shown to increase DspB's hydrolysis activity of fully acetylated sugars, and a DspB glycosynthase would be operating on fully acetylated substrates, so it was hypothesised that a double mutant of a successful glycosynthase with the E248Q would be more effective.

### 1.3.3. Which substrate is suitable for studying a PNAGase or a PNAGsynthase?

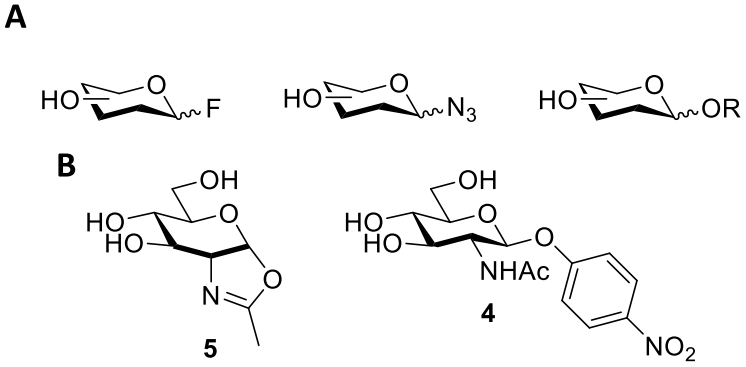



Figure 1.3.2: A) General structures of different substrates that have been used to study various GH enzymes from multiple families. (Left to right) glycosyl fluoride, glycosyl azide, aryl glycoside. B) Substrates more specific to GH families that utilise the NGP mechanism (Left to right) GlcNAc oxazoline **5**, *p*NP-GlcNAc **4**

Many different substrates have been proposed throughout the years studying the activity and mechanisms of CAZy enzymes. Glycosyl fluorides (Figure 1.3.2A) are often used as fluoride is an excellent leaving group[139], leading to substrates with high $\left(\frac{k_{cat}}{K_M}\right)$ values which makes these substrates excellent probes for kinetic studies for glycoside hydrolases[142,143], and for glycosynthases[157,158]. Glycosyl fluorides have proven to be a useful substrate in many previous studies of glycosynthase activity[1,159,160], however as previously mentioned, it would be likely be ineffective as a substrate for GH20 enzymes due to the NGP observed in the mechanism of the GH20 family. Any substrate that is being used to study the activity of a NGP GH enzyme, or its corresponding glycosynthase, would be required to form an oxazolium intermediate. The oxazoline formed during the reaction is an important transition state, and whilst oxazolines have been made using glycosyl fluorides chemically, these require strong basic conditions[161] that would likely hinder enzyme activity.

Azide containing substrates are also a promising avenue for the study of glycosynthases[162,163] (Figure 1.3.2A), benefitting from increased stability over the glycosyl fluoride

substrates, whilst also being sufficiently reactive as donors to be able to act as a substrate. Additionally, if the azide is incorporated at a position separate from the reactive centre, Click chemistry[164] can be used to analyse the products of the reaction[165,166] providing useful insight into the mechanism of the protein. However, caution must be taken with substrates containing azides, a study by Liu *et al.*[167] demonstrated the drastic effect that azide group incorporation can have on GH enzymes activity, and therefore likely on the glycosynthases as well, which they attribute to the size of the azide group causing steric hinderance within the active site.

Another potential substrate class that has found plenty of applications with all GH enzymes[160,168-170], including within the GH20 family[151,169,170], is aryl glycosides, as seen in Figure 1.3.2A where R is an aromatic group of some kind. These substrates have two main advantages, one, the aromatic ring can be used to mimic the 6 membered ring of the next saccharide monomer in a polysaccharide chain, and two, these can be modified to ensure the hydrolysis by-product of the glycosidic bond formation is a chromophore with known absorption properties. This leads to an effective substrate for glycosynthase enzymes which can simultaneously be used to monitor transition state formation via monitoring UV/Vis absorbance at a particular wavelength, which can give an insight into enzyme reaction kinetics. A common monosaccharide that is used for this purpose *p*NP-GlcNAc **4** (Figure 1.3.2B), where R is a *p*-nitrophenol group (*p*NP **6**). This group gives a distinct absorbance at 405 nm which can be monitored as the reaction proceeds allowing for accurate measurement of the amount of substrate turn over. Another positive of this substrate is that it is stable in solution for a suitable amount of time to monitor enzyme reactions, meaning background hydrolysis is not a big issue. Finally, *p*NP-GlcNAc **4** is also commercially available and does not require synthesising, which makes it an ideal substrate for any glycosynthase experiments.

DspB specifically, however, catalyses the hydrolysis of 1,6-linkages, and therefore, pNP-GlcNAc **4** may not be the most effective mimic of the forward reactions substrate. This is due to the fact that the oxygen atom connecting the *p*NP moiety to the sugar ring connects directly to the phenyl ring, whilst a 1-6-linkage has an extra carbon separating the oxygen and the next sugar ring. Chibba *et al.*[171] decided to investigate this further, through the design and synthesis of a carbamate substrate that better mimicked the 1,6-linkage (Figure 1.3.3).

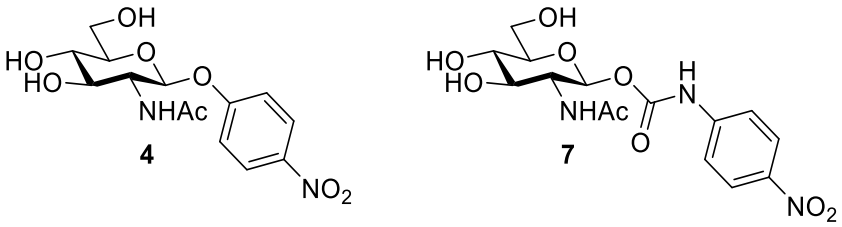



Figure 1.3.3: pNP-GlcNAc **4** (left) in comparison to the substrate Chibba *et al.* designed for DspB, carbamate **7** (right). The extra two atoms between the GlcNAc anomeric oxygen and the aromatic ring better mimic the 1,6-linkage hydrolysed by DspB

When WT DspB was screened with carbamate **7**, they observed a 200-fold improvement in the $\frac{k_{cat}}{K_m}$ value for the new substrate compared to the pNP-GlcNAc **4**, indicating the enzyme preferentially hydrolyses substrates that better mimic the 1,6-linkage.

Finally, as previously mentioned, oxazolines (Figure 1.3.2B), which act as a transition state mimic, work effectively as a glycosynthase substrate. The major downsides of the oxazoline **5** substrate is the inability to measure the kinetics of the reaction as unless you are able to tag this molecule, it is hard to observe its depletion during the course of the reaction, and that the oxazoline is sensitive to acid hydrolysis during synthesis, storage, analysis, and the glycosynthase reaction.

The effect of PNAG substrate length on hydrolysis kinetics has also been investigated for DspB. Fazekas *et al.*[120] compared the hydrolytic efficiency of DspB at hydrolysis of pentamers, tetramers, trimers, and dimers. They observed DspB that the enzyme is more efficient the longer the polysaccharide with a $\frac{k_{cat}}{K_m}$ value of 8.60 $M^{-1}s^{-1}$ in comparison to 7.11 $M^{-1}s^{-1}$ for the tetramer, and 6.07 $M^{-1}s^{-1}$ for the trimer. This indicates that to achieve maximal efficiency of on-protein PNAG polymerisation using a DspB glycosynthase in the future, PNAG oligosaccharides primers may be required. As discussed previously, chemical synthesis of pentasaccharides or longer for such activity is potentially extremely challenging in-solution due to multiple steps, and therefore diminishing returns. AGA synthesis of oligosaccharides, developed by Kröck *et al.*[99], is a potentially attractive solution to circumnavigate this issue, however, as also previously discussed this method also has its drawbacks, and it was decided against for this project. It was therefore also decided that initially this project would utilise only monosaccharide substrates for the investigation of turning DspB into a glycosynthase.

For this purpose three substrate targets were chosen for chemical synthesis in this thesis. Those were GlcNAc carbamate **7**, GlcNAc oxazoline **5**, and an alkyl-GlcNAc derivative herein referred to as $N_3$-propyl-GlcNAc **8** or acceptor **8** (Figure 1.3.4).

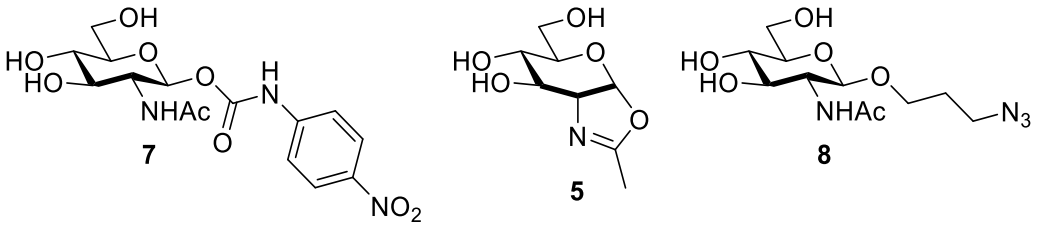



Figure 1.3.4: The synthetic targets selected for screening of DspB mutants for glycosynthase activity. (Left to right) GlcNAc carbamate **7**, GlcNAc oxazoline **5**, $N_3$-propyl-GlcNAc **8**.

*p*NP-GlcNAc **4** was also sourced commercially to be used as a positive control substrate. $N_3$-propyl-GlcNAc **8** was chosen to allow for the incorporation of an azide moiety at the reducing terminus of any oligo/polysaccharide chain that was produced in a way that removes the azide from the vicinity of the active site so it cannot hinder protein's activity.

# 2. Project Outline and Aims

The overall aim of this project was to expand the scope of knowledge surrounding poly-*N*-acetylglucosamine, both through furthering the understanding of its biosynthetic and biodegradation pathways, and through improving methods for the production of the polysaccharide.

The proposed methods that were to be utilised to achieve these aims were:

i. To use synthetic carbohydrate chemistry to create a library of monosaccharide substrates for the PNAGase DspB. These substrates could then be used to screen a group of novel DspB mutants for glycosynthase activity. Combined, this would create a new method for the *in vitro* production of PNAG oligosaccharide of known length and composition to allow for the use of these oligosaccharides to further the scope of knowledge surrounding PNAG.

ii. To isolate and purify PNAG that was produced *in vivo* to create a method for the characterisation of the biomodifications that PNAG undergoes during its biosynthesis, specifically the detection of *O*-succinylation in bacteria containing IcaC. This new method could then be used to probe the role of IcaC in *O*-succinylation via the utilisation of a class of succinyl-CoA mimics that would be synthesised alongside the monosaccharide DspB substrates.

iii. To investigate and characterise a novel group of putative PNAGases found in *staphylococci* species via the use of enzyme assay with PNAG mimicking substrates such as *p*NP-GlcNAc **4**. Doing so would increase the number of PNAG degrading enzymes substantially and may allow for the production of other glycosynthases with different substrate specificities.

iv. To chemically modify PNAG isolated from biological sources allowing for the incorporation of an azide at the reducing terminus to allow for functionalisation via Click chemistry.

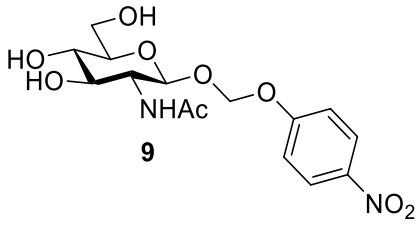
v. To utilise an inactive mutant of DspB that still displays PNAG binding activity to create a novel PNAG specific blotting agent. This would allow for the identification of PNAG in blotting experiments, as well as improving the identification of PNAG in the biofilms produced by bacterial cultures *in vitro*.

# 3. Results and Discussion

## 3.1. Chemical Synthesis

### 3.1.1. Introduction

As described in section 1.3.3, three synthetic substrate targets were identified at the beginning of this project for use in glycosynthase assays. These were carbamate **7**, oxazoline **5**, and acceptor **8** (Figure 1.3.4). Carbamate **7** was identified by Chibba *et al*[171] as a more effective DspB substrate than the commercially available, and more traditional glycoside hydrolase substrate, *p*NP-GlcNAc **4** (Figure 1.2.4). This is because DspB catalyses the hydrolysis of β-1,6-linked GlcNAc, and therefore its natural substrate contains an extra carbon between the anomeric oxygen, and the sugar ring of the second GlcNAc. Carbamate **7** better mimics this linkage and so was found to be significantly better substrate for DspB, an over 200 fold improvement in $\frac{k_{cat}}{K_M}$. Interestingly, an additional substrate was tested in this investigation that mimicked the β-1,6-linkage exactly (Figure 3.1.1), but DspB showed no activity when incubated with this substrate, so this substrate was not selected as a synthetic target, carbamate **7** was selected for synthesis.



**9**

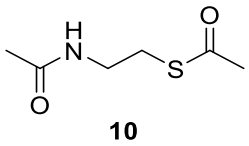
Figure 3.1.1: Aryl glycoside **9** that was proposed by Chibba *et al*[171] as a potential substrate for DspB. The alkyl linkage between the anomeric oxygen and the oxygen of the *p*NP group exactly mimics the β-1,6-linkage of PNAG.

Oxazoline **5** (Figure 3.1.3) is a transition state mimic of the neighbouring group participation mechanism displayed by DspB, as discussed in section 1.2.2. GlcNAc oxazolines have been used successfully in the literature as substrates for enzymatic transglycosylations catalysed by chitinase mutants[172], as well as Endo A and Endo M[173,174]. More recently, mutated GH20 enzymes have been shown to be capable of performing as glycosynthases utilising an oxazoline substrate to catalyse the formation of a glycosidic bond[151-154]. It was for these reasons that oxazoline **5** was also targeted as a synthetic substrate for use in the glycosynthase assays. However, as discussed in section 1.3.3, oxazoline **5** doesn't produce an easily detectable signal when it is hydrolysed like the aromatic containing glycosides like *p*NP-GlcNAc **4** or carbamate **7** do. It is also difficult to detect via techniques such as LC-MS as it is liable to degrade during the analysis into GlcNAc **1**. This is problematic as the presence of GlcNAc **1** in a glycosynthase assay could be indicative of enzymatic

hydrolysis activity, which could be causing the breakdown of any successfully synthesised oligosaccharides, resulting in a false negative. Larger oligosaccharides are also less likely to ionise as well, and therefore less likely to be seen amongst large amounts of monosaccharides. Furthermore, literature studies that utilise oxazoline substrates tend to use a suitable glycosidic acceptor molecule during the glycosynthase reaction. It was therefore decided that a suitable acceptor molecule was required, one that could be used to attach a MS tag onto any oligosaccharides produced to improve ionisation and therefore detection. Acceptor **8** (Figure 3.1.3) was chosen for this purpose.

Acceptor **8** is an azide containing aryl glycoside. The incorporation of the azide allows for the utilisation of Click chemistry to attach a MS tag, such as an imidazolium based tag like the one developed by Sittel *et al*[175] or a fluorescence tag as demonstrated by Richards *et al*[176], while ensuring the azide is far enough removed from the active site to avoid the disruption to enzyme activity described by Lui *et al*[167]. Acceptor **8** was therefore selected as the final glycosynthase synthetic target for this project.

Finally, one of the other aims of this project was to probe the function of IcaC within the biosynthetic pathway. IcaC had previously been identified as a member of the large family of acyltransferases (Acyl_transf_3 family (PF01757), InterPro family IPR002656[72]) by Atkin *et al*[59]. One of the most studied proteins within this family has been shown to utilise acetyl-CoA as a cofactor in its catalysis of the O-acetylation of peptidoglycan[177], which has led to the hypothesis that IcaC utilises a similar molecule, such as succinyl-CoA in its acyltransferase mechanism. Previous work by Keenan *et al*[178] has shown that PseH, an enzyme that is a member different family of very similar enzymes known as the Gcn5-related N-acetyltransferases[179] (GNAT), can utilise synthetically generated acetyl-CoA mimic known as SNAc **10** (Figure 3.1.2) to incorporate an acetyl group into a sugar known as pseudaminic acid, and that SNAc **10** could be chemically modified to allow for incorporation of an azide into the sugar to allow for Click chemistry to be performed.




**10**

Figure 3.1.2: Structure of SNAc **10**

It was hypothesised that IcaC, given the likelihood it utilises succinyl-CoA as its cofactor, could also utilise a similar molecule to perform its succinyltransferase reaction, and therefore an azide tag, or a fluoro tag for $^{19}F$ NMR, could potentially be smuggled into the PNAG chain via the succinyl groups. This would perform two tasks, one, it would prove the hypothesis that IcaC is a

succinyltransferase protein, and two, it would generate PNAG molecules with Click-able tags present in the chain for further study. For this purpose, the remaining synthetic targets of the project were selected, thioesters **11 – 15** (Figure 3.1.3).

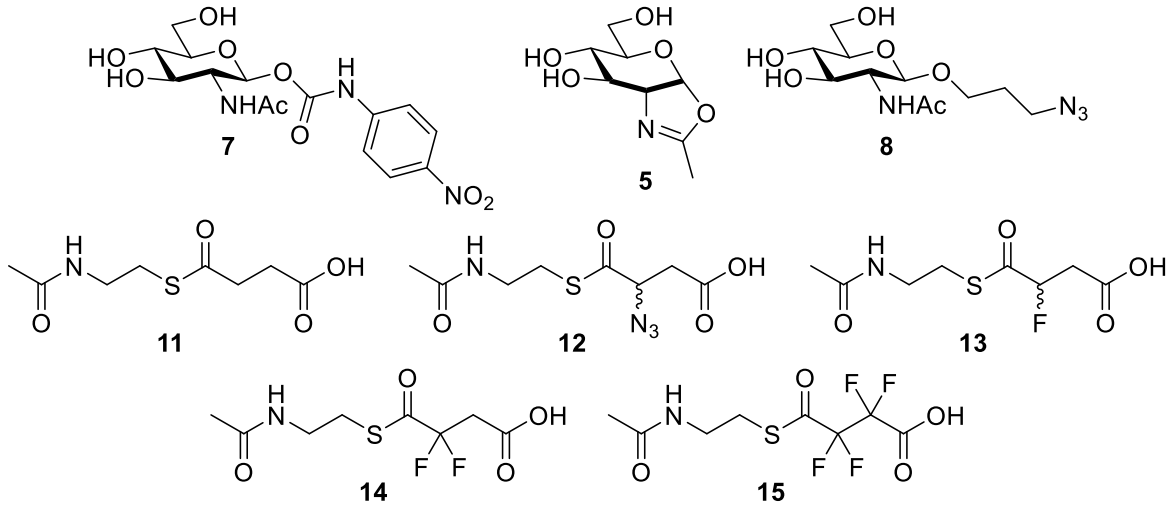
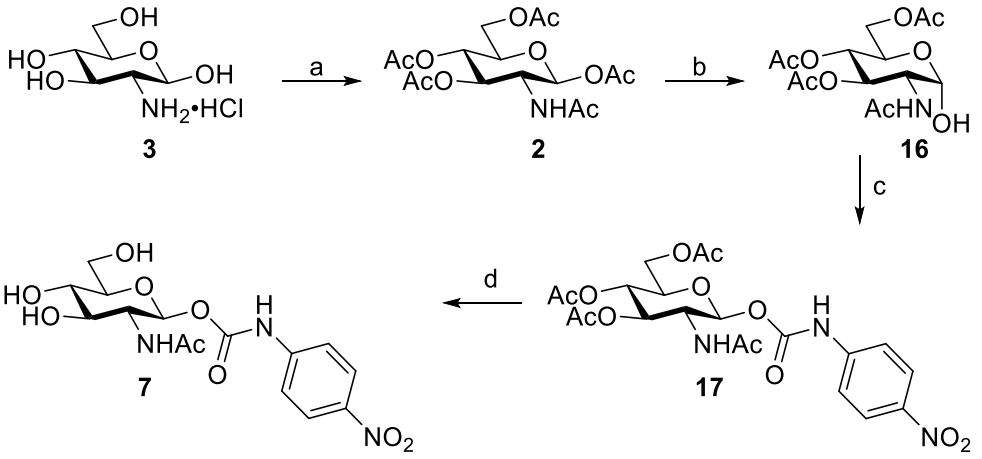



Figure 3.1.3: The complete selection of target molecules selected for synthesis during this project. (Left to right, top to bottom) Carbamate **7**, Oxazoline **5**, Acceptor **8**, Thioester **11**, Azido Thioester **12**, Monofluoro Thioester **13**, Difluoro Thioester **14**, Tetrafluoro Thioester **15**.

### 3.1.2. Carbamate Substrate **7**

The synthesis of the carbamate substrate **7**, adapted from Chibba *et al.*[171], is depicted in Scheme 3.1.1.



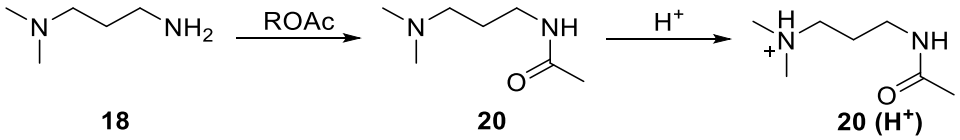

Scheme 3.1.1: a) $Ac_2O$ (excess), cat. DMAP, Pyr., 16 hr, R.T., 52%. b) Dimethylaminopropylamine **18** (5 eq.) THF, 2 hr, R.T., 65%. c) *p*-nitrophenyl isocyanate **19**, cat. TEA, Tol., 5 hr, R.T., $N_2$, 49%. d) $NH_2NH_2^{\cdot}\bullet H_2O$, MeOH, 6 hr, R.T., 42%.

Initially commercially sourced glucosamine hydrochloride **3** was acetylated by reaction with acetic anhydride in the presence of catalytic 4-dimethylaminopyridine (DMAP), yielding the peracetylated GlcNAc **2**. This was then selectively deacetylated at the anomeric position by the addition of 5 equivalents of dimethylaminopropylamine (DMAPA). Hemiacetal **16** was then reacted

with 2 equivalents of *p*-nitrophenyl isocyanate in the presence of catalytic triethylamine (TEA) yielding the carbamate **17**. This was then deacetylated by addition of 3.3 equivalents of hydrazine monohydrate in MeOH affording the final GlcNAc carbamate substrate **7** as a pale yellow solid.

The first step of the reaction gave a low yield of the monosaccharide **2** after being reacted for 16 hr, despite the TLC of the reaction mixture indicating no starting material remained in solution. This likely indicates that the starting material was poorly soluble in pyridine. Heating the reaction mixture above room temperature could potentially improve solubility of the compound, and therefore improve yield, however as enough of monosaccharide **2** was obtained from this step, this was not attempted.

The second step of the reaction proved challenging. Initially, the selective deprotection of the anomeric hydroxy group was attempted with benzylamine in dry THF as reported by Sadamoto *et al.*[180], and successfully performed by Chibba *et al.*[171] in their synthesis of this compound. However, when this was performed on monosaccharide **2**, a large impurity was present in the NMR that could not be separated from the product by column chromatography. This was likely a side product formed from the benzylamine as the NMR showed a multiplet in the region of around 7.2-7.3 ppm, which would indicate an aromatic group was present. As it was unclear if the presence of this impurity would affect the next step of the reaction, a different approach to this step was taken. First reported by Andersen *et al.*[181] in 2015, anomeric deacetylation of monosaccharides can also be achieved through the addition of DMAPA **18** to a solution of the sugar in THF. The benefit of this is twofold, one, the reaction no longer needs to be done under anhydrous conditions, and two, this reaction does not need column chromatography for its purification (Scheme 3.1.2).

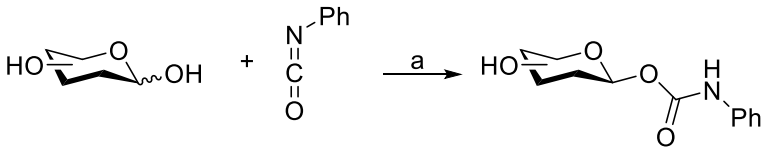



**18** **20** **20 ($H^+$)**

Scheme 3.1.2: Generation and subsequent protonation of the by-product acetylated amine **20**

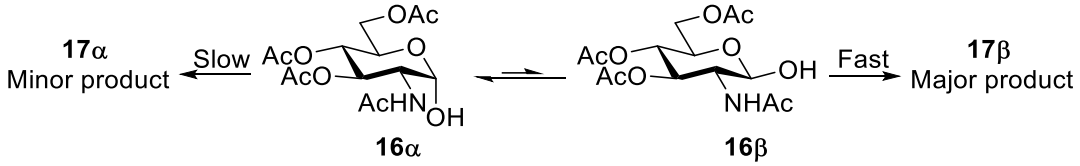
Excess DMAPA **19**, and the by-product acetylated amine **20**, can be easily protonated by washing the reaction mixture with 1M HCl, and once protonated, both will be washed away into the aqueous layer leaving the hemiacetal in the organic layer. NMR analysis of the final product confirmed this, as little to no impurity could be seen in the spectra. However, NMR also confirmed what Jensen *et al.* reported about the stereoselectivity of this reaction, as purely the α-anomer of hemiacetal **16** was formed, as confirmed by the coupling constant of the anomeric proton in the $^1$H NMR (Appendix 2), as opposed to the β anomer reported by Chiba and Sadamoto *et al*[171,180].

Natural PNAG polymers are purely β-linked, and so it is therefore necessary that carbamate **7** is synthesised as a pure β-anomer. Leenders *et al.*[182] previously investigated stereoselectivity of the reaction between reducing monosaccharides with isocyanates, and observed complete stereoselectivity for β-phenyl isocyanate anomers when using toluene as the solvent, with a catalytic amount of TEA (Scheme 3.1.3).




Scheme 3.1.3: Leenders *et al.* synthesis of β-carbamates from glycosyl donor with complete stereoselectivity. a) Tol., TEA (0.3 eq), 20 °C, 2 hr.

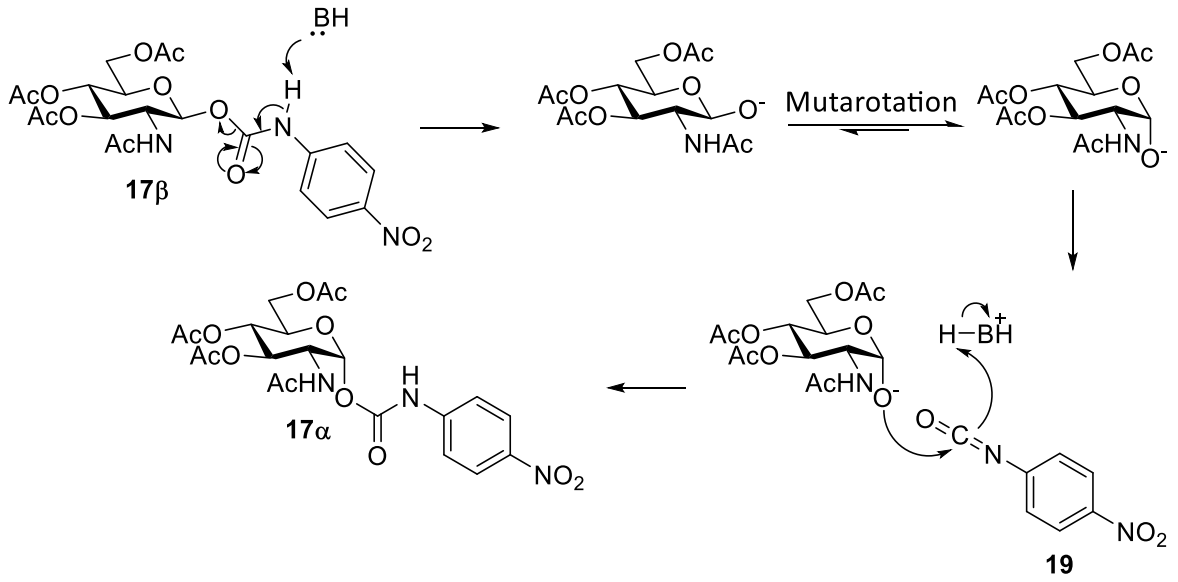
Leenders *et al.* hypothesised this stereoselectivity was a result of TEA promoted mutarotation of a racemic mixture of starting material into the α-anomer. They state that the α-anomer is too sterically hindered to react with the cyanate, and so β-carbamate formation is promoted as the two anomers are in equilibrium, and therefore some β-anomer will be present. Whilst NMR data showing mainly α-hemiacetal upon addition of TEA of the hemiacetal in various solvents supports this hypothesis, it is likely just exaggerating the difference in the nucleophilic character between the α-anomeric hydroxy group and β-anomeric hydroxy group. Due to the anomeric effect, the β hydroxy is more nucleophilic than the α[183], and so is already more reactive. The transition state for β attack is also likely of lower energy due to potential unfavourable diaxial interactions in the α-transition state. This is an example of a Curtin-Hammett kinetic scenario[184] (Scheme 3.1.4).




Scheme 3.1.4: Demonstration of the Curtin-Hammett kinetic scheme that leads to the β-carbamate **17** being formed in preference

The toluene solvent likely helps to exaggerate this effect by π stacking with the phenyl group of the isocyanate during the reaction, providing more steric hinderance to the α-anomer during the reaction, increasing selectivity for β-carbamate formation. This hypothesis is reinforced in a solvent screen performed by Leenders *et al.* where toluene afforded β-anomers with the highest selectively. Chibba *et al.* used these conditions in the synthesis of carbamate **17** and they reported complete stereoselectivity in the formation of the carbamate **17β**, when starting from hemiacetal **16β**.

To investigate if the stereoselectivity of this step would promote complete stereoselectivity to the β-anomer when the α-anomer of hemiacetal **16** was used, the reaction was initially performed on a small scale using 80 mg of hemiacetal **16**, yielding only the β-carbamate **17** in a 15% yield. Whilst the yield for the reaction was poor, the stereoselectivity was promising, so the reaction was scaled up to 915 mg and attempted again. Upon scale up, however, it became apparent that hemiacetal **16** was of low solubility in toluene, which likely lead to the low yield on a small scale. This led to a large amount of unreacted starting material being left in the reaction vessel, lowering the yield. Unfortunately, as was noted in a previous attempt at this synthesis within the group, if the reaction was allowed to run for longer times to allow more of the starting material to enter into solution, the stereoselectivity of the reaction began to favour α-carbamate **17** formation. This is likely due to the lability of the carbamate in the basic environment leading to generation of deprotonated hemiacetal **16**, which can mutarotate and re-react with the *p*-nitrophenyl isocyanate **19**, yielding the α-carbamate **17** (Scheme 3.1.5).




Scheme 3.1.5: Potential mechanism for the formation of the α-carbamate **17** via the TEA catalysed degradation of β-carbamate **17**, subsequent mutarotation of hemiacetal **16β** yielding hemiacetal **16α**, and final TEA catalysed re-reaction with isocyanate **19** yielding α-carbamate **17**.

Therefore, after 6 hours the solvent must be filtered off to avoid this occurring. This led to lower yield than was reported in the literature, approximately 56% compared to 86%. $^{1}$H NMR analysis revealed that the α-anomer had formed alongside the β-anomer in almost an ≈2:3 ratio in favour of the β. Why this was observed when others have reported complete stereoselectivity towards the β is unclear, but may be related to the low solubility of hemiacetal **16**. If there is no hemiacetal **16** in solution for the TEA to react with, it will attack the carbamate, as depicted in Scheme 3.1.5, yielding the α-carbamate. Initially, in order to separate the two anomers, column

chromatography was attempted, however, this was unsuccessful, so a different approach was attempted. The sugar was suspended in toluene, and the suspension heated to 35°C, and left with constant stirring for an hour. The rationale was that the α-anomer is slightly more soluble in toluene, as implied by Chiba[185], so the α-anomer might potentially enter solution at this temperature, whilst the β-anomer might not. This initial suspension was then filtered by gravity to avoid precipitation of any solubilised α-anomer due to a vacuum. This initial wash was successful, improving the α:β ratio to ≈1:2.2, as determined by $^{1}$H NMR, so this was repeated, being left for 5 hours this time. This successfully removed the α peaks from the $^{1}$H NMR spectra, leaving just the β-anomer in a 6% yield. Knowing the aforementioned issues, the synthesis was repeated at 35°C, giving significantly improved yield of 49%. Future attempts at this synthesis might benefit from reducing the amount of TEA to avoid α-anomer formation, as TEA is currently being used in vast excess of hemiacetal **16α.**

Finally, a global deprotection was performed on carbamate **17**. Due the carbamate lability, this could not be performed under standard Zemplén conditions (NaOMe/MeOH), and acidic conditions would also likely trigger the degradation of carbamate **17**, yielding hemiacetal **16**. Therefore, the reaction had to be performed by suspending carbamate **17** in MeOH, and carefully adding 3.3 eq. of hydrazine monohydrate, whilst monitoring the reaction by $^{1}$H NMR. This yielded the deprotected carbamate **7** as a yellow solid, however, further analysis of the NMR spectra revealed contamination with *p*-nitrophenyl isocyanate from the third step of the reaction. To remove this contamination, Sephadex LH20 resin was used for size exclusion chromatography (SEC) and the resulting fractions were analysed using negative mode HILIC-LC-MS (Figure 3.1.4).

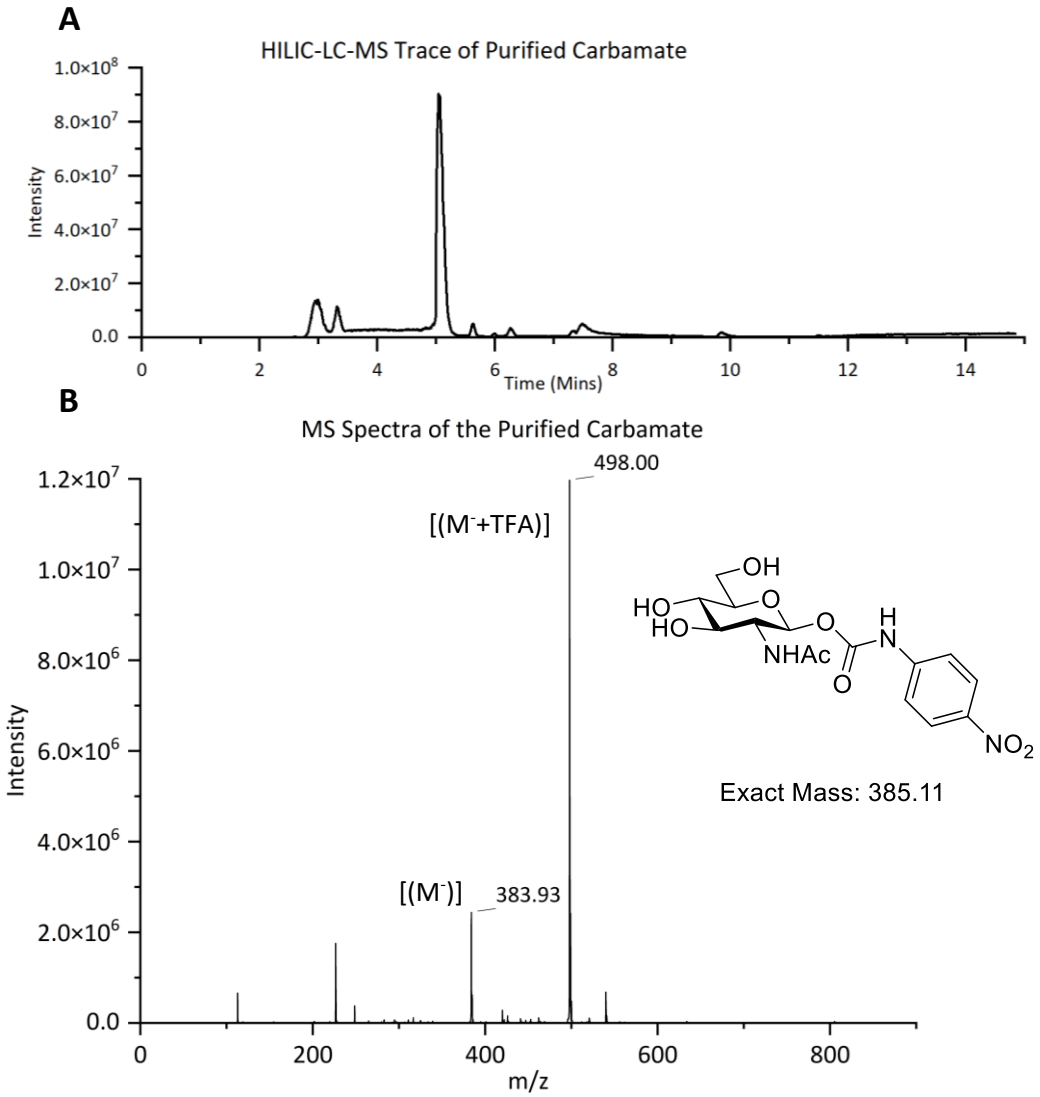
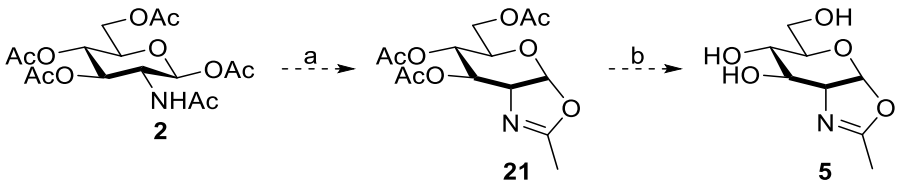



Figure 3.1.4: A) HILIC-LC-MS chromatogram of the purified carbamate **7**. B) Negative mode MS spectra of the purified carbamate **7**.

The resulting fractions were combined, and lyophilised yielding carbamate **7** as a pale yellow solid in a 42% yield. This yield could be improved by performing the SEC immediately after the glycosylation in step 3, but this was not attempted as enough of carbamate **7** was synthesised for the planned experiments.

### 3.1.3. Oxazoline Substrate **5**

Oxazoline **5** had been synthesised in a synthetic route previously described by La Paiotta *et al.*[186] starting from monosaccharide **2** (Scheme 3.1.6).




Scheme 3.1.6: a) TMSOTf (1.6 eq), dry DCM, overnight, 65 °C b) NaOMe, dry MeOH, 3 hr, R.T., $N_2$.

Monosaccharide **2** was dissolved in dry DCM, placed under $N_2$ and reacted with trimethylsilyl trifluoromethanesulfonate (TMSOTf) at reflux overnight, yielding peracetylated oxazoline **21**. Following purification, a global deacetylation of oxazoline **21** was proposed, to afford oxazoline **5.** However, this purification proved to be challenging as oxazoline **21** appeared to degrade during column chromatography. Considering the challenging purification, two alternative one-step reaction schemes for the synthesis of oxazoline **5**, described by Noguchi *et al.*[187,188], using either 2-chloro-1,3-dimethylimidazolinium chloride **22** (DMC **22**) (Scheme 3.1.8), or 2-chloro-1,3-dimethyl-1H-benzimidazol-3-ium chloride **23** (CDMBI **23**)(Scheme 3.1.10), was considered.

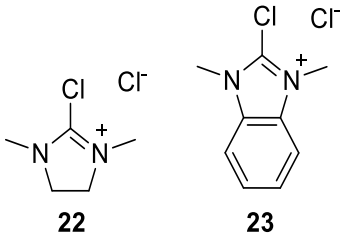
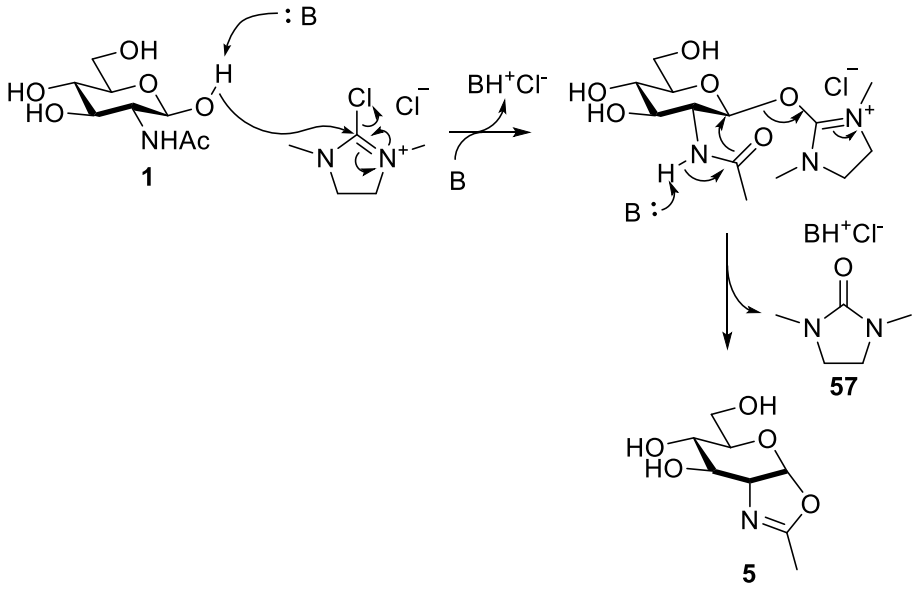


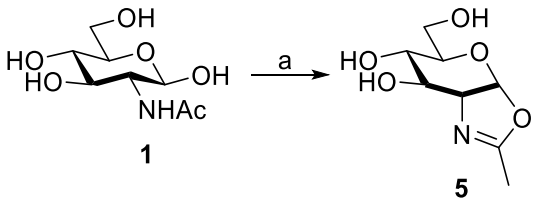

Figure 3.1.5: Structure of DMC vs CDMBI

DMC **22** and CDMBI **23** both act as an anomeric activating agent by substituting for the hydroxy group at the anomeric position, and then acting as a good leaving group that allows for the *N*-acetyl group to perform a neighbouring group participation mechanism to form the oxazoline ring (Scheme 3.1.7).




Scheme 3.1.7: Mechanism for the formation of anomeric activation and oxazoline ring formation when GlcNAc **1** is treated with DMC **22**, as proposed by Noguchi et al[187]

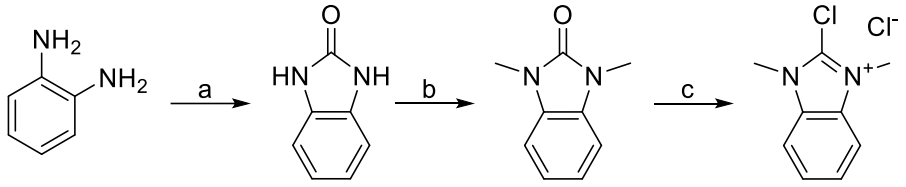
CDMBI **23** is a benzene derivative of DMC **22** (Figure 3.1.5) that is more stable in air, one of the major drawbacks to working with DMC **22**. Noguchi *et al*. also speculate that DMC **22** hydrolyses into 1,3-dimethyl-2-imidazolidin-2-one, DMI **57**, which is soluble in water, and a potentially potent inhibitor of enzymatic transglycosylation. CDMBI **23** can also hydrolyse into a diamide, diamide **24**, which could potentially also have an inhibitory effect, however, this compound is insoluble in water and so can be removed before any enzyme is added. Despite this proposed inhibition however, when WT DspB hydrolysis activity was assayed using *pNP*-GlcNAc **4** in the presence of a DMC **22** solution in water, activity via detection of released *p*NP **6** at 405 nm was still observed (Figure 3.2.6). CDMBI **23** is also not available commercially, unlike DMC **22**, and requires a 3-step synthesis with an overall yield of only 13%. Therefore, it was initially decided to proceed with the use of DMC **22** over CDMBI **23**.




Scheme 3.1.8: a) DMC **22** (3 eq.), TEA (9 eq.), $H_2O$, 15 mins, 0 °C, 78%.

Commercially sourced GlcNAc monosaccharide **1** was dissolved in $D_2O$ and TEA at 0 °C, and reacted with DMC **22**, yielding oxazoline **5** in a 78% yield. The yield was determined via addition of a 1M sodium formate solution in $D_2O$ to act as an internal standard. This yielded some promising results as oxazoline **5** was successfully synthesised from monosaccharide **1** using DMC **22** as a

reagent. Attempts to purify oxazoline **5** via both column chromatography, and gel filtration chromatography resulted in hydrolysis of oxazoline **5** into monosaccharide **1**, so it was decided to directly use the reaction mixture in enzyme assays. However, upon attempting this, it was discovered that the TEA used during the synthesis of oxazoline **5** was overpowering the buffer, leading to a basic reaction solution. According to Manuel *et al*[119], this will lead to complete loss of enzyme activity for DspB. A further alternative was therefore required. The synthetic route proposed by La Paiotta *et al*[186] was briefly reconsidered, but it was decided to attempt to synthesis CDMBI **23** following the synthetic pathway described by Noguchi *et al*[188] (Scheme 3.1.9).




**25** **26** **24** **23**

Scheme 3.1.9: a) Urea (1.31 eq.), ethylene glycol, 48 hr, 135 °C, 88%. b) MeI (2.20 eq.), NaOH, $Bu_4NBr$, 48 hr, 60 °C, 39%. c) $(COCl)_2$ (4.02 eq.), Tol., 168 hr, 80 °C, $N_2$, 29%.

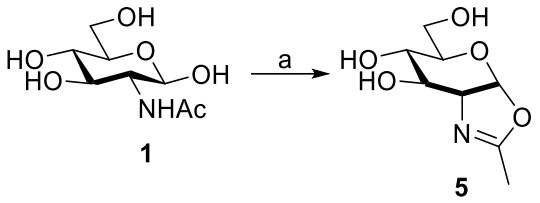
Diamine **25** was converted to diamide **26** via the transamination reaction with urea. Diamide **26** was then methylated via addition of methyl iodide (MeI) in the presence of NaOH and catalytic tetrabutylammonium bromide ($Bu_4NBr$) yielding diamide **24**. This was then chlorinated via addition of oxalyl chloride ($(COCl)_2$) to afford the CDMBI **23** as a beige solid.

The first step of this synthesis proceeded with little difficulty, however the reaction was quenched before complete conversion was seen via TLC due to time restraints. It is likely that with additional time the yield could be increased, but given the already high yield, this was not deemed necessary. The second step was also quenched early for the same reasons, which likely also contributed to the low yield, however, Schmölzer *et al*[151] demonstrated that replacing NaOH with KOH also leads to an increase in yield. If this step were to be repeated, this could help with improving the yield, but as with the previous step, enough product was obtained to proceed to the next step of the synthesis.

The final step of the synthesis was more challenging than the previous two steps. The reaction initially proceeded as expected, and a precipitate formed after 72 hr. However, a TLC of the reaction mixture revealed a significant amount of starting material remained, so the reaction was given another 48 hr. After the 48 hr period, there was still a significant amount of starting material remaining. At this point the reaction was filtered and the precipitate collected, however NMR revealed significant amounts of toluene remained on the product which could not be removed under reduced pressure. This was solved by dissolving the product in water, flash freezing to

prevent hydrolysis, and then lyophilising. This worked to remove the toluene, and there did not appear to be any noticeable hydrolysis product in the NMR, but the process likely caused some loss of product.

Following successful synthesis of CDMBI **23**, the compound was then used in the synthesis of oxazoline **5** (Scheme 3.1.10).




**1** **5**

Scheme 3.1.10: a) CDMBI **23** (3 eq.), $Na_3PO_4$ (7.5 eq.), $D_2O$, 1 hr, 0 °C, 56%.

Commercially sourced GlcNAc was dissolved in $D_2O$ with 3 eq. CDMBI **23**. To this solution, 0.3 M $Na_3PO_4$ was added and the reaction mixture was cooled to 0°C for 1 hour. After 1 hour, NMR confirmed formation of oxazoline **5** in a 56% yield. Noguchi *et al*[188] suggest that before addition of the enzyme, the reaction solution should be neutralised via addition of HCl, whilst Schmölzer *et al*[151] suggest that you need only separate the oxazoline from the reaction mixture via filtering with C18 absorbent. Recently, C18 reversed phase chromatography columns had been acquired from Teledyne, so an attempt was made to try and separate the oxazoline from the reaction mixture via reverse phase chromatography (Figure 3.1.6).

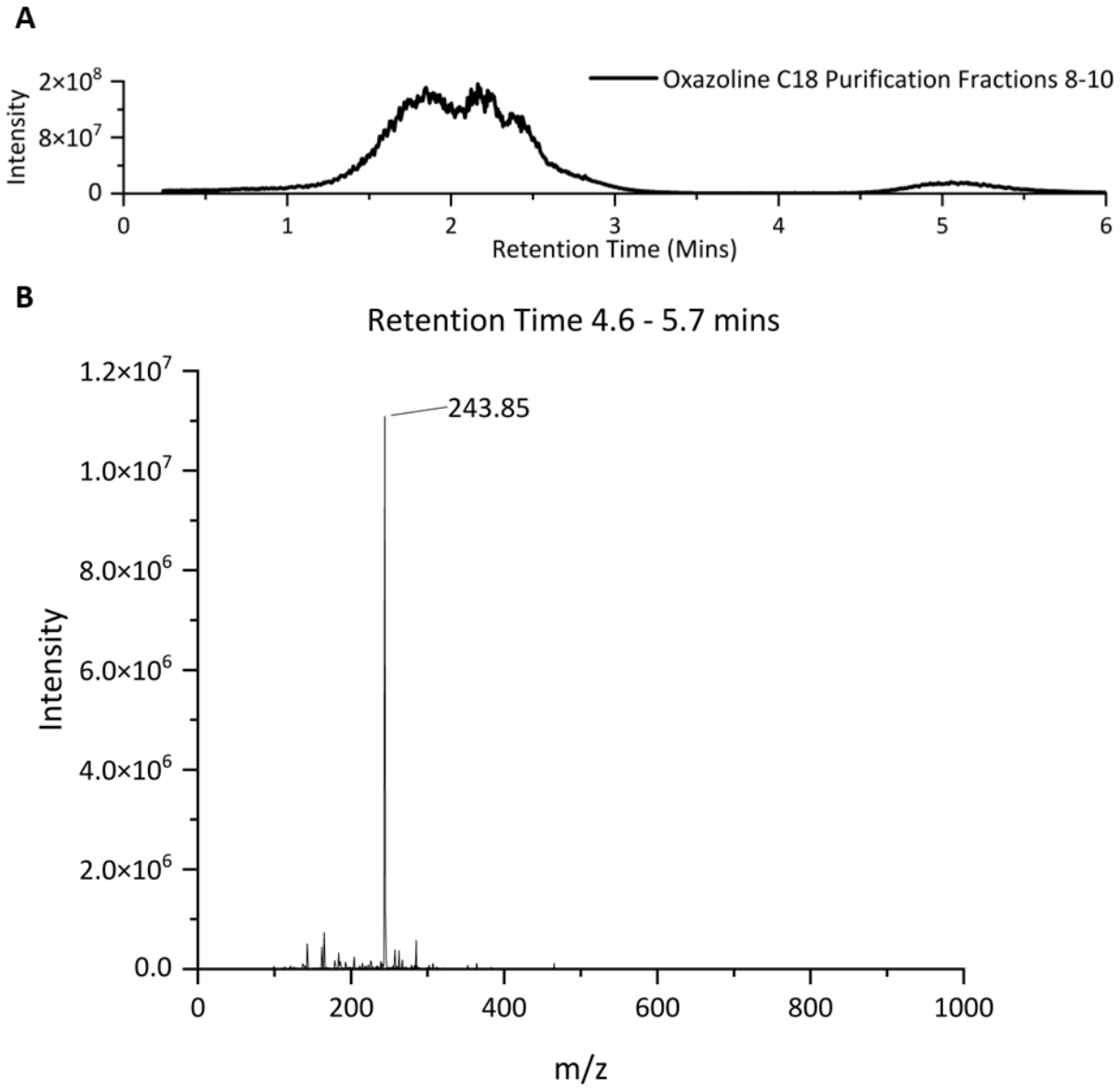



Figure 3.1.6: A) HILIC LC-MS chromatogram of fractions 8-10 of the purification via reverse phased HPLC of oxazoline **5**. B) MS trace of retention time 4.6-5.7 minutes

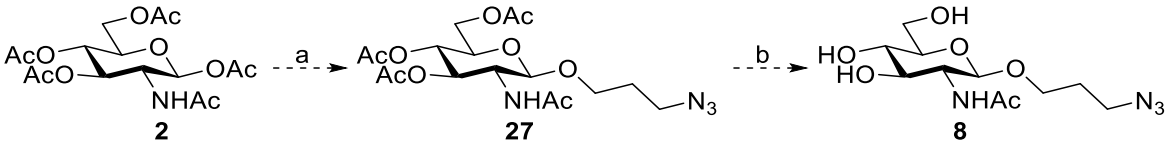
LC-MS analysis of the fractions of the purification revealed two major peaks. Upon further investigation into the large peak that elutes at an approximate retention time 2 mins, no peaks of recognisable mass appear in the MS trace, whilst the second peak contains only one peak of mass 243.85, which corresponds to the hydrolysed oxazoline, as confirmed by $^{1}$H NMR which showed that the sample contained no left over starting material which would have the same mass. The initial peak was therefore put down to artifacts from the LC-MS rather than impurities, which was also later confirmed by $^{1}$H NMR as no observable impurities were present in the sample. This purification was performed on a small scale and yielded little of the pure oxazoline, but these initial results are promising, and suggest that scaling up the production could lead to a pure oxazoline for

use in enzyme assays. For this purpose, a 100 mg scale reaction was preformed, and this yielded 51 mg of pure oxazoline **5** which could be used in glycosynthase assays.

### 3.1.4. Synthesis of Acceptor **8**

Due to its lack of UV/Vis absorbance, oxazoline **5** is not a suitable substrate for traditional plate reader assays, therefore a different assay needed to be designed to screen oxazoline **5** against the DspB mutants. Previous work in our lab had designed an assay system that could be monitored via TLC (see section 3.2.3). This assay required a suitable enzyme acceptor substrate, and for this purpose, $N_3$-propyl-GlcNAc acceptor **8** was selected.

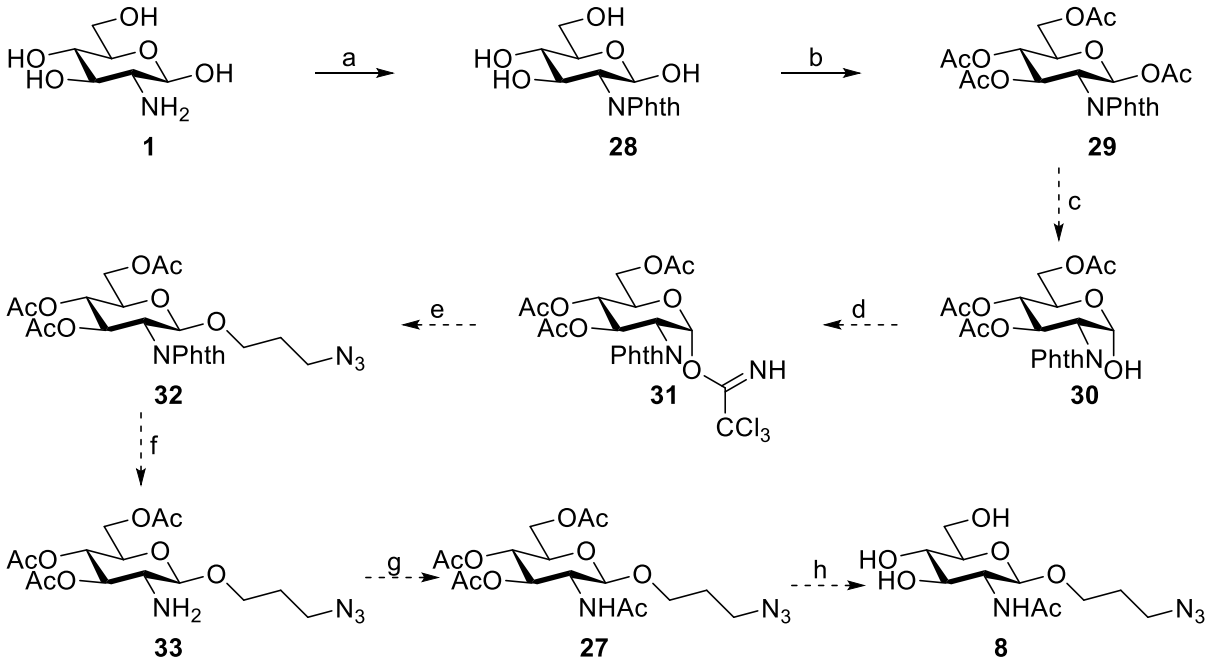
Acceptor **8** was selected for two reasons, 1) to allow for post glycosylation modification via Click chemistry at the azide group; and 2) as it was hypothesised that the azide group could potentially partially mimic the size of GlcNAc monomer in a PNAG polymer. Initially, it was proposed a glycosylation could be performed using either scandium (III) triflate ($Sc(OTf)_3$), or bismuth (III) triflate ($Bi(OTf)_3$), as proposed by Christensen *et al*[189], and Rasmussen *et al*[190] (Scheme 3.1.11).




Scheme 3.1.11: a) 3-azidopropan-1-ol (1.69 eq.), $Sc(OTf)_3$ (0.15 eq), dry DCM, $N_2$, 45 °C, 24 hr b) NaOMe, dry MeOH, $N_2$, overnight

This method, whilst initially appearing promising, when tested proved to be exceptionally slow. MS analysis revealed product formation after 24 hours, however, the conversion was very low. The reaction was left until the conversion appeared to be just above 50%, at which point, the reaction appeared to halt, and so the reaction was quenched. This experiment was also repeated with 3-bromopropan-1-ol with the view to perform a nucleophilic substitution on the resulting bromo glycoside, but no product was able to be recovered from this reaction. Therefore, a different approach was required.

When GlcNAc undergoes a glycosylation, it proceeds via a neighbouring group participation mechanism, where the *N*-acetyl group reacts with the anomeric carbon, creating an oxazoline ring, and kicking out the group at the anomeric position. The oxazoline ring that forms is quite stable and can slow the rate of product formation[190], leading to slow and arduous reactions. A method to combat this is to perform an *N*-protection of the GlcNAc molecule before proceeding with the glycosylation. The proposed method would proceed via an *N*-phthalimido GlcNAc derivative, depicted in Scheme 3.1.12.

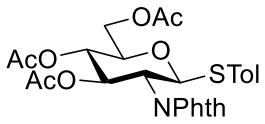


Scheme 3.1.12: a) Phthalic anhydride (2.07 eq.), NaOH (1.29 eq.), $NaHCO_3$ (2.57 eq.), $H_2O$:MeOH:Acetone (1:2:4), R.T. – 50 °C, 48 hr. b) Acetic anhydride, DMAP (cat.), Pyr., R.T., 24 hr, 29%. c) DMAPA **18** (5 eq.), THF, R.T., 2 hr. d) DBU (0.06 eq.), $Cl_3CCN$ (4.9 eq.), dry DCM, 0 °C - R.T., $N_2$. e) 3-azido-propan-1-ol **34** (10 eq.), TMSTOf (0.12 eq), dry DCM, $N_2$, RT - -10 °C, 25 mins. f) $H_2NNH_2{\cdot}H_2O$, MeOH, R.T., 24 hr. g) Acetic anhydride, DMAP (cat.), Pyr., R.T., 24 hr. h) NaOMe, dry MeOH, $N_2$, R.T., 24 hr.

Commercially sourced glucosamine hydrochloride **2** was N-protected via addition of phthalic anhydride in basic conditions giving phthalimido glycoside **28**, which was then O-acetylated via addition of acetic anhydride with a DMAP catalyst yielding phthalimido glycoside **29**[191,192]. Glycoside **29** would then be anomerically deacetylated by the addition of 5 equivalents of DMAPA **18** giving hemiacetal **30**[181]. Hemiacetal **30** would then undergo base catalysed glycosylation using $Cl_3CCN$ in the presence of catalytic DBU giving acetimidate **31**[193]. Acetimidate **31** would then be glycosylated using 3-azido-propan-1-ol **34** yielding azido glycoside **32**[194]. Glycoside **32** would then be N-deprotected using hydrazine monohydrate in MeOH giving amine **33**[195]. Amine **33** would then be N-acetylated via addition of acetic anhydride with a DMAP catalyst, yielding azide **27**, which could then be de-O-acetylated using NaOMe in MeOH resulting in acceptor **8**.

The *N*-protection, and subsequent global *O*-acetylation was successfully performed yielding glycoside **29** as a pale red solid. Following an initial aqueous workup some acetic anhydride remained but this was able to be removed by placement of the product under vacuum for one hour. Attempts at the anomeric deprotection lead to issues. DMAPA reacts with the phthalimido group of glycoside **29** resulting in N-deprotection rather than anomeric deprotection. An alternative method for this step is described by Jung *et al*[196] using ethylene diamine in acidic conditions that gives hemiacetal **30**. A further alternative would be to convert glycoside **29** into thioglycoside **35**

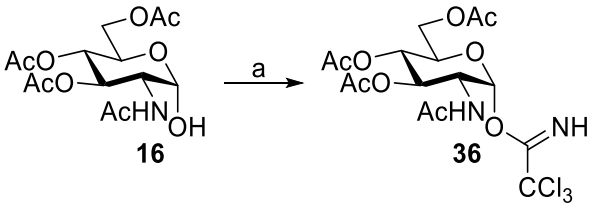
(Figure 3.1.7) via a thioglycosylation reaction with 4-methylbenzenethiol in the presence of $BF_3 \cdot OEt_2$[197]. This can then be used instead of the imidate **31** for the glycosylation, or the STol group can be removed, giving hemiacetal **30**.




**35**

Figure 3.1.7: Structure of thio glycoside **35**.

However, neither of these methods were required as simultaneously to these attempts, a different method was being explored utilising trichloroacetimidate GlcNAc derivative known as TCA-GlcNAc, or acetimidate **36**. Acetimidate **36** is well known to act as an efficient donor for glycosylation reactions. The TCA group acts as an excellent leaving group, which under acidic conditions is rapidly cleaved, leading to a quick build up of oxazoline **21.** This rapid increase in the concentration of oxazoline **21** helps overcome the slow reaction time caused by the oxazoline's stability, allowing for a glycosylation reaction to proceed. Acetimidate **36** can be synthesised from hemiacetal **16** as depicted in Scheme 3.1.13.

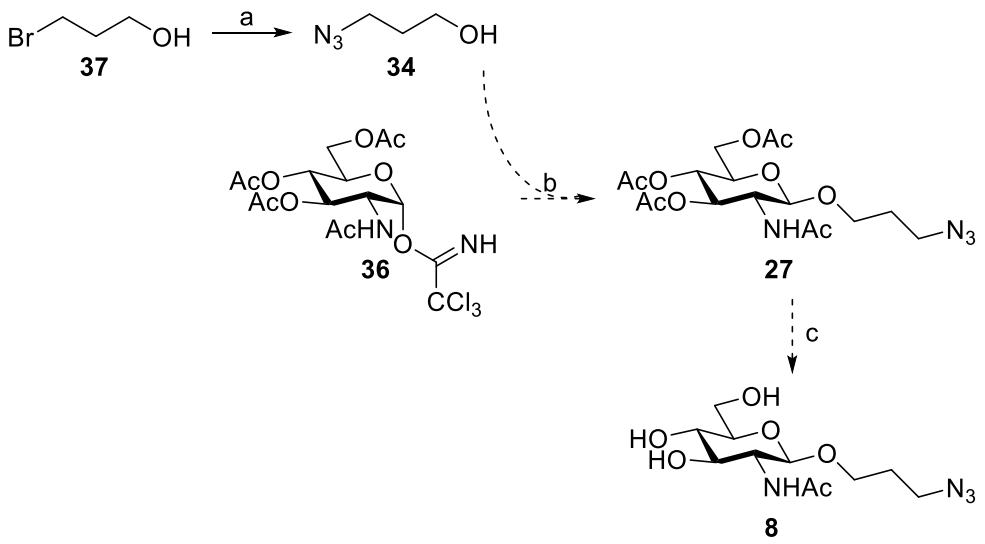



**16** **36**

Scheme 3.1.13: a) DBU (0.06 eq.), $Cl_3CCN$ (4.9 eq.), dry DCM, 0 °C - R.T., $N_2$, 90%.

Hemiacetal **16** underwent base catalysed glycosylation using trichloroacetonitrile ($Cl_3CCN$) in the presence of catalytic 1,8-diazobicycloundec-7-ene (DBU) giving acetimidate **36**[193].

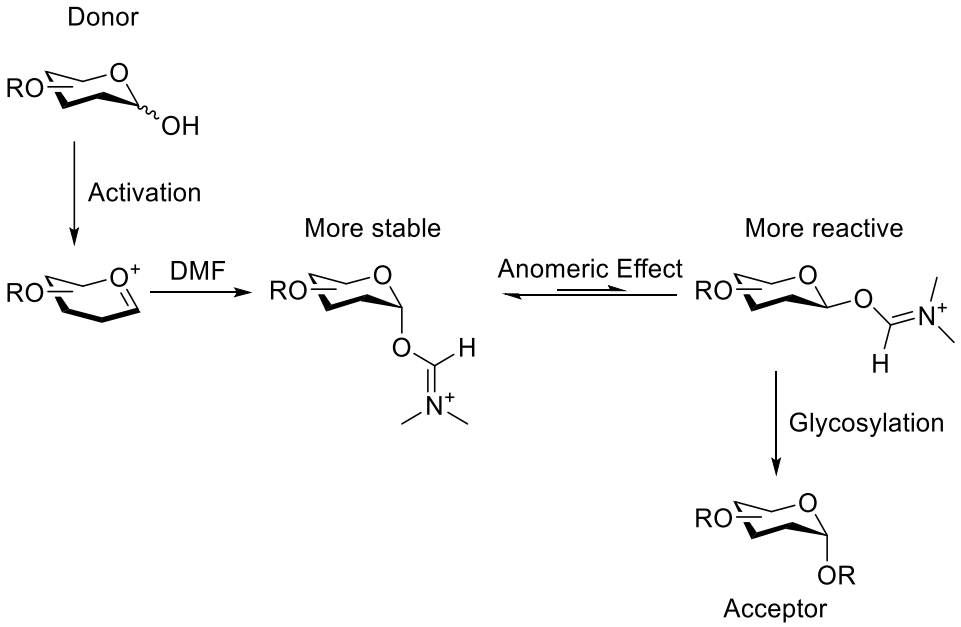
The reaction mixture had to be cooled to 0 °C before the $Cl_3CCN$ was added to avoid a runaway reaction occurring upon addition, however once added, the reaction was warmed to room temperature, and it proceeded without complication. Following purification via column chromatography (2:1 $Et_2O$:EtOAc) acetimidate **36** was yielded as a brown foam. Acetimidate **36** can then be used for a glycosylation reaction to form acetylated acceptor **27**. The initial synthesis involved reacting acetimidate **36** directly with 3-azidopropan-1-ol **34** (alcohol **34**), however, before this could be attempted, the supplier of this compound discontinued the product. This led to the proposed route depicted in Scheme 3.1.14, wherein alcohol **24** is synthesised from 3-bromo-propan-1-ol **37** (alcohol **37**).




Scheme 3.1.14: a) $NaN_3$ (excess), DMF, R.T., 4d, 83%. b) TMSTOf (0.12 eq), dry DCM, $N_2$, RT - -10 °C, 25 mins. c) NaOMe, dry MeOH, $N_2$, R.T., 24 hr.

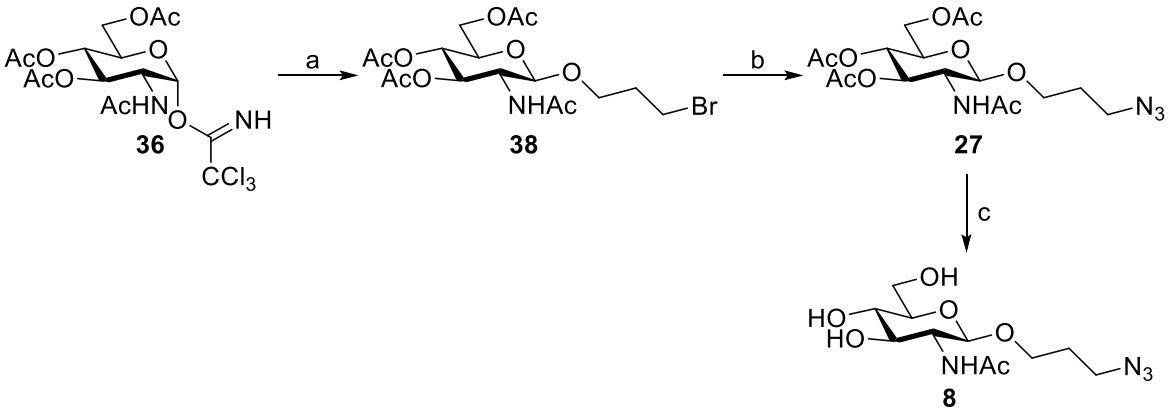
Commercially sourced alcohol **37** was converted to alcohol **34** via nucleophilic substitution of the bromine by sodium azide[198]. Alcohol **34** would then be used to perform a glycosylation of acetimidate **36** yielding azide **27**. This would then be globally deprotected using Zemplén conditions, yielding acceptor **8**[199].

The nucleophilic substitution of alcohol **37** was successfully performed affording alcohol **34** as a yellow oil in an 83% yield. This was left to react over 72 hr as TLC analysis showed there was still starting material remaining after two days, and this appears to have lead to higher than expected yield. This reaction had also been attempted previously, but the product was never used due to contamination with DMF leading to concern over the effect this could have on the stereochemistry of the glycosylation. DMF is well known to have an effect on the stereochemistry of glycosylations, promoting α formation over β[200,201] via the formation of a glycosyl intermediate (Scheme 3.1.15).




Scheme 3.1.15: Theoretical justification for why DMF causes α formation over β formation.

The Curtin-Hammett kinetic scenario states the β is the more reactive stereoisomer of the intermediate, and will therefore react faster than the α. In this case, the anomeric group is acting as a leaving group, leading to α product. As the glycosylation step of the reaction requires stereoinversion to form the β, this contaminant would be required to be removed before the next step was attempted. However, this time, the wash with $H_2O$ appeared to have removed all the DMF, and therefore the product can be used for glycosylations with no concerns over stereochemistry. This was not attempted, as simultaneously, the glycosylation was also attempted with alcohol **37**, as depicted in Scheme 3.1.16.




Scheme 3.1.16: a) Alcohol **37** (10 eq.), TMSTOf (0.12 eq), dry DCM, $N_2$, RT - -10 °C, 25 mins, 24%. b) $NaN_3$ (5 eq.), DMF, 60 °C, 3 hr, 44%. c) NaOMe, dry MeOH, $N_2$, R.T., 24 hr, 83%.

Acetimidate **36** was glycosylated using alcohol **37** yielding bromo glycoside **38**[194]. Bromo glycoside **38** was then converted to azide **27** via nucleophilic substitution. Azide **27** was then globally deprotected utilising Zemplén conditions yielding acceptor **8**. The first step proceeds with few

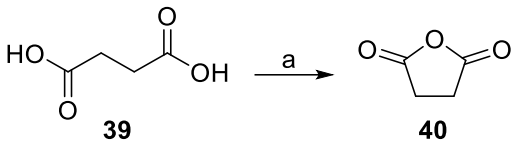
complications yielding bromo glycoside **38** as a colourless solid, however a large amount of product appeared to be lost during purification by column chromatography.

The nucleophilic substitution of bromo glycoside **38** into azide **27** proceeded with few complications, yielding azide **27** as a colourless solid. The aqueous wash of azide **27** appeared to have successfully removed the DMF from the product. It is possible that leaving the reaction for longer could have improved the yield as observed with the synthesis of alcohol **34**, however, this was deemed unnecessary as enough product was made to azide **27** was synthesised to proceed to the next step.

The global deprotection was then successfully performed on azide **27** to yield acceptor **8**. This required no further purification and was concentrated *in vacuo* leaving 15 mg of acceptor **8** as a red solid that could be used for glycosynthase assays.

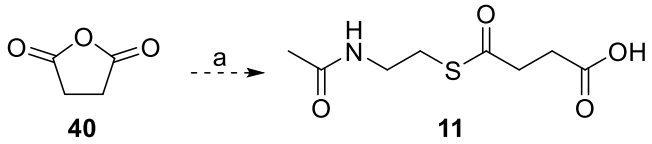
### 3.1.5. Succinyl-CoA Mimics

The synthesis of the succinyl CoA mimics depended on a route being developed for the formation of the thioester bond. For this purpose, anhydride **40** was targeted initially. The anhydride structure is important as it forms a key intermediate species in the synthesis of both azido thioester **12**, and the fluoro thioesters. Therefore, a reliable method for synthesising the succinic anhydrides was required. Manoni *et al*[202] described the method depicted in Scheme 3.1.17.




Scheme 3.1.17: a) $Ac_2O$ (excess), 80 °C, 2 hr, 97%.

Commercially sourced acid **39** was ring closed to yield anhydride **40**. This was achieved using an excess of acetic anhydride and proceeds with no complications, and anhydride **40** was produced in a 97% yield. Next, a method for producing the thioester bond from the anhydride was required. A patent filed in 2015[203] described a method for synthesising thioester **11** from anhydride **40** (Scheme 3.1.18).




Scheme 3.1.18: a) *N*-acetylcysteamine **41** (1 eq.), TEA (1 eq.), DMAP (0.1 eq.), THF, 70 °C, 16 hr.

In this method, anhydride **40** undergoes base catalysed thioesterification with *N*-acetylcysteamine **41** to produce thioester **11**. Initially in our hands, this appeared to produce some

positive results, as a peak corresponding to the sodiated mass of thioester **11** (m/z = 242.0463) could be seen in MS analysis of the reaction mixture (Figure 3.1.8). However, the peak was dwarfed by a much larger peak at m/z = 259.0542, and the compound could not be identified in an NMR of the crude reaction mixture.

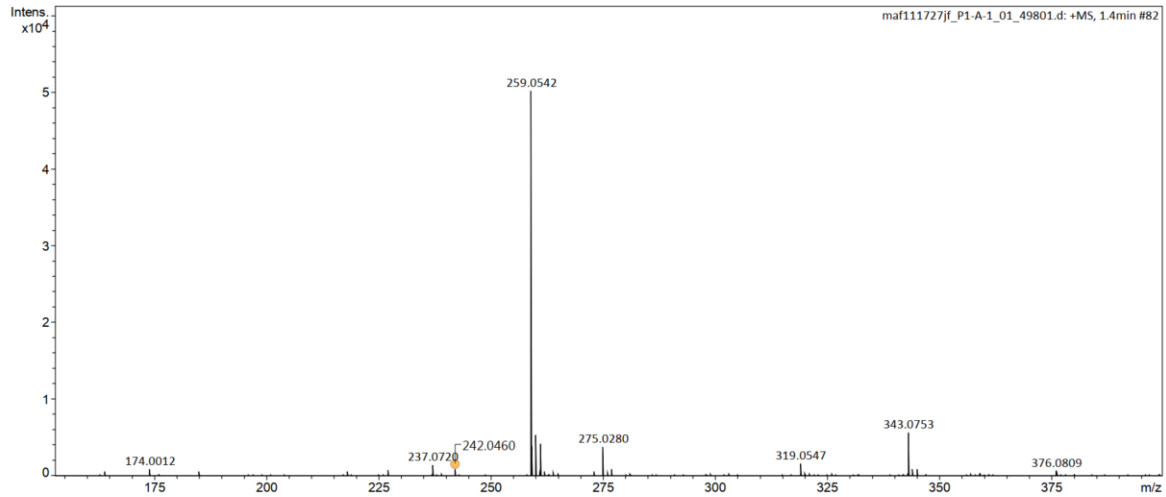
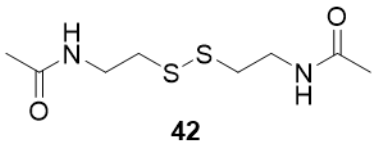



Figure 3.1.8: ESI-MS spectra of the reaction mixture for the synthesis of thioester **11**. The potential product mass is marked with a yellow circle.

This peak could not be assigned to any sort of adduct of thioester **11**, so the reaction mixture was purified via column chromatography, but this resulted in the loss of the product peak. The identity of the peak at m/z = 259 was determined to be a *N*-acetylcysteamine dithiol, dithiol **42** (Figure 3.1.9A), via ESI-MS analysis of the sample of *N*-acetylcysteamine **41** used for the reaction (Figure 3.1.9B)

A



42

B

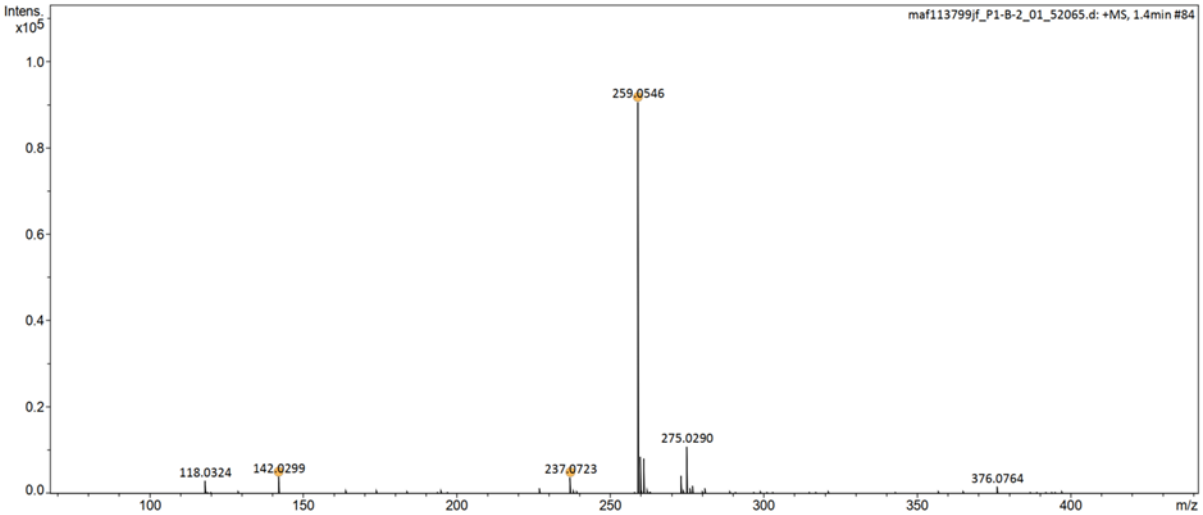
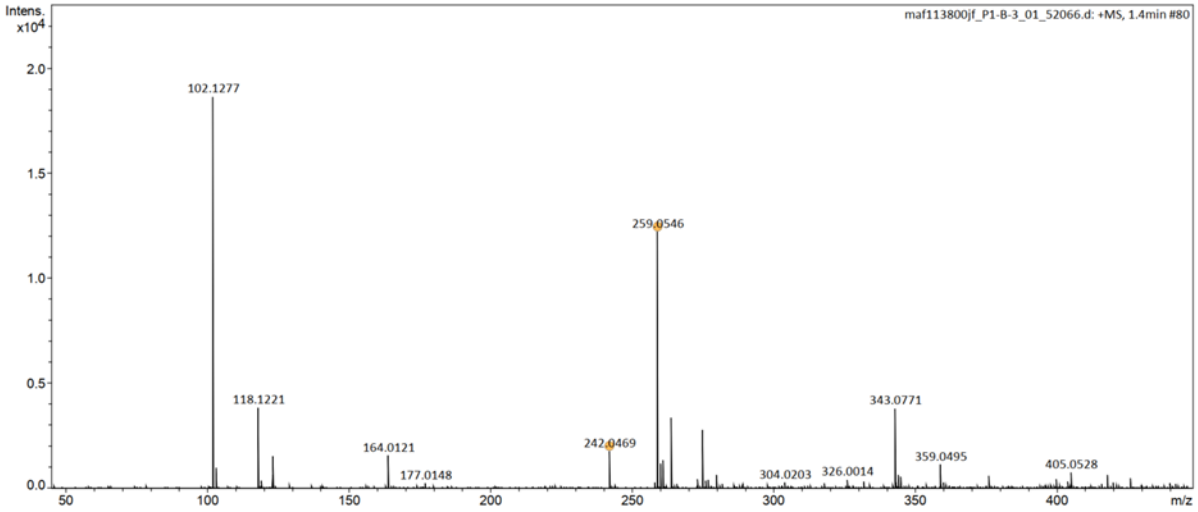


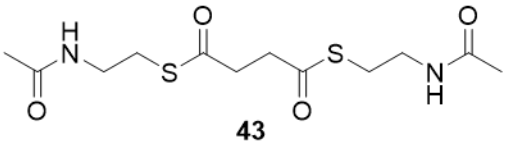

Figure 3.1.9: A) Structure of dithiol **42**, the impurity found in the *N*-acetylcysteamine **41** used during the thioesterification reaction. B) ESI-MS spectra of the *N*-acetylcysteamine **41** sample showing peaks corresponding to *N*-acetylcysteamine **41** (m/z = 142.0299), and dithiol **42** (m/z = 237.0723). The mass that appeared in the ESI-MS analysis of the thioesterification at m/z = 259 can also be seen and this corresponds to the sodiated dithiol **42**.

This would also explain why there was so little product formation, as dithiol **42** is unable to undergo thioesterification. In an attempt to counter this, a new bottle of *N*-acetylcysteamine **41** was purchased and placed under $N_2$ to prevent oxidation. However, MS analysis of this new batch also revealed large amounts of dithiol **42**, so in an effort to counter this issue, DMAP was added to the reaction mixture to act as a nucleophilic catalyst to accelerate the thioester formation, and the reaction was repeated. Again, initial ESI-MS analysis of the reaction mixture showed a large peak at 259, and an additional peak m/z = 343 (Figure 3.1.10A), which suggested a potential double esterification giving dithioester **43** (Figure 3.1.10B)

A



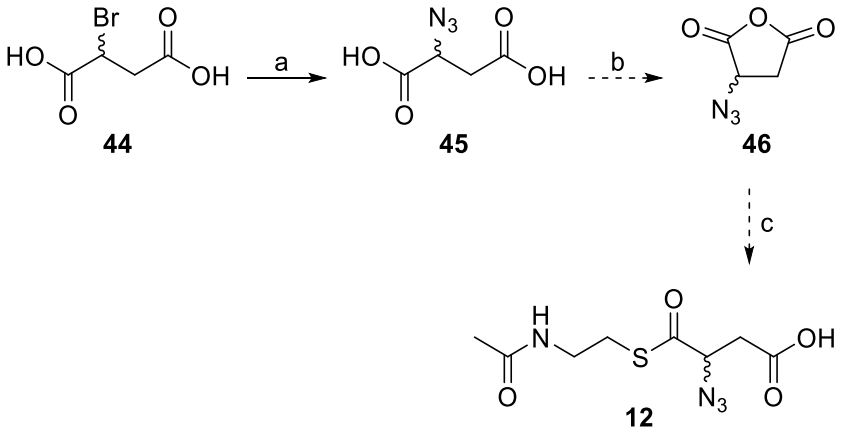

B



43

Figure 3.1.10: A) ESI-MS spectra of the reaction mixture for the synthesis of thioester **11** utilising DMAP to break the disulfur bond in dithiol **42**. Whilst a peak still exists for dithiol **42** can be still be seen, a peak for thioester **11** is also present (m/z = 242.0469), alongside a peak at m/z = 343.0771 which corresponds to a potential thiodiester **43**. B) Structure of the potential thiodiester **43**.

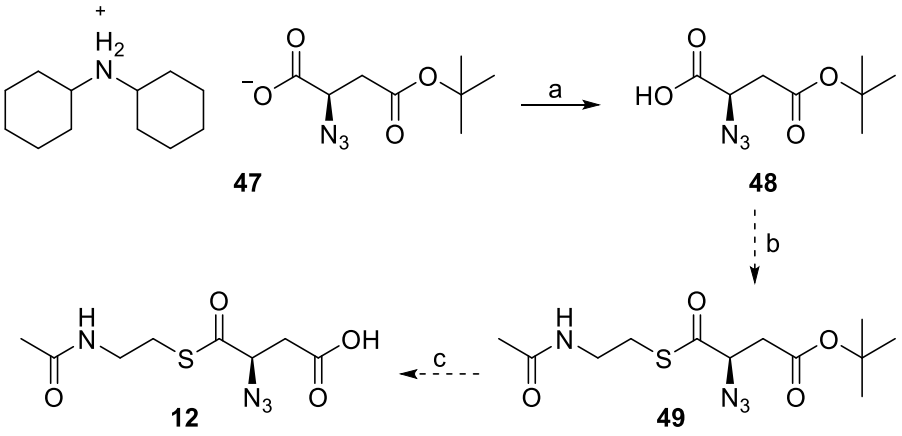
The presence of this mass alongside that of thioester **11** would suggest that the DMAP was able to successfully accelerate the reaction, but in doing so also catalysed the thioesterification of the second carboxylic acid group, resulting in the dithioesterification of thioester **11**. NMR analysis of the reaction mixture confirmed product formation, and the reaction mixture was purified utilising column chromatography. NMR of the purified fractions appeared to have significantly better defined peaks, however, ESI-MS analysis showed that dithiol **42** and dithioester **43** were still present after columning, so an accurate yield could not be determined. At this point, purification attempts were halted in favour of pursuing the synthesis of the other succinyl-CoA mimics. It was reasoned that the addition of functional groups to the succinic acid would increase the reactivity of the carboxylic acid group that they are in the α-position of, and therefore, improve the conversion of each acid to its corresponding thioester, while decreasing the amount of thioesterification on the second carboxylic acid. It was also hypothesised that any contaminating thiodiesters or dithiols would be unlikely to interfere with the incorporation of the modified succinyl groups into PNAG. First, azido-thioester **12** was targeted (Scheme 3.1.19).




Scheme 3.1.19: a) $NaN_3$ (excess), $H_2O$, 60 °C, 3d, 84%. b) $Ac_2O$ (excess), 80 °C, 2 hr. c) *N*-acetylcysteamine **41** (1 eq.), TEA (1 eq.), DMAP (0.1 eq.), THF, 70 °C, 16 hr.

In this proposed route, commercially sourced bromo acid **44** was converted into azido acid **45** via nucleophilic substitution[204]. Azido acid **45** then undergoes ring closure to yield azido anhydride **46**[202]. Azido anhydride **46** is then converted to azido thioester **12** via thioesterification[203]. The first step of this reaction is based off of work performed by Freudenberg *et al*[204] in 1939. This work does not state the length of time the reaction is expected to take, so the reaction was allowed to react over the weekend and checked with TLC to monitor progress. After 3 days, the TLCs showed no starting material remaining so the reaction mixture was lyophilised and analysed via NMR. $^1$H NMR analysis showed an approximately 84% conversion from bromo acid **44** to azido acid **45**. However, due to the polar nature of the two compounds, purification attempts proved difficult. Therefore, it was decided to take the crude product forward to the next step without purification.

The ring closure of azido acid **45** to anhydride was also described in Freudenberg *et al*'s work, wherein they utilise thionyl chloride to achieve the ring closure. However, as acetic anhydride was successfully able to perform the conversion of acid **39** to anhydride **40**, it was decided to use this method for azido acid **45**. This proved to be unsuccessful. No evidence could be found in either ESI-MS analysis, or in $^1$H-NMR analysis. It is unclear what caused this, however, as this reaction step was being attempted, another method for the synthesis of azido thioester **12** was being developed utilising dicyclohexylamine salt **47** (Scheme 3.1.20).



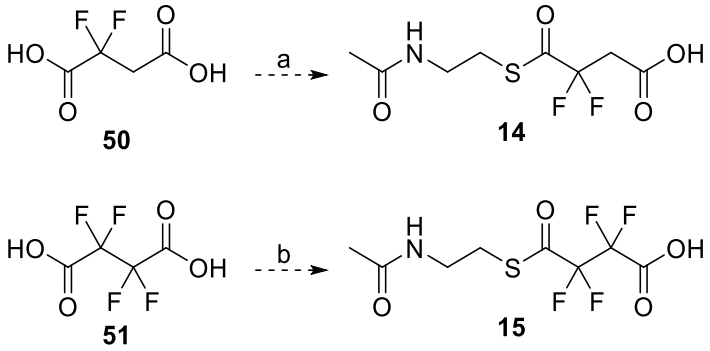

Scheme 3.1.20: a) Amberlite IR170, R.T., $H_2O$, qual.. b) *N*-acetylcysteamine **41** (1.1 eq.), DCC (1.1 eq.), DMAP (1 eq.), DCM, R.T., 16 hr. c) 85% w/v $H_3PO_4$, DCM, R.T., 16 hr.

In this proposed route, commercially sourced dicyclohexylamine salt **47** was converted into azido acid **48** via protonation of the carboxylic acid functional group. Azido acid **48** can then undergo a Steglich thioesterificiation to yield the *t*-butyl protected thioester **49**[178,205]. *t*-butyl thioester **49** can then be deprotected via acid cleavage of the *t*-butyl group to yield azido thioester **12**[206]. This method avoided the need to perform the cyclisation that caused the issues with the pervious method, and this also circumnavigates the issue with dithioesterification as the second carboxylic acid group is protected and therefore cannot undergo any reaction.

The first reaction step is fairly trivial, NMR confirmed the lack of the dicyclohexylamine following protonation and subsequent extraction using diethyl ether. Conversion was assumed to be 100% and no further purification was performed with the resulting white solid being using directly in the second step of the synthesis.

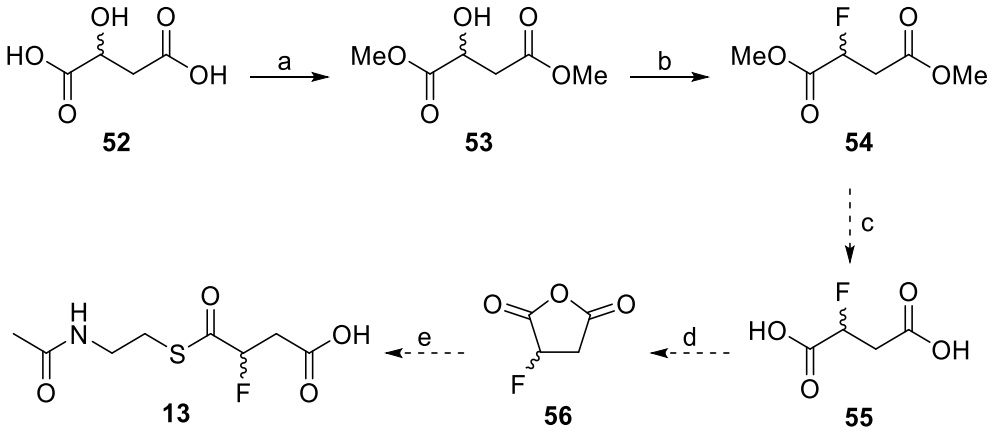
The second step differs from previous attempts to create this thioester bond in that it utilises a Steglich esterification protocol rather than the previously used TEA/DMAP method. The reason for this was that the DCC/DMAP protocol is a well known esterification procedure in the literature for producing thioesters, and works by improving the reactivity of the carboxylic acid towards the alcohol or thiol. It was reasoned that given the low conversion into the thioester in previous attempts, this increased reactivity could help push the reaction towards the formation of the thioester bond. The increased amount of DMAP utilised in this method could also increase the amount of available *N*-acetylcysteamine **41**, thereby also increasing conversion. However, several attempts at this synthesis showed this not to be the case. No evidence of thioesterfication could be found in ESI-MS analysis, and $^1$H-NMR analysis appeared to show that the two starting compounds, azido acid **48** and *N*-acetylcysteamine **41** had remained unreacted throughout the process. At this point synthesis attempts were abandoned due to the project coming to an end.

Finally, fluorosuccinyl-CoA mimics were also considered as targets, including the mono, di, and tetra fluoro versions. Difluoro succinic acid **50** and tetrafluoro succinic acid **51** can both be acquired commercially, and so a Steglich esterification was attempted with both (Scheme 3.1.21).




Scheme 3.1.21: a and b) *N*-acetylcysteamine **41** (1.1 eq.), DCC (1.1 eq.), DMAP (1 eq.), DCM, R.T., 16 hr.

The proposed route would see each acid undergo a Steglich thioesterification to yield the corresponding fluoro thioesters, **14** and **15**. However, as with the azido thioester attempts to use this method, no evidence of thioesterification could be detected using fluorosuccinic acids **50** and **51**, and further attempts at making the difluoro and tetrafluoro thioesters were abandoned in favour of the monofluoro thioester **13** (Scheme 3.1.22).




Scheme 3.1.22: a) Amberlite IR120H Resin, MeOH, R.T., 48 hr, 74%. b) DAST, DCM, -10 °C – R.T., 1 hr, 48%. c) 1M NaOH, MeOH, 0 °C, 2 hr. d) TFAA, *i*-propyl acetate, 60 °C, 2 hr. e) *N*-acetylcysteamine **41** (1 eq.), TEA (1 eq.), DMAP (0.1 eq.), THF, 70 °C, 16 hr.

In the proposed route, commercially sourced malic acid **52** was converted into dimethyl malate **53** via an acid catalysed double esterification. Dimethyl malate **53** then underwent fluorination of the hydroxy group via the fluorination reagent diethylaminosulfur trifluoride (DAST)[207] to yield fluoro dimethyl malate **54**. Fluoro malate **54** would then be deprotected via an acid catalysed hydroylsis to yield fluoro succinic acid **55**[207]. Fluoro acid **55** would then undergo a

cyclisation to produce fluoro anhydride **56**[208]. Fluoro anhydride **56** would then be ring opened and thioesterified to yield monofluoro thioester **13**[203].

The initial step proceeded with few complications and required no purification. Dimethyl malate **53** was synthesised in a good yield as a light red solid. The second step also proceeded to completion according to TLC analysis, and following neutralisation, the product required purification via column chromatography. This is likely the source of the lower yield, however, enough fluoro malate **54** was obtained to move forward to the next step of the synthesis as a light red solid.

The third step was attempted, however, NMR analysis showed that the methyl groups remained, suggesting that the reaction conditions were not basic enough. This could potentially be remedied by increasing the concentration of the NaOH, 1 M was chosen to mimic the literature reaction, however, that reaction was performed on a sugar molecule, rather than on a alkyl carboxylic acid, and therefore fluoro malate **54** might be able to withstand a more concentrated base. It is also possible a stronger base could be used, such as KOH, but neither option was attempted. At this point the synthesis of all succinyl-CoA mimics was abandoned due to the project coming to an end.

### 3.1.6. Conclusions and Future Work

In conclusion, carbamate **7**, oxazoline **5**, and acceptor **8** were successfully synthesised in a high enough yield to be taken forward for biological assays. The yield of carbamate **7** could potentially be improved via optimisation of step 3, as this step had some issues with stereoselectivity, and this could maybe be resolved by decreasing the amount of TEA in the reaction mixture.

Oxazoline **5** was successfully synthesised alongside CDMBI **23** in acceptable yield. Yield of oxazoline **5** could potentially be improved via increasing the reaction time, or increasing the amount of CDMBI **23** but this was deemed unnecessary.

Acceptor **8** was successfully synthesised via the use of acetimidate **36** as a glycosidic donor for the glycosylation with alcohol **37**. This reaction was low yielding but proved effective enough to proceed with the nucleophilic substitution and subsequent deprotection to yield acceptor **8**.

None of the proposed succinyl-CoA mimics could be synthesised successfully during the course of this project. Whilst some preliminary evidence would be produced to suggest that a small of amount of thioester **12** was produced via the thioesterification of anhydride **40**, no purification could be performed successfully, and this was unable to be repeated with any of the modified

thioesters that were proposed. Given more time, a method for the production of the thioester bond could be developed that avoids the issues that were encountered during this project, like that of dithiol **42**, or the thiodiesterification of succinic acid. However, this was not possible during this project, and the work has been handed over to another member of the Fascione group who may be more successful in the future.

## 3.2. Glycosynthase Assays

### 3.2.1. Introduction

Glycosynthases were first described in 1998 by Mackenzie *et al*[1] when the first one was created via a mutation of the active site of the glycoside hydrolase Abg that removed all hydrolysis activity, but allowed for transglycosylation. This new class of enzyme opened an exciting realm of possibilities for the generation of oligosaccharides and polysaccharides that previously had been unobtainable due to the nature of the synthesis involved, or the fact that the enzymes responsible for their production *in vivo* were too difficult to produce *in vitro*.

One such polysaccharide that has proven difficult to make *in vitro* is PNAG. The two characterised biosynthetic pathways of PNAG both contain mostly membrane bound proteins which makes their overexpression and purification difficult[42,46,66]. Therefore, it would advantageous if a glycosynthase could be developed that produced PNAG. In order to do so a glycoside hydrolase must first be able to break down PNAG, and within the literature there are 7 known proteins to be able to do this. Of these 7, only two have been fully characterised, DspB and PgaB[70,116]. For this project, DspB was chosen to attempt to be turned into a glycosynthase. There were a few reasons for this. One, at the time the project was being devised, the other 5 now known PNAGases were unknown to the field and so could not be selected for this purpose[118]. Two, DspB is solely a glycoside hydrolase protein that does not rely on other proteins for its activity, as opposed to PgaB, which is a dual function protein; also acting as a *N*-deacetylase[70]; which requires dimer formation for maximum *N*-deacetylase activity[71], and therefore could be reliant on said dimer for glycoside hydrolase activity as well. Finally, PgaB is a membrane protein[59] whilst DspB is excreted into the surrounding media. This makes DspB a much better candidate as, as mentioned before, membrane bound proteins have problems with their purification, often requiring strong detergents[209,210]. Additionally, the process of synthesising the PNAG oligosaccharides might be disrupted by the present of detergents, and therefore DspB represents a more suitable candidate for potentially creating a new glycosynthase.

Having chosen DspB as the candidate for creating a glycosynthase, the next question was which mutants would have the best chance of creating a glycosynthase. The classic method is to mutate one of the acidic residues in the active site into an alanine to remove the hydrolysis activity[1], and this method has previously been utilised to create a glycosynthase out of GH20 enzymes before when Schmölzer *et al* made glycosynthases from BdhI and LndB utilising this mutation[151,152]. It has also been shown that mutation of the equivalent residue to Y278 in DspB can create glycosynthases in GH20 enzymes when Slámová *et al* utilised Y→N and Y→F mutations to generate glycosynthase

activity in a fungal GH20 called TfHex[155]. These three mutations, D183A, Y278N, and Y278F, were therefore selected as the best candidates for producing glycosynthase activity in DspB and plasmids encoding for these proteins were purchased from GenScript. A plasmid encoding for the mutant E248Q was then also purchased in order to confirm the results of Breslawec *et al*[85] that show this mutant acts as a more efficient glycoside hydrolase. Once these results had been confirmed, the next step would be to create an E248Q double mutant of any mutant shown to act as a glycosynthase, which would hopefully have the same effect, creating a more efficient glycosynthase.

With these enzymes selected, and the substrates with which they were to be tested successfully synthesised, the hypothesis that DspB could be turned into a PNAG producing glycosynthase could be tested.

### 3.2.2. Overexpression and Purification of WT DspB and DspB Mutants

Having successfully synthesised the donor substrates for the glycosynthase reaction, we began focussing on the second initial goal of the project, the generation and purification of potential glycosynthase mutants of DspB. The chosen mutants, D183A, Y278F, and Y278N, were overexpressed in BL21 DE3 chemically competent cells. Overexpression was performed by addition of isopropyl-β-D-1-thioglacatopyranoside (IPTG) to a final concentration of 0.2 M. Successful overexpression was determined by performing a BugBuster protocol on a small portion of the cell pellet from the cell cultures. The resulting soluble fraction was analysed by 12% acrylamide SDS-PAGE gel (Figure 3.2.1)

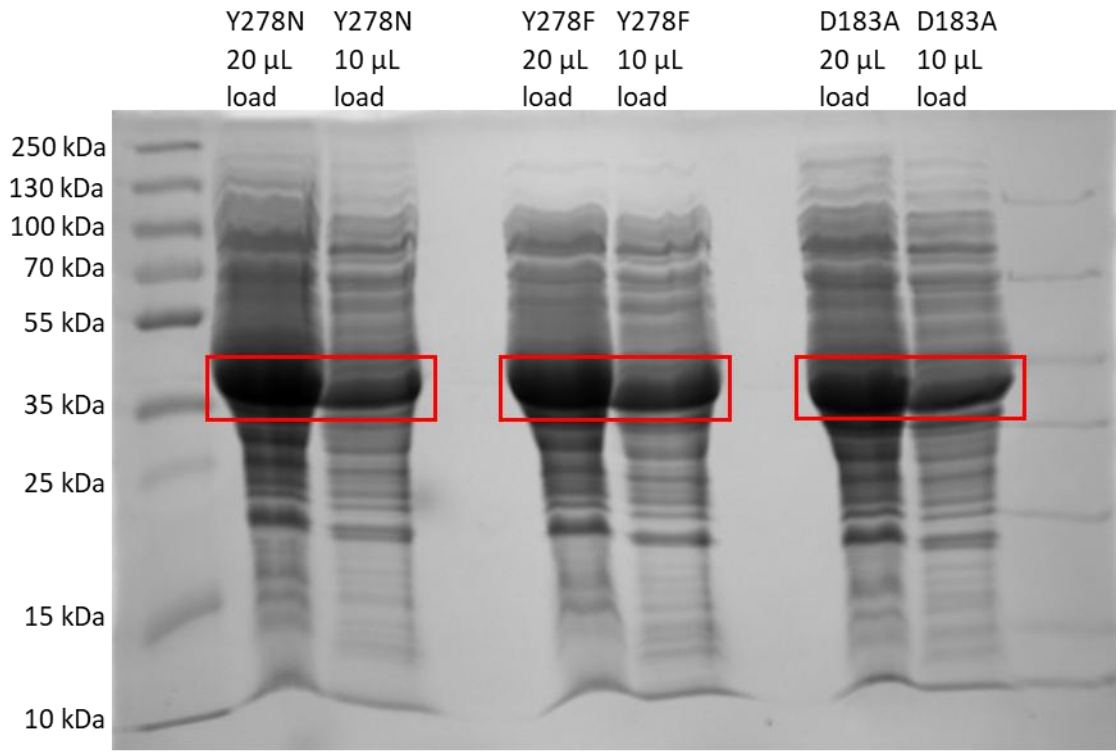



Figure 3.2.1: 12% acrylamide SDS-PAGE gel ran with a 20 µL and 10 µL aliquot of the soluble fraction of the BugBuster lysate of each of the three mutant protein cell pellets. All molecular weights are approximate.

DspB has a $M_w$ of 40 kDa, and the overexpressed band (red boxes) appears just after the 35 kDa band so this was judged to have been a success.

The cell pellets were lysed and sonicated to break open the cells, and centrifuged to pellet the debris. The supernatant was subjected to a 5 mL HisTrap column (Figure 3.2.2).

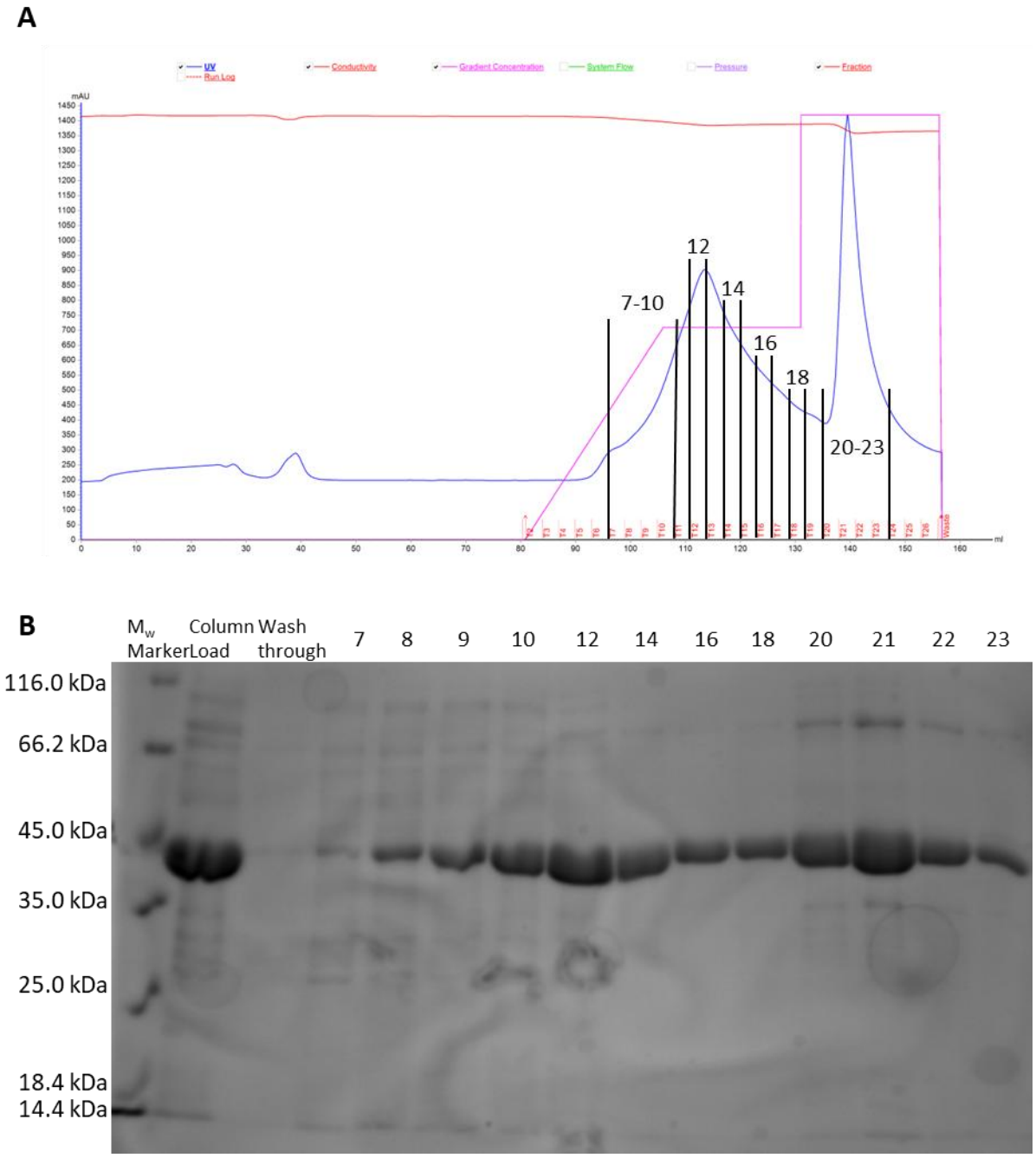



Figure 3.2.2: A) Chromatogram of the purification of DspB Y278N. The fractions selected for analysis by SDS-PAGE are marked. B) 12% acrylamide SDS-PAGE showing the fractions from the purification of DspB Y278N. All molecular weights shown are approximate

Fractions were then collated together, and the enzyme was dialysed into pH 5.9 100 mM phosphate buffer saline (PBS). This process was repeated for both D183A and Y278F (Appendix 25 and Appendix 27). Upon purchasing the plasmids for WT DspB and E248Q, this process was repeated for these enzymes too. E248Q proceeded exactly as the other mutants had (Appendix 28), however the WT produced an interesting result (Figure 3.2.3).

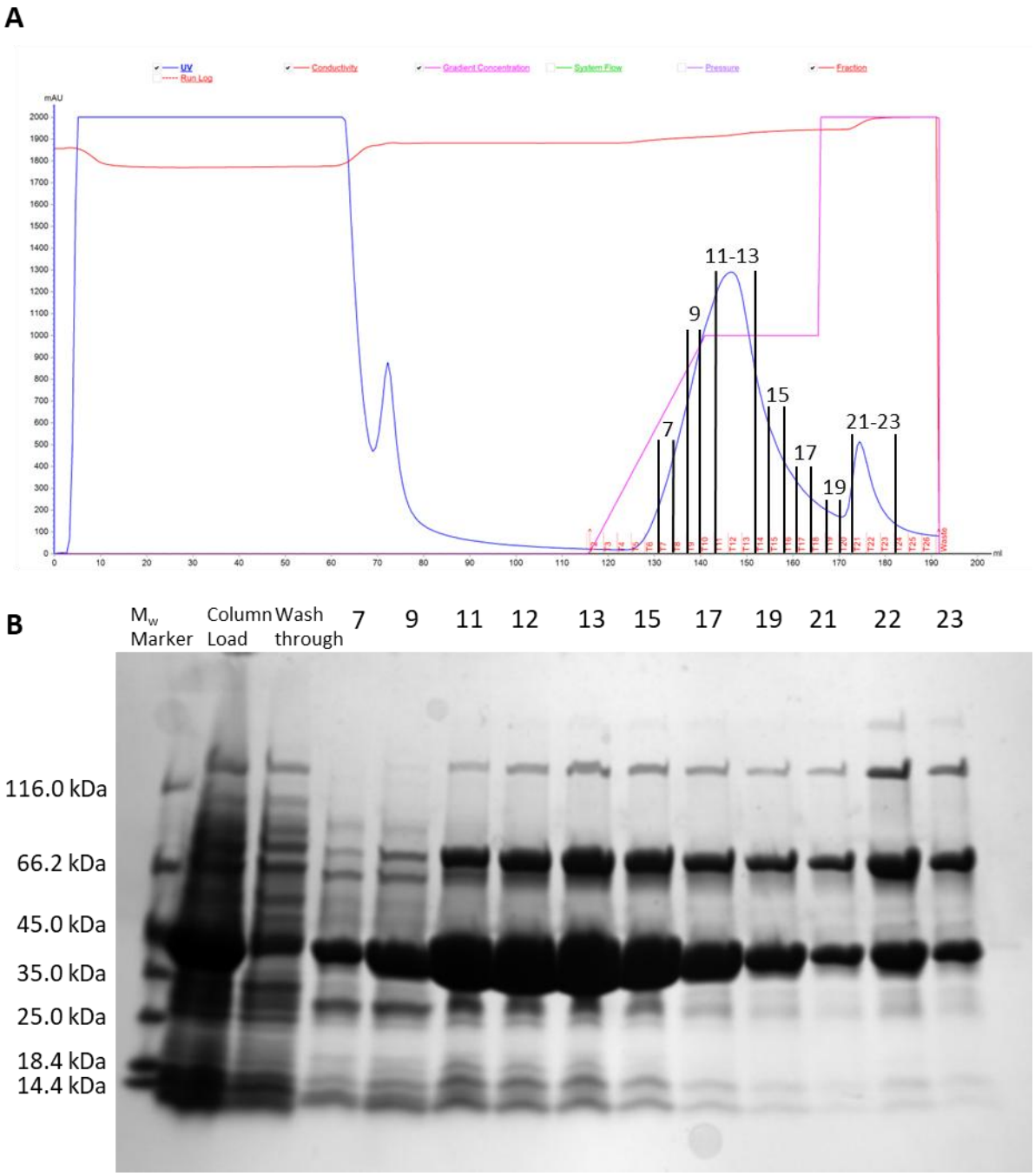



Figure 3.2.3: A) *Chromatogram* of the purification of WT DspB. The fractions selected for analysis by SDS-PAGE are marked. B) 12% acrylamide SDS-PAGE showing the fractions from the purification of WT DspB. All molecular weights shown are approximate.

In the gel of the purification, each fraction of the purification contained two other large bands that were potential contaminants. However, upon inspecting the sequence of DspB it was noticed that a N-terminus construct had been left in the sequence (Appendix 22-24). This construct contained a Cys-Cys pair which it was theorised could be forming disulfide bonds. This would explain the two bands as each band is approximately 40 kDa apart, and so would be the correct size for a

dimer or a trimer respectively. To test this, a small sample of the protein was incubated with 100 mM DTT to break down any potential disulfide bonds (Figure 3.2.4)

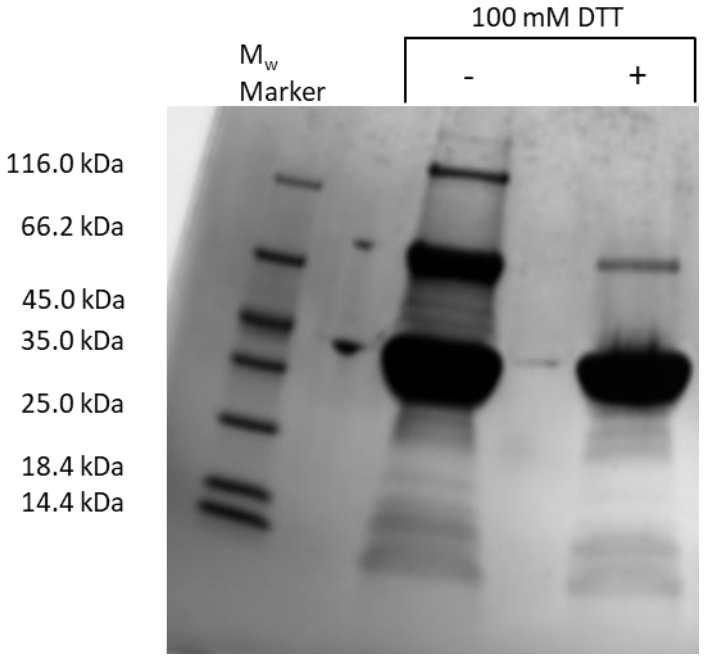
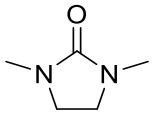



Figure 3.2.4: 12% acrylamide SDS-PAGE comparing a sample from the WT DspB purification attempt with the same sample after being treated with 100 mM DTT.

This caused the potential dimer and trimer bands to dramatically decrease in size, or disappear completely. Whilst it was unclear why the β-mercaptoethanol (BME) in the reducing buffer used in the gel preparation was unable to break these disulfides bonds, it was concluded that the purification had been successful, and the potential contaminants were just DspB dimers and trimers. The purification fractions were then collated, and dialysed into 100 mM PBS as with the previous purifications.

### 3.2.3. Determination of Glycosynthase Activity via Enzyme Assays

In order to determine if the selected mutants were active as glycosynthases, several assays were designed and performed utilising the various substrates synthesised in section 3.1, as well as commercially obtained *p*NP-GlcNAc **4**. The first thing that needed to be determined was whether or not 1,3-dimethyl-2-imidazolinone (DMI **57**, Figure 3.2.5), the byproduct product of DMC **22** catalysed oxazoline formation, would act as an inhibitor of DspB hydrolysis activity. Noguchi *et al*[188] state in their 2012 paper that DMI **57** strongly inhibits transglycosylation activity when attempting to perform a one-pot transglycosylation utilising DMC **22** for oxazoline formation. This was important as at this initial stage, oxazoline **5** hadn't been successfully purified, and so in order to ustilise oxazoline **5** as a donor substrate, the unpurified reaction mixture containing DMI **57** would have to be used.



**57**

Figure 3.2.5: Structure of DMI **57**, a potential inhibitor of glycosynthase activity

It was rationalised that if WT DspB could still successfully hydrolyse *p*NP-GlcNAc **4** in the presence of water solubilised 15 mM DMC **22,** which was allowed 1 hour to be hydrolysed into DMI **57**, then the potential glycosynthase mutants would be able to perform the enzymatic glycosylation reaction with this impurity present. For the purpose of this assay, WT DspB was donated by Dr Alexandra Males from the Davies Lab (University of York) which was truncated compared to the WT DspB overexpressed in section 3.2.2 and didn't contain the N-terminus construct (Figure 3.2.6). This construct was used as this was before the WT DspB had been successfully overexpressed and purified.

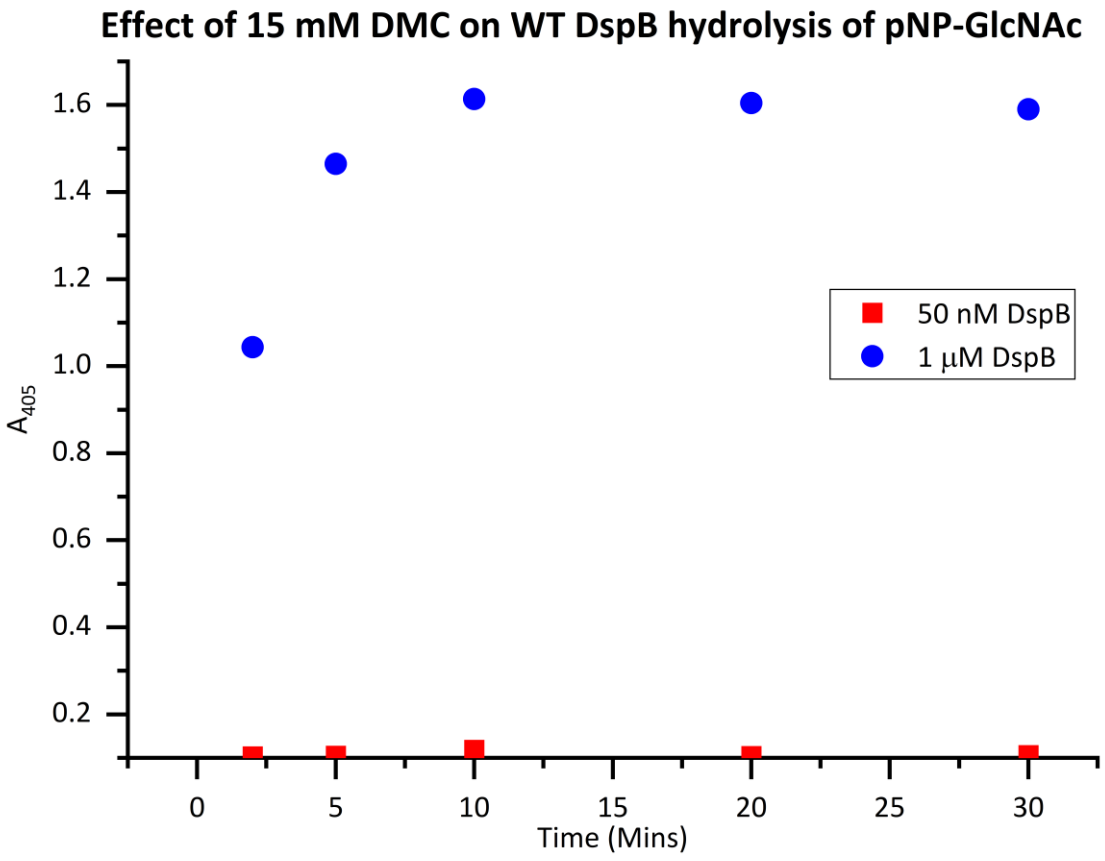



Figure 3.2.6: Enzyme activity assay of WT DspB at 50 nM (red squares) and 1 µM (blue circles) in the presence of 5 mM pNP-GlcNAc **4** and 15 mM DMC **22** in100 mM pH 5.9 PBS . 15 mM DMC **22** is the reaction concentration used synthesis of oxazoline **5** as detailed in section 3.1.3.

From this assay, it is clear that the WT DspB was still capable of performing the hydrolysis of *p*NP-GlcNAc, as an absorbance signal was still detected at $A_{405\ nm}$, indicating hydrolysed *p*NP-

GlcNAc **4**. With this knowledge, it was determined that unpurified oxazoline **5** could be utilised as a glycosidic donor for glycosynthase assays going forward. Acceptor **8** was selected as the glycosidic acceptor for these assays. Initially the reactions were performed at a 2:1 equivalence of oxazoline **5** (10 mM in $dH_2O$) to acceptor **8** (5 mM in $dH_2O$). Each sample was kept at room temperature in 100 mM pH 5.9 PBS and a time point was taken and quenched in MeCN from each sample after 15 mins, 1 hour, as well as being left overnight. The samples were then analysed by thin layer chromatography (TLC)(Figure 3.2.7).

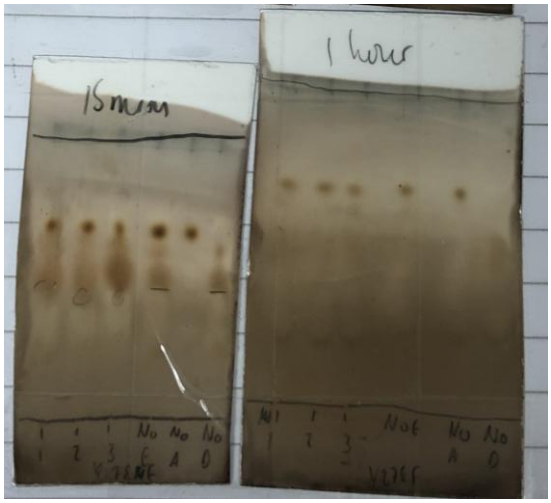



Figure 3.2.7: TLC plates run after 15 mins and 1 hour of the Y278F oxazoline **5** and acceptor **8** assay. The TLC plates were run using a 2:1:1, n-Butanol:$H_2O$:AcOH, solvent system, and stained using anisaldehyde stain.

TLC analysis proved to be ineffective for these assays, none of the three enzymes showed any sign of glycosynthase activity, and the sugars themselves did not stain particularly well. This led to some initial speculation that perhaps none of the mutants were in fact glycosynthases. Upon HILIC-LC-MS analysis of the overnight samples, some potential evidence for glycosynthase activity was discovered (Figure 3.2.8).

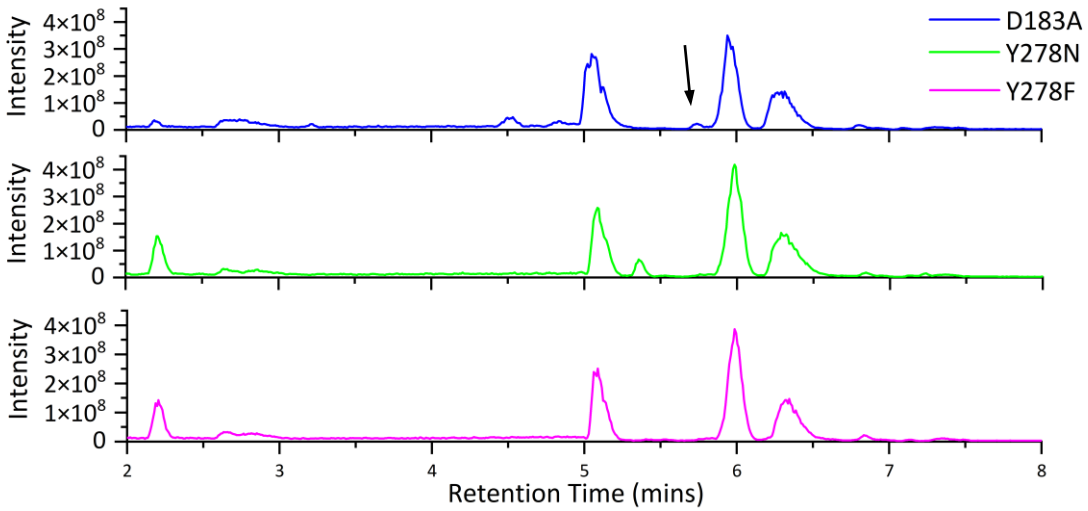



Figure 3.2.8: Full MS trace of all three DspB mutants after being left overnight with 5 mM acceptor **8** and 10 mM oxazoline **5**. Further analysis revealed a mass that matches that of a potential disaccharide in this peak.

These LC-MS chromatograms appear to mostly be a combination of the chromatograms for acceptor **8** and donor **5** (Appendix 31), however, in the LC-MS chromatogram for the overnight D183A sample (blue line), a small peak elutes (5.6 – 5.9 mins) before the larger peak at ≈6 mins (Figure 3.2.8, marked by a black arrow). Extracted ion chromatogram (EIC) analysis of the D183A MS trace revealed that this peak contained two peaks at 508 and 530 (Figure 3.2.9).

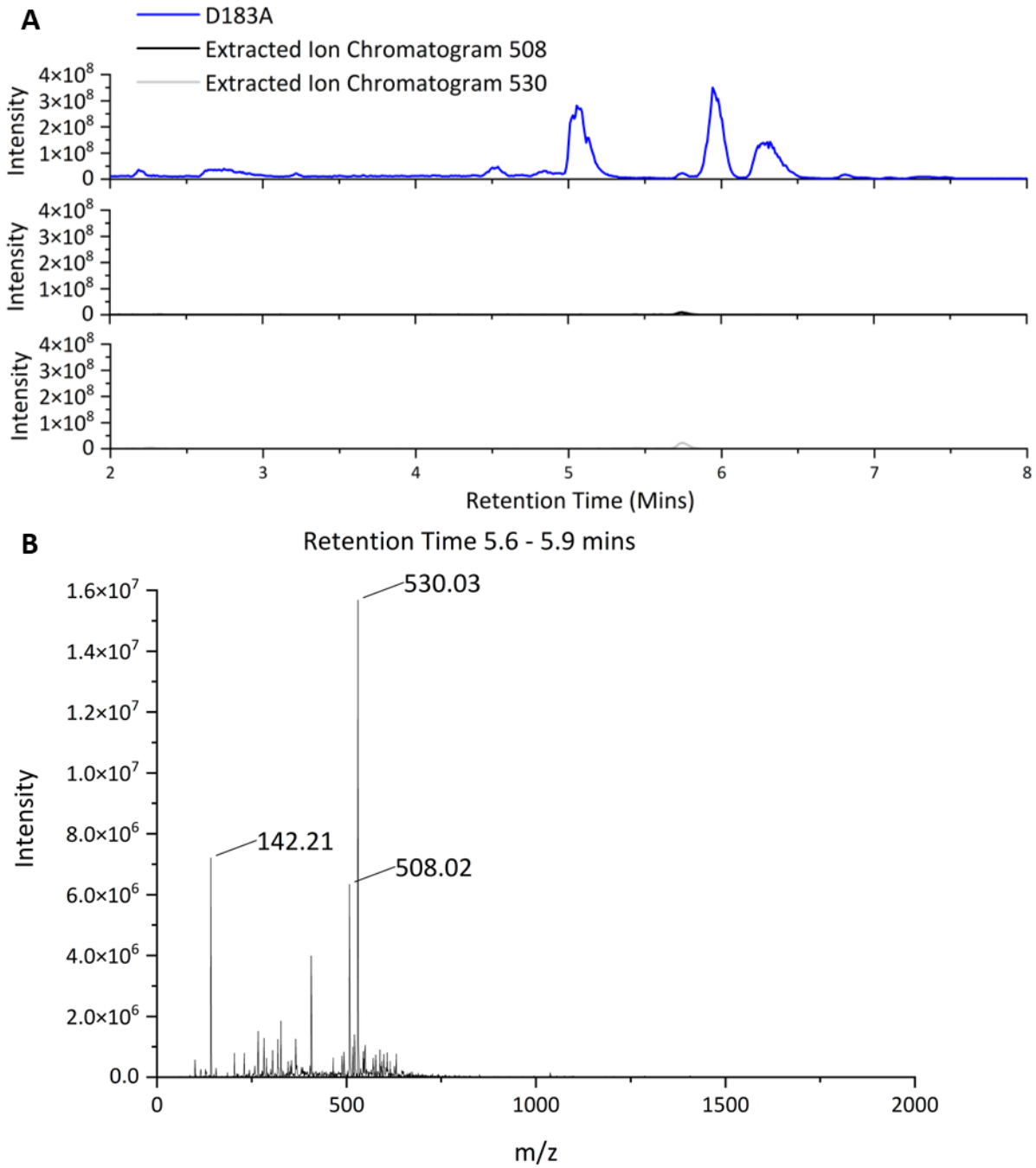
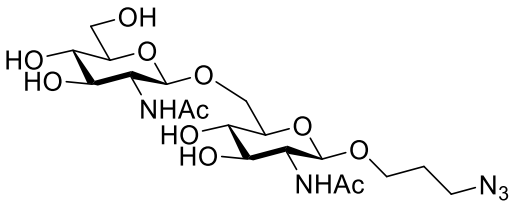



Figure 3.2.9: A) Extracted ion chromatograms of the overnight D183A sample showing masses 508 and 530. B) Mass spectra for retention time 5.6-5.9 mins of the D183A overnight sample.

These two masses correspond to the [M+($H^+$)] and [M+($Na^+$)] adducts of the disaccharide (Figure 3.2.10).




**58**

Exact Mass: 507.2177

Figure 3.2.10: Structure and exact mass of disaccharide **58** formed from the glycosylation of acceptor **8** and oxazoline **5**.

These results were repeated for both the Y278N and Y278F mutants (Appendix 29 and Appendix 30). There is a small potential peak in the same place in one of the three no enzyme control reactions (Figure 3.2.11), however, this was not present in the two repeats of the no enzymes controls (Appendix 31) so this was considered anomalous. No evidence of trisaccharide formation was found, nor any longer oligosaccharide, suggesting that perhaps more donor is required to push this reaction in favour of longer oligosaccharide formation. However, due to the unknown effect of increasing the DMI **57** concentration past the tested 15 mM, the oxazoline **8** concentration was left at 2 eq. while other variables could be tested.

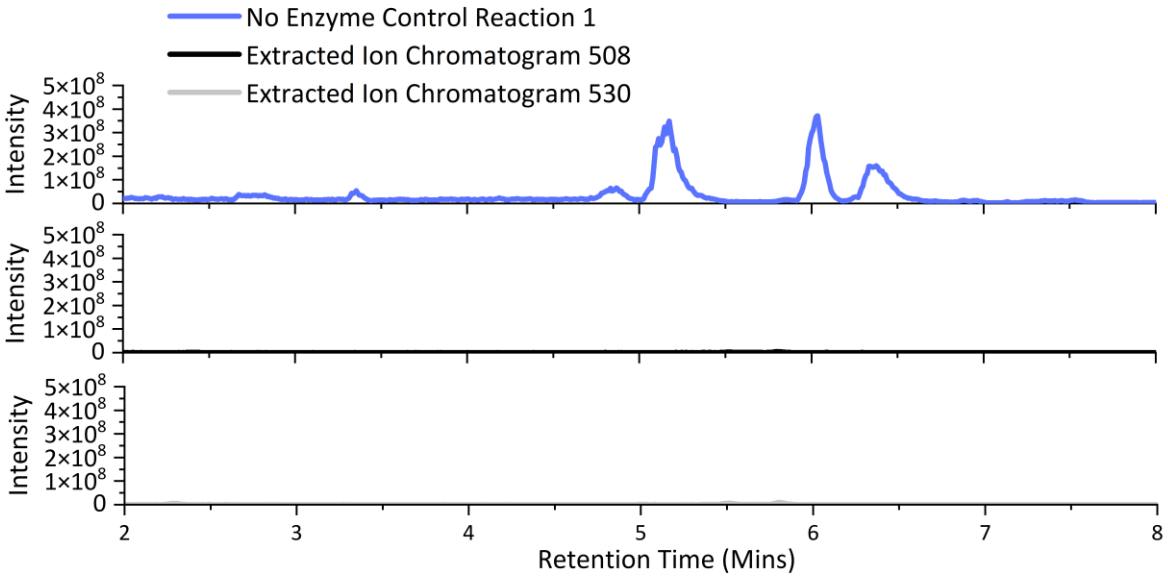



Figure 3.2.11: A) Extracted ion chromatograms of the overnight no enzyme sample showing masses 508 and 530. B) Mass spectra for retention time 5.6-5.9 mins of the no enzyme overnight sample.

Temperature was the first variable that was changed in an attempt to improve any potential disaccharide formation. The reaction was repeated but the samples were incubated at 37 °C (Figure 3.2.12). For these assays, D183A was focussed on as it appeared to have the most disaccharide **58** formation in the previous assays.

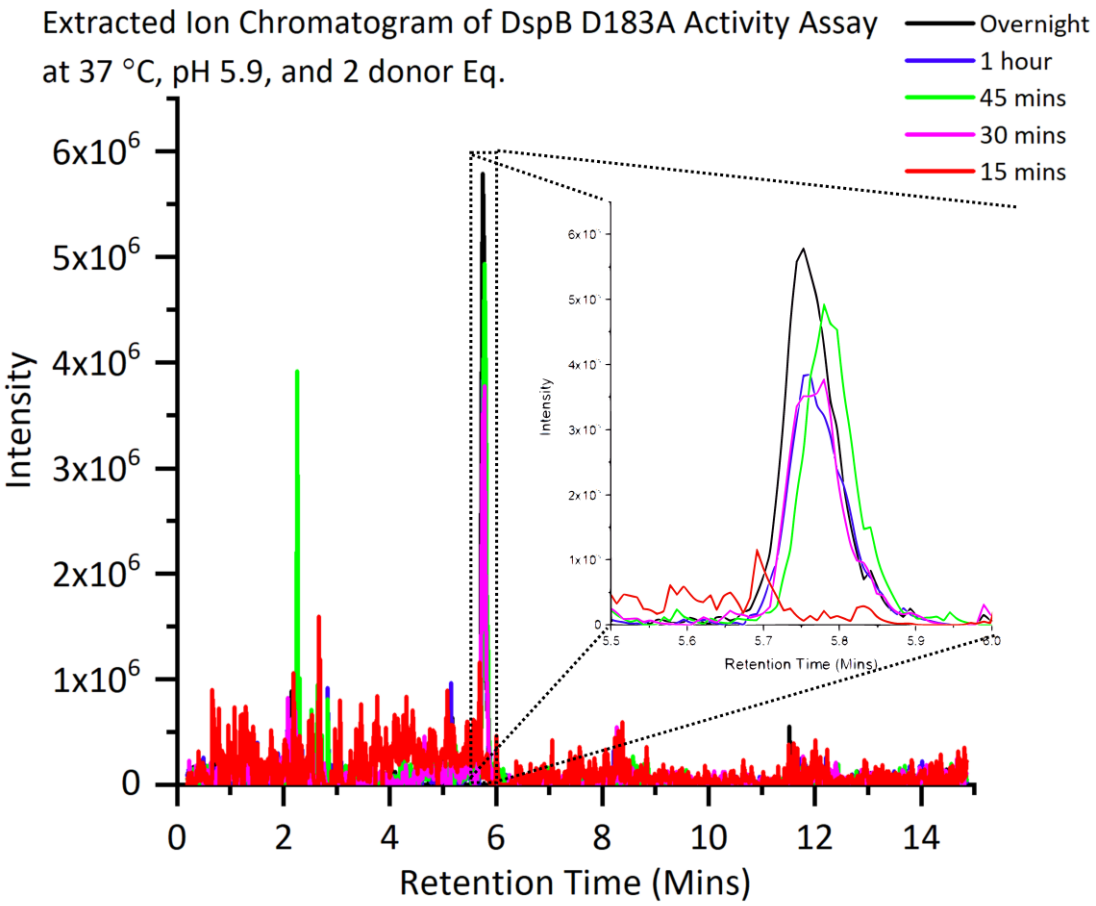



Figure 3.2.12: Overlaid EIC for the mass 530 for the 37 °C D183A glycosynthase activity assay.

EIC analysis of these reactions shows a peak at the same retention time that contains the same masses that could potentially correspond to disaccharide **58**, however, the relative size of the peak did not increase, perhaps suggesting that the amount of disaccharide **58** produced during the reaction was unaffected by the increase in temperature. Oxazoline **5** is quite acid sensitive and can break down into GlcNAc **1** under acidic conditions. It was therefore speculated that perhaps changing from pH 5.9 to pH 7 could potentially allow for more oxazoline **5** to be available during the reaction. Manuel *et al* demonstrated that while pH 5.9 is the optimum pH for DspB, it can perform enzymatic hydrolysis at pH 7[119], and so therefore it was assumed that any glycosynthase activity would also be possible at this pH. This hypothesis was tested by repeating the enzyme assay in pH 7 100 mM PBS (Figure 3.2.13). The temperature was left at 37 °C for comparison purposes.

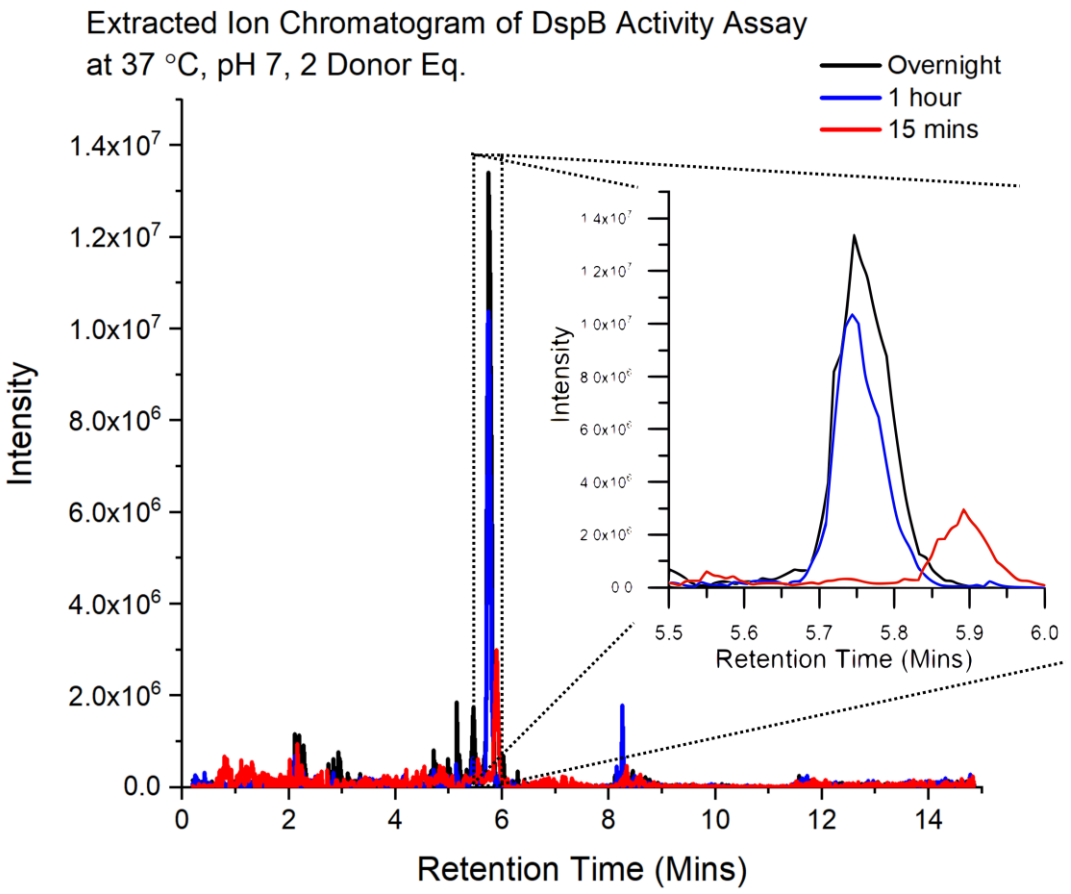



Figure 3.2.13: Overlaid EIC for the mass 530 for the pH 7 D183A glycosynthase activity assay.

From these results, it can be seen that, again, the mass of the potential disaccharide **58** could be found within the LC-MS trace, and therefore, it can be concluded that D183A is also potentially capable of performing as a glycosynthase at pH 7. However, when the intensity of this peak is compared to the previous assays at pH 5.9, the size of this peak has actually decreased from the original pH 5.9 room temperature assay (Figure 3.2.14).

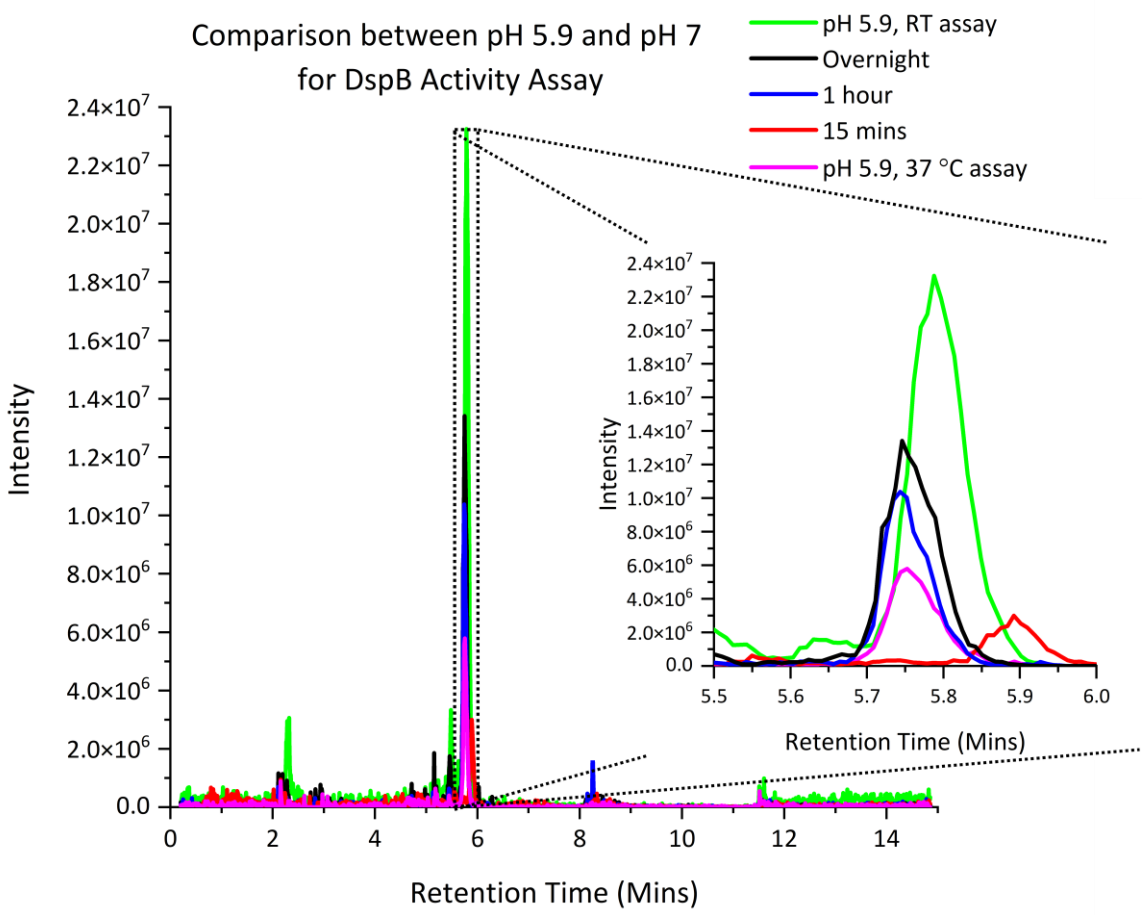



Figure 3.2.14: Overlaid EIC analysis of the pH 7 assay and the previous two pH 5.9 assays

This might suggest that the best conditions for any glycosynthase assays is pH 5.9 and room temperature. However, while these initial results are promising, they are neither conclusive nor quantitative. As disaccharide **58** contains an azido group, Click chemistry can be used to attach a reliably ionisable group to make the results more quantitative. As Click chemistry can only be performed with molecules containing an azide, anything that gets tagged during the process of performing the Click chemistry must also therefore contain an azide, and therefore if the mass seen throughout the previous assays is in fact disaccharide **58** it will be tagged and will be visible in LC-MS. Similarly to work performed by Sittel *et al.*[175], imidazolium tag (ITAG) **59** is attached via Copper (I) catalysed azide-alkyne cycloaddition (CuAAC), as described by Kolb *et al*[164] (Figure 3.2.15A).

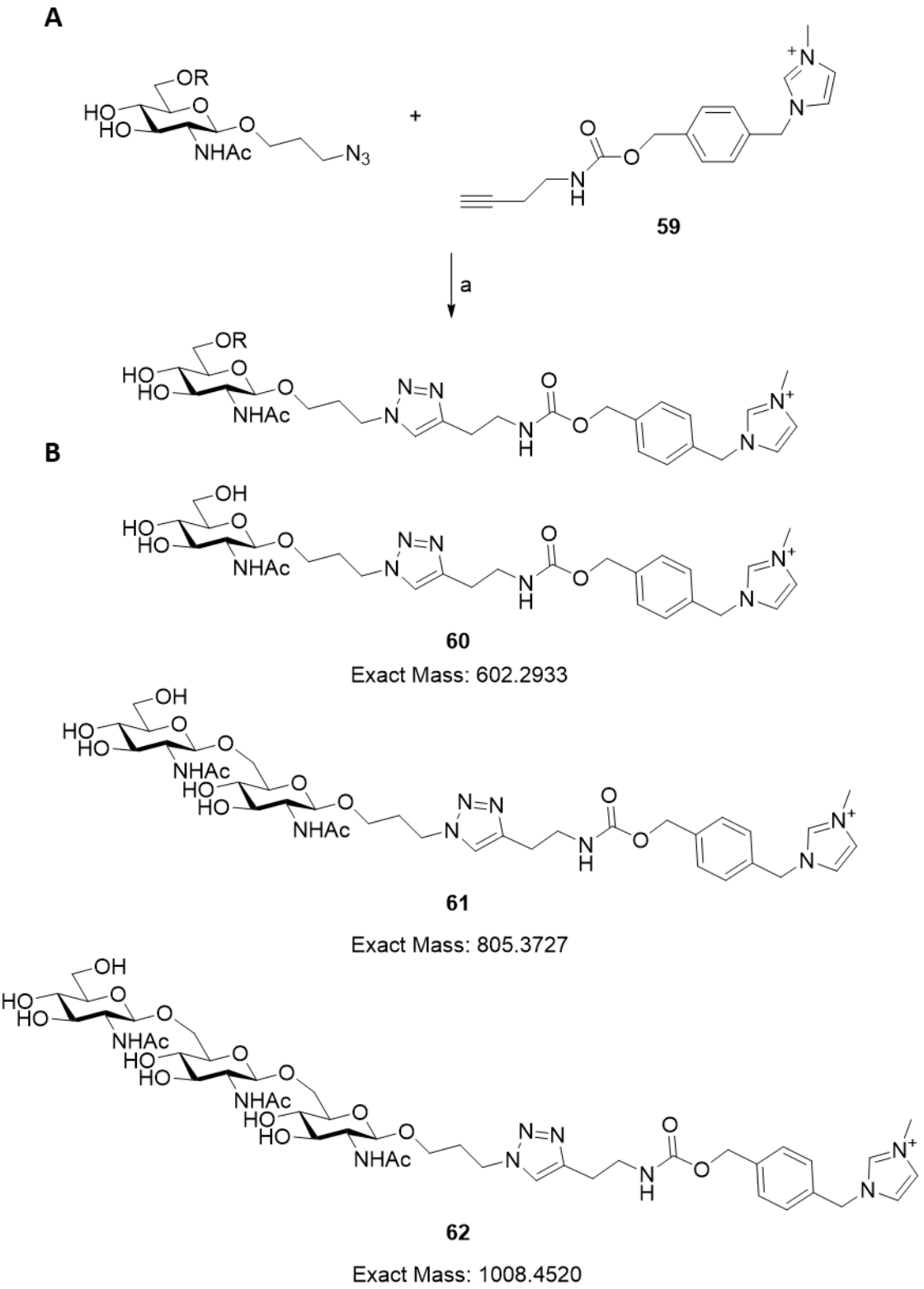
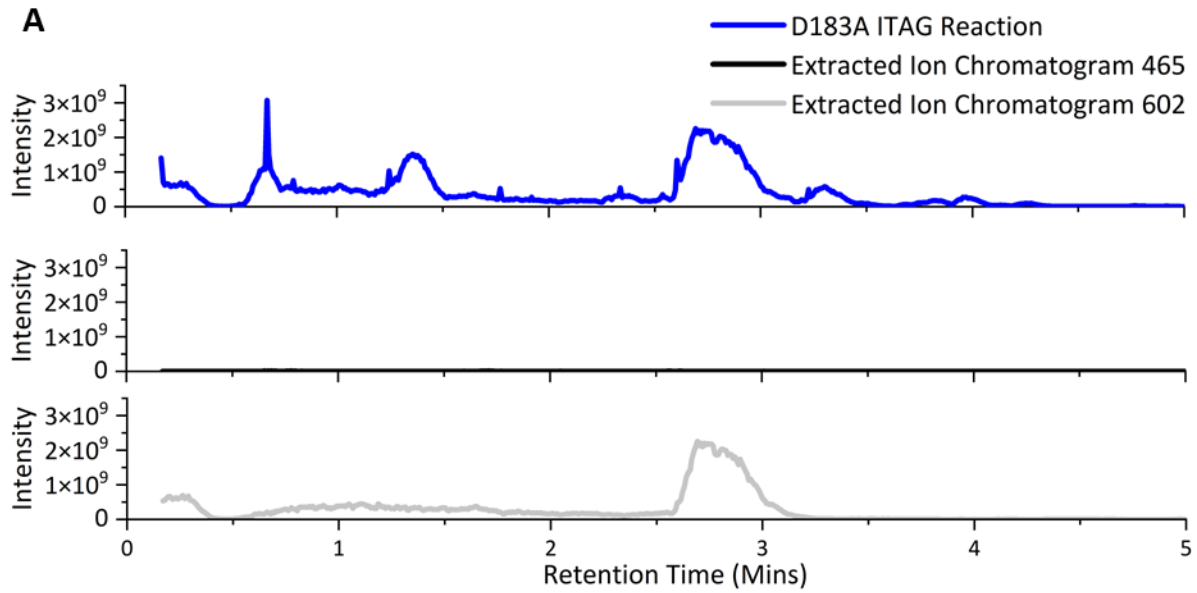
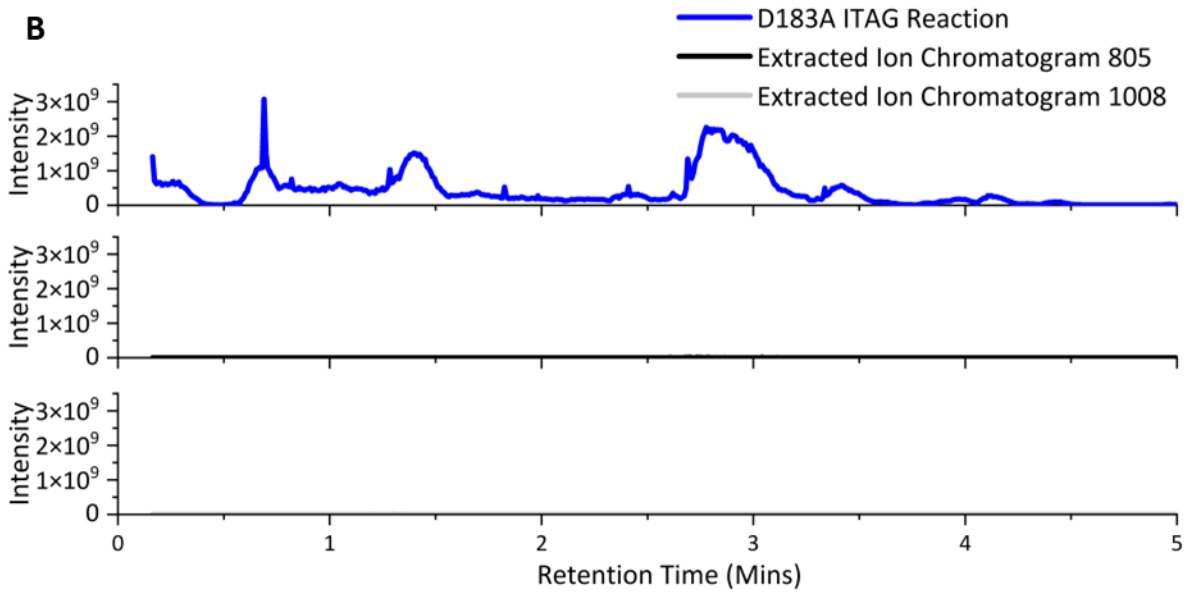



Figure 3.2.15: A) General scheme for the CuAAC ITAG reaction. a) $CuSO_4.5H_2O$ (10 eq), THPTA (10 eq), Na ascorbate (20 eq.), R.T., 2 hr. B) The structure and exact masses of the mono-, di-, and trisaccharide version of the ITAGed oligosaccharides

Any tagged molecule should be clearly visible due to how readily ionised the tag is. When this was performed on the assay solutions from the TLC assays however, no peak resembling the

ITAGed disaccharide **58** could be found for any of the three mutants when analysed by reverse-phased LC-MS (Figure 3.2.16).


A
D183A ITAG Reaction
Extracted Ion Chromatogram 465
Extracted Ion Chromatogram 602
Intensity
3×10^9
2×10^9
1×10^9
0
0
1
2
3
4
5
Retention Time (Mins)


B
D183A ITAG Reaction
Extracted Ion Chromatogram 805
Extracted Ion Chromatogram 1008
Intensity
3×10^9
2×10^9
1×10^9
0
0
1
2
3
4
5
Retention Time (Mins)

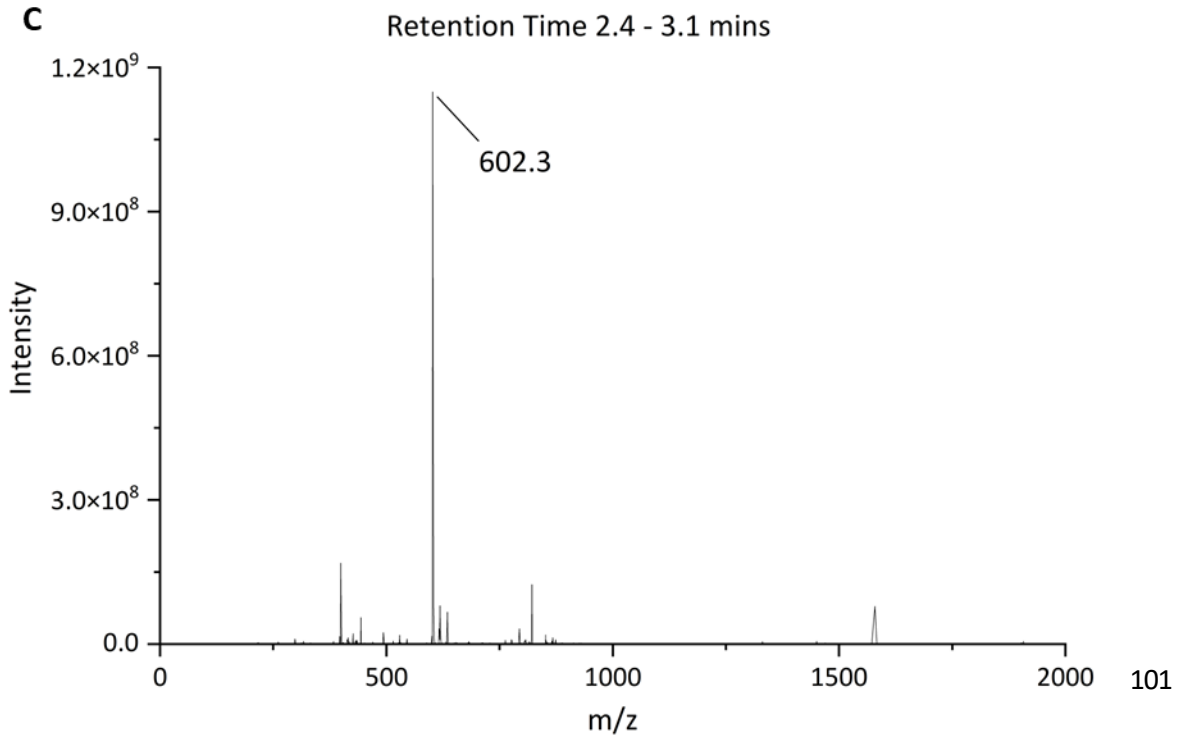
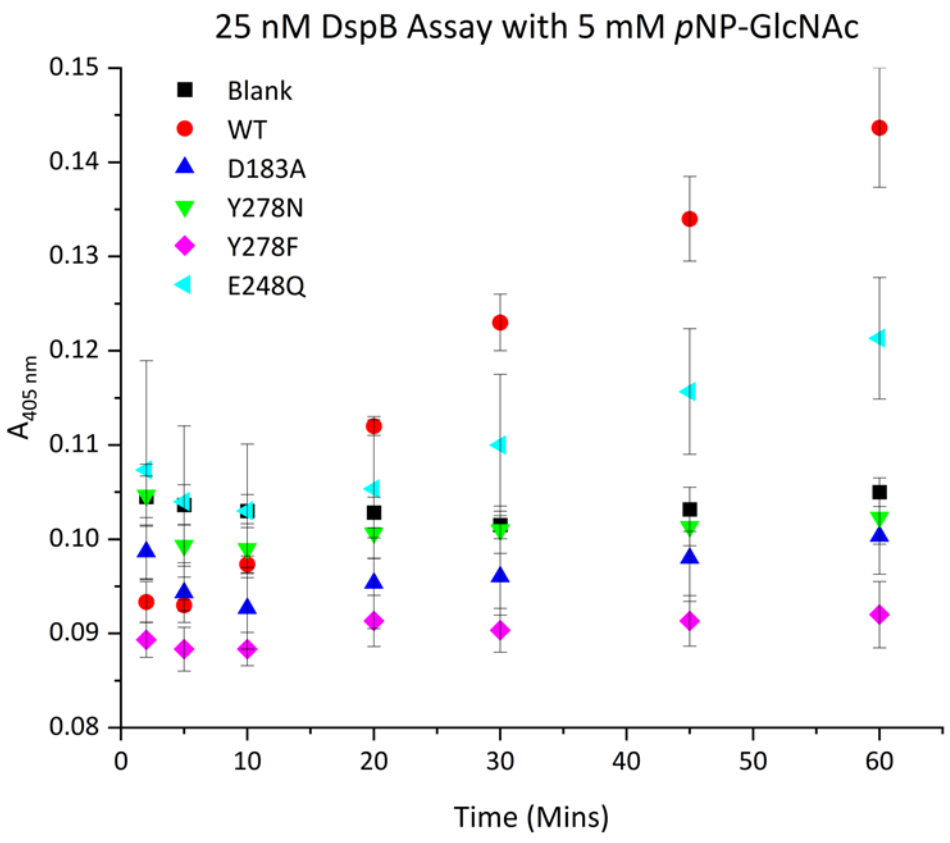

C
Retention Time 2.4 - 3.1 mins
1.2×10^9
9.0×10^8
6.0×10^8
3.0×10^8
0.0
Intensity
602.3
0
500
1000
1500
2000
m/z

Figure 3.2.16: A) Reverse-phase LC-MS chromatogram for the ITAGed TLC assay for D183A including two EIC chromatograms for masses 465 (the mass of the GlcNAc **1** [2M+Na$^+$] adduct) and 602 (monosaccharide **60**). B) Reverse-phase LC-MS chromatogram for the ITAGed TLC assay for D183A including two EICs for masses 805 (disaccharide **61**) and 1008 (trisaccharide **62**). C) Mass spectra from 2.4-3.1 mins of D183A's EIC revealing little to no disaccharide **58** present

The ITAGed samples reveal a large amount of unreacted acceptor **60** (m/z: 602.39, Figure 3.2.16A,C), but show no signs of the potential disaccharide at m/z: 805 (Figure 3.2.16B). This led to the conclusion that the mass seen throughout the previous assay was in fact not disaccharide **58**, but an artifact of in the LC-MS, perhaps coming from the oxazoline **5** reaction mixture. Further investigation into the potential reasons these assays failed to produce any glycosynthase activity revealed that the small amount of TEA used during the synthesis of oxazoline **5** (Section 3.1.3) had a much greater impact on the pH than previously thought. The TEA was able to overpower the buffer solution, leading to a solution far too basic for DspB to active, and so any further attempts to use oxazoline **5** would require a purified compound, so these assays were temporarily halted and attention was turned towards assays utilising the other glycosynthase substrates that at this point had been successfully purified. At this point, the WT DspB gene was also purchased in an expression plasmid to be utilised as a positive control in all future assays, alongside E248Q which was to be tested to confirm Breslawec *et al*'s results that show that this mutant is a much more efficient glycoside hydrolase[85] with the hope that once a suitable glycosynthase mutant was identified, a double mutant utilising E248Q could be made to hopefully improve the efficiency of that glycosynthase too.

The first substrate to be tested was *p*NP-GlcNAc **4**. 5 mM of *p*NP-GlcNAc **4** was incubated with two different concentrations of DspB (25 nM and 1 µM) at 37 °C, in pH 5.9 100 mM PBS, for 1 hour and the reaction was tracked via UV/Vis measurements of the absorbance at 405 nm, the $\lambda_{max}$ of the hydrolysis product of *p*NP-GlcNAc **4**, *p*-nitrophenol **6** (Figure 3.2.17).

**A**


25 nM DspB Assay with 5 mM pNP-GlcNAc
Blank
WT
D183A
Y278N
Y278F
E248Q
A405 nm
0.15
0.14
0.13
0.12
0.11
0.10
0.09
0.08
0
10
20
30
40
50
60
Time (Mins)

**B**

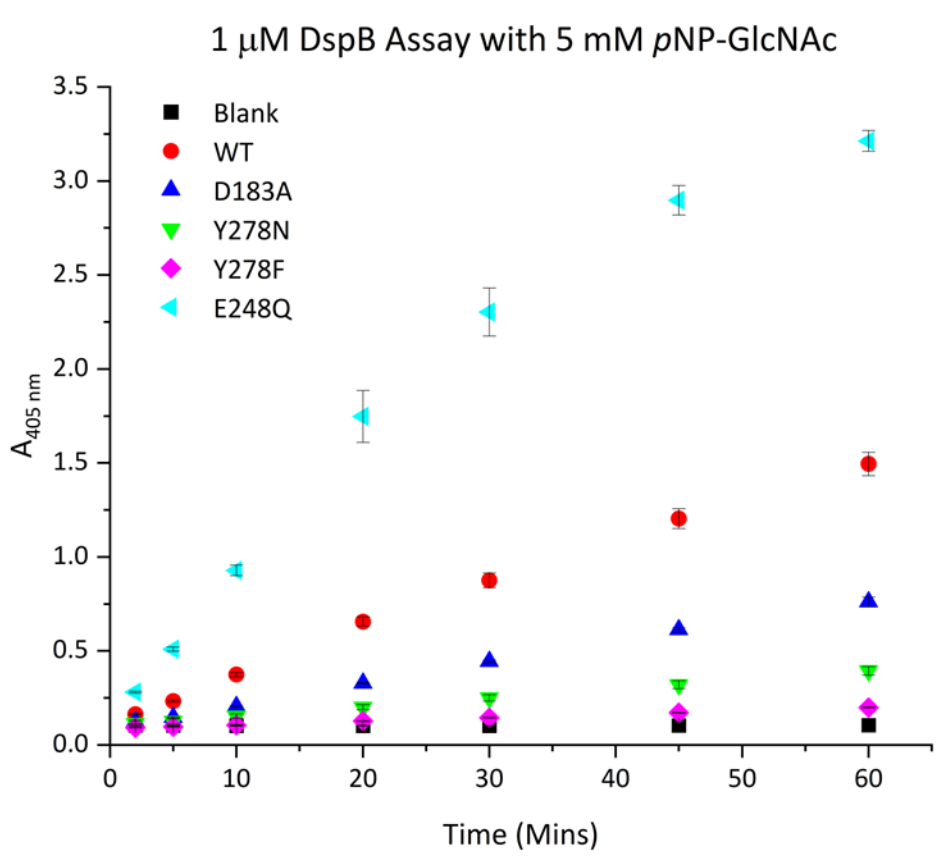

1 µM DspB Assay with 5 mM pNP-GlcNAc
Blank
WT
D183A
Y278N
Y278F
E248Q
A405 nm
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
60
Time (Mins)

Figure 3.2.17: A) UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 25 nM DspB at 37 °C, pH 5.9, for 1 hour. B) UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM DspB at 37 °C, pH 5.9, for 1 hour.

These initial results confirm the observations of Breslawec *et al* with E248Q acting as a much more efficient glycoside hydrolyase than WT DspB, although interestingly they seem to suggest that this is concentration dependent as the WT is more efficient at the lower concentrations. They also show that two of the three glycosynthase DspB mutants do in fact catalyse the hydrolysis of *p*NP-GlcNAc **4**, as an increase in absorbance can be seen for both D183A and Y278N. D183A was the most effective of the two, but neither mutant shows comparable hydrolysis activity to WT DspB. Y278F did not appear to show any significant hydrolysis activity above the background.

Hydrolysis activity alone though is not enough to conclude that the mutants are acting as glycosynthases, so further tests were required to see if this hydrolysis was the result of glycosidic bond formation, or if the two mutants were just acting as less effective glycoside hydrolases. This was initially performed via LC-MS (Figure 3.2.18).

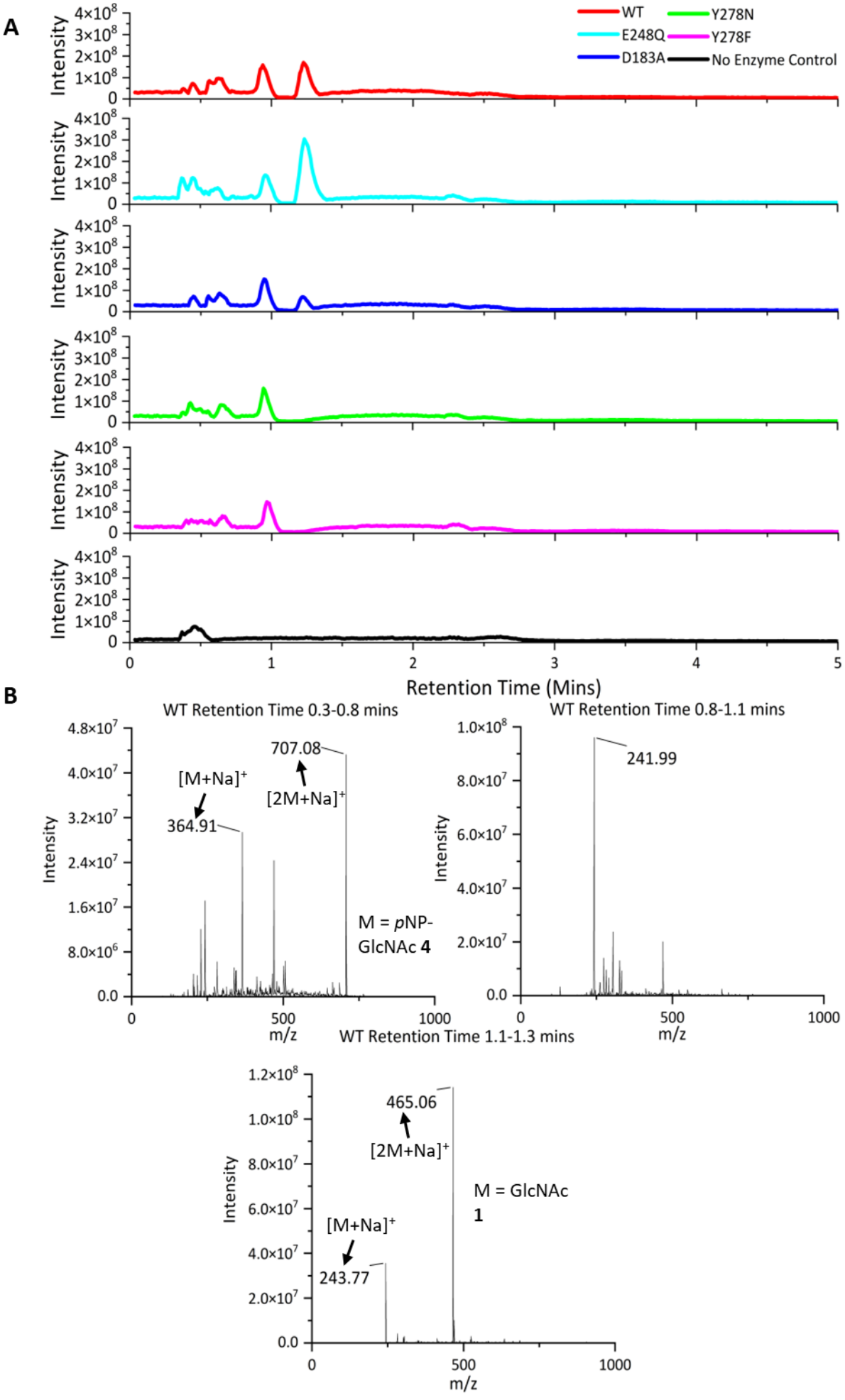

A
WT
Y278N
E248Q
Y278F
D183A
No Enzyme Control
Intensity
Retention Time (Mins)
B
WT Retention Time 0.3-0.8 mins
707.08
[2M+Na]+
[M+Na]+
364.91
M = pNP-GlcNAc 4
WT Retention Time 0.8-1.1 mins
241.99
WT Retention Time 1.1-1.3 mins
465.06
[2M+Na]+
M = GlcNAc 1
[M+Na]+
243.77
m/z

Figure 3.2.18: A) LC-MS Chromatograms of the 1 µM DspB *p*NP-GlcNAc **4** assay. B) MS spectra of the major peaks that appear in the WT chromatogram

The LC-MS analysis appeared to show that no glycosynthase activity had occurred. The three mutants had no new peak that was not in the WT chromatogram or the no enzyme control (Figure 3.2.18A), and none of the peaks that appear, but were not present in the no enzyme control, could be assigned to an oligosaccharide of any length (Figure 3.2.18B). In fact, these results seem to prove the hypothesis that the three mutants are just less effective glycoside hydrolases, as the two new peaks that appear in the WT chromatogram correspond to GlcNAc **1**, and these peaks are either smaller, or not present for the mutants, and so that suggests that the mutants are less effective than the WT. These results were further confirmed via TLC analysis ( Figure 3.2.19)

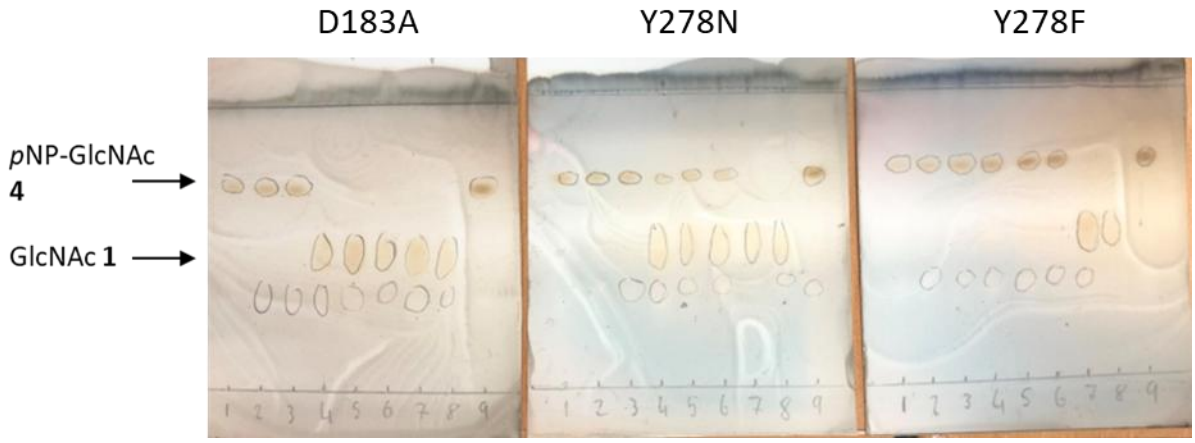



Figure 3.2.19: TLCs of the 1 µM DspB *p*NP-GlcNAc **4** assay. (Left to right) D183A, Y278N, Y278F. For all three TLC plates: Lanes 1-3: Reaction mixture after 1 hour. Lanes 4-6: Reaction mixture after being left overnight. Lane 7: WT positive control after 1 hour. Lane 8: GlcNAc **1**. Lane 9: *p*NP-GlcNAc **4**.

Again in the TLCs, no new spots appear at a lower $R_f$ than *p*NP-GlcNAc **4**, other than GlcNAc **1**, and for all three enzymes, *p*NP-GlcNAc **4** remained in the reaction mixture after 1 hour in comparison to the WT lane where no spot is observed. This would appear to suggest that the mutants are just less effective glycoside hydrolases, not glycosynthases, but in an effort to conclusively prove this, further tests were carried out. Firstly, the pH of the *p*NP-GlcNAc **4** assay was changed to pH 7.4 (100 mM pH 7.4 PBS) in an effort to see if this would promote glycosynthase activity from the three mutants (Figure 3.2.20). From this assay onwards, continuous monitoring was used for all assays due to new software for the plate reader providing a better ability to create assay methods with continuous monitoring.

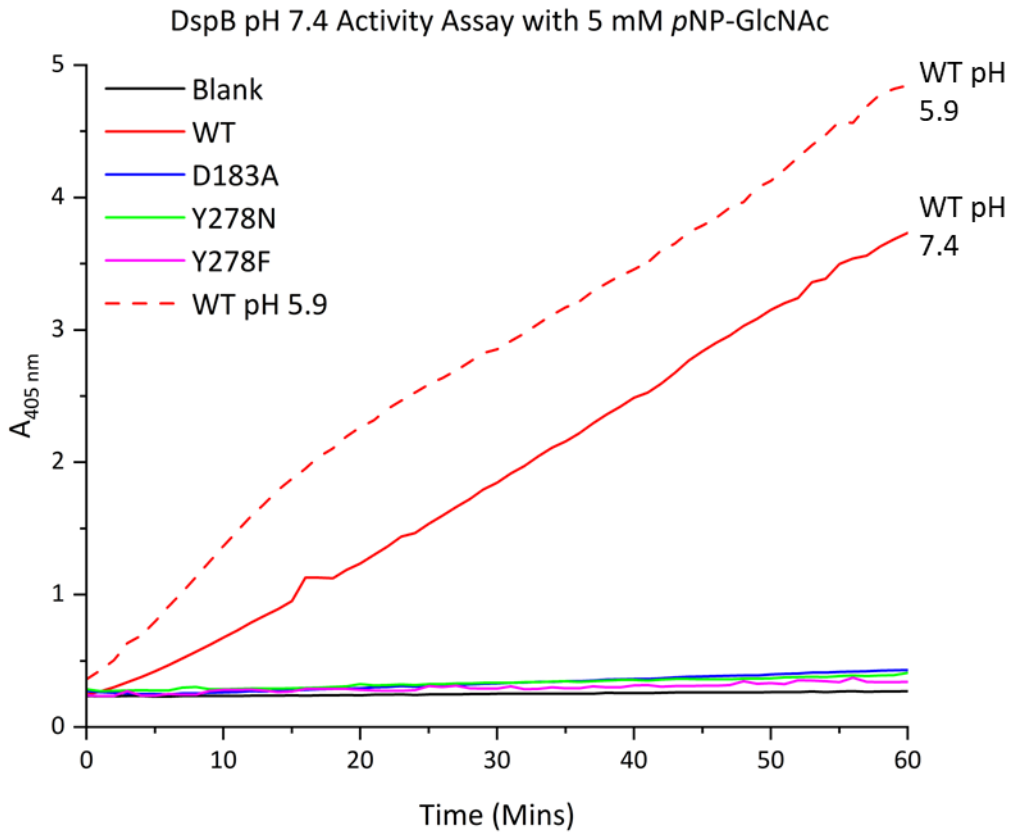



Figure 3.2.20: UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM DspB at 37 °C, in 100 mM PBS pH 7.4, for 1 hour. UV/Vis absorbance measurements for WT DspB in 100 mM PBS pH 5.9 included for comparison (dashed red line).

However, this appeared to have just completely halted any hydrolysis activity of the three mutants. This would suggest that the mutants were not performing any glycosynthase reactions as this would also generate *p*-nitrophenol during the formation of the glycosidic bond. Another potential method that was tested for promoting glycosynthase activity was the introduction of azido-propyl acceptor **8**, as performed with oxazoline **5**. The rationale being that acceptor **8** was unlikely to be hydrolysed by the mutants, but it could potentially act as a nucleation point for the beginning of an oligosaccharide chain. In order to test this, the initial conditions for the *p*NP-GlcNAc **4** were repeated, however this time, the enzymes were also incubated with 1 mM of acceptor **8** (Figure 3.2.21).

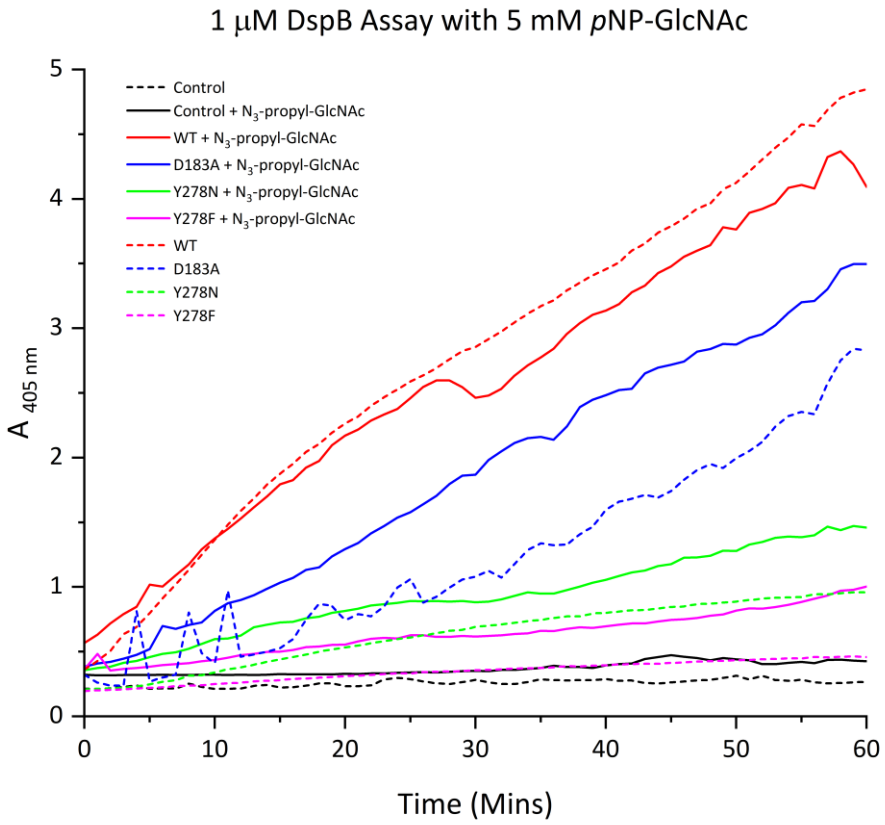



Figure 3.2.21: (Solid lines) UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM DspB and 1 mM acceptor **8** at 37 °C, pH 5.9, for 1 hour. (Dashed lines) UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM DspB at 37 °C, pH 5.9, for 1 hour.

Interestingly, when incubated with 5 mM *p*NP-GlcNAc **4** and 1 mM acceptor **8** the three enzyme mutants all saw an increase in absorbance at 405 nm indicating an increase in hydrolysis activity caused by the addition of acceptor **8**, and notably, this is not observed for WT DspB. To see if this pattern would continue the reaction mixtures were allowed to incubate overnight and the absorbance was measured the next day (Figure 3.2.22)

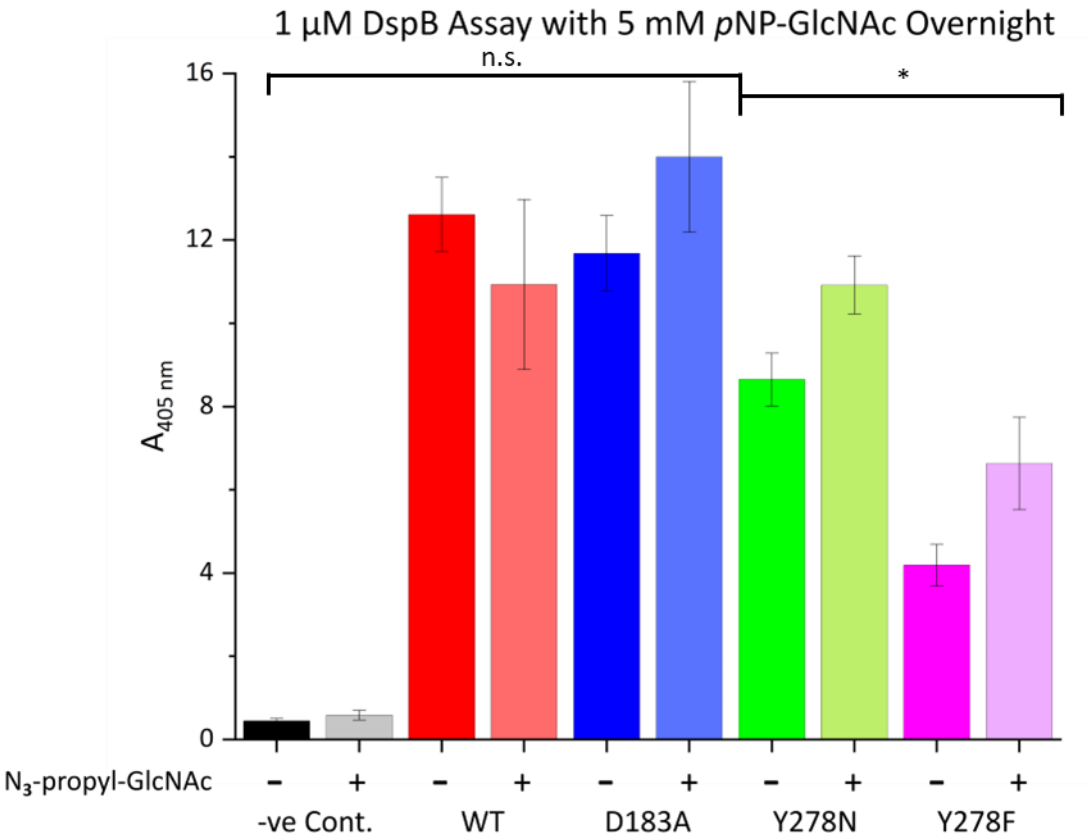



Figure 3.2.22: UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM DspB and 1 mM acceptor **8** at 37 °C, pH 5.9, overnight.n=3.

When the enzyme reaction mixtures were left overnight, it was observed that for two of the three mutants there was a noticeable increase in the hydrolysis activity when acceptor **8** was present, while again, this result was not observed for the WT enzyme. In fact, the WT enzyme appeared to be being inhibited slightly by the presence of acceptor **8**. This could suggest that acceptor **8** is acting as a nucleation point for oligosaccharide production by the glycosynthase mutants as was initially hypothesised, but in order to test this, another TLC analysis was performed (Figure 3.2.23).

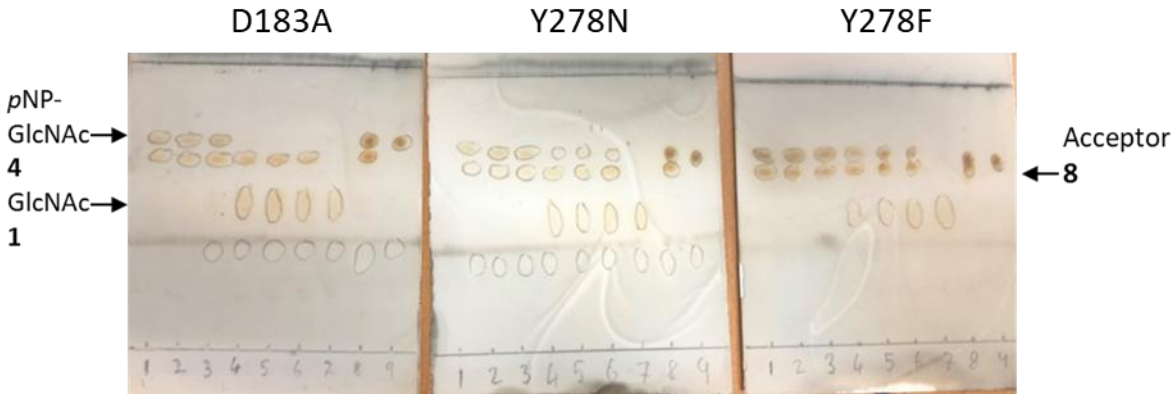



Figure 3.2.23: TLCs of the 1 µM DspB *p*NP-GlcNAc **4** + acceptor **8** assay. (Left to right) D183A, Y278N, Y278F. For all three TLC plates: Lanes 1-3: Reaction mixture after 1 hour. Lanes 4-6: Reaction mixture after being left overnight. Lane 7: WT positive control after 1 hour. Lane 8: *p*NP-GlcNAc **4** + acceptor **8**. Lane 9: *p*NP-GlcNAc **4**.

Unfortunately, the TLC analysis appeared to disprove the hypothesis as, as was the case with the previous TLCs, no new spots appeared in the reaction mixture lanes that would suggest glycosynthase activity. With the evidence gathering that suggested that the attempt to create a glycosynthase had failed, the third glycosynthase substrate was tested to see if this would change the end result. Initially, the substrate was tested with just WT DspB to confirm the results of Chibba *et al*[171] where they had shown carbamate **7** to be a much more effective substrate for DspB than *p*NP-GlcNAc **4**. WT DspB was incubated with 1 mM of carbamate **7** at 37 °C, in 100 mM pH 5.9 PBS, for 1 hour, and absorbance measurements were taken at $A_{410\ nm}$, the characteristic absorbance of the carbamate hydrolysis product, *p*-nitroaniline (Figure 3.2.24).

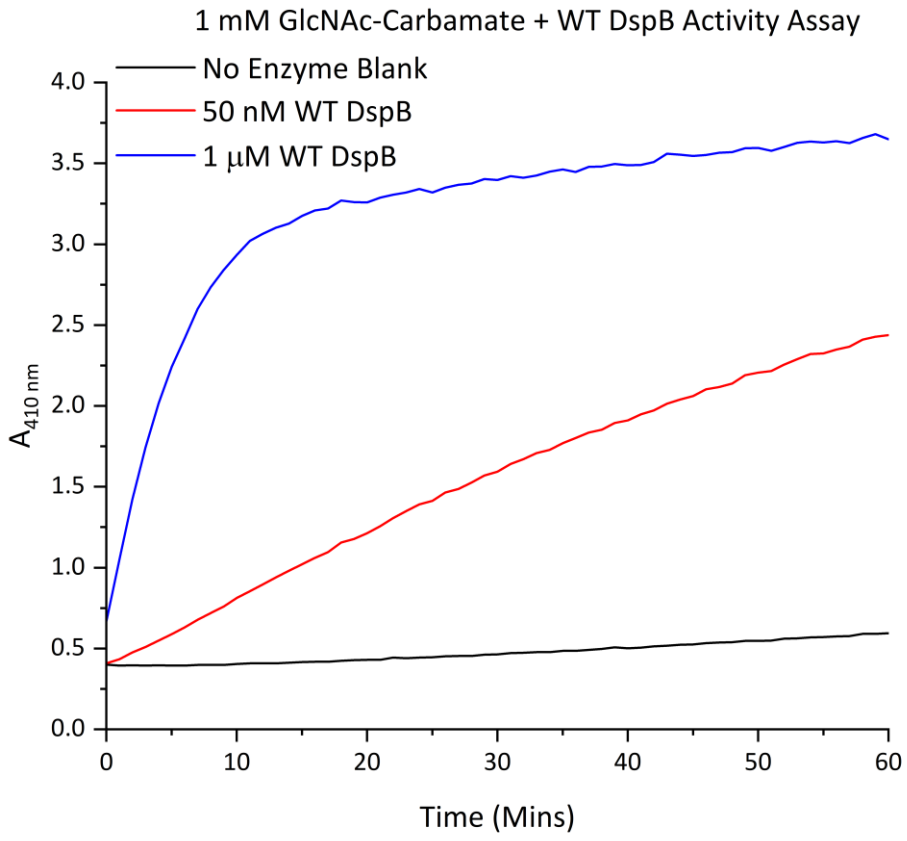



Figure 3.2.24: UV/Vis absorbance measurements at 410 nm of 1 mM of carbamate **7** incubated with 50 nM and 1 µM WT DspB at 37 °C, pH 5.9, for 1 hour.

Carbamate **7** was found to be difficult to solubilise, hence the lower reaction concentration than *p*NP-GlcNAc **4**, however, initial experiments seemed to confirm Chibba *et al*'s results that carbamate **7** is indeed a better substrate for DspB. In these initial experiments, the higher enzyme concentration managed to reach substrate depletion, which it had not been able to do with *p*NP-GlcNAc **4**, and the lower enzyme concentration had a noticeable higher signal than it did with *p*NP-GlcNAc **4**. These results however could be due to the lower substrate concentration, so to fully

confirm the results, a kinetic analysis had to be performed. This, however, ran into a problem, the aforementioned solubility issue. Normally, to determine Michaelis-Menton enzyme kinetics with a substrate, the substrate concentration is varied to above and below the $K_M$.

Wang *et al.*[88] designed a similar compound to carbamate **7**, opting to utilise a disaccharide instead of a monosaccharide. They also found issues with solubility, however, they were able to use substrate depletion kinetics to circumnavigate this issue. Substrate kinetic depletion kinetics describe the relation between $V_0$ and $\frac{k_{cat}}{K_M}$ when $K_M \gg [S]$ (Equation 3.2.1).

$$V_0 = \frac{V_{max}[S]}{K_M + [S]}$$

$$V_{max} = k_{cat}[E]_T$$

$$\therefore$$

$$V_0 = \frac{k_{cat}[E]_T[S]}{K_M + [S]}$$

$$If: K_M \gg [S], then: K_M + [S] \approx K_M$$

$$\therefore$$

$$V_0 = \frac{k_{cat}[E]_T[S]}{K_M} = \left(\frac{k_{cat}}{K_m}\right)[E]_T[S]$$

Equation 3.2.1: Derivation of the 1st order rate equation of substrate depletion kinetics

Equation 3.2.1 is a 1st order rate equation. Therefore, by measuring the initial rate of reaction at a known substrate concentration, where that concentration is much smaller than $K_M$, at different enzyme concentrations, you can obtain a 1st order plot for the reaction. In such a plot, the following is true:

$$V_0 = \left(\frac{k_{cat}}{K_M}\right)[E]_T[S] = m[E]_T$$

$$m = \left(\frac{k_{cat}}{K_M}\right)[S]$$

$$Where\ m = gradient$$

$$\therefore$$

$$\frac{k_{cat}}{K_M} = \frac{m}{[S]}$$

Equation 3.2.2: Equation to determine $\frac{k_{cat}}{K_M}$ from a plot of $V_0$ vs $[E]_T$

Therefore, a value for $\frac{k_{cat}}{K_M}$ can be determined without having to calculate both individually by measuring the initial reaction velocity of several enzyme concentrations. To do so, the WT DspB was assayed as before, but the absorbance was converted to concentration of *p*-nitroaniline (*p*NA) via the Beer-Lambert Law, using ε = 5450 $M^{-1}$ $cm^{-1}$ at 410 nm[171] (Figure 3.2.25).

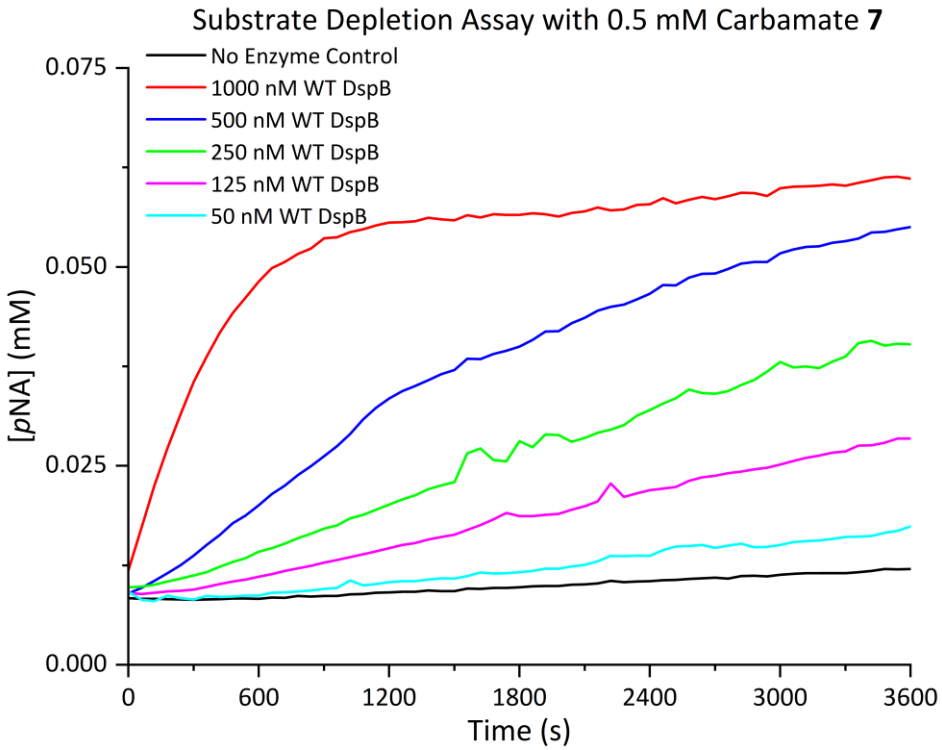



Figure 3.2.25: Substrate Depletion Assay for WT DspB using 0.5 mM carbamate **7**.

Having assayed varying concentrations of WT DspB with 500 µM carbamate **7**, $V_0$ can be determined for each concentration of WT DspB by monitoring the first 5 minutes of the assay and applying a linear fit to each data set (Figure 3.2.26).

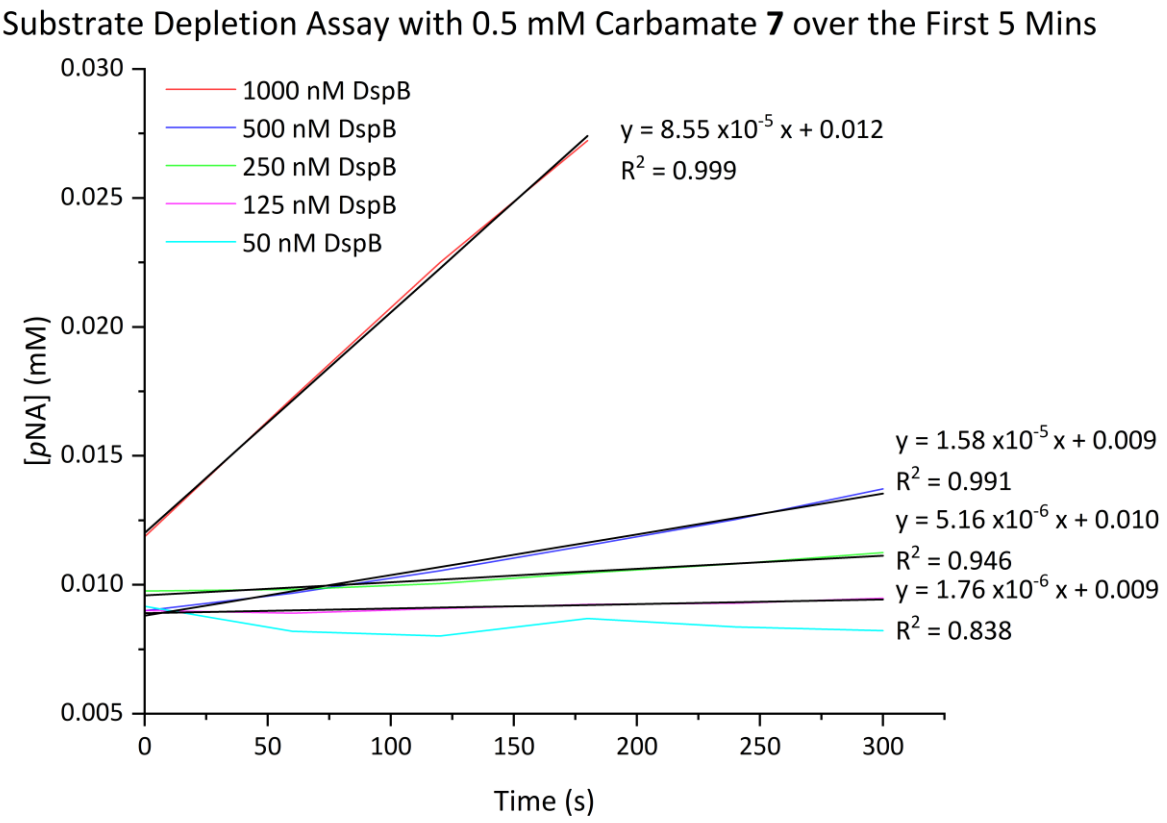



Figure 3.2.26: First 5 mins of the data sets for each enzyme concentration of the WT DspB substrate depletion assay. The Solid black lines are the lines of best fit plotted by Origin. The equations to the right are the equations for each line of best fit.

Having done this, the equation for each line of best fit gives the value for $V_0$ for each enzyme concentration except 50 nM which had too low of an initial rate to be detectable after 5 minutes. Then using Equation 3.2.2, the $\frac{k_{cat}}{K_M}$ can be determined by plotting these values against the corresponding enzyme concentrations (Figure 3.2.27).

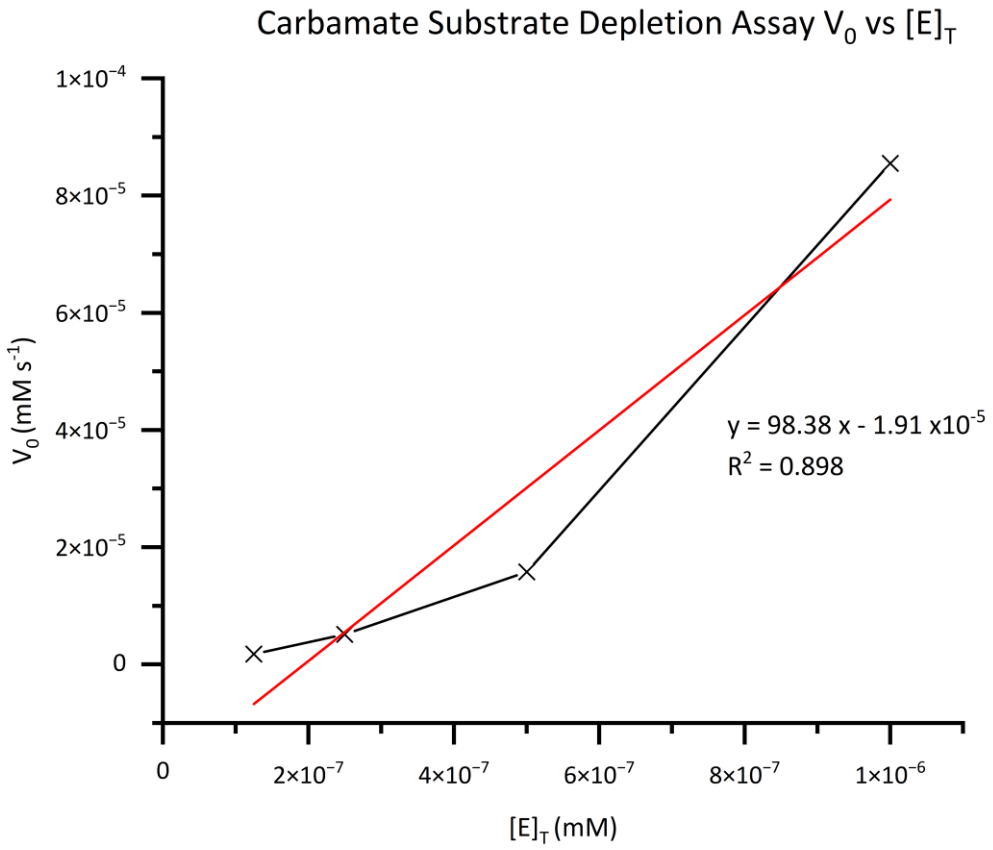



Figure 3.2.27: A plot of the $V_0$ values determined experimentally during the substrate depletion assay for carbamate **7** plotted against enzyme concentration. The red line is the line of best fit plotted by Origin.

Plotting a line of best fit on this graphs gives a value for *m* of 98.38 $s^{-1}$. Dividing this number by the concentration of carbamate **7** in mM (0.5 mM) gives a value of 190.76 $mM^{-1}s^{-1}$. This is two orders of magnitude higher than the literature value given by Chibba *et al* of 6.1 $mM^{-1}s^{-1}$ [185] and several orders of magnitude higher than for the $\frac{k_{cat}}{K_M}$ of *p*NP-GlcNAc **4**, which Manuel *et al*[119] found to be 0.021 $mM^{-1}s^{-1}$. This assay confirmed the findings of Chibba *et al*, and demonstrates how efficient of a substrate of DspB carbamate **7** is. Having experimentally determined that carbamate **7** is a better substrate for DspB, it was then assayed against the three potential glycosynthase mutants using the same conditions as the WT assay (Figure 3.2.28), however the concentration of carbamate **7** was increased to 2 mM as some success had been had in getting more of it into solution following the SEC described in section 3.1.2. (Figure 3.1.4).

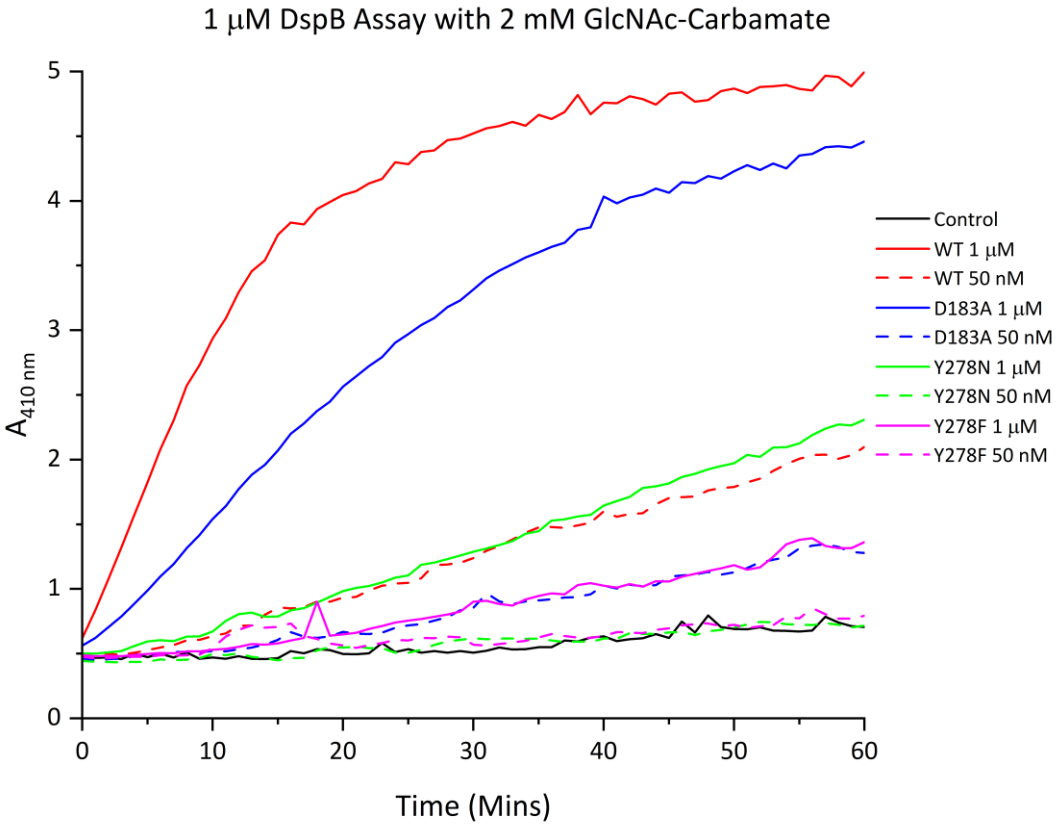



Figure 3.2.28: UV/Vis absorbance measurements at 410 nm of 2 mM of carbamate **7** incubated with 50 nM and 1 µM of the DspB mutants and WT DspB at 37 °C, pH 5.9, for 1 hour.

This produced some promising results, as D183A and Y278N showed significant increase in activity with carbamate **7** over *p*NP-GlcNAc **4**, with D183A almost reaching substrate depletion. To see if this increase in activity was due to any glycosynthase activity, another set of TLC assays were ran (Figure 3.2.29).

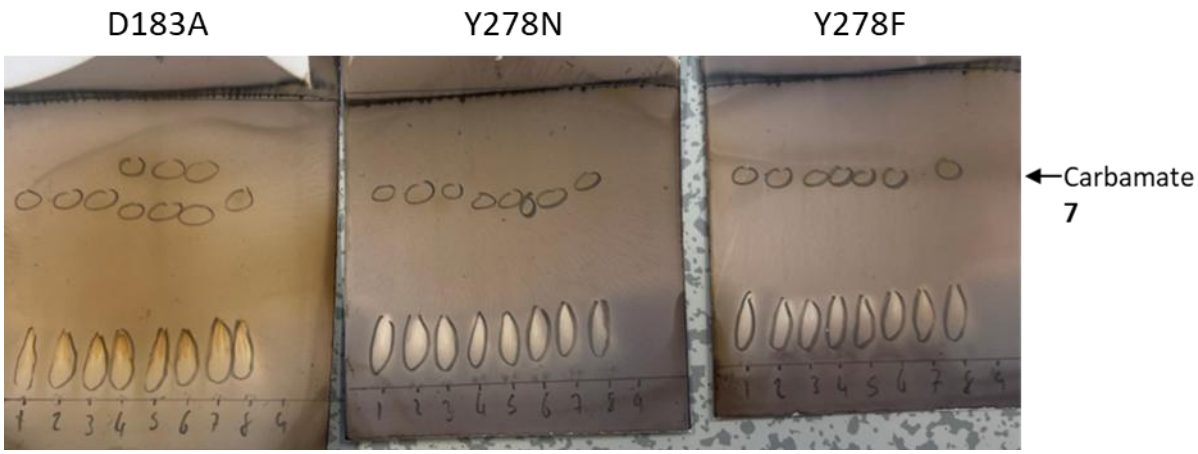



Figure 3.2.29: TLCs of the 1 µM DspB carbamate **7**. (Left to right) D183A, Y278N, Y278F. For all three TLC plates: Lanes 1-3: 50 nM enzyme reaction mixture. Lanes 4-6: 1 µM enzyme reaction mixture. Lane 7: WT positive control after 1 hour. Lane 8: Carbamate **7**. Lane 9: *p*-nitroaniline.

However, as happened with the previous TLC assays, no evidence could be found to support the hypothesis that the mutants were acting as glycosynthases. A new spot did appear in the 1 µM D183A reactions, however, this spot had a higher $R_f$ than carbamate **7** suggesting a less polar substance, and given that each additional monomer that is added to an oligosaccharide dramatically increases its polarity, it is unlikely this spot is an oligosaccharide. At this point, oxazoline **5** had been successfully purified, and so, in a final effort to produce glycosynthase activity from the three mutants, the previous assays utilising oxazoline **5** and acceptor **8** were repeated using a higher concentration of oxazoline **5**. The three mutants and WT DspB were incubated at 37 °C, pH 5.9, overnight with 5 mM of acceptor **8** and 25 mM of oxazoline **5** (Figure 3.2.30).

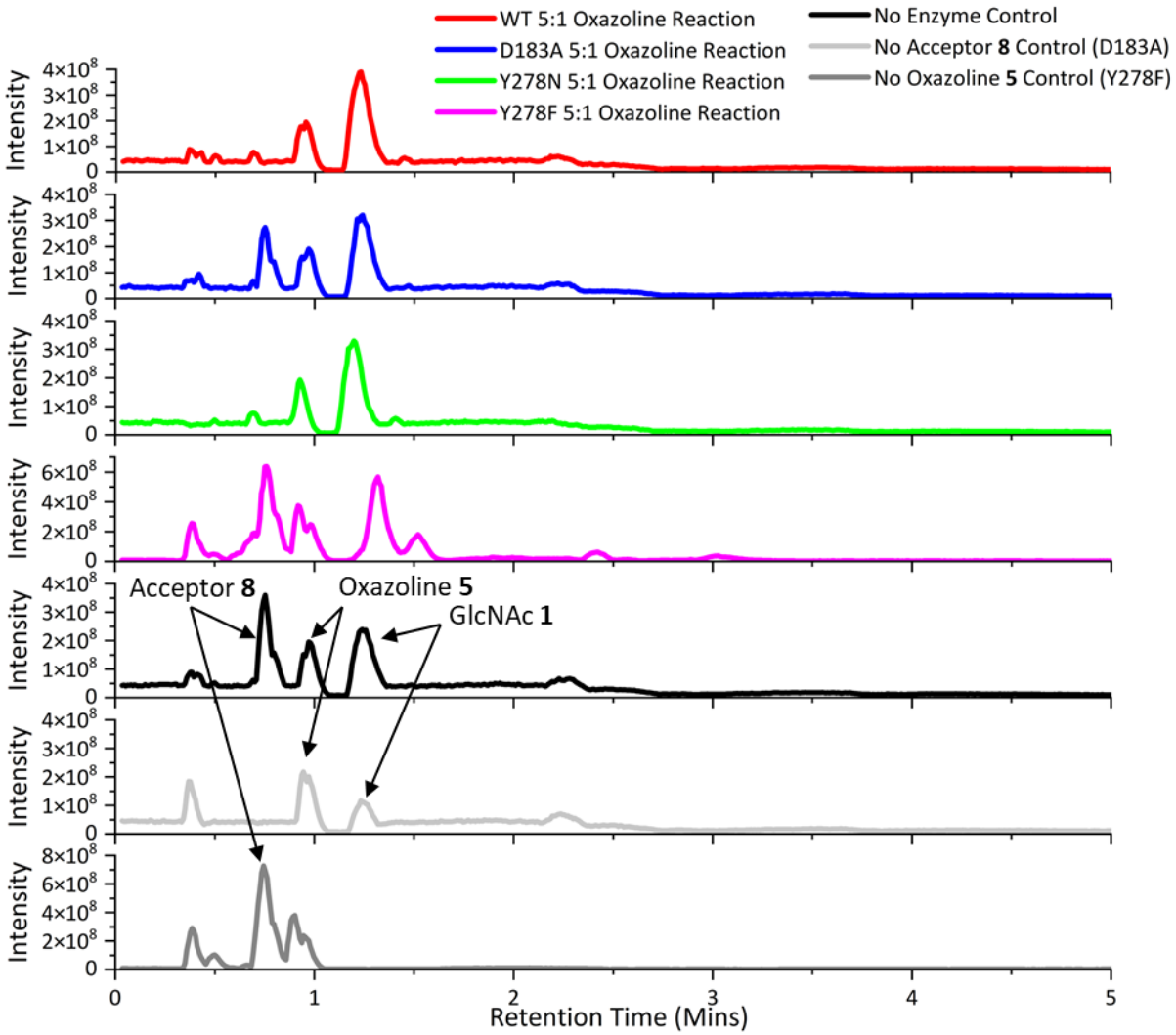



Figure 3.2.30: Full MS trace of all three DspB mutants and WT DspB after being left overnight with 5 mM acceptor **8** and 25 mM oxazoline **5**.

Upon analysis, only one new peak appears in comparison to the WT and no enzyme across the three mutants at 1.4 – 1.6 mins in the Y278F chromatogram. MS analysis of this however reveals this to just be GlcNAc **1** (Figure 3.2.31).

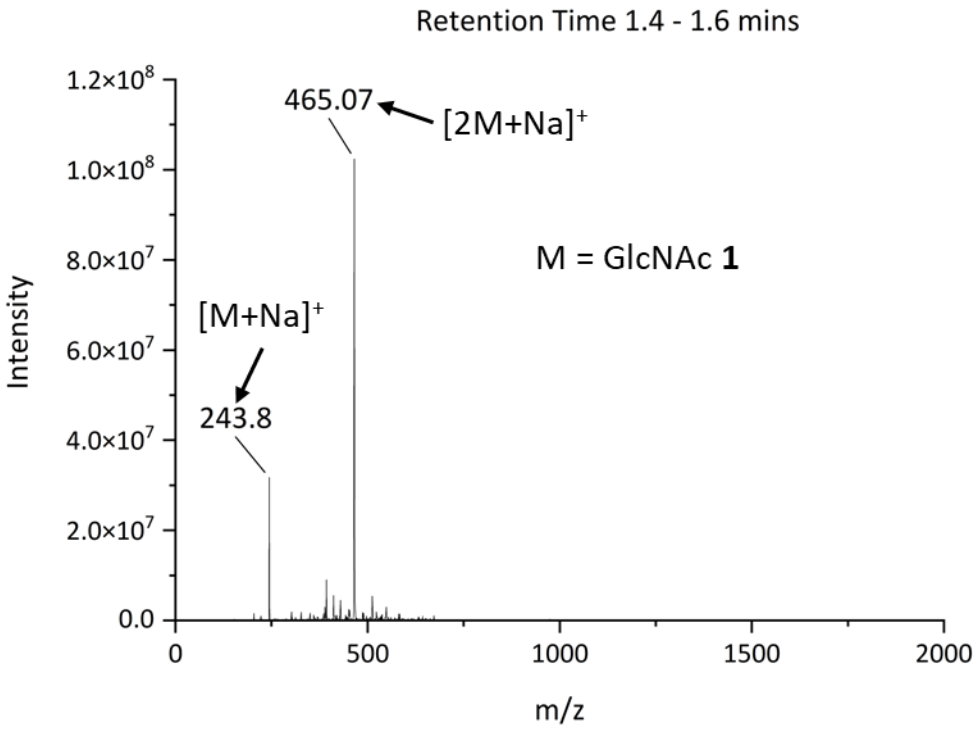



Figure 3.2.31: MS Spectra for retention time 1.4 – 1.6 mins of the Y278F 5:1 oxazoline **5** reaction chromatogram.

In fact, D183A and Y278F both showed no activity at all as their chromatograms are just repeats of the no enzyme control, and Y278N is a repeat of the WT, suggesting no glycosynthase activity either. This was confirmed via extracted ion chromatogram analysis which showed no evidence for any disaccharide formation in any of the enzyme's LC-MS traces (Appendix 34). Furthermore, upon increasing the amount of oxazoline **5** to 20 equivalents, Y278N also lost all activity (Figure 3.2.32).

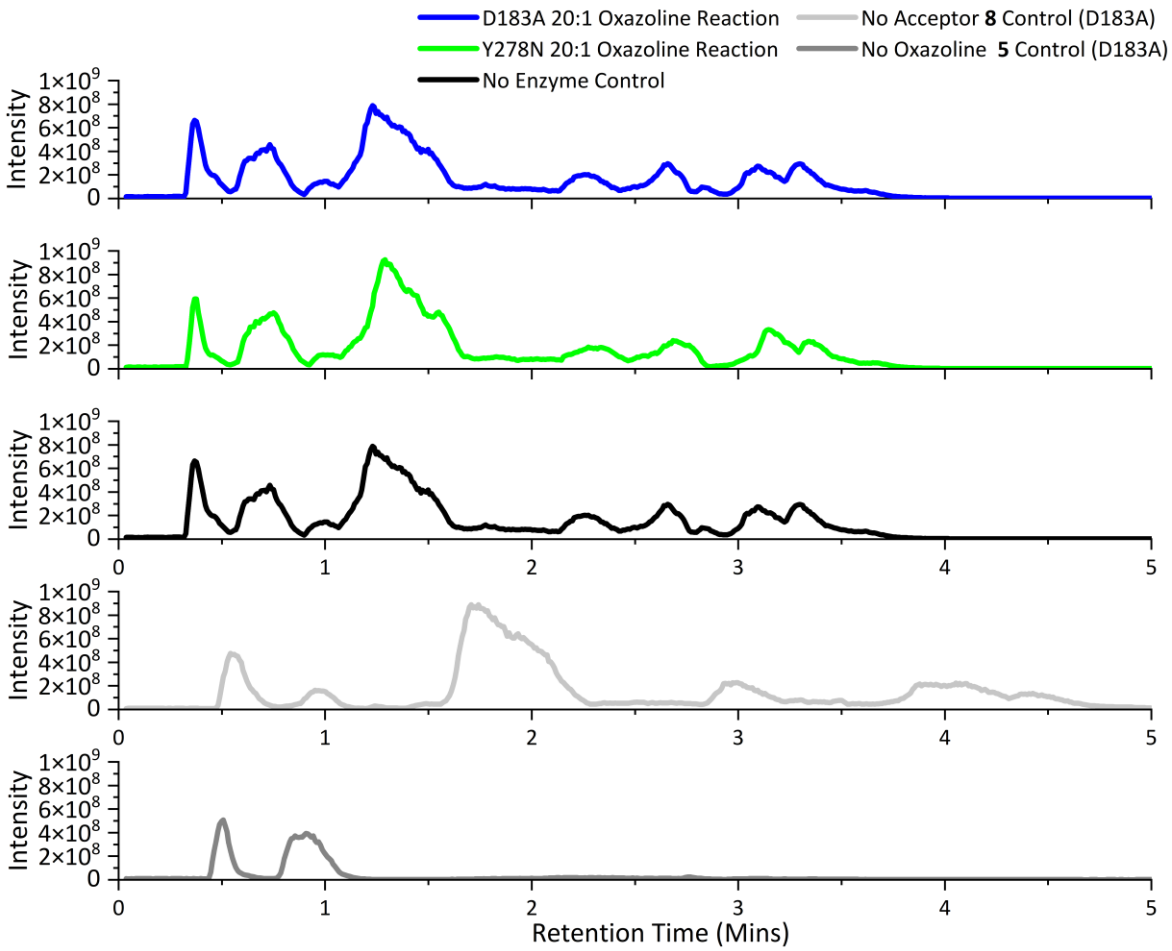



Figure 3.2.32: Full MS trace of all three DspB mutants and WT DspB after being left overnight with 5 mM acceptor **8** and 100 mM oxazoline **5**.

This would suggest that oxazoline **5** is acting as an inhibitor of the DspB mutants. To test this, D183A was incubated with 5 mM *p*NP-GlcNAc **4** alongside 25 mM oxazoline **5** at 37 °C, pH 5.9, for 1 hour to determine if the reduction in activity would be repeated with a different substrate (Figure 3.2.33).

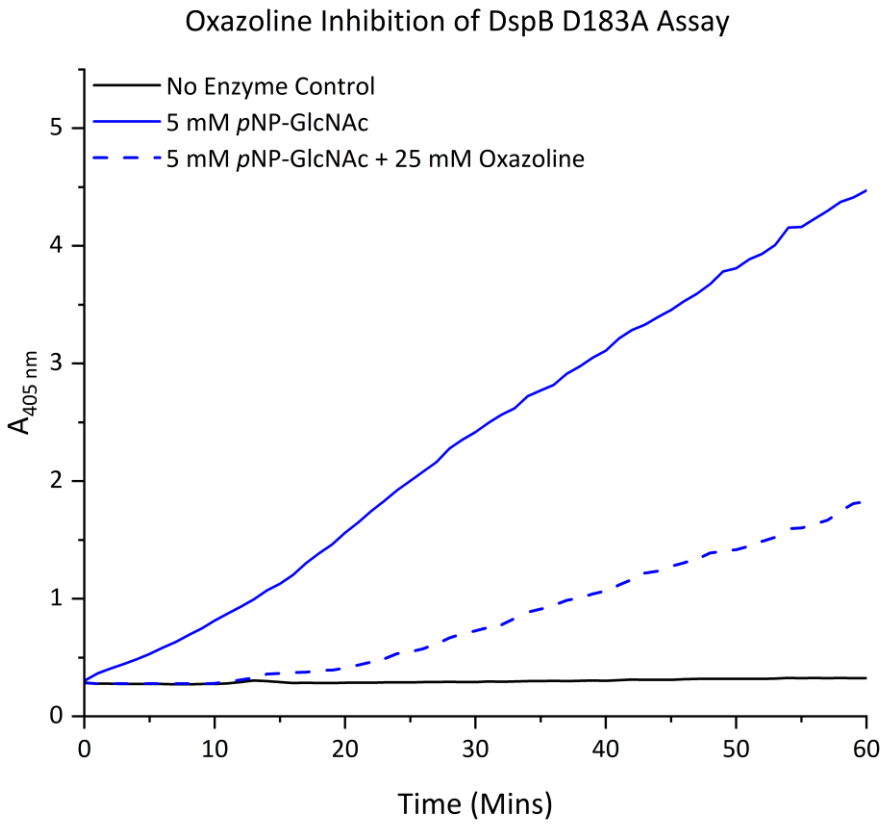



Figure 3.2.33: UV/Vis absorbance measurements at 405 nm of 5 mM of *p*NP-GlcNAc **4** incubated with 1 µM of D183A at 37 °C, pH 5.9, for 1 hour with and without 25 mM oxazoline **5**.

This assay provided preliminary confirmation the fact that oxazoline **5** was acting as an inhibitor of the DspB mutants. It is likely most prominent in D183A due to the fact that this mutant has an altered active site, and so cannot break down the oxazoline ring as easily as the WT or Y278N. This, however, was the final attempt at utilising these mutants as glycosynthases. Given all the evidence gathered over the course of the project, it was concluded that these mutants do not act as glycosynthases, and further study would be required to find a mutant that could.

### 3.2.4. Conclusions and Future Work

Over the course of the project, a large amount of evidence was gathered that suggested that none of the three mutants were capable of utilising the four chosen substrates to act as glycosynthases. In fact, the gathered evidence would suggest that the three mutants are just less efficient, or even non-active glycoside hydrolases. This led to the direction of the project moving away from DspB and glycosynthases in general, however, these results do reveal an interesting avenue for further research.

Consistently throughout the enzyme assays, Y278N was more efficient than Y278F. Manuel *et al*[119] surmised from their results that Y278 acts to orient the substrate into the active site in WT DspB, however, how it does this was left unanswered. Given that tyrosine and asparagine are both capable of forming hydrogen bonds, whereas phenylalanine is not, this perhaps suggests that hydrogen bonding is important to substrate orientation. To investigate this, a fluorinated version of tyrosine could be incorporated into DspB using unnatural amino acid mutagenesis[211]. Fluorine is an electron withdrawing substituent, and so its presence on the phenyl ring would increase the acidity of the -OH group of the tyrosine, potentially increase its hydrogen bonding efficiency. This would cause a noticeable increase in enzyme efficiency that could be measured utilising the methods used during this study.

There are also a few other methods that could be attempted to produce glycosynthase activity from DspB. Schmölzer *et al*[151] were able to generate glycosynthase activity utilising a DE active site mutant, while Wang *et al*[88] showed that a disaccharide version of carbamate **7** is a highly effective substrate of DspB. These two findings could be combined in order to potentially generate a glycosynthase from DspB, however, that was beyond the scope of this project, and so was not attempted.

## 3.3. Putative PNAGase enzymes in *staphloccocci*

### 3.3.1. Introduction

Given the lack of an apparent PNAGase in the *staphylococcus* genus, one of the aims of this project was the to identify, express, purify, and characterise a PNAGase from a *staphylococci* species. This would address a gap in the current literature, and could create additional avenues for the generation of new glycosynthases that could be used to generate PNAG oligosaccharides. Frank *et al* initiated the work towards identifying a new *staphylococci* PNAGase in 2007[131] when they discovered an ORF that describes a protein within *S. lugdunensis* isolates with high sequence identity with DspB. Utilising this sequence as the query in 2022, Higginson[132] was able to perform a BLASTp search in the UniProtKB and NCBI databases, which revealed a further 16 *staphylococci* proteins that could be PNAGases, alongside proteins in a variety of other species that were revealed when utilising DspB's sequence as the query (Figure 3.3.1).

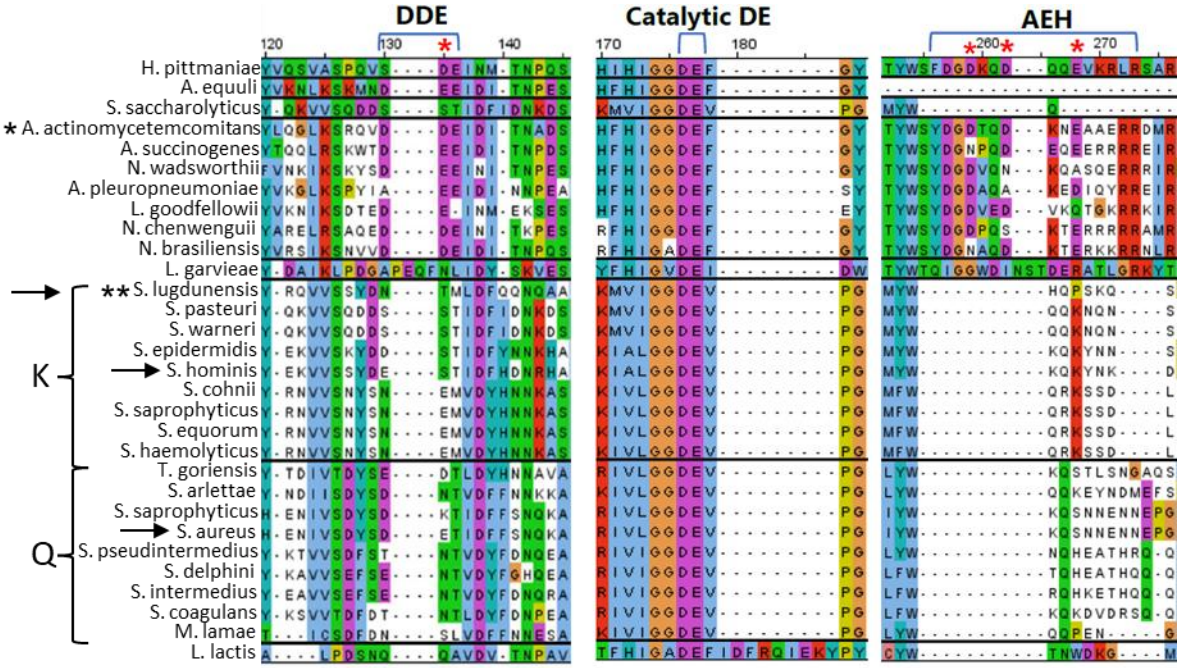



Figure 3.3.1: The results of the two BLASTp searches performed by Higginson. The sequence of DspB (*) revealed the sequences not in the K or Q clade.. The search using the sequence discovered by Frank *et al* (**) revealed the sequences of the putative staphylococci PNAGases. The black arrows denote the proteins selected for expression and characterisation.

Recent work from Males *et al*[118] also corroborates this work as their phylogenetic analysis identified several potential GH20 PNAGases within the same *staphylococci* species as Higginson's work did. This evidence prompted the selection of three of the identified *staphylococci* proteins, SlHex, ShHex, and SaHex, for expression and characterisation to determine whether or not these proteins are PNAGases.

### 3.3.2. AlphaFold Predictions of Putative PNAGase Structure Lack α-Helix of the Anionic Bonding Groove

One of the key features of DspBs structure is α-helix that makes up the anionic bonding groove leading into its active site. Across multiple different papers, the Poulin group have shown that residues in this anionic bonding groove make up binding sites that give DspB a preference for dPNAG[85-87]. The sequence alignments of the putative PNAGases seems to suggest that these new PNAGases lack this α-helix structure, and therefore may not possess the bonding groove, or two of the four binding sites present in DspB. To test this, AlphaFold2 was used to predict the structure of SlHex, and this predicted structure was then aligned with DspB's structure (Figure 3.3.2)

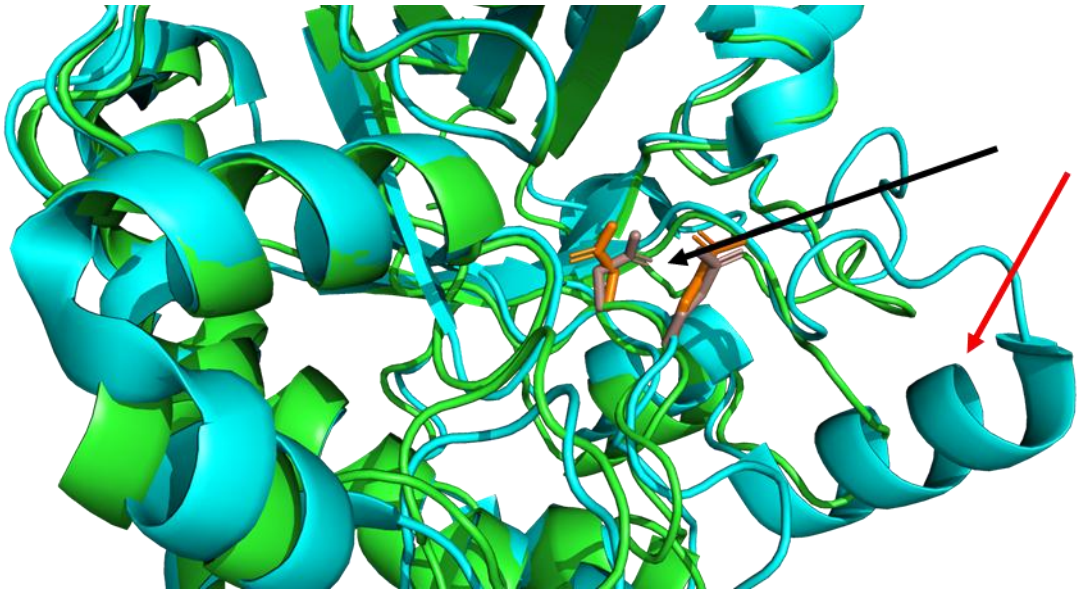


Figure 3.3.2: Structure of DspB (cyan) overlaid with structure of the putative PNAGase from *S. lugdunensis* (SlHex)(green) showing the active site of both proteins (black arrow, DspB residues are in orange, SlHex residues are in grey) in relation to the location of the α-helix that makes up the anionic bonding groove in DspB (red arrow). This structure is absent in SlHex. SlHex structure prediction produced using Colabfold: Alphafold2[73,74], structure of DspB obtained from PDB entry 1HYT[116], image produced using PyMol Molecular Graphics System[125].

The predicted AlphaFold structure lacks the α-helix, as predicted. Furthermore, looking at the electrostatic surface charge of SlHex in comparison with DspB, we can see a complete loss of the anionic bonding groove discovered in DspB by Breslawec *et al*[85] (Figure 3.3.3).

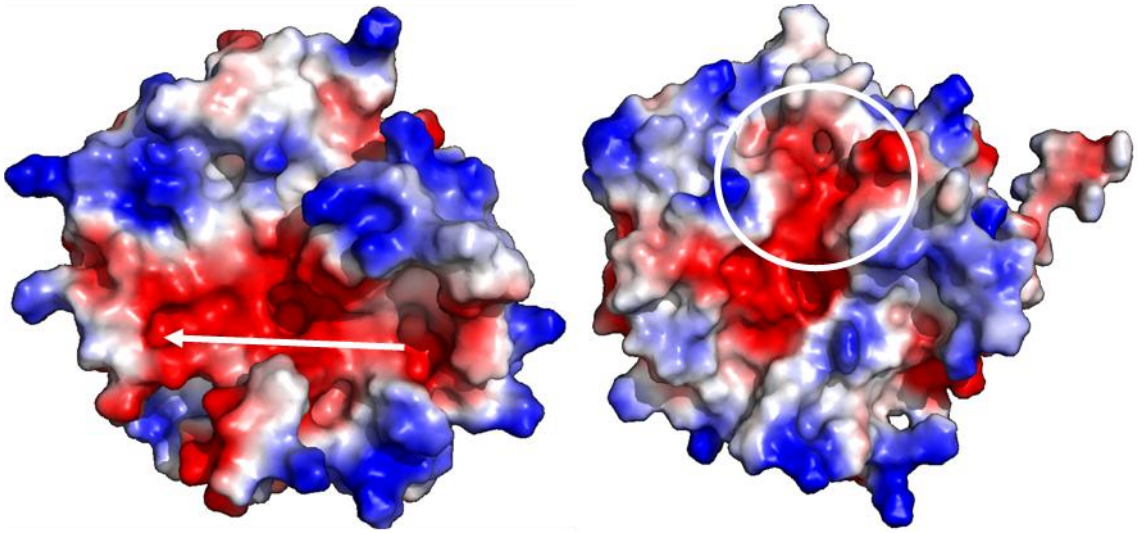


Figure 3.3.3: Electrostatic surface charge of DspB (left) compared to the predicted electrostatic surface charge of SlHex (right) showing the anionic bonding groove found in DspB (white arrow). This structure is absent in SlHex which matches the predictions based off of the sequence alignments of the putative PNAGases, however a new area of negative electrostatic charge does appear in the predicted structure (white circle). SlHex structure prediction produced using Colabfold: Alphafold2[73,74], structure of DspB obtained from PDB entry 1HYT[116] image produced using PyMol Molecular Graphics System[125].

However, there is an area of negative electrostatic charge (Figure 3.3.3, white circle) slightly above the active site that could be performing the same function, but with out further investigation into the residues forming this negative charge, it is unclear as to whether or not this area does contain any binding sites. These findings confirm the hypothesis that the putative PNAGases are structurally distinct to DspB, and provide the beginnings of the evidence required to prove one the three hypotheses outlined in Section 1.2.4.

### 3.3.3. Experimental Evidence of PNAGase Activity

Expression plasmids encoding the genes for SaHex, ShHex, and SlHex were purchased from GenScript. The plasmids were overexpressed in BL21 DE3 chemically competent cells. Overexpression was performed by addition of IPTG to a final concentration of 0.2 mM. Successful overexpression was determined by performing a BugBuster protocol on a small portion of the cell pellet from the cell cultures. The resulting soluble fractions were run in 12% acrylamide SDS-PAGE gel (Figure 3.3.4).

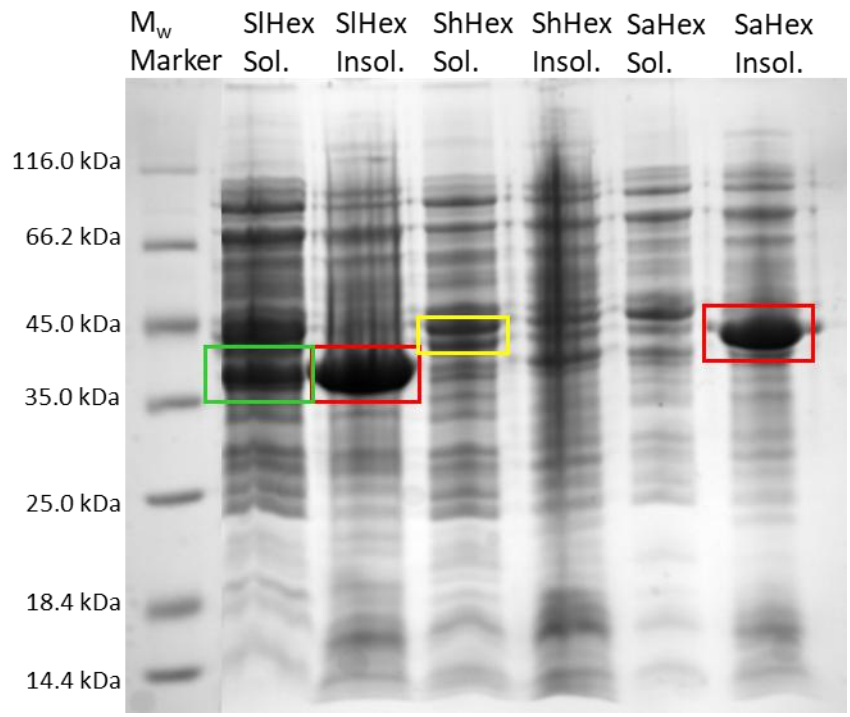



Figure 3.3.4: 12% acrylamide SDS-PAGE gel ran with a 20 µL aliquot of the soluble and insoluble fraction of the BugBuster lysate of each of the three *staphylococci* protein cell pellets. All molecular weights are approximate. SlHex and SaHex appear to show high levels of insoluble expression (red boxes), whilst SlHex does appear to have some soluble expression (green box). ShHex shows little to no expression in both the soluble and insoluble fractions, however a small band around 40 KDa could indicate some soluble expression (yellow box).

SlHex and SaHex appeared to express well but were highly insoluble (red boxes), No attempt was made to purify the soluble fraction of SaHex as there appeared to be little to no soluble protein, and this protein would therefore require expression trials to determine the correct expression conditions. Whilst highly insoluble, SlHex does appear to have a large band in the soluble fraction (green box) as well so the cell pellet was treated with lysis buffer, sonicated, and purified by HisTrap in the same way ShHex was (Figure 3.3.5).

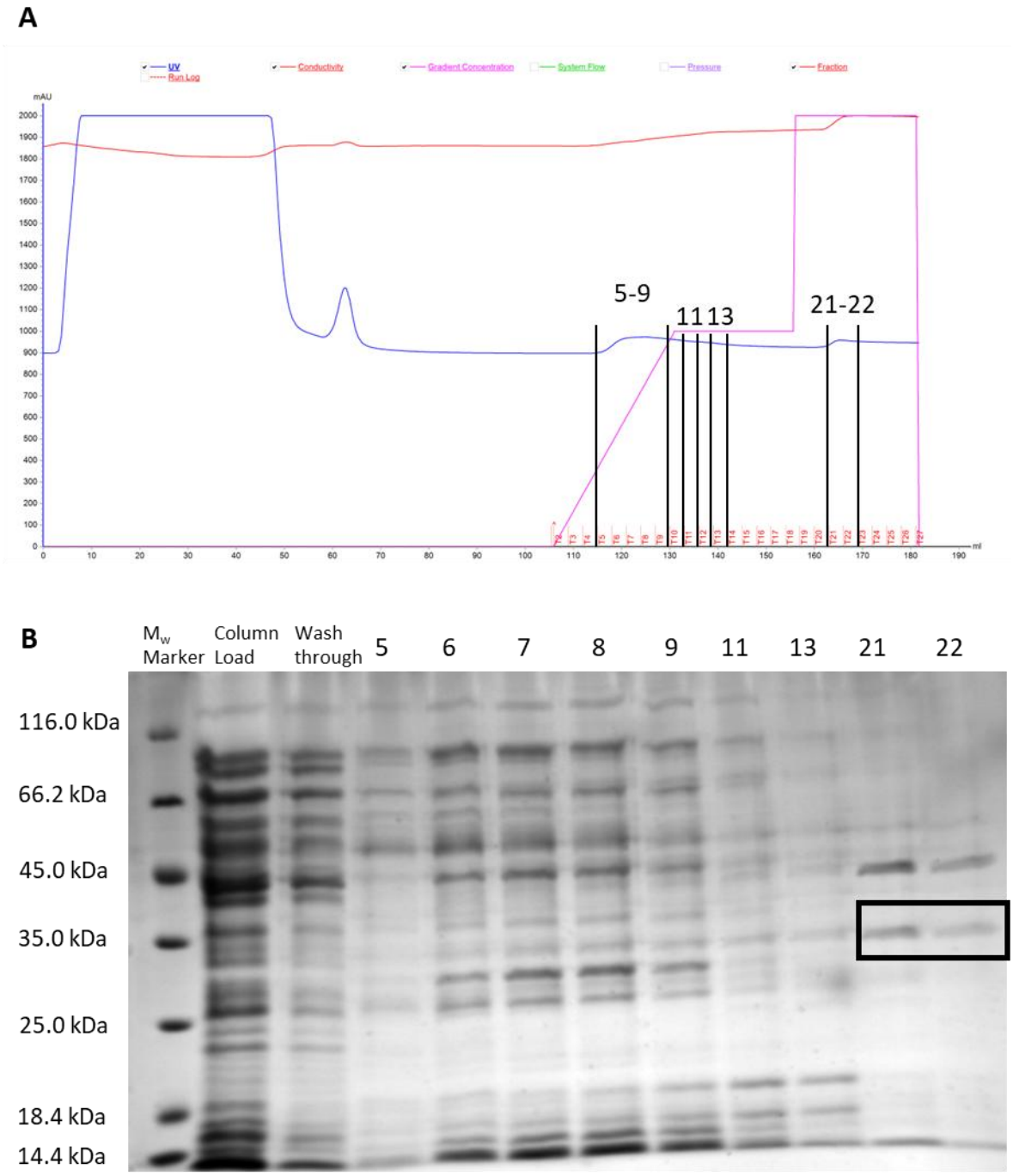



Figure 3.3.5: A) Chromatogram of the purification of SlHex. B) 12% acrylamide SDS-PAGE showing the fractions from the purification of SlHex. All molecular weights shown are approximate.

This proved to be more successful, and after confirmation of protein identity by MALDI-MS/MS on the bands seen at around 35 kDa in fractions 21 and 22 (black box), an activity test was performed using combined fractions and *p*NP-GlcNAc **4** (Figure 3.3.6).

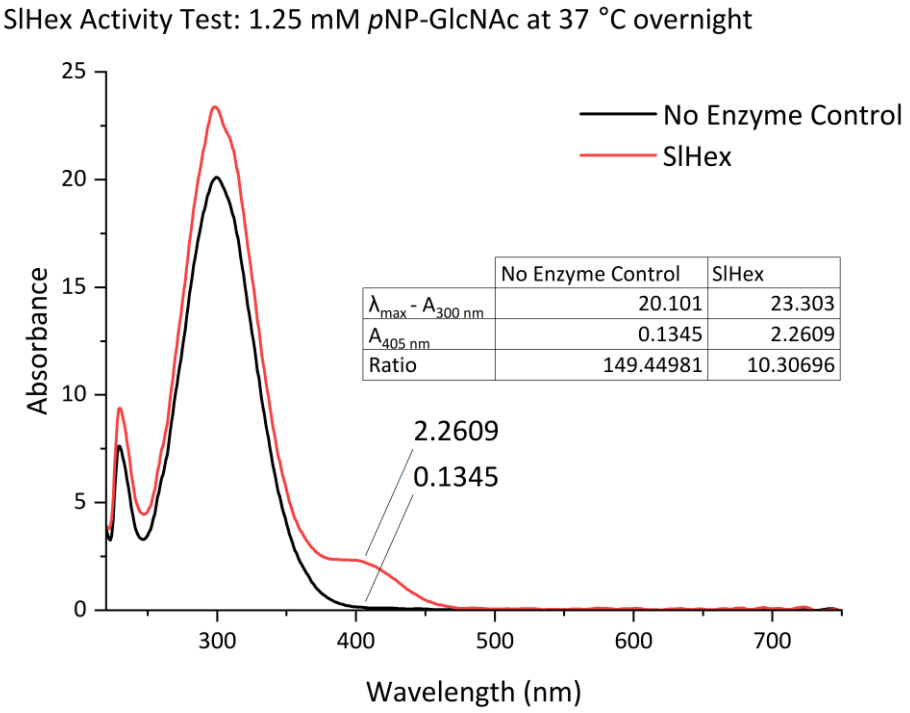



Figure 3.3.6: UV/Vis spectra of the concentrated fractions 21-22 from the SlHex HisTrap purification incubated with 1.25 mM *p*NP-GlcNAc **4** at 37 °C overnight, overlaid with incubated a no enzyme control under the same conditions. The table shows the ratio between the $\lambda_{max}$ of *p*NP-GlcNAc **4**, 300 nm, and the characteristic absorbance at 405 nm of *p*-nitrophenol, the hydrolysis product of *p*NP-GlcNAc **4**.

These initial results would appear to suggest that SlHex is capable of turning over PNAG, however, the sample was not pure enough for kinetic analysis. Therefore, the decision was made to pass the sample down a S200 size exclusion chromatography (SEC) column (Figure 3.3.7).

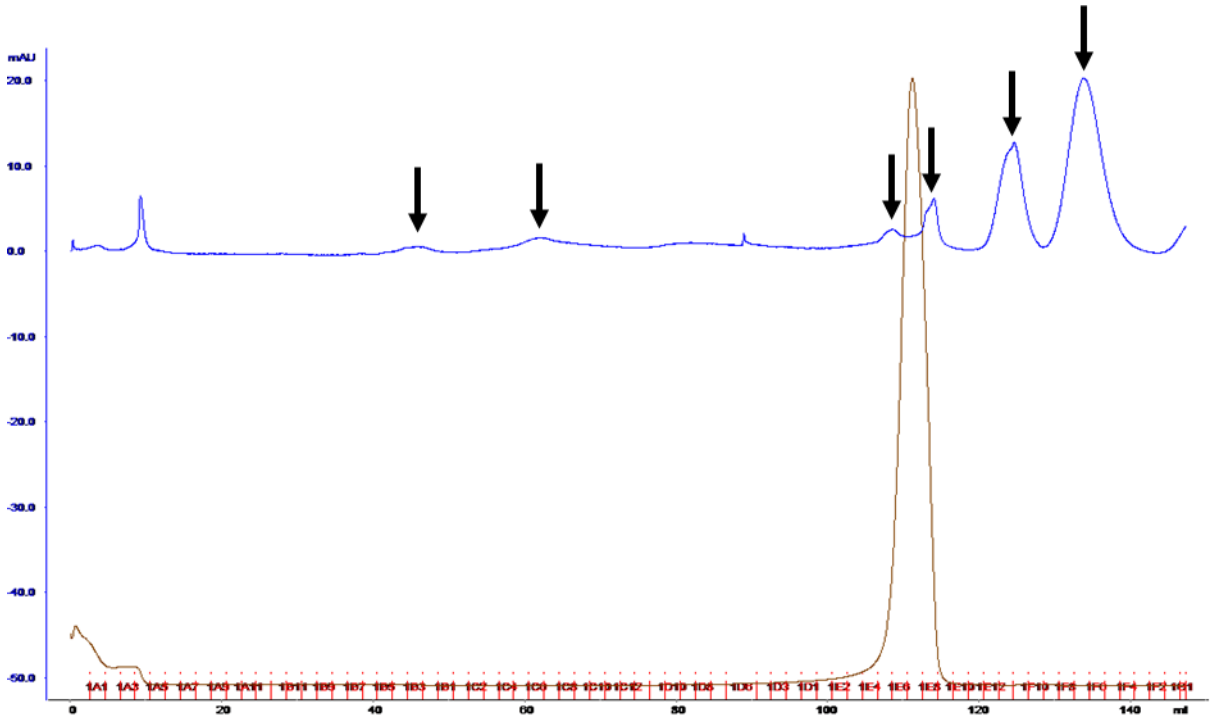
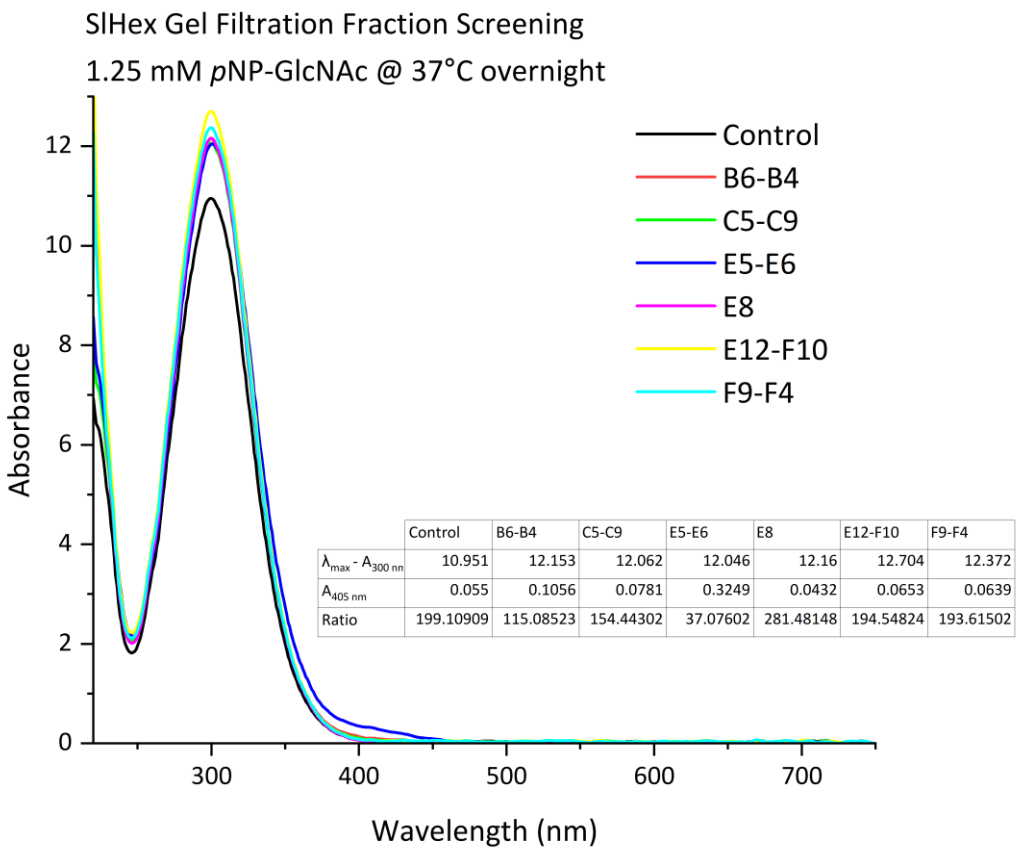



Figure 3.3.7: Chromatogram of the S200 SEC purification of the active SlHex fractions. The black arrows indicate which fractions were taken forward for screening. (Blue) UV/Vis, (Brown) Conductivity.

An attempt was made to run the selected fractions from the SEC purification on an 12% acrylamide SDS-PAGE gel, however, no bands were visible after staining. Therefore, in order to determine which fractions the protein had eluted in, the activity test was repeated under the same conditions for each fraction (Figure 3.3.8).




| | Control | B6-B4 | C5-C9 | E5-E6 | E8 | E12-F10 | F9-F4 |
|---|---|---|---|---|---|---|---|
| $\lambda_{max}$ - $A_{300\ nn}$ | 10.951 | 12.153 | 12.062 | 12.046 | 12.16 | 12.704 | 12.372 |
| $A_{405\ nm}$ | 0.055 | 0.1056 | 0.0781 | 0.3249 | 0.0432 | 0.0653 | 0.0639 |
| Ratio | 199.10909 | 115.08523 | 154.44302 | 37.07602 | 281.48148 | 194.54824 | 193.61502 |

Figure 3.3.8: Overlaid UV/Vis spectra of the collated fractions from the S200 SEC purification of SlHex after incubation overnight with 1.25 mM *p*NP-GlcNAc **4**. The table shows the ratio between the $\lambda_{max}$ of *p*NP-GlcNAc **4**, 300 nm, and the characteristic absorbance at 405 nm of *p*-nitrophenol **6**, the hydrolysis product of *p*NP-GlcNAc **4**.

This result appears to suggest that the protein was successfully purified and eluted in fractions E5-E6, however, the concentration of the protein in these fractions was too small to be quantified by UV/Vis analysis. It was therefore decided to scale up the production of SlHex to 8 L of cell culture. However, upon attempting to purify the cell lysate via HisTrap (Appendix 39), it became apparent that scaling up the cell culture did not increase the amount of protein produced, and only caused an increase in the amount of contaminating protein. Attempts were made to use different detergents to resolubilise the protein from the insoluble cell debris, but no detergent was successfully able to solubilise the protein (Figure 3.3.9).

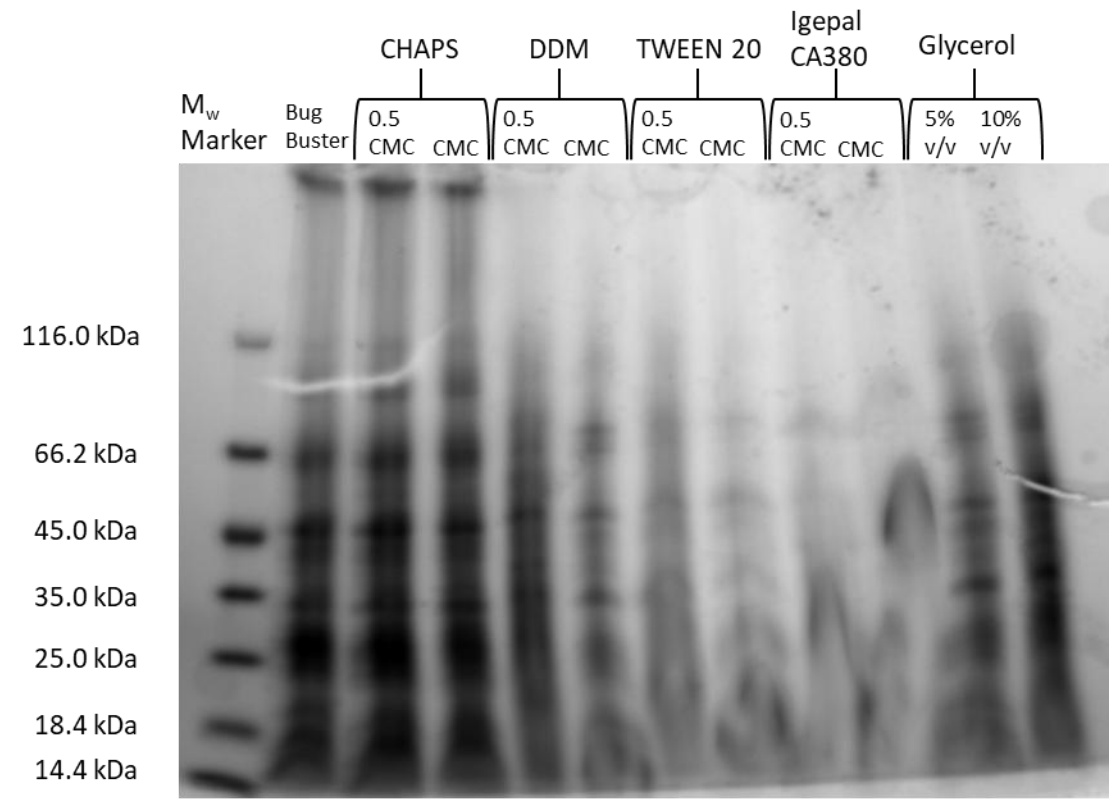



Figure 3.3.9: 12% SDS-PAGE gel of the SlHex detergent screen. Cell lysis was performed with BugBuster lysis solution with addition of the detergent at 0.5 CMC and CMC.

Due to this, it became apparent that SlHex too would require expression trials and so attempts were made to purify ShHex. ShHex did not show as high levels of expression, but the soluble fraction of ShHex did appear to contain a small band around ≈ 40 kDa (Figure 3.3.4, yellow box) ($M_w$ of ShHex = 38,841.6 g $mol^{-1}$), so the cell pellet was subjected to lysis and sonication and passed down a 5 mL HisTrap column (Appendix 40). This proved to be unsuccessful, as whilst there appeared to be a protein band of the correct molecular weight in the column load, this appears to have completely failed to bind to the column. This again suggested that expression trials would be required, however, as SlHex is also a member of the K clade, it was decided to focus on the purification of this protein over ShHex.

### 3.3.4. Expression Trials

With the detergent screen failing to solubilise the SlHex further, it was decided to attempt expression trials for both SlHex and SaHex to determine if an optimised expression protocol could be found. The conditions that were chosen for screening were temperature (16 °C, 25 °C, and 37 °C), and induction time (4 hours, and overnight), whilst the IPTG concentration, $OD_{600}$, and cell line would remain the same. Trials were initially performed on SlHex (Figure 3.3.10).

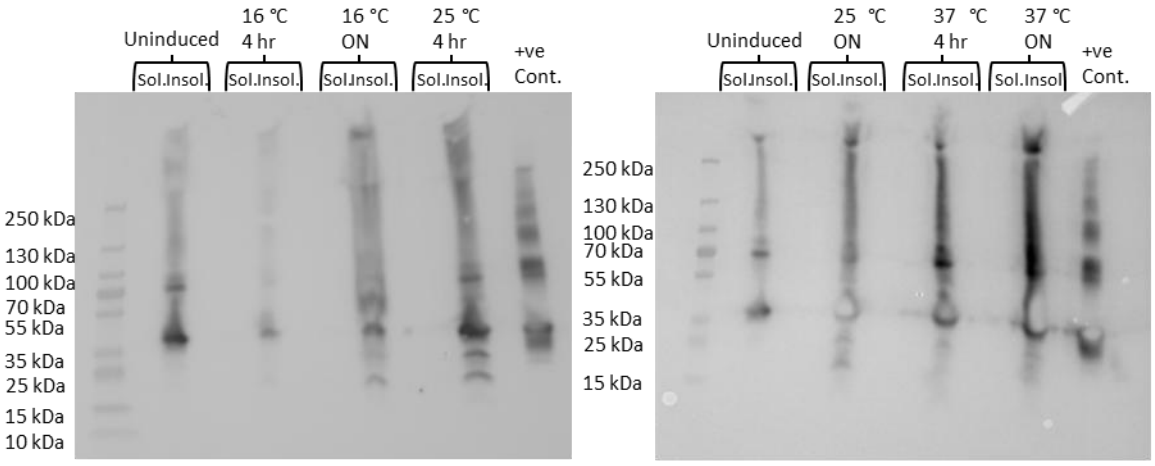



Figure 3.3.10: Anti-His Western Blot of the expression trial of SlHex in Bl21 DE3 cells, induced by addition of 0.2 mM IPTG at $OD_{600}$ = 0.6.

The initial Western Blots were inclusive due to a large amount of insoluble protein overpowering the signal of any soluble protein. These Western Blots also to indicate that the expression system of for SlHex is leaky, potentially leading to cell toxicity, which would explain the longer growth times that were observed for this protein during overexpression attempts. A second Western Blot was performed with just the soluble fractions to see if any bands could be seen ( Figure 3.3.11).

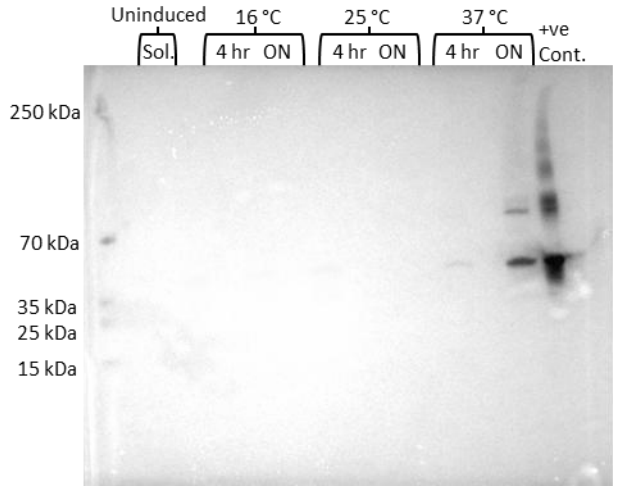



Figure 3.3.11: His-Tag Antibody Western Blot of the soluble fractions of the SlHex expression test.

This revealed two potential protein bands from the 37 °C cultures, with the largest band being in the overnight expression. This prompted a larger scale of expression to be attempted using both conditions, whilst also trialling to use BL21 PLysS cells in an attempt to address the leaky expression (Figure 3.3.12).

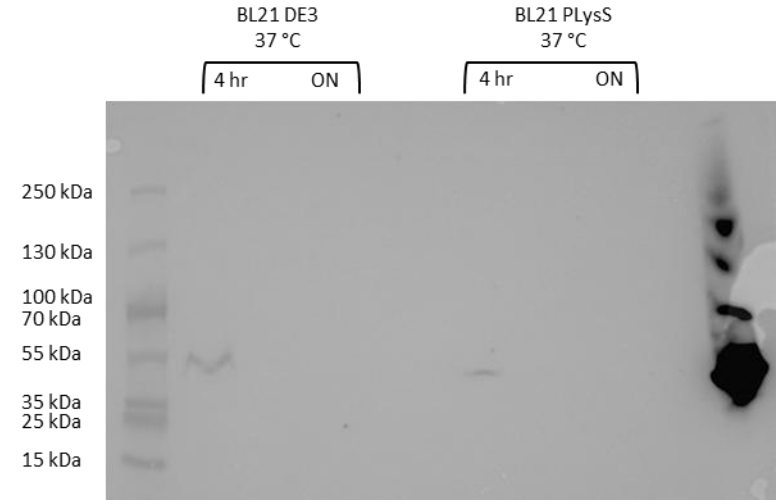


Figure 3.3.12: His-Tag Antibody Western Blot of BL21 DE3 expression trial vs BL21 PLysS expression trial for SlHex.

Unfortunately, this trial appears to contradict the previous results for the expression trials of SlHex, leading to some confusion as to which result should be trusted. However, as a 1.6 L culture of cells had already been made for the SlHex BL21 DE3 cells at 37 °C with overnight induction, it was decided to attempt to purify some protein out of this culture (Figure 3.3.13).

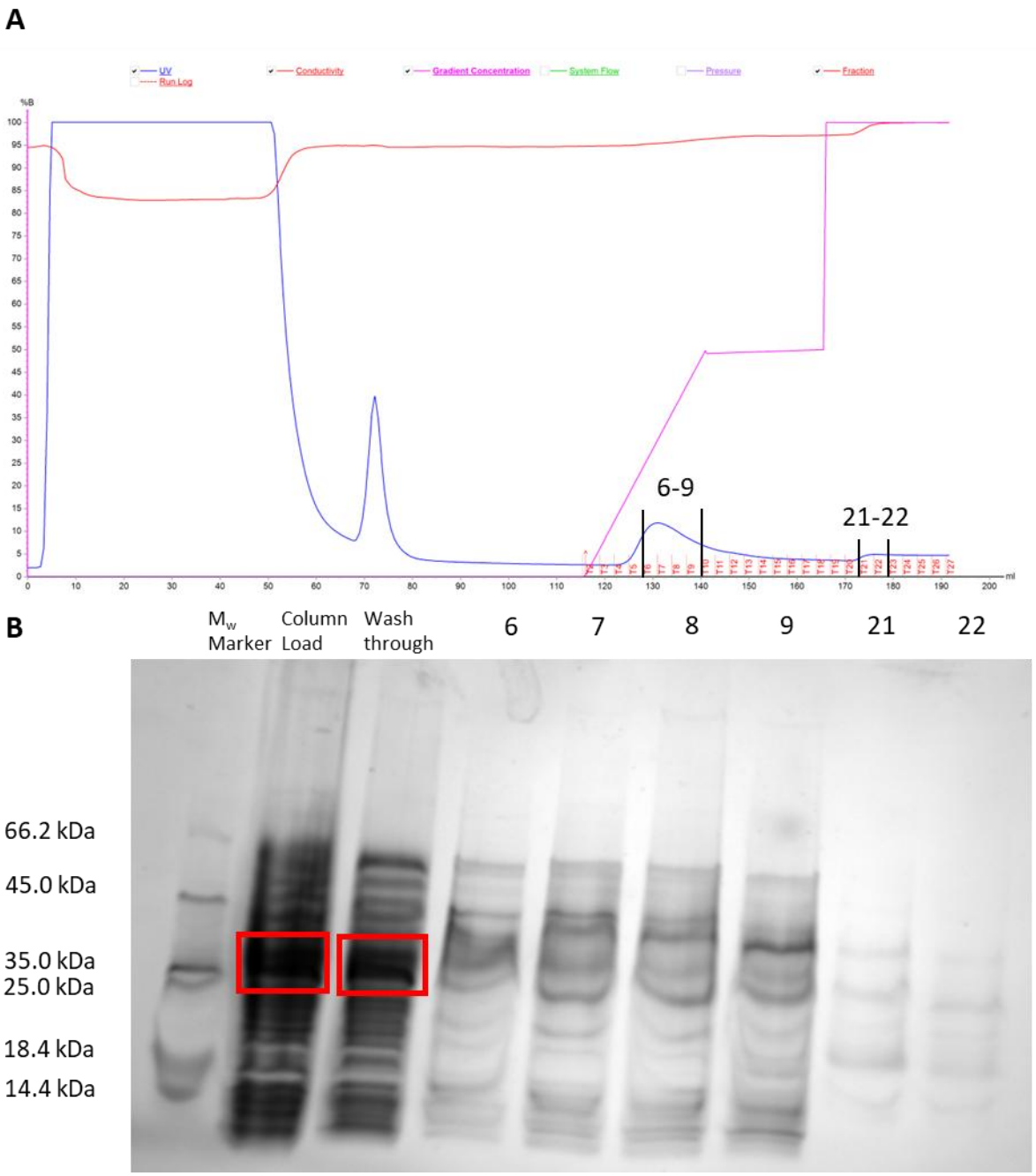



Figure 3.3.13: A) Chromatogram of the purification of the 37 °C ON 1.6L SlHex culture. B) Gel of the fractions from the purification of the 37 °C ON 1.6L SlHex culture.

Protein MS/MS confirmed no protein was present in either of the dark bands (Figure 3.3.13B, red boxes) in the column load or the wash through. An expression trial was also conducted for SaHex testing for the same conditions as SlHex (Figure 3.3.14), however, this expression trial used BL21 PLysS cells due to the cells failing to grow when BL21 DE3 cells were transformed with the plasmid.

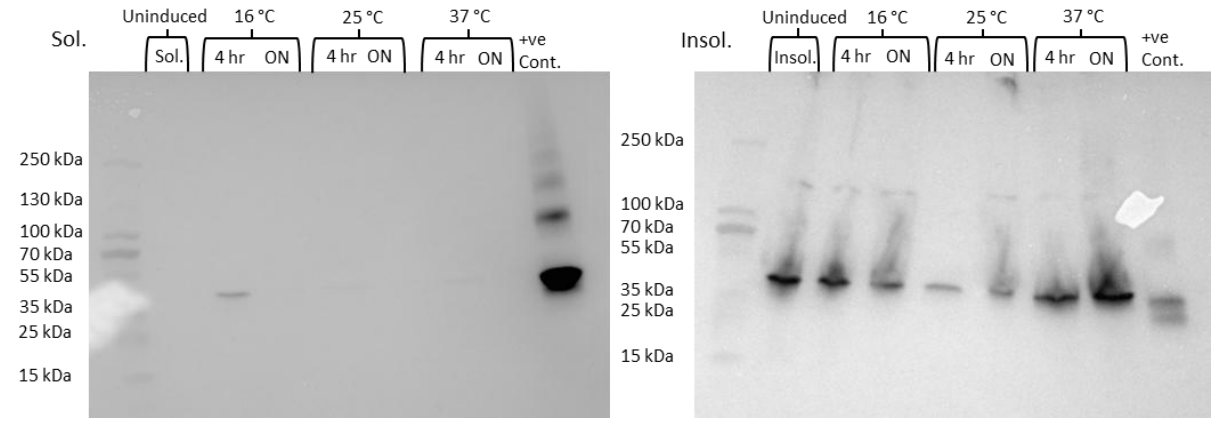



Figure 3.3.14: (Left) His-Tag Antibody Western Blot of the soluble fractions of the expression trial for SaHex. (Right) His-Tag Antibody Western Blot of the insoluble fractions of the expression trial of SaHex. Expression trial performed using BL21 PLysS cells.

This trial suggests that the best conditions for expressing soluble SaHex would be 16 °C with a 4 hour induction time, however, this also shows that SaHex also suffers from a leaky expression system. At this point the decision was made to stop work towards purifying both proteins as it became clear the task would require more substantial optimisation. This work has been since been taken over by Conor Munroe who is working in the Thomas Lab (University of York).

### 3.3.5. Conclusions and Future Work

The preliminary evidence gathered during this project would suggest that the enzymes identified in the *staphylococci* species are in fact PNAGases, and that these PNAGases are structurally distinct from DspB. The AlphaFold2 predictions of the structure of SlHex reveal a lack of the key α-helix, and -1 and -2 binding site found in DspB, suggesting a preference towards fully acetylated PNAG, rather than the dPNAG preferred by DspB. The ability of fractions containing SlHex to break down *p*NP-GlcNAc **4** giving rise to an absorbance signal at 405 nm mirrors that of DspB, a known PNAGase, and given that the signal was much larger than the background signal, it can be concluded this was due to activity of something that was present within the purification fraction.

Unfortunately, the purification of these enzymes proved to be challenging. Expression trials performed on this enzyme gave conflicting results upon being repeated, and the amount of protein purified each time was minimal in comparison to DspB, which the initial purification protocol was based off. These expression trials, however, were fairly minimal in comparison to the extent of variables that could be tested, such as changing the concentration of IPTG during induction, or by changing the $OD_{600}$ at which the induction takes place. These variables have as much of an impact on the protein expression, and future work could reveal these to be the more important variables.

Another method that could be used to help improve protein expression is solubility tags, a method that has been used in the Fascione lab to great success before [212]. It was at this point that the scale of the work required became apparent, so this part of the project was written into its own project, which has been taken on by Connor Munroe in the Thomas Lab (University of York). This will hopefully allow for the eventual scope of PNAGases to be expanded significantly.

## 3.4. *In vitro* modification of PNAG

### 3.4.1. Introduction

With the attempt to create a glycosynthase using DspB failing to produce any PNAG oligosaccharides, the direction of the project had to be changed and a new method for producing PNAG for functionalisation needed to be explored. It has previously been shown that PNAG can be purified from biological sources as this PNAG was what was used to perform ELISA experiments to test the immune response towards PNAG[55,104]. These same studies also showed that *in vitro* modification of PNAG was possible, as dPNAG was generated during these studies via chemical deacetylation of the PNAG samples procured from *in vivo*. Therefore it was reasoned that other chemical methods could be used to produce different types of PNAG for use in further studies.

One such target for this was a functionalised PNAG with a Click handle integrated into the polysaccharide chain. A Click handle would allow for the easy conjugation of PNAG to any molecule of choice, or for the easy insertion of a further functionalised molecule into the PNAG chain. In order to integrate a Click handle, either an alkyne or an azide needed to be incorporated into the PNAG structure somewhere. Of the two options, the azide presented the more realistic option. Incorporation of an azide into GlcNAc molecules can be easily achieved via the use of either DMC **22**[213], and by extension CDMBI **23**, or by the use of a reagent developed by Lim *et al*[2] called 2-azido-1,3-dimethylimidazolinium hexafluorophosphate, or ADMP **63**. This reagent, which is based upon DMC **22**, works both as an anomeric activating reagent in a similar way to DMC **22** during the synthesis of oxazoline **5** (Scheme 3.1.7), and as also as a source of azido ions to be incorporated into the now activating anomeric centre of the GlcNAc molecule. This is beneficial, as the previously mentioned routes using DMC **22** or CDMBI **23** require large excesses of sodium azide, and take place over two reaction steps, which complicates the purification process and decreases yield, neither of which are true for the ADMP **63** route. For this reason, this method was chosen as a potential way to functionalise PNAG that had been purified *in vitro*.

Another modification that would be beneficial is the generation of dsPNAG. As mentioned previously, dPNAG has been previously generated and utilised in studies investigating the immunological properties of PNAG and whether or not the lack of *N*-acetyl groups caused an increase or decrease in the human immune response to PNAG. However, what was not reported in these studies was the effect of *O*-succinylation. dPNAG is generated via the base catalysed *N*-deacetylation, usually using NaOH or something similarly strongly basic. These conditions would also cause *O*-desuccinylation, and therefore it is unknown whether the presence of the succinyl side groups would have an effect on the immune response. It is unknown why some PNAG producing

species perform this modification, and some do not, and it was theorised that, like with the deacetylation modification, the introduction of charged side chains would have a noticeable effect on the strength of the immune response. Generating dsPNAG for use in ELISA studies of PNAG was therefore identified as a goal of this project in order to fill in this gap in the current knowledge. To do so, a less basic reaction would be required. During their studies to identify the source of the negatively charged PNAG fractions initially described by Mack *et al*[54], Sadovskaya *et al*[57] describe a method for the selective *O*-desuccinylation of PNAG utilising 10% $NH_4OH$. This method, confirmed by $^1H$ NMR was able to successfully remove the succinyl groups from the PNAG polysaccharide whilst leaving the N-acetyl groups intact. This method was therefore identified as suitable for the generation of dsPNAG for the purposes of performing ELISA analysis.

### 3.4.2. Purification and Analysis of *Staphylococci* PNAG

Based off of the work of Sadovskaya *et al*[57], a method for the purification of PNAG from *in vivo* sources was developed in the Thomas Lab by Reyme Herman. *S. carnosus* was transformed with a plasmid containing a pTX vector encoding for *ica*ADBC. This transformed *S. carnosus* pTX*ica*ADBC was used to inoculate Tryptic Soy Broth (TSB), and the bacteria was allowed to grow overnight. This growth produces a colourless film that coats the inside of the flask, and this film contains the PNAG that is to be isolated. In order to isolate the PNAG, the cells must be removed from the biofilm so purification via SEC can be performed. This is achieved by resuspending the biofilm in the media and centrifuging to remove the cells that are free in the media. The remaining suspension is then sonicated at a low amplitude. This takes advantage of the fact that gram-positive bacteria are resistance to low levels of sonication[214] to ensure that the cells that are attached to the biofilm matrix are knocked free, whilst remaining intact so as to not release large amounts of contaminants into the media that would complicate the purification process. These cells can also then be centrifuged away. Any residual protein is precipitated out of solution via addition of trichloroacetic acid (TCA), and then the resulting solution is concentrated, and passed down a S200 SEC column monitoring absorbance at 206 nm and 280 nm (Figure 3.4.1).

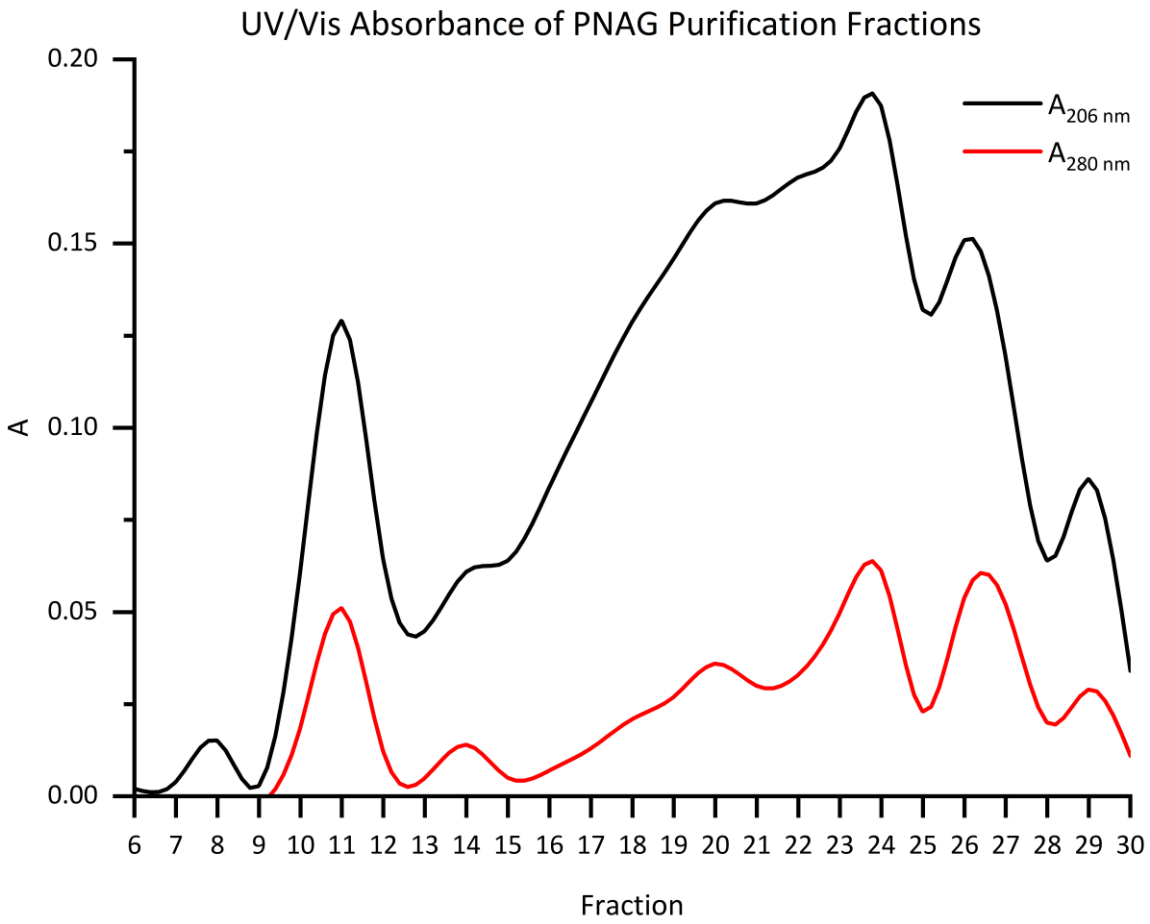



Figure 3.4.1: UV/Vis Chromatogram of the SEC purification of PNAG isolated from *S. carnosus* pTX*ica*ADBC. Purification performed in collaboration with Reyme Herman, Connor Munroe and Beth Kinniment-Williams.

This revealed the potential PNAG containing fractions which can be screened via a Dot Blot using 2% BSA to block the nitrocellulose membrane, and biotinylated succinylated WGA-agglutinin (Vector LabsB-1025S) as a GlcNAc binding reagent[215], and streptavidin-HRP as the blotting reagent (Figure 3.4.2).

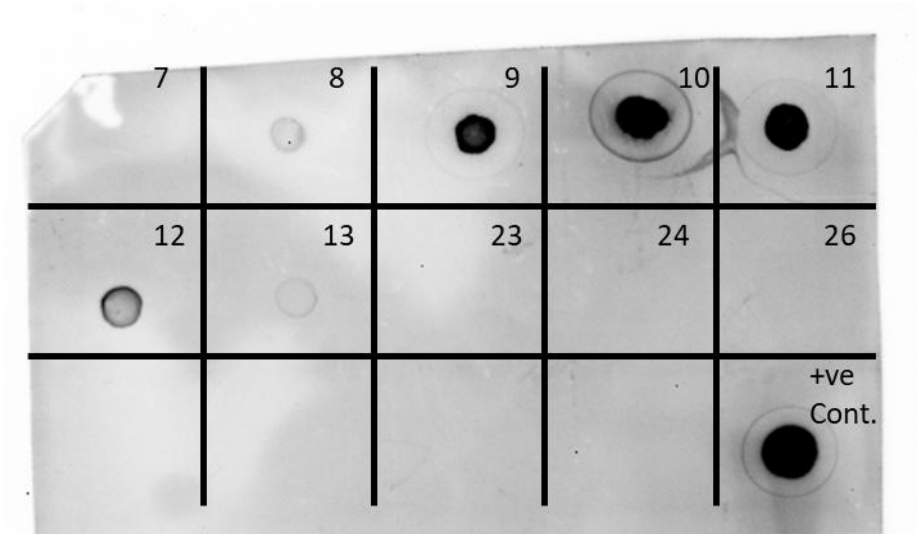



Figure 3.4.2: Dot Blot of the PNAG purification fractions using biotinylated succinylated WGA-aglutinin as a PNAG binding reagent, streptavidin-HRP as the blotting reagent, and 2% BSA solution as a blocking solution.

This revealed the initial peak of the purification chromatogram to contain the purified PNAG, so the fractions could be collated, and lyophilised leaving behind PNAG as a solid powder. This solid powder can then be redissolved in $D_2O$, allowing for analysis via $^1H$ NMR. Initially, attempts were made to utilise both 400 MHz and 500 MHz NMR analysis, however, these magnets proved not to be strong enough to detect any PNAG signals so analysis was performed utilising a 700 MHz spectrometer (Figure 3.4.3).

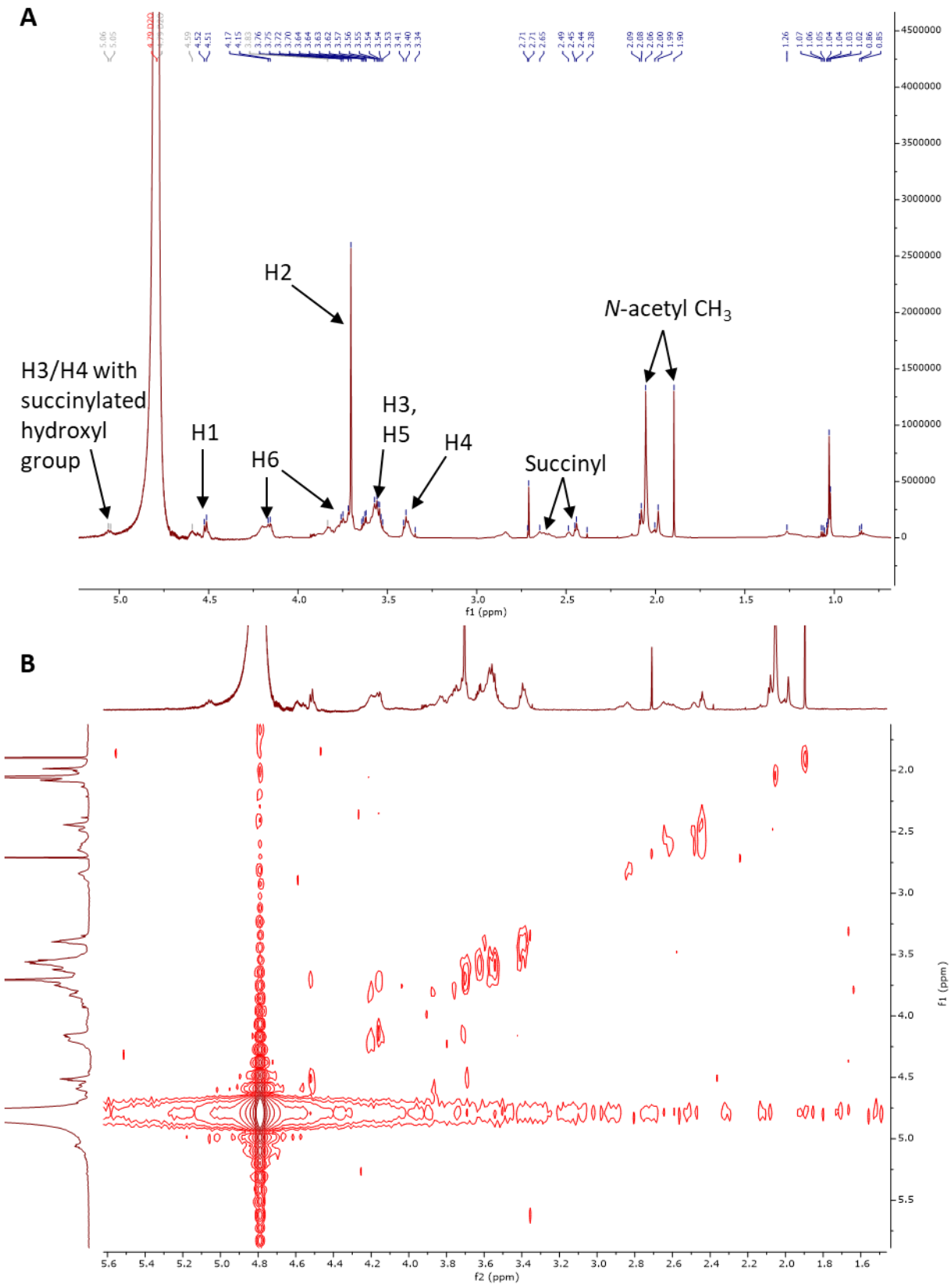
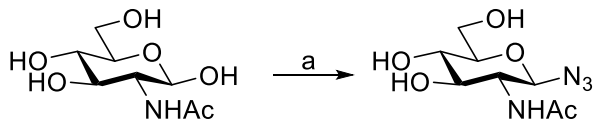



Figure 3.4.3: A) 700 MHz $^{1}$H NMR analysis of PNAG isolated from *S. carnosus* pTX*ica*ADBC. B) 700 MHz COSY analysis of PNAG isolated from *S. carnosus* pTX*ica*ADBC zoomed into sugar peaks region.

Comparing these two spectra to the spectra obtained by Joyce *et al*[58], it is clear that the purification was successful. In addition, peaks corresponding to the *O*-succinylation found in *staphylococci* PNAG can be seen in the region of 2.25-2.75 ppm. These also match the literature and this means that generation of dsPNAG is possible with this PNAG. This purified PNAG could then be used to attempt to create both azido-functionalised PNAG and the aforementioned dsPNAG.

### 3.4.3. Azide Incorporation into PNAG

Initially, the method used by Tanaka *et al*[213] to incorporate an azide into a glycoside was repeated, using GlcNAc **1** as a test bed (Scheme 3.4.1).




**1** **64**

Scheme 3.4.1: a) DMC **22** (3 eq.), $NaN_3$ (10 eq.), 2,6-lutidine (6 eq.), $D_2O$, 4 °C, 48 hr.

After 48 hr the reaction mixture was analysed via $^1H$ NMR (Figure 3.4.4A).

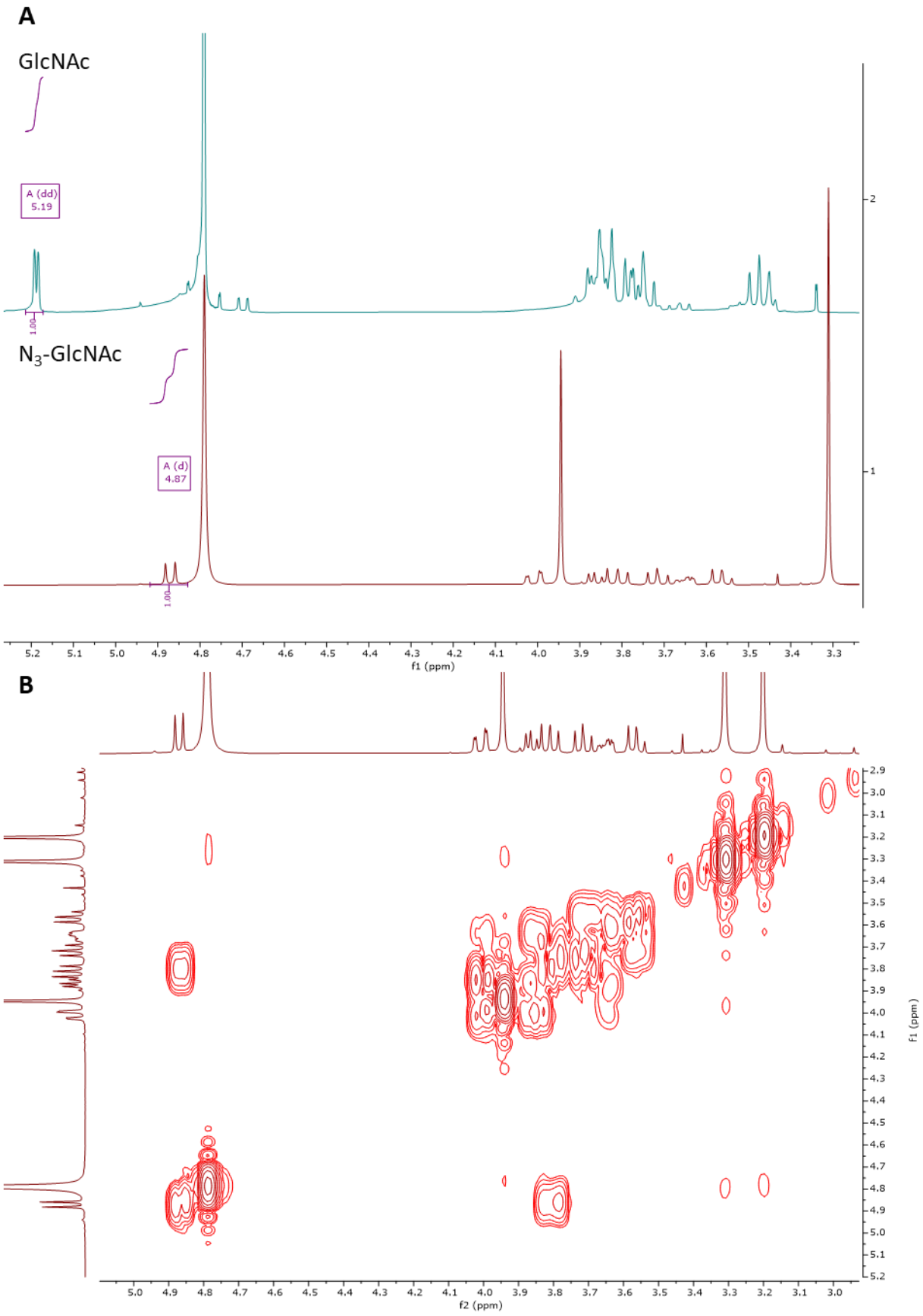
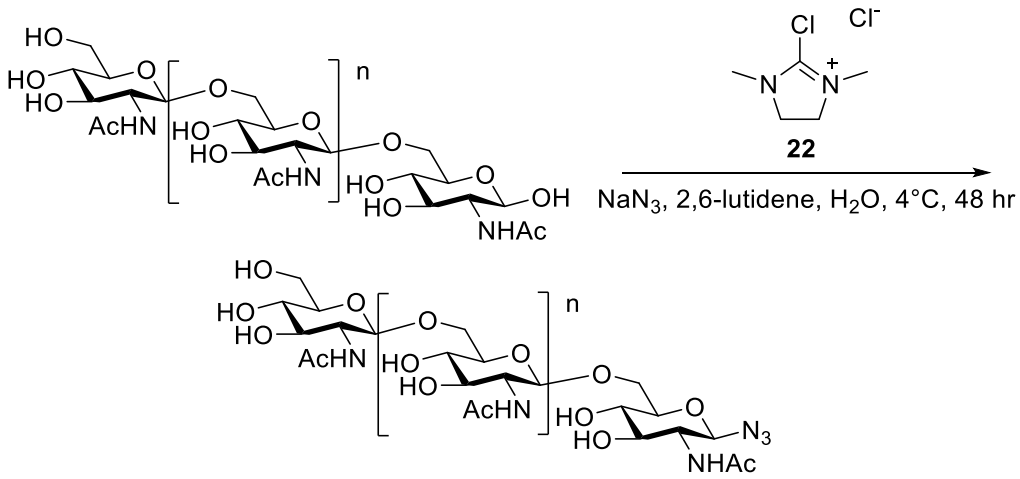

A
GlcNAc
A (dd)
5.19
1.00
N3-GlcNAc
A (d)
4.87
1.00
2
1
5.2 5.1 5.0 4.9 4.8 4.7 4.6 4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3
f1 (ppm)
B
2.9 3.0 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9 4.0 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 5.0 5.1 5.2
f1 (ppm)
5.0 4.9 4.8 4.7 4.6 4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2 3.1 3.0
f2 (ppm)

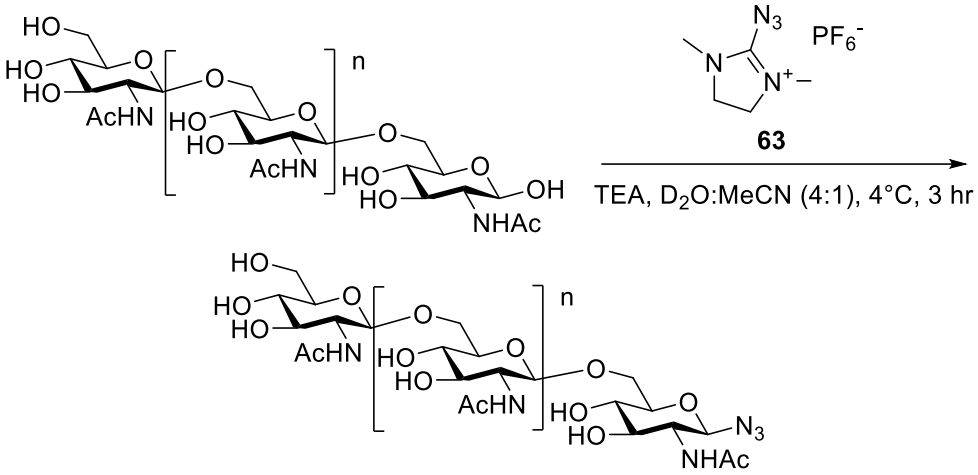
Figure 3.4.4: A) (Top) 400 MHz $^1$H NMR analysis of GlcNAc. (Bottom) 400 MHz $^1$H NMR analysis of the reaction mixture of Scheme 3.4.1. Both spectra have the the anomeric peak highlighted. B) COSY analysis of the reaction mixture of Scheme 3.4.1.

Comparing the anomeric peaks of GlcNAc **1** to the reaction mixture shows that the azide was successfully incorporated as no peak can be seen at 5.19 ppm in the reaction mixture, instead being replaced by a peak at 4.87 ppm. COSY analysis (Figure 3.4.4B) reveals that this new peak is the new anomeric peak, and MS analysis contained a peak that corresponds to $N_3$-GlcNAc **64** (Appendix 41). Whilst it is unclear why the anomeric peak observed does not match the literature value for the anomeric peak of $N_3$-GlcNAc **64**[2,213] (≈ 4.6 ppm), it was concluded that the reaction had been successful, and so the reaction was repeated replacing GlcNAc **1** with the purified PNAG (Scheme 3.4.2).




Scheme 3.4.2: Scheme for the incorporation of an azide moiety into PNAG via the use of DMC **22** inspired by the work of Tanaka *et al*[213]

However, the large amount of $NaN_3$ present in the reaction mixture meant that $^1$H NMR analysis proved difficult as the salt interfered with the signal and therefore, no PNAG peaks could be defined, even when using the 700 MHz spectrometer. The large amount of $NaN_3$ also meant that purification proved difficult, with attempts at dialysis resulting in the complete loss of product. It is possible that SEC purification could be used to remove the large amount of $NaN_3$, however, it was decided to change the azide incorporation reagent to ADMP **63**. ADMP **63** replaces DMC **22** as the oxazoline formation reagent in the reaction mechanism (Scheme 3.1.7), and in doing so releases an azido ion into solution that can then break open the oxazoline to incorporate the azide at the anomeric carbon[2]. As such, ADMP **63** can be used at lower equivalents (Scheme 3.4.3), which should hopefully mean the purification has less issues.




Scheme 3.4.3: Scheme for the incorporation of an azide moiety into PNAG via the use of ADMP **63** inspired by the work of Lim *et al*[2].

Rather than using NMR to attempt to track the progress of the reaction, a tricine SDS-PAGE (TSDS-PAGE) gel can be run on a sample of the reaction mixture that has undergone a Click reaction with a fluorescent dye (Alexa Fluoro 488 (AF488)). This way, if any of the PNAG successfully had an azide incorporated into it, it would appear as a band in the gel, as opposed to the ADMP **63** which would run off the bottom of the gel due to its small size. This experiment was performed using both copper catalysed Click (CuAAC), and strain promoted Click (SPAAC), as it was unknown which method would work best for azido-PNAG (Figure 3.4.5), with the assistance of Dr Joe Nabarro.

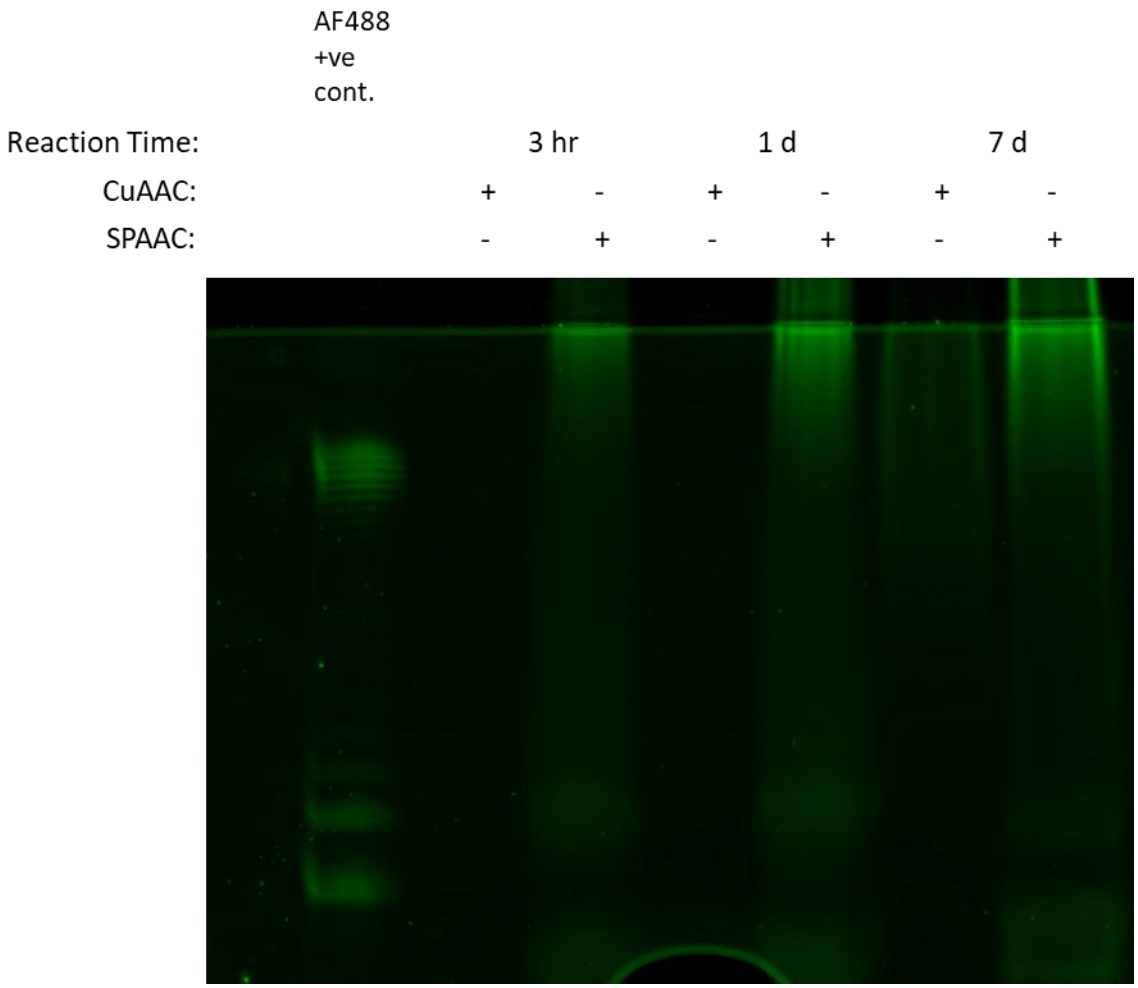



Figure 3.4.5: TSDS-PAGE gel showing samples taken from the reaction mixture of Scheme 3.4.3 having undergone either CuAAC or SPAAC. Fluoresence was measured at 488 nm. Fluoresence analysis performed by Dr Joe Nabarro (Fascione Lab).

These initial results looked promising, and appear to suggest that some of the PNAG was successfully converted into azido-PNAG. These initial results also suggest that the best method for Click chemistry with PNAG is SPAAC, rather than CuAAC. It was speculated that the Cu(II) ions might be coordinated by the deacetylated PNAG monomers, thus stopping them from performing the catalysis of the azide-alkyne coupling. It was therefore decided to only utilise SPAAC going forward.

Whilst these initial results seem to suggest that the reaction was a success, it is unclear if that is indeed the case due to the large amount of smearing seen in the gel making it difficult to tell if the fluorescence seen is actually PNAG, or if it is left over dye. In order to determine whether or not the signals being seen were actually azido-PNAG, the samples were lyophilised, redissolved at 4 mg/mL, and combined with 4 mg/mL Cyclooctyne-Green Fluoresence Protein (CycloOct-GFP) made by Dr Robin Brabham as part of his PhD in the Fascione Lab[216]. CycloOct-GFP contains a strained alkyne, and as such can undergo SPAAC when in the presence of an azide, and as GFP is a protein, it will have a distinct band on an SDS-PAGE gel. Therefore, if the signal seen in the TSDS-

PAGE gel was azido-PNAG, then the SPAAC reaction with CycloOct-GFP would cause the $M_w$ of the GFP to increase, and so the band would move up the gel. After 1 hour, the samples can then be treated with 2mg/mL DspB for 1 hour, as this would degrade the PNAG, causing the GFP band to return to its original position on the gel, proving the successful incorporation of an azide into PNAG. These samples were prepared and run simultaneously on a 12% acrylamide SDS-PAGE gel. Additionally, a sample of the azido-PNAG was dialysed into $dH_2O$ to see if this would have an effect on the SPAAC reaction (Figure 3.4.6).

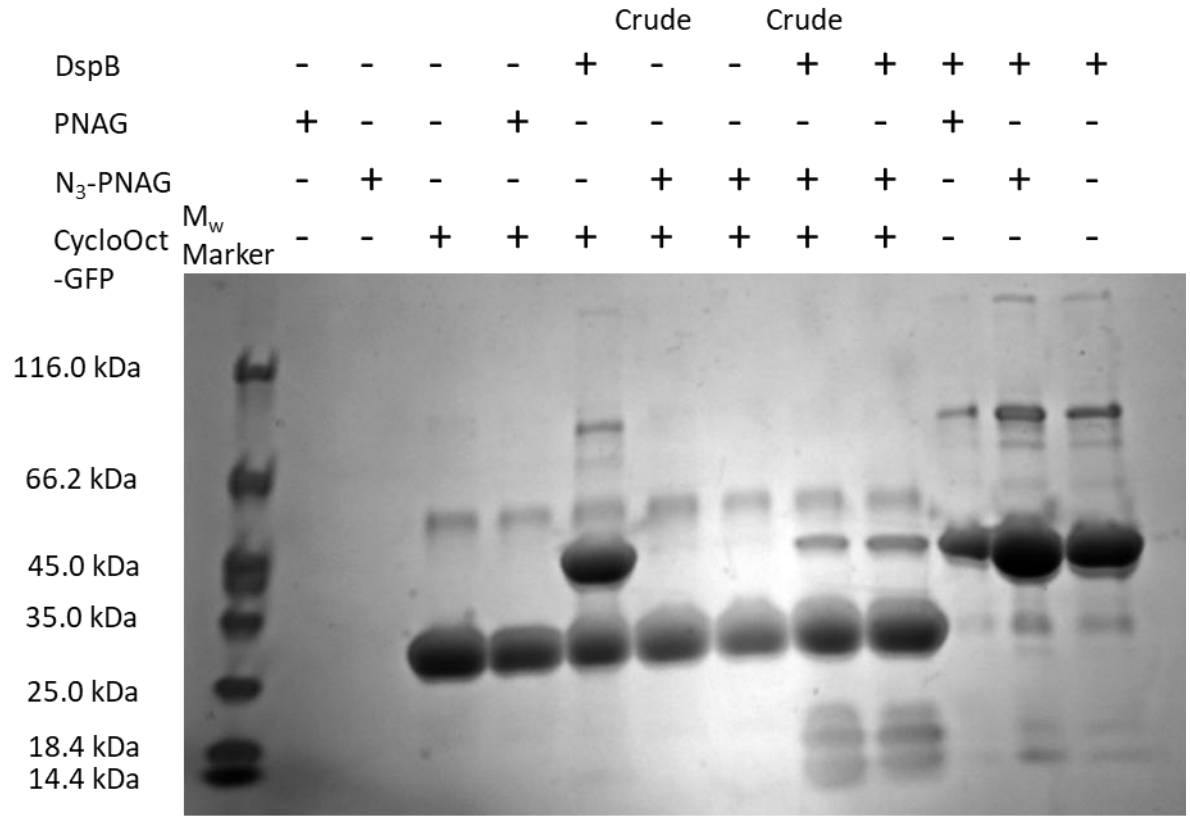



Figure 3.4.6: 12% acrylamide SDS-PAGE gel of the SPAAC reaction between azido-PNAG and CycloOct-GFP, and of the subsequent DspB digest of the reaction mixture.

Unfortunately, this experiment would suggest that very little, if any, CycloOct-GFP reacted with azido-PNAG as no new bands could be observed in the reaction mixture. However, SDS-PAGE is not a sensitive technique, so it is possible that some reacted but just could not be seen in the gel. To address this, the reaction was repeated, and analysed using a 12% acrylamide native PAGE gel. This technique does not involve denaturing samples, and therefore the fluorescence of the CycloOct-GFP can be used instead of Coomassie Stain. This is more sensitive and hopefully would reveal any reacted CycloOct-GFP. The equivalents of CycloOct-GFP were also lowered in an attempt to push the equilibrium of the reaction towards the product (Figure 3.4.7).

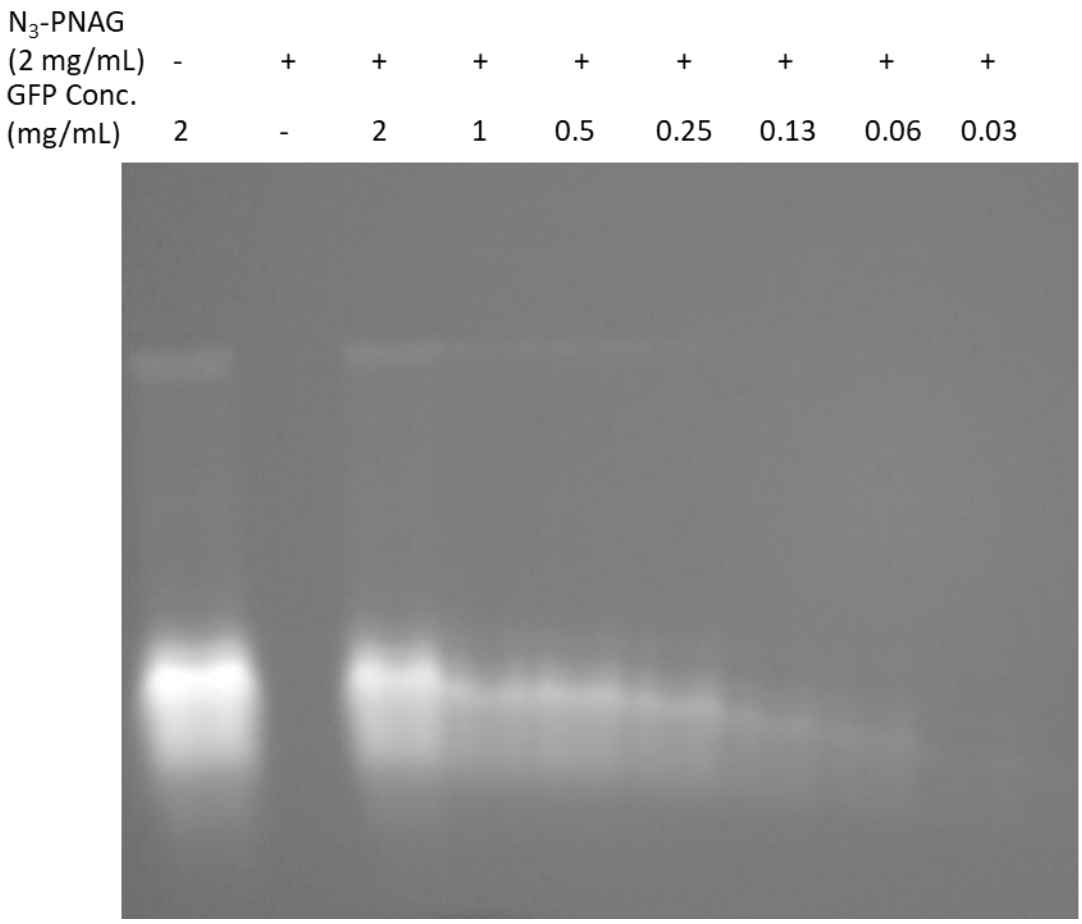



Figure 3.4.7: 12% acrylamide native PAGE gel of SPAAC reaction between azido-PNAG and CycloOct-GFP. Fluorescence measured at 509 nm.

Again, however, no new bands can be seen in the reaction mixtures in comparison to the GFP band. It was therefore concluded that one of three things had happened. One, the reaction had failed and no azide had been incorporated into the PNAG molecules, two, the Click reaction failed, or three the CycloOct-GFP had degraded in storage, as it had been at -80 °C since 2016. If one of the first two reasons is true, then a new methods for the incorporation of an azide into PNAG, or the Click reaction, would need to be explored, however, if the latter were true, then new CycloOct-GFP would need to be produced using amber stop codon suppression as described by Dr Brabham[216]. However, due to time constraints on the project, this was never attempted, and the work towards azide incorporation into PNAG concluded at this point.

### 3.4.4. Attempted Generation of dPNAG and dsPNAG

Another goal of the project was to generate dPNAG and dsPNAG with the aim of filling in the current hole in the immunological research into PNAG surround succinylation. It is known that levels of acetylation have an effect on the human immune response to PNAG, but it is currently unknown if succinylation has the same effect. Both dPNAG and dsPNAG have been generated

previously in the literature so these methods were utilised in this study as well. Firstly, in order to generate dPNAG, the PNAG isolated from pTX*ica*ADBC was dissolved at 2 mg/mL in 5M NaOH in water and incubated at 37 °C overnight. Following neutralisation, the PNAG solution was dialysed into $dH_2O$ and lyophilised giving a white solid. Analysis of this white solid however, revealed significant contamination (Figure 3.4.8).

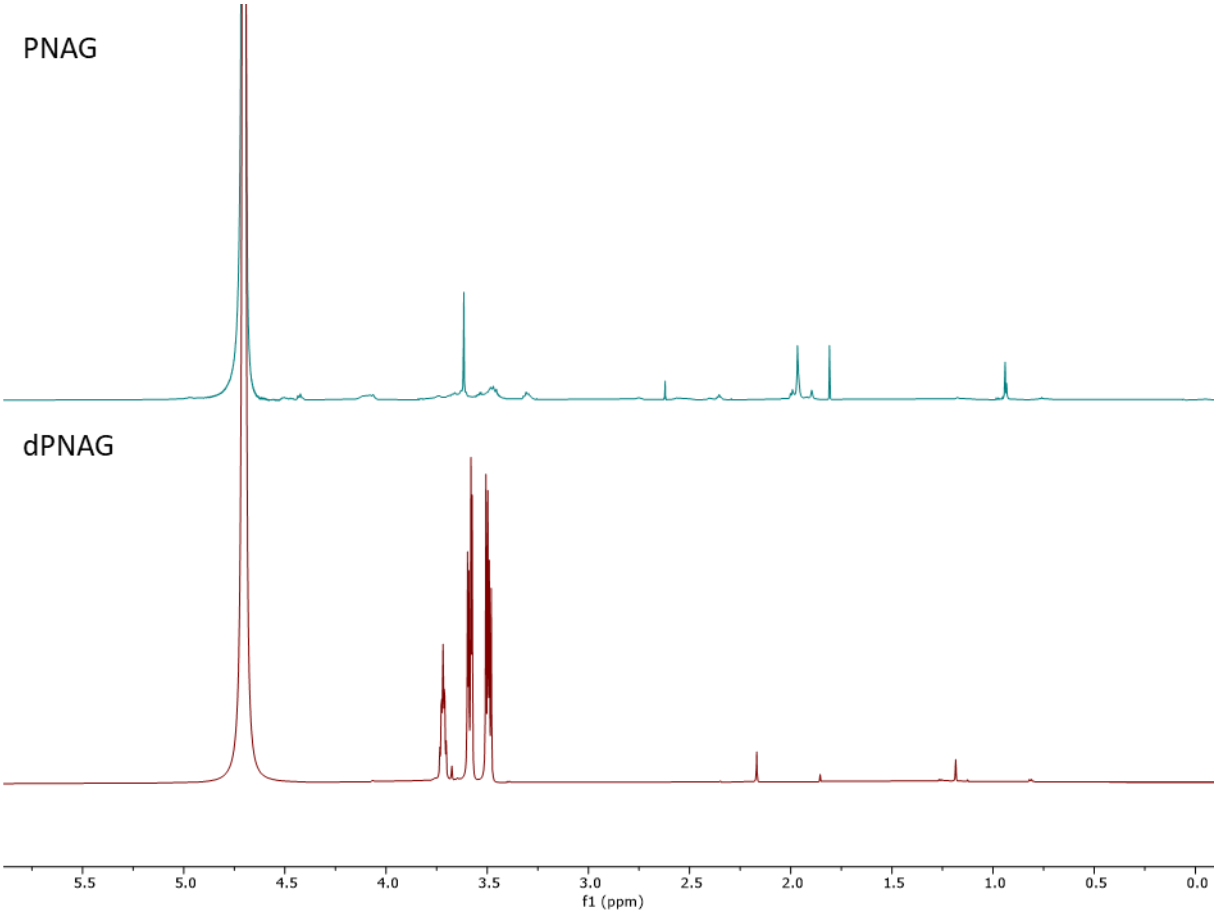



Figure 3.4.8: 700 MHz spectra of PNAG (Top) compared with the spectra of the attempt to generate dPNAG (Bottom).

Large peaks can be seen in the region of 3.5-4 ppm, which is indicative of PEG contamination. This likely occurred during the dialysis process, as these peaks were not present in the previous PNAG spectra. The same issue occurred during the attempted generation of dsPNAG (Figure 3.4.9), which was performed using the same procedure but replacing the 5M NaOH with 10% $NH_4OH$ solution in water, as this should be a strong enough base to remove any *O*-succinyl groups without removing any *N*-acetyl groups.

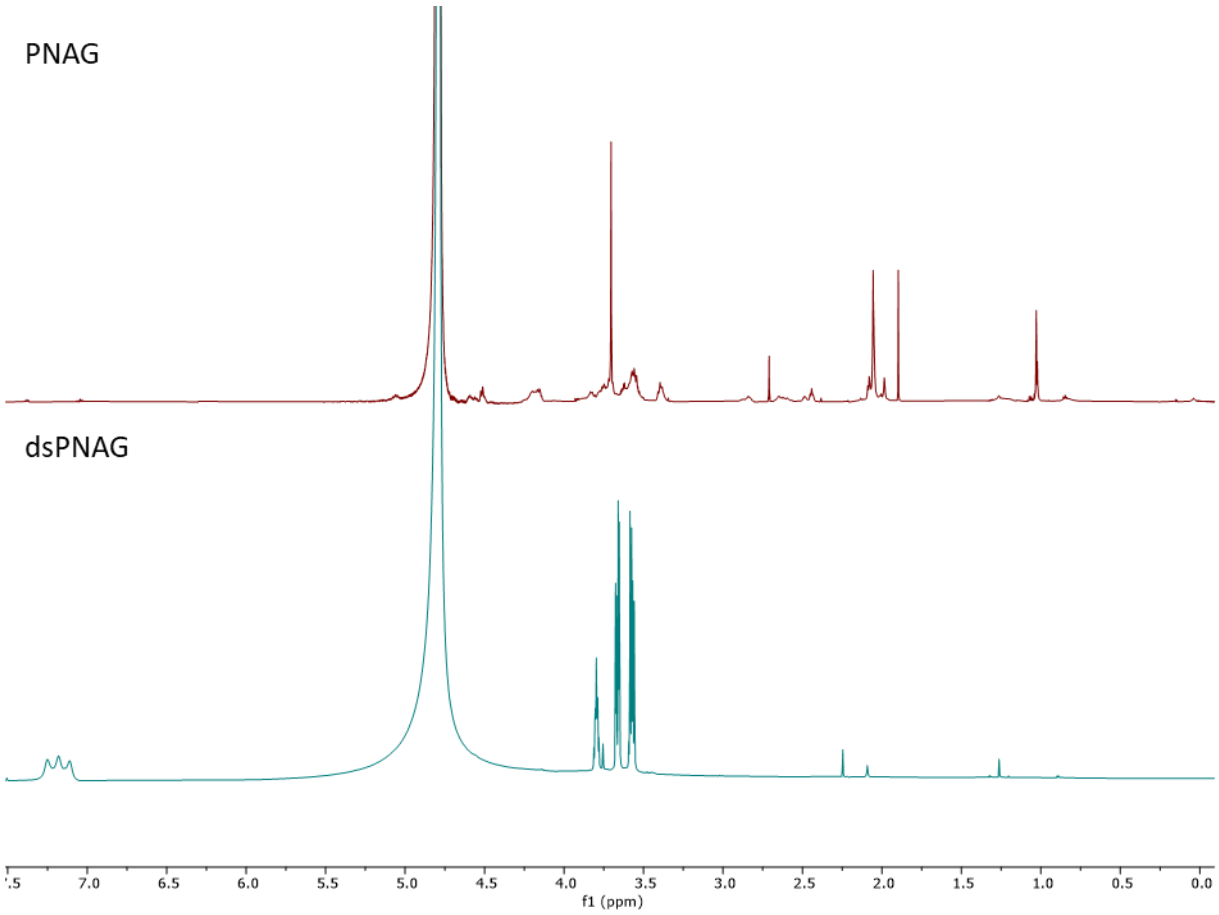
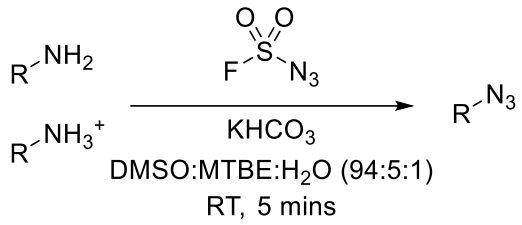


Figure 3.4.9: 700 MHz spectra of PNAG (Top) compared with the spectra of the attempt to generate dsPNAG (Bottom).

In order to resolve this, the SEC purification could be repeated from Section 3.4.2, however, this was unable to be performed during this project due to time constraints, and so, no dPNAG or dsPNAG could be determined to have been generated.

### 3.4.5. Conclusions and Future Work

Attempts to modify PNAG *in vitro* were unfortunately unsuccessful. Whilst PNAG was successfully isolated from *in vivo* sources and characterised as containing succinyl groups, and some initial evidence was gathered showing signs of azide incorporation into PNAG, further investigation led to the conclusion that utilising ADMP **63** to incorporate an azide into PNAG was not an effective strategy. Whilst this avenue of research appears to be closed, an alternative method for the incorporation of an azide, or multiple azides, into PNAG would be to utilise the chemistry developed by Meng *et al*[217] (Scheme 3.4.4).




Scheme 3.4.4: Scheme for the conversion of primary amines into azides via the application of fluorosulfuryl azide

This chemistry allows for the conversion of primary amines into azides using a diazotizing species known as fluorosulfuryl azide. This chemistry also works on amines carrying a charge, and this is important, as in PNAG, it is well known that a percentage of the *N*-acetyl groups have been *N*-deacetylated, leaving a primary amine group that at biological pH carries a positive charge. This chemistry, therefore, could allow for the conversion of up to 15% of the GlcNAc residues in PNAG into azides, potentially circumnavigating the issues encountered during the course of this project.

The generation of dPNAG and dsPNAG were also unsuccessful, however, this is due to time constraints rather than experimental failure. If the samples had undergone SEC purification, then it is likely that these experiments would have been successful, and ELISA experiments with dsPNAG could've been carried out. It should also be noted that the use of dialysis tubing in this experiment is not recommended, as it is the most likely source of the contaminating PEG peaks seen in the NMR.

## 3.5. **Creating a new PNAG detection assay**

### 3.5.1. Introduction

One of the challenges presented in working with PNAG is the difficulty in detecting it within *in vivo* samples. As demonstrated in section 3.4.2, PNAG can be isolated and analysed via NMR analysis, however, this is only possible with a purified sample. If, for example, the PNAG being analysed is found in bacterial media, or cell lysate, this type of analysis is not possible. Currently, there are a number of methods to approach this analysis. One such method is crystal violet staining, which utilises crystal violet dye to stain biofilms within cell cultures[218]. This works as biofilms often contain protein and DNA that can bind to the crystal violet stain[219,220]. This therefore means that while this method is effective if you know that the primary component of the biofilm of the studied species is PNAG, if you are looking specifically to confirm the presence of PNAG in a sample, this method is ineffective as it does not bind to PNAG itself.

A different method that has been previously employed in this type of analysis is to use wheat-germ agglutinin (WGA). WGA is well known to bind to GlcNAc[215], and therefore is capable of binding to PNAG. If the WGA is biotinylated therefore, you can perform a blot detection, either a dot blot or a Western blot[221,222] to detect the PNAG via blotting with streptavidin-HRP, as demonstrated in Figure 3.4.2. This technique suffers from the same drawback as the crystal violet stain however, in that, whilst it is a more PNAG specific method than crystal violet staining, it still does not allow for definitive PNAG identification in *in vivo* samples, as GlcNAc is extremely common in biological settings, and so false positives would be extremely common. However, the fact that PNAG can be detected via blotting with WGA suggests that, if a more PNAG specific enzyme could be found, then it could be possible to create a method for PNAG detection that is specific enough to be used on unpurified biological samples.

Whilst there are not any known enzymes that can bind to PNAG without breaking it down, it has previously been shown that the use of inactive DspB probes can allow for PNAG detection *in vivo*. Eddenden *et al* used DspB-GFP E184Q to detect PNAG formation in cell cultures by monitoring the fluorescence of the conjugated GFP protein[223]. In this study, an inactive DspB mutant was also inadvertently created in DspB Y278F. DspB Y278F was demonstrated to have little to no activity with GlcNAc analogues throughout the assays conducted in section 3.2, and therefore might not breakdown PNAG should it be incubated with it. Y278 is also not a residue that forms a part of the binding sites found within DspB, and therefore a mutation of this residue would hopefully leave the

binding sites intact and available to bind to PNAG. If both these things were true, then DspB Y278F would be the ideal enzyme for a new PNAG detection blot. However, since this protein is not conjugated to GFP, a method was needed to convert DspB Y278F into a detectable blotting reagent. This could be achieved via biotinylation, as a biotinylated protein could bind to streptavidin-HRP allowing for detection via dot blotting or Western blotting. DspB is not naturally biotinylated, so this would require a way to biotinylate DspB *in vitro*.

There are numerous ways that proteins can be modified *in vitro*[224] but perhaps the easiest method is to utilise an NHS ester to modify any primary amine found within the structure of DspB Y278F. This method has been known since 1972 when it was first described by Becker *et al*[225], and it is popular due to its ease, requiring only a slightly basic pH (typically pH 7-9) to work under aqueous conditions, and the abundance of surface exposed lysine residues that are typically found in proteins[226]. This popularity has also led to NHS-biotin esters being easy to obtain commercially, and so this method was chosen for this study, leading to the following proposal for the development of a new PNAG specific blot:

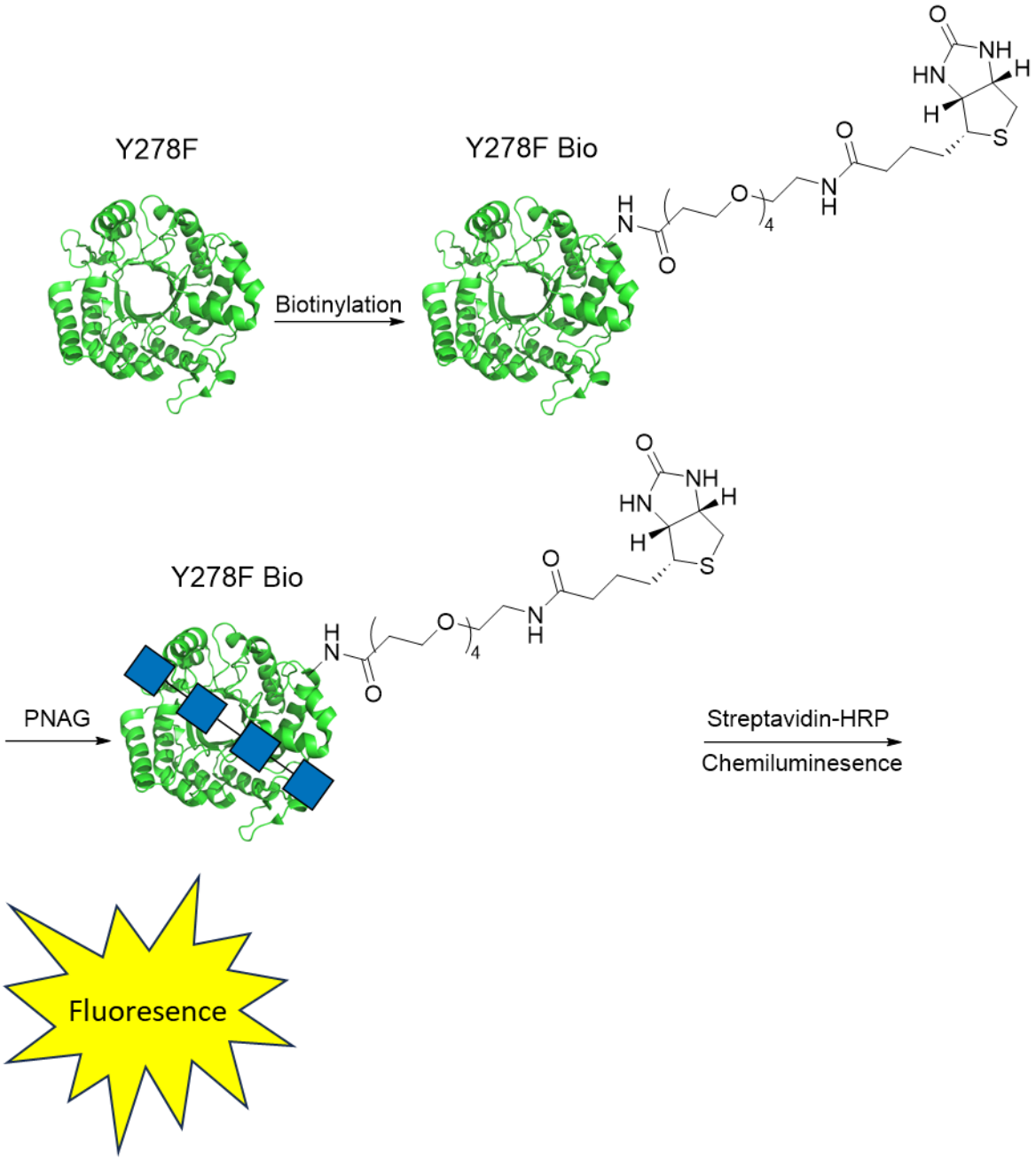


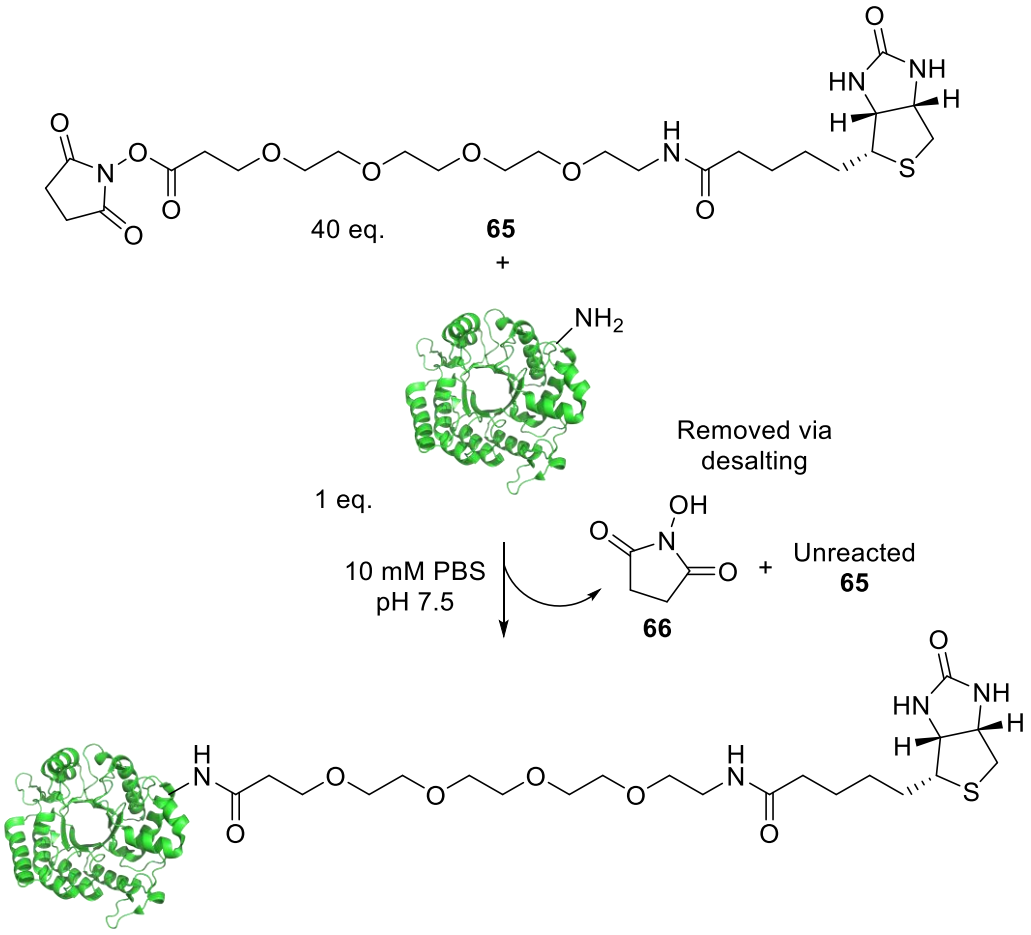

Figure 3.5.1: The proposed method for the creation of a new PNAG blotting reagent via the biotinylation of DspB Y278F.

Should this be successful, this would mean that any inactive version of DspB could be converted into a PNAG blotting reagent easily via biotinylation. It would also mean that should an enzyme be discovered that naturally binds to PNAG, it too could be converted into a PNAG blotting reagent, increasing the scope for this technique, and allowing for easy PNAG detection *in vivo* and *in vitro*.

### 3.5.2. Biotintylation of DspB Y278F

Before biotinylation of DspB was performed, the unmodified mutant was tested by Connor Munroe from the Thomas Lab for PNAG binding. Using Quartz Crystal Microbalance with Dissipation

monitoring (QCM-D), PNAG was layered onto the sensor and then DspB was washed over the PNAG layer and binding kinetics were monitored. Both WT DspB and DspB Y278F showed evidence of binding to the PNAG, but whilst the WT DspB appeared to show subsequent dissipation of the PNAG layer, likely the PNAG being broken down by the WT DspB, the DspB Y278F showed no further activity. This confirmed the hypothesis that DspB Y278F would be a good candidate for a new PNAG blotting reagent, as this data suggests that the PNAG binding is unaffected by the mutation of Y278, but the hydrolysis activity is abrogated. Therefore, the next step was to show that DspB Y278F could be successfully biotinylated. As mentioned previously, this is a popular method of *in vitro* protein modification, and as such, kits are available commercially to perform the reaction. For this study, the Pierce™ Antibody Biotinylation Kit for IP from ThermoFischer Scientific. This utilises NHS-$PEG_4$-Biotin **65** to perform the biotinylation (Scheme 3.5.1).




Scheme 3.5.1: The scheme for the biotinyation of DspB Y278F using the Pierce™ Antibody Biotinylation Kit for IP.

This biotinylation was performed on two different stocks of DspB Y278F, an older purification, and a fresher purification used for the later assays in section 3.2 (Figure 3.5.2).

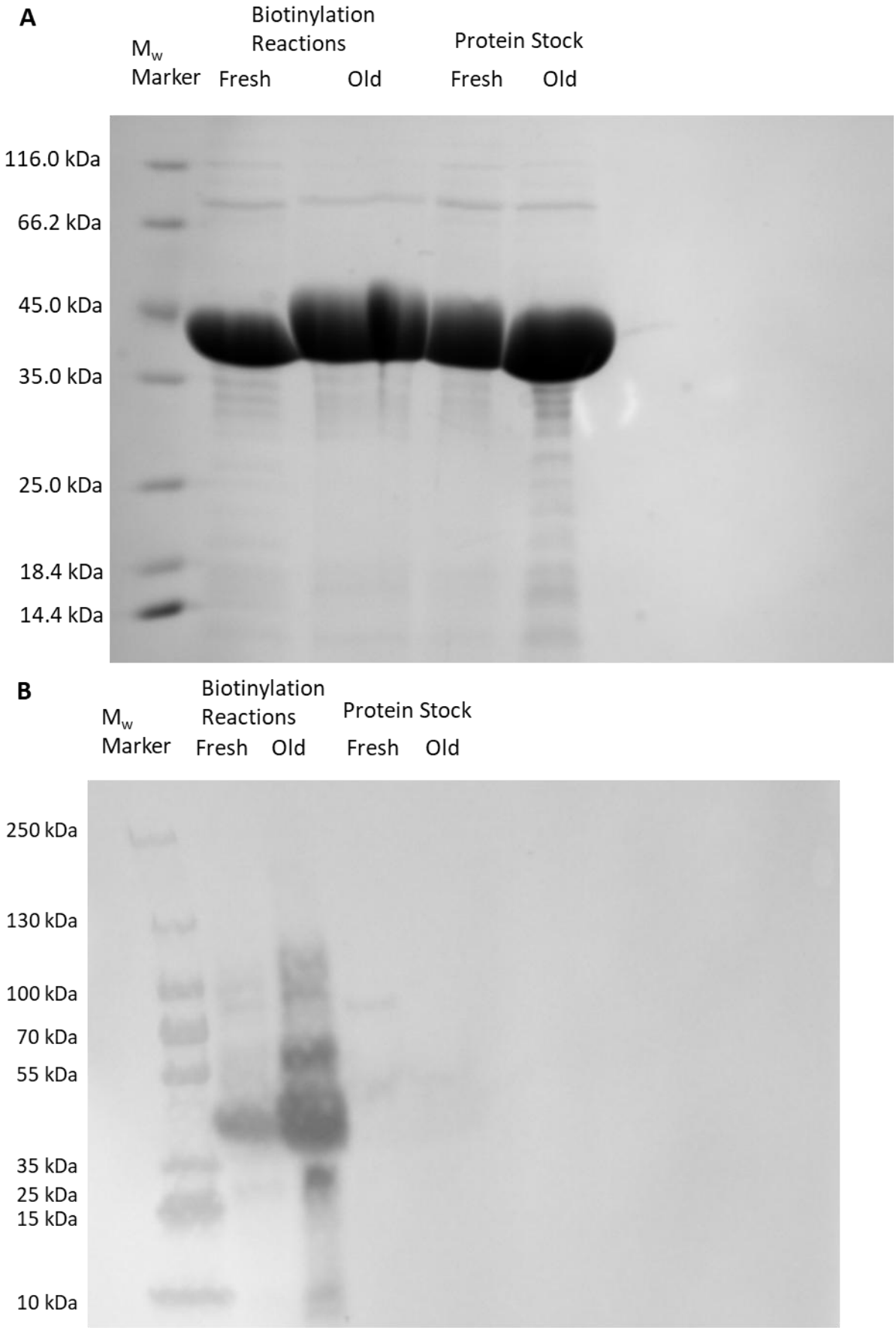



Figure 3.5.2: A) 12% acrylamide SDS-PAGE gel of the biotinylation of DspB Y278F. B) Biotin Western Blot of the biotinylation of DspB Y278F using Streptavidin-HRP as the blotting reagent and 2% BSA solution as a blocking solution.

The Western Blot shows that the biotinylation was performed successfully with both stocks. At this point it was decided to only use the newer stock going forward to avoid any issues with degradation. With the biotinylated DspB Y278F successfully generated, the next step was to begin testing to see if this enzyme could now be used as a blotting reagent.

### 3.5.3. Investigation of Blotting Capability of Biotin-DspB Y278F

The first test that was carried out was a dot blot test to see if biotinylated DspB Y278F could be used as a blotting reagent. To test this, 5 µL each of 2 mg/mL PNAG 2 mg/mL GlcNAc **1**, 2 mg/mL non-biotinylated DspB Y278F, and 2 mg/mL biotinylated DspB Y278F was dotted onto the membrane. The membrane was then blocked with 2% BSA solution, and blotted twice, once with a 0.07 µg/mL solution of biotinylated DspB Y278F to match the concentration of succinylated biotinylated WGA used in Figure 3.4.2, followed by streptavidin-HRP (Figure 3.5.3).

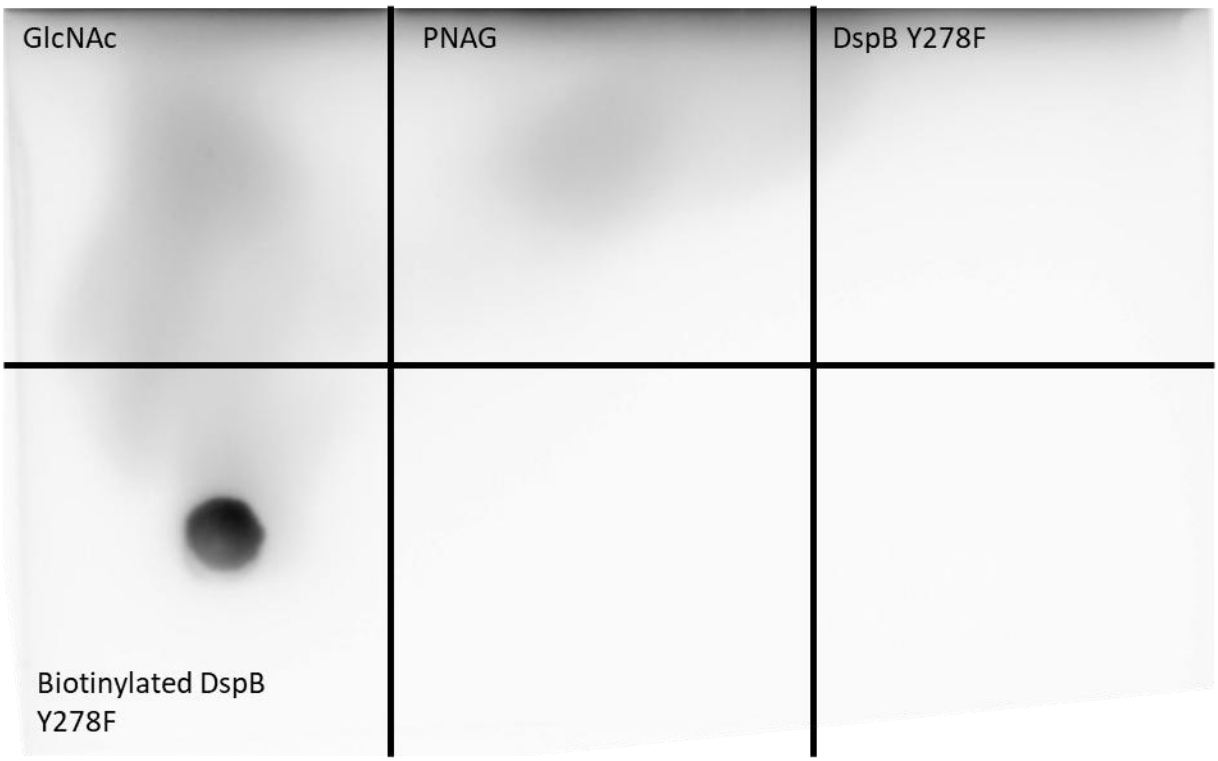



Figure 3.5.3: Dot Blot showing GlcNAc, PNAG, unbiotinylated DspB Y278F and biotinylated DspB Y278F. Biotinylated DspB Y278F and Streptavidin-HRP used as the blotting reagents and 2% BSA solution used as the blocking solution.

This dot blot would suggest that perhaps the enzyme concentration used for the blotting solution was too low for any effective binding to take place. As such, the blot was repeated with a 100 times more concentrated biotinylated DspB Y278F solution (Figure 3.5.4)

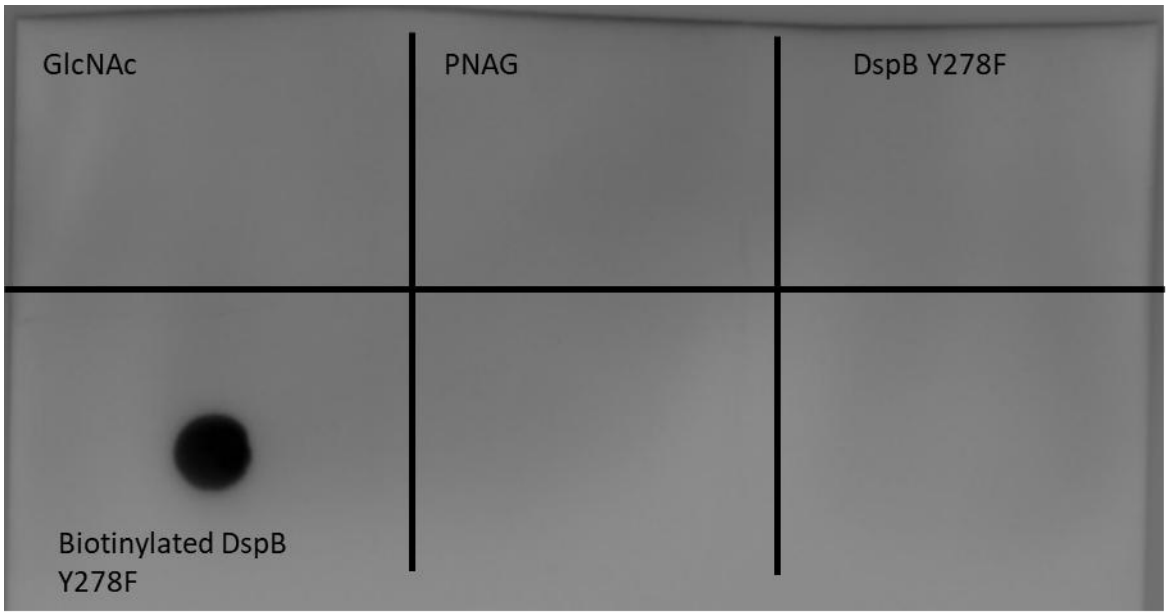



Figure 3.5.4: Dot Blot showing GlcNAc, PNAG, unbiotinylated DspB Y278F and biotinylated DspB Y278F. 100X Biotinylated DspB Y278F and Streptavidin-HRP used as the blotting reagents and 2% BSA solution used as the blocking solution.

Increasing the concentration of biotinylated DspB Y278F by 100 times did not cause a spot to appear in the PNAG grid, suggesting that the biotinylated DspB Y278F had failed to bind to the PNAG on the membrane. To potentially circumnavigate this issue, biotinylated DspB Y278F was premixed with PNAG in order to allow the PNAG and biotinylated DspB Y278F to properly bind together before running the sample down a 12% acrylamide native PAGE gel, and then performing Western Blot analysis using Streptavidin-HRP (Figure 3.5.5).

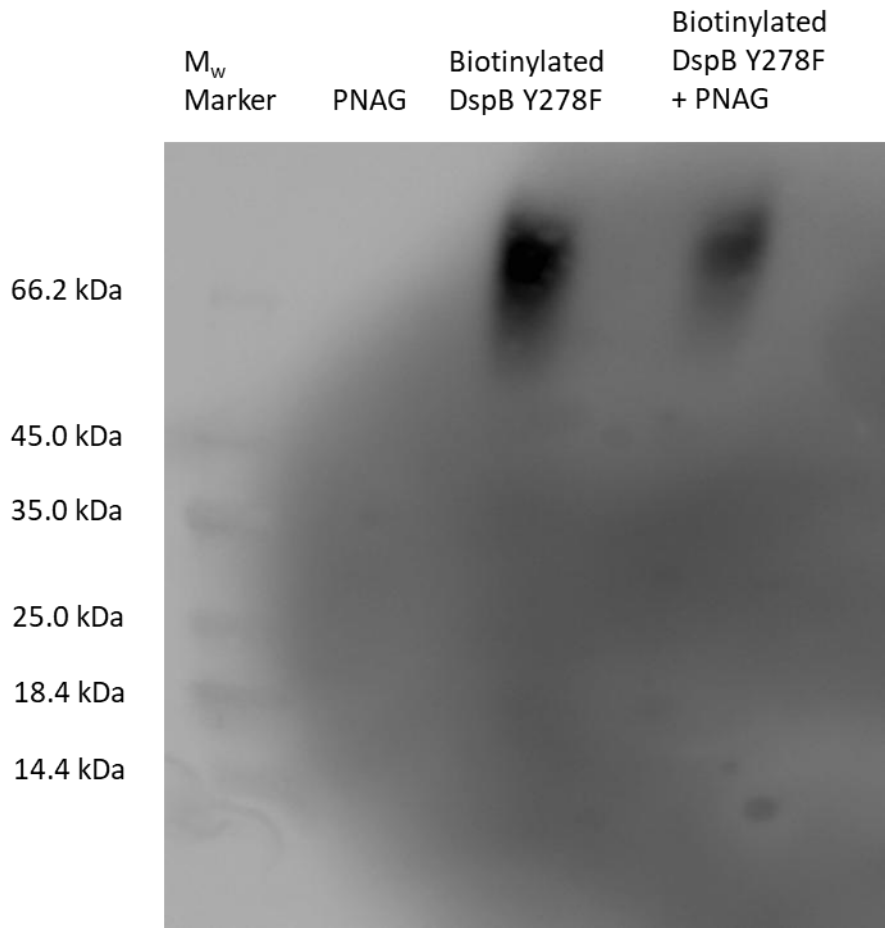



Figure 3.5.5: Biotin Western Blot of a 12% acrylamide native PAGE gel comparing PNAG and biotinylated DspB Y278F to a premixed sample of the two. Streptavidin-HRP used as the blotting reagent and 2% BSA solution used as the blocking solution

Interestingly, the band for the DspB mutant does not move from the loading gel, whilst the ladder does move from the loading gel, whilst having a similar $M_w$ to DspB. This perhaps suggests that this method is not the best for monitoring the binding of biotinylated DspB Y278F to PNAG, however, taking the results at face value suggest that the binding did not work. To test this, the experiment was repeated using a TSDS-PAGE gel instead (Figure 3.5.6)

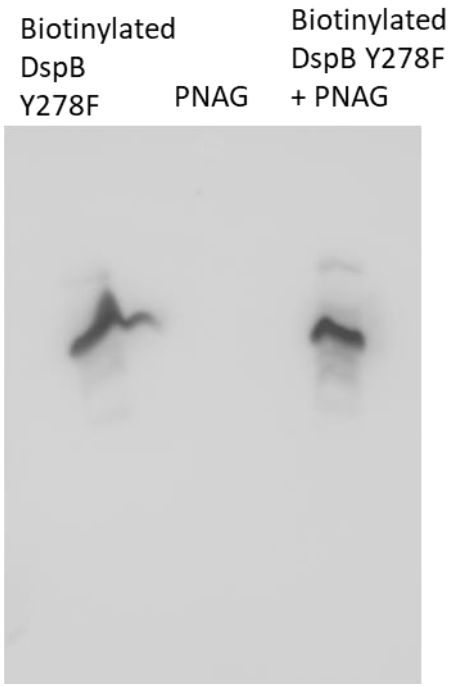



Figure 3.5.6: Biotin Western Blot analysis of 10% TSDS-PAGE gel comparing PNAG and biotinylated DspB Y278F to a premixed sample of the two. Streptavidin-HRP used as the blotting reagent and 2% BSA solution used as the blocking solution.

The TSDS-PAGE analysis once again shows that the DspB Y278F band remains in the same position after being premixed with PNAG, again suggesting that the biotinylated DspB Y278F mutant is ineffective as a PNAG binding reagent, and this was further confirmed by Ponceau Red staining of the Western Blot membrane (Figure 3.5.7).

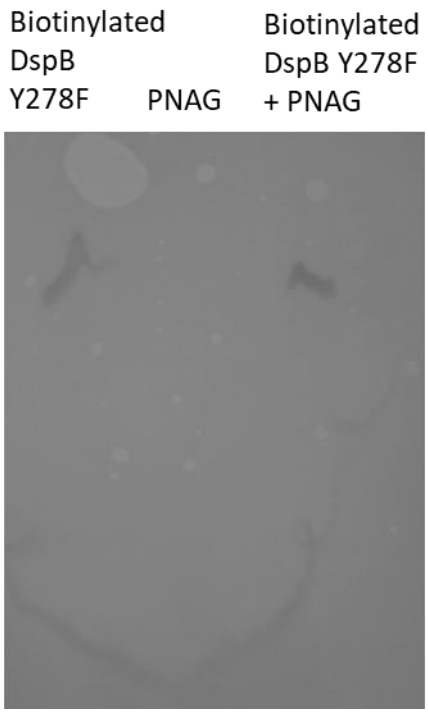



Figure 3.5.7: Ponceau Red stain of 10% TSDS-PAGE gel comparing PNAG and biotinylated DspB Y278F to a premixed sample of the two.

This once again confirmed that no extra bands were present in the Western Blot membrane adding to the mounting evidence that biotinylated DspB Y278F is an ineffective PNAG binding reagent. With these new pieces of evidence, it was concluded that the biotinylated DspB Y278F was not a viable PNAG blotting reagent, and this work was ceased.

### 3.5.4. Conclusions and Future Work

Biotinylated DspB Y278F was successfully produced, with Western Blot analysis confirming the presence of the biotin group on the protein. However, once this newly biotinylated protein was tested to see if it could function as a PNAG blot, it failed to produce any signal via a standard Dot Blot of GlcNAc and PNAG, or via Western Blot analysis of samples of PNAG and biotinylated DspB Y278F premixed together. This suggested that biotinylated DspB Y278F was incapable of acting as a PNAG specific blotting reagent. This contradicts the initial QCM-D studies performed by Connor Munroe that suggested that the PNAG binding ability was unaffected by the mutation of the Y278 residue. However, it should be noted that these studies were performed on the unbiotinylated DspB Y278F, whilst all subsequent experiments utilised the biotinylated mutant. This might suggest that the biotinylation is interfering with the binding in some way. Looking at the structure of DspB, there are several lysine residues in close proximity to the binding sites that are surface exposed (Figure 3.5.8).

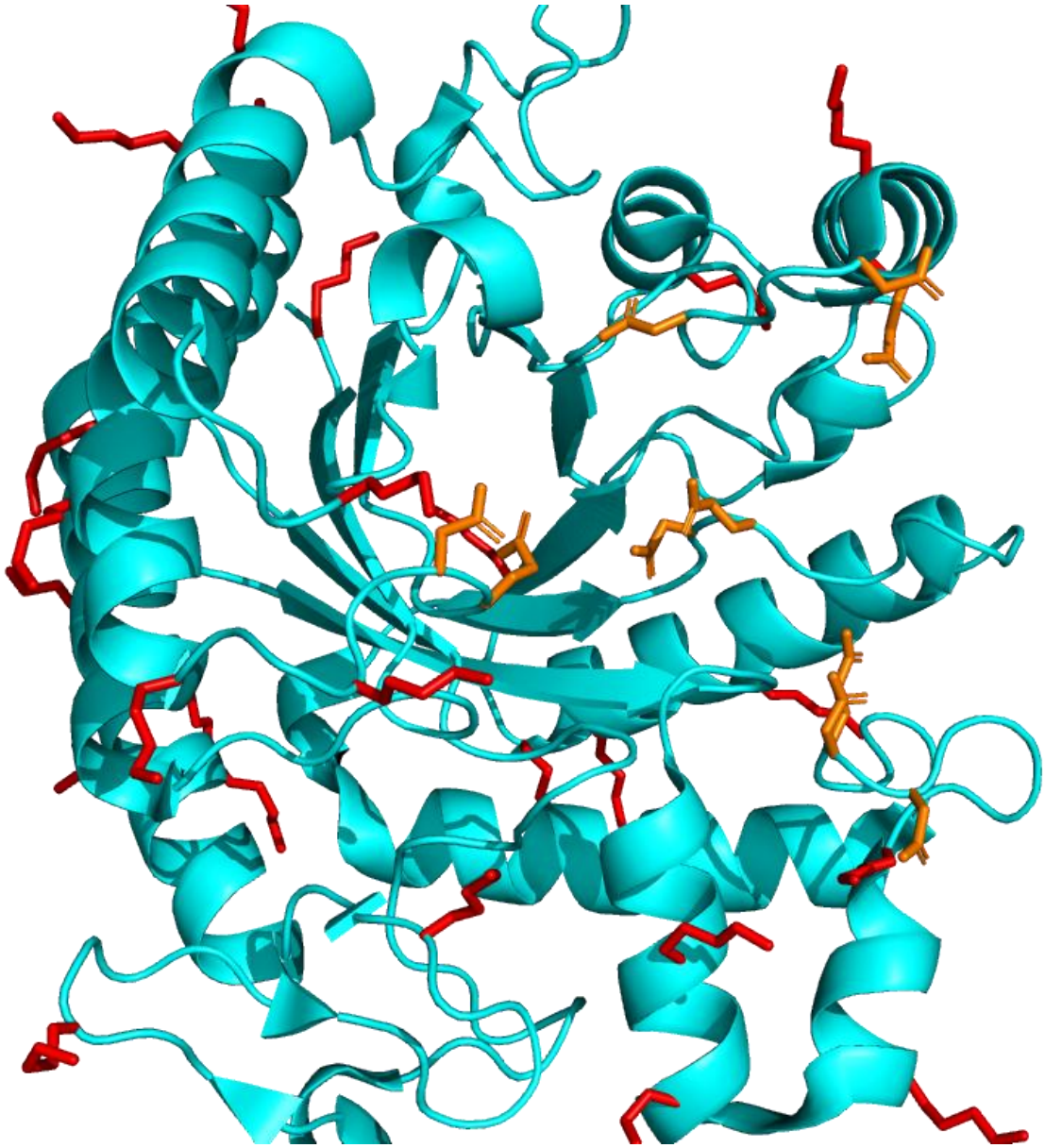


Figure 3.5.8: The structure of DspB showing the residues that make up the binding site (orange) versus the lysine residues found throughout the structure (red)

Given that the NHS-$PEG_4$-Biotin is a fairly large molecule, the close proximity of these lysine residues to the binding sites could mean that the biotinylation is causing steric hinderance of the binding to PNAG. Therefore, if this work were to be continued, the first thing to try would be a new QCM-D study using the biotinylated DspB Y278F to see if binding is still observed. Alternatively, a different mutant could be tried, such as Y278N which showed activity with the GlcNAc substrates in section 3.2.3, and so therefore, might be more likely to bind to PNAG successfully, but not as likely as the WT or D183A mutant to cause the PNAG to be hydrolysed. Alternatively, a mutant of residue E184 could be tried, as this residue is also involved in the mechanism of WT DspB, or a D183N mutant could be generated, as Manuel *et al*[119] show this mutant had the lowest $\frac{k_{cat}}{K_M}$ value of all the mutants they tested, and this mutant would also have three unmutated binding sites with

which to bind to PNAG. Finally, a different conjugate could be utilised to detect the DspB, such as the DspB-GFP conjugate utilised by Eddenden *et al*[223]. This conjugate has been shown to be able to bind to PNAG in a media, and therefore it could be utilised in the solid phase as well for blotting purposes. This conjugate also utilises a mutation of the aforementioned E184 residue, in this case E184Q, and could therefore be generated alongside a biotinylated version for comparison purposes. For this project however, the conclusion that can be drawn from the work presented here is that DspB Y278F cannot be used as a PNAG blotting reagent.

# 4. Conclusions and Future Prospectives

A greater understanding of poly-*N*-acetylglucosamine (PNAG), its structure, synthesis, and degradation would lead to the creation of several new avenues of research that could improve treatment outcome of PNAG presenting infections, as well as a better general understanding of the role large polysaccharides can play in biological systems. This project aimed to achieve this through several methods. Initially, the project aimed to create a new way to generate PNAG polysaccharides *in vitro* for further study via utilisation of the enzymatic chemistry developed by Mackenzie *et al*[1] to turn DspB, a PNAG degrading glycoside hydrolase, into a glycosynthase. To do this, several substrates were successfully synthesised, and four different mutants were generated to test for glycosynthase activity. Unfortunately, upon testing of these four mutants, none presented glycosynthase activity and so this work was concluded to be unsuccessful. Further mutants of DspB could be tested, or a different type of substrate, such as a disaccharide or larger polysaccharide. It is also possible that a different PNAGase could hold the key to unlocking a PNAG producing glycosynthase. When this project began, there were only two known PNAGase enzymes, the aforementioned DspB, and PgaB, however, thanks to the work by Males *et al*[118], five more PNAGases have been identified that could be used for this purpose.

Further to the work conducted by Males *et al*, this study found some preliminary evidence to suggest that some new enzymes identified in several *staphylococci* species could also act as PNAGases. Unfortunately, these enzymes proved difficult to overexpress and purify, and therefore the only evidence that was generated was quite weak, however, it is a promising first step, and the expression trials that followed were able to rule out a number of purification conditions that should make future attempts to purify these enzymes easier. This project has been taken over by Connor Munroe, but remains unpublished at the time of writing.

Attempts to synthesise probes to elucidate the function of IcaC were unsuccessful. Formation of multiple side products hampered the chosen synthetic for this project, leading to the work being handed over to Matthew Warnes in the Fascione group. At the time of writing, no further progress has been made on this front, and so the role of IcaC remains uncharacterised.

Upon realising that the generation of a PNAG producing glycosynthase had failed, the attention of the project turned to gaining further understanding into the structure of PNAG, and whether the structure of PNAG could be modified *in vitro* to unlock further avenues of study. Attempts were made to generate $N_3$-PNAG, dPNAG, and dsPNAG from PNAG that had been purified from *in vivo* sources. Initial evidence suggested that $N_3$-PNAG was successfully generated using

chemistry developed by Lim *et al*[2], however, further investigation into this provided no additional evidence for this. One possible explanation for this lack of evidence is that the CycloOct-GFP being used had degraded, so any future work on this area of the project should start by generating some fresh CycloOct-GFP and repeated the experiments detailed here. However, failing that, it is possible that an azide could be incorporated into PNAG utilising chemistry developed by Meng *et al*[217] to convert all the primary amines on the deacetylated monomers in the PNAG into azides. The attempted to generate dPNAG and dsPNAG for use in ELISA analysis were curtailed by contamination of PEG from the dialysis tubing used. SEC purification could remove this contamination, and then the ELISA could be performed as was initially planned.

Finally, efforts to convert DspB Y278F into a PNAG blotting reagent also met with failure. Whilst the initial biotinylation of DspB Y278F was successful, attempts to use this newly biotinylated protein to detect PNAG via Dot Blot analysis, and Western Blot analysis both produced no signal for the PNAG containing samples. This could be due to the choice of mutant, or it could be due to the choice to biotinylate all primary amines in the DspB structure causing steric hinderance. Any further attempts should initially try and utilise a different mutant, such as a mutant of E184, or should attempt to reproduce the results from Eddenden *et al* that utilise a GFP conjugate to monitor PNAG production *in vivo*[223].

In conclusion, the aims that were set at the outset of this project remain unmet. However, it is this author's hope that anyone who takes up this project from this point can use this body of work to inform and optimise any future experiments.

# 5. Experimental

## 5.1. General Methods

### 5.1.1. Lysogeny Broth (LB) Recipe

NaCl (1%% w/v), Tryptone (1% w/v), Yeast extract (0.5% w/v) in $dH_2O$, autoclaved.

### 5.1.2. LB Agar Recipe

NaCl (1% w/v), Tryptone (1% w/v), Yeast extract (0.5% w/v), Agar (1.5% w/v) in $dH_2O$, autoclaved.

### 5.1.3. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) procedure

SDS-PAGE Stacking Gel Buffer: SDS (0.4% w/v), 0.5 M Tris-HCl, $dH_2O$, pH 6.8.

SDS-PAGE Resolving Gel Buffer: SDS (0.4% w/v), 1.5 M Tris-HCl, $dH_2O$, pH 8.8.

SDS-PAGE gels were made using the following recipes:

Resolving Gel:

| **% Acrylamide** | **5** | **7.5** | **8.75** | **10** | **12** | **15** |
|---|---|---|---|---|---|---|
| **$dH_2O$ (mL)** | 5.7 | 4.9 | 4.4 | 4.0 | 3.2 | 2.4 |
| **SDS-PAGE Resolving Gel Buffer (mL)** | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |
| **Acrylamide/Bis Acrylamide (30% v/v) (mL)** | 1.7 | 2.5 | 2.9 | 3.3 | 4.2 | 5.0 |
| **20% w/v ammonium persulfate (APS)** | 50 µL | 50 µL | 50 µL | 50 µL | 50 µL | 50 µL |
| **Tetramethylethylenediamine (TEMED)** | 10 µL | 10 µL | 10 µL | 10 µL | 10 µL | 10 µL |

Recipe makes 2 gels, each gel is poured and allowed to set before the stacking gel was poured on top.

Stacking Gel: 3.2 mL $dH_2O$, 1.3 mL SDS Stacking Buffer, 0.5 mL Acrylamide (30% v/v), 25 µL 20% APS, 8 µL TEMED.

The stacking gel has either a 10-well comb or a 15-well comb placed into it before it can set.

5 x Reducing Buffer: SDS (10% w/v), Glycerol (20% w/v), Bromophenol Blue (0.05% w/v), 200 mM Tris-HCl, 10 mM β-mercaptoethanol, $dH_2O$, pH 6.8.

Running Buffer: 0.1% w/v SDS, 40 mM Glycine, 25 mM Tris-HCl.

5 x reducing buffer was added to each sample in a 1:5 dilution. Each sample was then denatured at 95 °C for 5 mins before being loaded into the stacking gel. The gel was then placed into the electrophoresis tank and submerged in running buffer, and then 30 mA was passed through the gel until the gel front reached the bottom of the gel. The gel was removed from the tank for either further processing or staining.

Staining Procedure:

Fixing Solution: 40% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Staining Solution: Coomassie Brilliant Blue R-250 (0.1% w/v), 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Destaining Solution: 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

The gel was placed in fixing solution for 1 hour. After 1 hour the fixing solution was decanted, and the gel is placed in staining solution for 1 hour. After 1 hour the staining solution was decanted, and the gel was placed in destaining solution until the protein bands are visible against the background. The gel was then imaged using G:Box imager from Syngene.

### 5.1.4. Western Blot Procedure

Transfer Buffer: 192 mM Glycine, 25 mM Tris-HCl, 20% MeOH, 80% $dH_2O$, pH 8.3.

A PAGE gel was run as described in sections 5.1.3, 5.1.6, or 5.1.7 until the fixing step. 6 layers of blotting paper are soaked in transfer buffer and placed in the transfer machine followed by a nitrocellulose membrane. The PAGE gel was removed from the electrophoresis tank and placed on top of the membrane, followed by another 6 layers of blotting paper soaked in transfer buffer. Protein transfer was then performed using a Power Blotter-Semi-dry transfer system from ThermoScientific™.

His-Tag Antibody Western Blot Procedure:

The membrane was removed and soaked in 5% w/v milk powder, 0.1% Tween ® 20 in 1 x Tris Buffered Saline (TBS) solution. The membrane was placed on a rocking platform at room temperature and allowed to incubate for 1 hour. After an hour the blocking solution was decanted, and the membrane was soaked in 1% milk powder in 1 x TBS with HRP-conjugated anti-His antibody

in a 1:2000 dilution. The membrane was incubated in this solution overnight. The blotting solution was then decanted, and the membrane was washed 3 times for 5 minutes in 0.1% Tween ® 20 in 1 x TBS on the rocking platform at room temperature. The membrane was then developed using Amersham ECL Prime Western Blotting Detection Reagent, and imaged using G:BOX imager from Syngene.

Biotin Western Blot Procedure:

The membrane was removed and soaked in 2% w/v BSA, 0.1% Tween ® 20 in 1 x TBS solution. The membrane was placed on a rocking platform at room temperature and allowed to incubate for 1 hour. After an hour the blocking solution was decanted, and the membrane was soaked in 1% BSA in 1 x TBS with Streptavidin-HRP in a 1:5000. The membrane was incubated in this solution overnight. The blotting solution was then decanted, and the membrane was washed 3 times for 5 minutes in 0.1% Tween ® 20 in 1 x TBS on the rocking platform at room temperature. The membrane was then developed using Amersham ECL Prime Western Blotting Detection Reagent, and imaged using G:BOX imager from Syngene.

### 5.1.5. Dot Blot Procedure

A section of nitrocellulose membrane was cut out and split into 1 cm x 1 cm blocks. Into each block 5 µL of each sample was loaded and allowed to dry. The membrane was then blocked overnight at 4 °C using the appropriate blocking solution. The blocking solution was discarded, and the membrane was then incubated with the appropriate blotting solution(s), washed with blocking solution, and then $dH_2O$. The membrane was then developed using Amersham ECL Prime Western Blotting Detection Reagent, and imaged using G:BOX imager from Syngene.

### 5.1.6. Tricine SDS-PAGE (TSDS-PAGE) Procedure

3 x TSDS-PAGE Gel Buffer: SDS (0.3% w/v), 3 M Tris-HCl, $dH_2O$, pH 8.45.

TSDS-PAGE gels were made using the following recipes:

Resolving Gel (15% w/v acrylamide): 2.62 mL $dH_2O$, 10 mL 3 x TSDS-PAGE Gel Buffer, 3 g Glycerol, 15 mL Acrylamide (30% v/v), 150 µL APS, 15 µL TEMED.

Recipe makes 2 gels, each gel was poured and allowed to set before the stacking gel was poured on top.

Stacking Gel (6% w/v acrylamide): 5.6 mL $dH_2O$, 4 mL 3 x TSDS-PAGE Gel Buffer, 2.4 mL Acrylamide (30% v/v), 110 µL APS, 11 µL TEMED.

The stacking gel has either a 10-well comb or a 15-well comb placed into it before it can set.

4 x Sample Buffer: SDS (8% w/v), Glycerol (40% w/v), Bromophenol Blue (0.02% w/v), 250 mM Tris-HCl, β-mercaptoethanol (3% v/v), $dH_2O$, pH 6.8.

Anode Buffer: 0.1 M Tris-HCl, $dH_2O$, pH 8.9.

Cathode Buffer: 0.1 M Tris-HCl, 0.1 M Tricine, $dH_2O$, pH 8.25.

4 x sample buffer was added to each sample in a 1:4 dilution. Each sample was then heated to 45 °C for 1 hour before being loaded into the stacking gel. The gel was then placed into the electrophoresis tank, and cathode buffer was poured into the reservoir in the tank. The tank was then submerged in running buffer, and then 30 mA was passed through the gel until the gel front reached the interface between the stacking and resolving gel. Then 70 mA was passed through the gel until the gel front reaches the bottom of the gel. The gel was removed from the tank for either further processing or staining.

Staining Procedure:

Fixing Solution: 40% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Staining Solution: Coomassie Brilliant Blue R-250 (0.1% w/v), 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Destaining Solution: 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

The gel placed in fixing solution for 1 hour. After 1 hour the fixing solution was decanted, and the gel was placed in staining solution for 1 hour. After 1 hour the staining solution was decanted, and the gel was placed in destaining solution until the protein bands are visible against the background. The gel was then imaged using G:Box imager from Syngene.

### 5.1.7. Native PAGE Procedure.

Native PAGE gels were made using the following recipes:

Resolving Gel:

| % Acrylamide | 6 | 8 | 10 | 12 | 15 |
|---|---|---|---|---|---|
| **0.375M Tris-HCl, pH 8.8 (mL)** | 7.89 | 7.29 | 6.49 | 5.89 | 4.89 |
| **Acrylamide/Bis Acrylamide (30% v/v) (mL)** | 2 | 2.6 | 3.4 | 4 | 5 |
| **20% w/v APS** | 100 µL | 100 µL | 100 µL | 100 µL | 100 µL |
| **TEMED** | 10 µL | 10 µL | 10 µL | 10 µL | 10 µL |

Recipe makes 2 gels, each gel was poured and allowed to set before the stacking gel was poured on top.

Stacking Gel: 4.275 mL 0.375 M Tris-HCl pH 8.8, 0.67 mL Acrylamide (30% v/v), 50 µL 20% APS, 5 µL TEMED.

The stacking gel has either a 10-well comb or a 15-well comb placed into it before it can set.

2 x Sample Buffer: Glycerol (25% w/v), Bromophenol Blue (1% w/v), 62.5 mM Tris-HCl, $dH_2O$, pH 6.8.

Running Buffer: 192 mM Glycine, 25 mM Tris-HCl

2 x sample buffer was added to each sample in a 1:2 dilution. Each sample was then loaded into the stacking gel. The gel was then placed into the electrophoresis tank and submerged in running buffer. The tank was placed on ice and then 30 mA was passed through the gel for until the gel front reached the bottom of the gel. The gel was removed from the tank for either further processing or staining.

Staining Procedure:

Fixing Solution: 40% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Staining Solution: Coomassie Brilliant Blue R-250 (0.1% w/v), 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

Destaining Solution: 50% EtOH, 10% glacial AcOH, 40% $dH_2O$.

The gel was placed in fixing solution for 1 hour. After 1 hour the fixing solution was decanted, and the gel was placed in staining solution for 1 hour. After 1 hour the staining solution was decanted, and the gel was placed in destaining solution until the protein bands are visible against the background. The gel was then imaged using G:Box imager from Syngene.

### 5.1.8. Size Exclusion Chromatography (SEC) Procedure

Sephadex LH20 SEC:

The column was packed with Sephadex LH20 resin in MeOH. A sample of the compound in MeOH was then applied to the top of the resin and allowed to flow into the resin. The column was then left overnight under MeOH at 30 mL/hr. Fractions were collected and analysed via TLC and LC-MS analysis. Fractions containing the desired compound were collated and concentrated *in vacuo*.

S200 SEC:

The S200 column was preequilibrated with the appropriate buffer. The sample was loaded into the sample loop and injected into the column. The sample is then eluted through the column by application of the appropriate buffer. The sample was detected by UV/Vis absorbance, and fractions containing product were taken for further analysis by SDS-PAGE, dot blot, or assay.

### 5.1.9. General LC-MS Procedure

High Performance Liquid Chromatography-Electrospray Ionisation Mass Spectrometry (LC-MS) was accomplished using a Dionex UltiMate® 3000 LC system (ThermoScientific) equipped with an UltiMate® 3000 Diode Array Detector (probing 250-400 nm) in line with a Bruker HCTultra ETD II system (Bruker Daltonics), using Chromeleon® 6.80 SR12 software (ThermoScientific), Compass 1.3 for esquire HCT Build 581.3, 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) at The University York Centre of Excellence in Mass Spectrometry (CoEMS). Data analysis was performed using ESI Compass 1.3 DataAnalysis, Version 4.1 software (Bruker Daltonics).

For HILIC-LC-MS, analyses were carried out using a SeQuant ZIC®-HILIC HPLC column. Water, 0.1% formic acid by volume (solvent A), and acetonitrile, 0.1% formic acid (solvent B) were used as the mobile phase at a flow rate of 0.3 mL/min at room temperature. A multi-step gradient of 9 min was programmed as follows: 95% B for 0.5 min, followed by a linear gradient to 5% B over 4.5 min, followed by 5% B for an additional 0.5 min. A linear gradient to 95% B was used to re-equilibrate the column.

For analyses carried out by reverse-phased LC-MS, samples were analysed using a Symmetry C18 5 µm 3.0 x 150 mm reverse-phase column. Water + 0.1% formic acid by volume (solvent A) and acetonitrile + 0.1% formic acid (solvent B) were used as a mobile phase at a flow rate of 300 µL $min^{-1}$ at room temperature. A multi-step gradient of 7.5 min was programmed as follows: 95% A for 1.0 min, followed by a linear gradient to 95% B over 6.5 min, followed by 95% B for an additional 1.0 min. A linear gradient to 95% A was used to equilibrate the column.

### 5.1.10. Plate Reader Assay General Procedure

Plate Reader Assays were obtained using a CLARIOstar Plate Reader. Samples were loaded into a 96-well plate and incubated at the appropriate temperature. UV/Vis absorbance measurements were made over the course of the reaction at the appropriate wavelength.

### 5.1.11. Nuclear Magnetic Resonance Spectroscopy (NMR) Procedure

$^{1}$H 400 MHz Nuclear Magnetic Resonance Spectroscopy data was obtained on a JOEL 400. Chemical shifts are reported in ppm using the following references: $D_2O$: δ: 4.80; $(CD_3)_2CO$: δ: 2.05; $CDCl_3$: δ: 7.26; DMSO-$d_6$: δ: 2.50; MeOD: δ: 3.31.

$^{13}$C 101 MHz Nuclear Magnetic Resonance Spectroscopy data was obtained on a JOEL 400. Chemical shifts are reported in ppm using the following references: $(CD_3)_2CO$: δ: 29.00; $CDCl_3$: δ: 77.16; DMSO-$d_6$: δ: 40.00; MeOD: δ: 49.00.

$^{1}$H 500 MHz Nuclear Magnetic Resonance Spectroscopy data was obtained on a Bruker AVIIIHD500 FT-NMR Spectrometer. Chemical shifts are reported in ppm using the following references: $CDCl_3$: δ: 7.26.

$^{13}$C 126 MHz Nuclear Magnetic Resonance Spectroscopy data was obtained on a Bruker AVIIIHD500 FT-NMR Spectrometer. Chemical shifts are reported in ppm using the following references: $CDCl_3$: δ: 77.16.

$^{1}$H 700 MHz Nuclear Magnetic Resonance Spectroscopy data was obtained on a Bruker Avance Neo 700 MHz Spectrometer. Chemical shifts are reported in ppm using the following references: $D_2O$: δ: 4.80.

Assignments were made using COSY, HSQC, and DEPT-135 experiments. The following abbreviations were used to describe signal multiplicities or appearances: singlet: s; doublet: d; doublet of doublets: dd; doublet of doublet of doublets: ddd; triplet: t; doublet of triplets: dt; doublet of doublet of triplets: ddt; triplet of doublets: td; quartet: q; quartet of doublets: qd; pentet: p; multiplet: m.

### 5.1.12. High Resolution Mass Spectroscopy (HRMS) Procedure

Small-molecule HRMS data were obtained at room temperature on a Bruker Daltonics micrOTOF at The University York Centre of Excellence in Mass Spectrometry (CoEMS).

### 5.1.13. Fourier Transformed Infra-Red Spectroscopy (FT-IR) Procedure

Infra-Red Spectroscopy data were obtained at room temperature on a PerkinElmer UATR 2 spectrometer.

### 5.1.14. Optical Rotation Spectroscopy ($[\alpha]_D$) Procedure

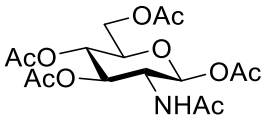
Optical Rotation Spectroscopy data were obtained at room temperature on a Bellingham and Stanley ADP 450 Automatic Digital Peltier Controlled Polarimeter equipped with a 589 nm LED.

### 5.1.15. Ultraviolet-Visible Spectroscopy (UV/Vis) Procedure

Ultraviolet-Visible Spectroscopy data were obtained at room temperature on a DeNovix nanodrop photospectrometer or a CLARIOStar Plate Reader.

## 5.2. **Chemical Synthesis**

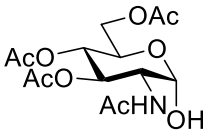
### 5.2.1. 2-acetamido-1,3,4,6-tetra-O-acetyl-2-deoxy-D-glucopyranose (**2**)




**2**

Commercially sourced glucosamine hydrochloride **1** (2.00 g, 9.27 mmol) was dissolved in pyridine (20 mL) under an inert atmosphere. A catalytic amount of DMAP was added to the solution, followed by dropwise addition of acetic anhydride (20 mL, excess). The solution was left to stir overnight at room temperature. The reaction mixture was concentrated *in vacuo*, and the resulting solid was redissolved in DCM (20 mL). The solution was washed with 20 mL aliquots of water (2x), 1M HCl (2x), sat. aq. $NaHCO_3$ solution (2x), and brine (1x). The organic layer was dried over $MgSO_4$ and filtered. The filtrate was then concentrated *in vacuo* giving monosaccharide **2** as a colourless crystal (1.91 g, 4.90 mmol, 53%). $^1$H NMR ($CDCl_3$, 500 MHz) δ: 6.20 (d, *J* = 3.63 Hz, 1H, H1), 5.55 (d, *J* = 9.00 Hz, 1H, NH), 5.30-5.20 (m, 2H, H3, H4), 4.51 (ddd, *J* = 3.63, 9.00, 10.65 Hz, 1H, H2), 4.28 (dd, *J* = 4.12, 12.49 Hz, 1H, H6'), 4.09 (dd, *J* = 2.29, 12.49 Hz, 1H, H6), 4.02 (ddd, *J* = 2.29, 4.12, 9.76 Hz, 1H, H5), 2.22 (s, 3H, $CH_3$), 2.18-2.01 (m, 9H, 3 $CH_3$), 1.97 (s, 3H, $CH_3$). $^{13}$C NMR (101 MHz, $CDCl_3$) δ: 170.83 (C=O), 170.80 (C=O), 170.05 (C=O), 169.19 (C=O), 168.72 (C=O), 90.77 (C1), 70.75 (C3), 69.78 (C5), 67.53 (C4), 61.60 (C6), 51.13 (C2), 23.13 ($CH_3$), 21.02 ($CH_3$), 20.80 ($CH_3$), 20.77 ($CH_3$), 20.64 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$: 3438.57 (N-H), 1738.63 (C=O), 1673.88 (C-N). $[\alpha]_D$ 96.22 – 96.54 (c 1.00 in DCM). HRMS (ESI): m/z calculated for $C_{16}H_{24}NaNO_{10}$ [(M+Na$^+$)]: 412.1214, found: 412.1220

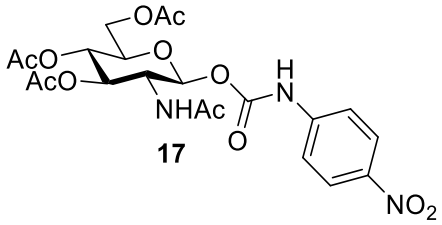
### 5.2.2. 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-glucopyranose (**16**)




**16**

DMAPA (3.578 mL, 28.11 mmol, 5 eq.) was added to a solution of monosaccharide **2** (2.127 g, 5.62 mmol) in THF (28 mL). The solution was left stirring at room temperature and was monitored by TLC. Once the reaction had proceeded to completion, the solution was diluted with $CHCl_3$ (28 mL), and washed with 1M HCl (56 mL), and brine (56 mL). The organic layer was dried over $MgSO_4$ and filtered. The filtrate was concentrated *in vacuo* giving hemiacetal **16** as a brown oil (1.270 g, 3.66 mmol, 65%). $^1$H NMR ($CDCl_3$, 400 MHz) δ: 5.94 (d, *J* = 9.42 Hz, 1H, NH), 5.27 (dd, *J* = 9.63, 10.83 Hz, 1H, H3), 5.23 (d, *J* = 3.54 Hz, 1H, H1), 5.11 (t, *J* = 9.63 Hz, 1H, H4), 4.27 (ddd, *J* = 3.54, 9.42, 10.83 Hz, 1H, H2), 4.23-4.15 (m, 1H, H6'), 4.15-4.04 (m, 1H, H6), 3.77-3.68 (m, 1H, H5), 2.07 (s, 3H, $CH_3$), 2.02-2.00 (d, *J* = 3.09 Hz, 6H, $CH_3$), 1.95 (s, 3H, $CH_3$). $^{13}$C NMR ($CDCl_3$, 101 MHz) δ: 171.56 (C=O), 171.04 (C=O), 170.58 (C=O), 169.52 (C=O), 91.65 (C1), 71.04 (C3), 68.30 (C4), 67.61 (C5), 62.18 (C6), 52.39 (C2), 23.24 ($CH_3$), 20.86 ($CH_3$), 20.84 ($CH_3$), 20.76 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$: 3346.47 (O-H), 1737.57 (C=O), 1658.64 (C-N). $[\alpha]_D$ 243.00– 251.78 (c 0.56 in DCM). HRMS (ESI): m/z calculated for $C_{14}H_{21}NaNO_9$ [(M+Na$^+$)]: 370.1109, found: 370.1113.

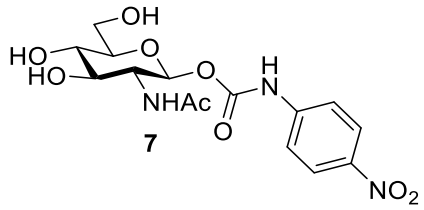
### 5.2.3. 4-nitrophenylcarbomyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-d-glucopyranoside (**17**)



**17**

Hemiacetal **16** (296 mg, 0.85 mmol) was dissolved in toluene (30 mL) under inert conditions. TEA (85.3 µL) was added to this solution, followed by addition of p-nitrophenyl isocyanate (280 mg, 1.7 mmol, 2 eq.). The reaction was left stirring at 35°C for 6 hours, and then filtered by gravity. The precipitate was then resuspended in 35 °C toluene and re-filtered by gravity. This was repeated until $^1$H NMR showed little to no α presence, leaving carbamate **17** as a yellow crystal (215 mg, 0.42 mmol, 49%). $^1$H NMR ($CDCl_3$, 400 MHz) δ: 8.19 (d, *J* = 9.12 Hz, 2H, ArH), 7.90 (s, 1H, O=C-NH-Ar), 7.56 (d, *J* = 9.12 Hz, 2H, ArH), 5.97 (d, *J* = 9.49 Hz, 1H, NH), 5.68 (d, *J* = 8.78 Hz, 1H, H1), 5.24 – 5.12 (m, 2H, H3, H4), 4.30 (q, *J* = 9.51 Hz, 1H, H2), 4.30 (dd, *J* = 2.25, 12.47 Hz, 1H, H6'), 4.15 (dd, *J* = 2.25, 12.47 Hz, 1H, H6), 3.92 – 3.84 (m, 1H, H5), 2.07 (s, 3H, $CH_3$), 2.06 (s, 6H, 2 x $CH_3$), 1.96 (s, 3H, $CH_3$). $^{13}$C NMR ($CDCl_3$, 101 MHz) δ: 171.45 (C=O), 170.88 (C=O), 170.80 (C=O), 169.47 (C=O), 151.04 (C=O), 143.66 (ArC-NO), 143.25 (ArC-NH), 125.33 (ArC-H), 118.34 (ArC-H), 94.26 (C1), 73.01 (C5), 72.58 (C3), 67.85 (C4), 61.78 (C6), 52.94 (C2), 23.48 ($CH_3$), 20.84 ($CH_3$), 20.78 ($CH_3$), 20.74 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$: 3367.01 (N-H), 1732.57 (C=O), 1544.33 (N-O). HRMS (ESI): m/z calculated for $C_{21}H_{25}NaN_3O_{12}$ [(M+Na$^+$)]: 534.1330, found: 534.1328.

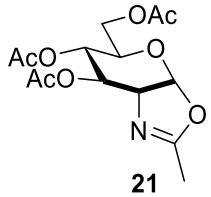
5.2.4. 4-nitrophenylcarbomyl-2-acetamido-2-deoxy-β-D-glucopyranoside (**7**)




**7**

A suspension of carbamate **17** (100 mg, 0.19 mmol) was made in MeOD (1 mL). Hydrazine monohydrate (31.7 µL, 0.65 mmol, 3.3 eq) was added to the suspension and left at room temperature for 5 hours. The reaction was monitored by $^{1}$H NMR every 30 minutes. After the reaction had proceeded to completion, the reaction mixture was concentrated *in vacuo*, and the resulting solid was purified by Sephadex LH20 Size Exclusion Chromatography giving carbamate **7** as a yellow solid (32 mg, 0.08 mmol, 42%). $^{1}$H NMR (MeOD, 400 MHz) δ: 8.21 (d, *J* = 9.15 Hz, 2H, ArH), 7.67 (d, *J* = 9.15 Hz, 2H, ArH), 5.59 (dd, *J* = 3.42, 9.06 Hz, 1H, H1), 3.99 – 3.83 (m, 2H, H2, H6'), 3.72 (d, *J* = 12.36 Hz, 1H, H6), 3.51 (d, *J* = 10.38 Hz, 1H, H3), 3.44 – 3.39 (m, 2H, H4, H5), 1.98 (d, *J* = 3.25, 3H, $CH_3$). $^{13}$C NMR (MeOD, 126 MHz) δ: 174.07 (C=O), 153.54 (C=O), 144.27 (ArC-NO), 127.13 (ArC-NH), 125.92 (ArC), 119.18 (ArC), 95.45 (C1), 78.81 (C5), 75.97 (C3), 71.54 (C4), 62.35 (C6), 55.97 (C2), 22.94 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$: 3323.06 (O-H), 1722.30 (C=O), 1644.59 (C-N), 1545.53 (N-O). HRMS (ESI): m/z calculated for $C_{15}H_{18}N_3O_9$ [($M^-$)]: 384.1049, found: 384.1047.

5.2.5. 2-methyl-(3,4,6-tri-O-acetyl-1,2-dideoxy-α-D-glucopyranoso)[1,2-d]-2-oxazoline (**21**)

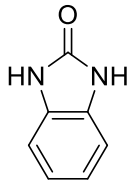



**21**

Monosaccharide **2** (1.002 g, 2.57 mmol) was dissolved in dry DCM (25 mL) under $N_2$. Solution was cooled to 0 °C and TMSOTf (1.5 mL, 8.24 mmol, 3.2 eq.) was added carefully. The solution was then heated to reflux and monitored by TLC until all starting material had reacted. Once all starting material had reacted, the reaction was quenched by addition of TEA until the solution became basic. The reaction mixture was concentrated *in vacuo* and purified by flash column chromatography (3:1 EtOAc:Hex + 1% TEA) giving oxazoline **21** was a brown oil (418 mg, 1.27 mmol, 49%). $^{1}$H NMR (DMSO-$d_6$, 400 MHz) δ: 6.02 (dd, *J* = 1.58, 7.36 Hz, 1H, H1), 5.03 (q, *J* = 2.13 Hz, 1H, H3), 4.76 (dq, *J* = 2.13, 9.23 Hz, 1H, H4), 4.14 – 4.02 (m, 3H, H2, H6, H6') 3.59 – 3.53 (m, 1H, H5), 2.06 – 1.90 (m, 12H, 4 x $CH_3$). $^{13}$C NMR (DMSO-$d_6$, 101 MHz) δ: 170.25 (C=O), 169.48 (C=O),

169.18 (C=O), 165.72 (N=C-O), 98.69 (C1), 69.92 (C3), 68.25 (C4), 66.75 (C5), 64.26 (C2), 63.00 (C6), 20.82 ($CH_3$), 20.71 ($CH_3$), 20.63 ($CH_3$), 13.60 ($CH_3$). $[\alpha]_D$ 11.66– 12.81 (c 0.65 in DCM). FT-IR $\nu_{max}/cm^{-1}$: 1740.93 (C=O), 1672.20 (C=N), 1220.71 (C-O), 1027.65 (C-O). HRMS (ESI): m/z calculated for $C_{14}H_{20}NO_8$ [(M+H$^+$)]: 330.1183, found: 330.1177. m/z calculated for $C_{14}H_{19}NO_8Na$ [(M+Na$^+$)]: 352.1003, found: 352.0997.

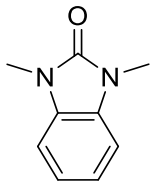
### 5.2.6. 1,3-dihydrobenzimidazol-2-one (**26**)



**26**

*o*-phenylenediamine **25** (25 g, 231.18 mmol) was dissolved in ethylene glycol (120 mL) and heated to 135 °C. Urea (16.697 g, 278.01 mmol) was added to the solution and the reaction was left stirring for 24 hr. After 24 hr another addition of urea (1.387 g, 23.10 mmol) was made and the reaction left for another 24 hr. The precipitate that formed was filtered by vacuum and washed with toluene (100 mL) and $H_2O$ (100 mL) giving diamide **26** as a dark brown crystal (27.236 g, 203.18 mmol, 88%). $^1$H NMR (DMSO-$d_6$, 400 MHz) δ: 10.57 (s, 2H, NH), 6.91 (s, 4H, ArH). $^{13}$C NMR (DMSO-$d_6$, 101 MHz) δ: 155.34 (C=O), 129.70 (ArC), 120.48 (ArC), 108.55 (ArC). FT-IR $\nu_{max}/cm^{-1}$: 3137.80 (O-H), 1630.01 (N-H), 1196.15 (C-N). HRMS (ESI): m/z calculated for $C_7H_7N_2O$ [(M+H$^+$)]: 135.0553, found: 135.0543. m/z calculated for $C_{14}H_{12}N_4NaO_2$ [(2M+Na$^+$)]: 291.0852, found 291.0858.

### 5.2.7. 1,3-dimethylbenzimidazol-2-one (**24**)

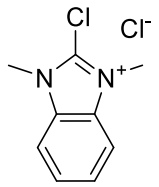


**24**

Diamide **26** (12 g, 89.52 mmol) was dissolved in toluene (60 mL). To this solution 40% w/v NaOH (22.920 mL, 229.22 mmol) and $Bu_4NBr$ (1.431 g, 4.44 mmol) was added, followed by MeI (12.24 mL, 196.61 mmol) dropwise and the solution was heated to 60 °C with stirring for 48 hr. After 48 hr, the reaction mixture was quenched via addition of 1M HCl (100 mL). The organic layer was collected, washed with sat. $NaHCO_3$ (100 mL), dried over $MgSO_4$, and concentrated *in vacuo*. The resulting solid was recrystallised from acetone and hexane to give diamide **24** as a light brown crystal (5.680 g, 35.04 mmol, 39%). $^1$H NMR (DMSO-$d_6$, 400 MHz) δ: 7.08 (m, 4H, ArH), 3.31 (s, 6H,

2 x $CH_3$). $^{13}C$ NMR (DMSO-$d_6$) δ: 153.81 (C=O), 129.64 (ArC), 120.85 (ArC), 107.55 (ArC), 26.88 ($CH_3$). FT-IR $\nu_{max}/cm^{-1}$: 3067.24 (C-H), 1694.71 (C=O). HRMS (ESI): m/z calculated for $C_9H_{10}N_2NaO$ [(M+$Na^+$)]: 185.0685, found: 185.0688.

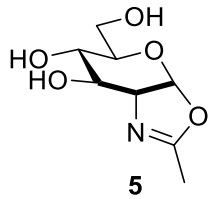
### 5.2.8. 2-chloro-1,3-dimethyl-1*H*-benzimidazol-3-ium (**23**)




**23**

Diamide **24** (4.156 g, 25.64 mmol) was dissolved in toluene (35 mL) under $N_2$. Oxalyl chloride (5.462 mL, 62.69 mmol) was added dropwise, and the reaction mixture heated to 80 °C with stirring. After 48 hr an additional aliquot of oxalyl chloride (2.038 mL, 39.51 mmol) was added and the reaction left for another 72 hr. After 72 hr the precipitate was isolated via vacuum filtration and washed with toluene (50 mL) giving CDMBI **23** as a beige solid (1.630 g, 10.05 mmol, 29%). $^1H$ NMR ($D_2O$, 400 MHz) δ: 7.83 (m, 2H, ArH), 7.70 (m, 2H, ArH), 4.06 (s, 6H, $CH_3$). $^{13}C$ NMR ($D_2O$, 101 MHz) δ: 140.55 (C=O), 131.31 (ArC), 127.19 (ArC), 112.62 (ArC) 32.38 ($CH_3$). FT-IR $\nu_{max}/cm^{-1}$: 761.08 (C-Cl). HRMS (ESI): m/z calculated for $C_9H_{10}ClN_2$ [(M-$Cl^-$)]: 181.0527, found: 181.0530.

### 5.2.9. 2-methyl-(1,2-dideoxy-α-D-glucopyranoso)-[2,1-d]-2-oxazoline (**5**)

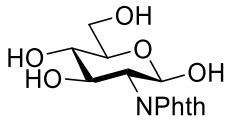



**5**

Commercially sourced N-acetyl-D-glucosamine **1** (100 mg, 0.45 mmol) was dissolved in $H_2O$ (4.520 mL). CDMBI **23** (293 mg, 1.36 mmol, 3 eq.) was added to the solution, and the solution was cooled to 0 °C. A solution of 3M $Na_3PO_4$ in $H_2O$ (11.301 mL, 3.39 mmol, 7.5 eq.) was added and the mixture was allowed to stir for 1 hr. The reaction mixture was centrifuged at 4000 xg for 5 minutes and the supernatant was decanted. The supernatant was purified by reverse phase column chromatography with a graduated solvent system of $H_2O$ to MeCN yielding oxazoline **5** as a colourless solid (51 mg, 0.25 mmol, 55%). $^1H$ NMR ($D_2O$, 400 MHz) δ: 6.01 (dd, *J* = 1.51, 7.38 Hz, 1H, H1), 4.04 (ddt, *J* = 1.51, 3.59, 7.41 Hz, 1H, H2), 3.90 (td, *J* = 1.42, 3.59 Hz, 1H, H3), 3.85 – 3.61 (m, 1H, H4), 3.65 – 3.40 (m, 2H, H6, H6'), 3.42 – 3.24 (m, 1H, H5), 1.97 (dt, *J* = 1.80, 6.50 Hz, 3H, $CH_3$). $^{13}C$ NMR ($D_2O$, 101 MHz) δ: 163.68 (N=CO), 100.48 (C1), 73.02 (C5), 71.77 (C3), 68.81 (C6), 66.04 (C2), 61.79 (C4), 13.20 ($CH_3$). FT-IR $\nu_{max}/cm^{-1}$: 3246.85 (O-H), 2924.32 (C-H), 1663.78 (C=N). $[\alpha]_D$ -

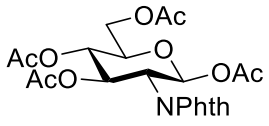
9.70 - - 15.24 (c 0.05 in DCM). HRMS (APCI): m/z calculated for $C_8H_{14}NO_5$ [(M+H$^+$)]: 204.0866, found: 204.0867.

5.2.10. 2-deoxy-2-(1,3-dioxo-1,3-dihydro-2H-isoindol-2-yl)-β-D-glucopyranose (**28**) + 1,3,4,6-tetra-*O*-acetyl-2-deoxy-2-phthalimido-β-D-glucopyranose (**29**)




**28**

Commercially sourced glucosamine hydrochloride **3** (5.000 g, 23.19 mmol) was dissolved in a 1:2 $H_2O$:MeOH mix (30 mL). NaOH (1.200 g, 30.00 mmol, 1.29 eq) was added slowly until it dissolved into a clear solution, to which a solution of phthalic anhydride (7.109 g, 48.00 mmol, 2.07 eq.) in acetone (40 mL) was added, and then $NaHCO_3$ (5.000 g, 59.52 mmol, 2.57 eq.). The solution was heated to 50 °C for 1 hour giving a pale yellow solution which was allowed to cool to room temperature and left overnight with stirring. The reaction mixture was acidified to pH 2, concentrated in vacuo, and left at 4 °C overnight, yielding phthalimido glycoside **28** as a colourless solid. The resulting solid was used directly in the next reaction step. $^1$H NMR (DMSO-$d_6$, 400 MHz) δ: 8.01 (d, *J* = 8.20 Hz, 1H, ArH), 7.76 – 7.69 (m, 1H, ArH), 7.62 – 7.45 (m, 2H, ArH), 6.37 (d, *J* = 4.27 Hz, 1H, OH), 5.07 (t, *J* = 3.75 Hz, 1H, H1), 4.94 (d, *J* = 5.47 Hz, 1H, OH), 4.53 (s, 1H, OH), 4.45 (s, 1H, OH), 3.79 – 3.47 (m, 5H, H3, H4, H5, H6, H6'), 3.18 (td, *J* = 4.88, 9.19, 9.24 Hz, 1H, H2). $^{13}$C NMR (DMSO-$d_6$, 101 MHz) δ: 168.86 (C=O), 138.61 (ArC), 131.43 (ArC), 129.75 (ArC), 129.48 (ArC), 128.57 (ArC), 91.11 (C1), 72.67 (C3), 71.37 (C5), 71.01 (C4), 61.67 (C6), 55.54 (C2). FT-IR $\nu_{max}$/cm$^{-1}$:. $[\alpha]_D$ -34.63 - -49.86 (c 0.05 in DCM). HRMS (ESI): m/z calculated for $C_{14}H_{14}NO_7$ [(M-H)]: 308.0776, found: 308.0769.

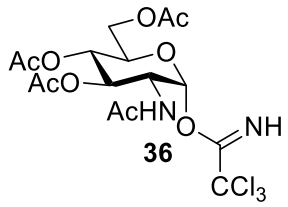



**29**

Phthalimido glycoside **28** was dissolved in pyridine (70 mL) along with a catalytic amount of DMAP. Acetic anhydride (52 mL) was added slowly to this solution, and the reaction mixture was left overnight at room temperature with stirring. The reaction mixture was concentrated in vacuo, and redissolved in DCM (50 mL), and washed with $H_2O$ (2x 50 mL), 1M HCl (2x 50 mL), sat. $NaHCO_3$ (2x 50 mL), and brine (1x 50 mL). The organic layer was collected, dried over $MgSO_4$, and concentrated in vacuo yielding phthalimido glycoside **29** as a pale red solid (3.196 g, 6.70 mmol, 29%). $^1$H NMR ($CDCl_3$, 400 MHz) δ: 7.85 (dd, *J* = 3.08, 5.49 Hz, 2H, ArH), 7.74 (dd, *J* = 3.08, 5.49 Hz,

2H, ArH), 6.56 (dd, *J* = 9.19, 11.58 Hz, 1H, H3), 6.29 (d, *J* = 3.39 Hz, 1H, H1), 5.17 (dd, *J* = 9.19, 10.24 Hz, 1H, H4), 4.72 (dd, *J* = 3.39, 11.58 Hz, 1H, H2), 4.41 – 4.27 (m, 2H, H6, H6'), 4.14 (dd, *J* = 2.06, 10.24 Hz, 1H, H5), 2.23 (s, 3H, $CH_3$), 2.12 (s, 3H, $CH_3$), 2.09 (s, 3H, $CH_3$), 2.06 (s, 3H, $CH_3$), 1.87 (s, 3H, $CH_3$). $^{13}$C NMR ($CDCl_3$, 101 MHz) δ: 171.14 (C=O), 170.24 (C=O), 170.00 (C=O), 169.78 (C=O), 167.87 (C=O), 134.91 (ArC), 131.56 (ArC), 124.18 (ArC), 90.96 (C1), 70.61 (C5), 69.80 (C4), 67.42 (C3), 61.94 (C6), 53.52 (C2), 21.43 ($CH_3$), 21.18 ($CH_3$), 21.10 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$:. $[\alpha]_D$ 97.35 - 98.93(c 0.35 in DCM). HRMS (ESI): m/z calculated for $C_{22}H_{23}NNaO_{11}$ [(M+Na)]: 500.1163, found: 500.1158.

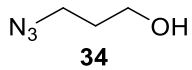
### 5.2.11. 2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-α-D-glucopyranosyl trichloroacetimidate (**36**)



**36**

Hemiacetal **16** (270 mg, 0.78 mmol) was dissolved in dry DCM (2 mL) and placed under $N_2$. The solution was cooled to 0 °C, and $Cl_3CCN$ (511 mg, 383 µL, 3.82 mmol, 4.91 eq) was added dropwise with stirring. The reaction mixture was then warmed to room temperature and allowed to stir for 3 hours. Th reaction mixture was then concentrated in vacuo, and purified by flash column chromatography (2:1 $Et_2O$:EtOAc) leaving acetimidate **36** as a colourless foam (343 mg, 0.70 mmol, 90%). $^1$H NMR (400 MHz, $CDCl_3$) δ: 8.79 (s, 1H, C=NH), 6.36 (d, *J* = 3.66 Hz, 1H, H1), 5.65 (d, *J* = 8.95 Hz, 1H, NH), 5.37 – 5.19 (m, 2H, H3, H4), 4.55 (ddd, *J* = 3.66, 8.95, 10.56 Hz, 1H, H2), 4.25 (dd, *J* = 4.64, 12.97 Hz, 1H, H6'), 4.18 – 4.06 (m, 2H, H5, H6), 2.14 – 2.01 (m, 9H, 3 x $CH_3$), 1.93 (s, 3H, $CH_3$). $^{13}$C NMR ($CDCl_3$, 101 MHz) δ: 171.74 (C=O), 170.77 (C=O), 170.15 (C=O), 169.32 (C=O), 160.40 (OC=NH), 94.89 (C1), 90.94 ($CCl_3$), 70.81 (C3), 70.37 (C5), 67.39 (C4), 61.55 (C6), 51.89 (C2), 23.21 ($CH_3$), 20.88 ($CH_3$), 20.82 ($CH_3$), 20.72 ($CH_3$). FT-IR $\nu_{max}$/cm$^{-1}$: 1743.65 (C=O), 1674.03 (C=N), 1221.63 (C-N), 1016.32 (C-O), 795.06 (C-Cl). $[\alpha]_D$ 59.25 – 66.33 (c 0.45 in DCM). HRMS (ESI): m/z calculated for $C_{16}H_{21}Cl_3N_2NaO_9$ [(M+Na$^+$)]: 513.0205, found: 513.0207.

### 5.2.12. 3-azido-propan-1-ol (**34**)

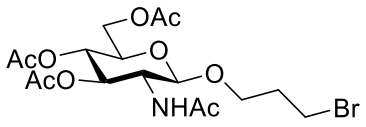


**34**

Commercially sourced 3-bromopropan-1-ol **37** (2.000 g, 1.301 mL, 14.39 mmol) was dissolved in DMF (12 mL). To this solution, sodium azide (3.274 g, 50.36 mmol, 3.5 eq.) was added slowly with stirring. The solution was then left stirring overnight. The reaction mixture was then diluted with $Et_2O$ (12 mL) and washed with $H_2O$ (1 x 12 mL) and brine (1 x 12 mL). The aqueous

layers were combined and extracted with $Et_2O$ (5 x 12 mL). The organic layers were combined and concentrated under airflow leaving alcohol **34** as a colourless oil (1.210 g, 11.97 mmol, 83%). $^1H$ NMR ($CDCl_3$, 400 MHz) δ: 3.75 (dt, *J* = 5.07, 9.85 Hz, 2H, O-$CH_2$), 3.45 (t, *J* = 6.62 Hz, 2H, $N_3$-$CH_2$), 1.83 (p, *J* = 6.62, 2H, R-$CH_2$-R), 1.76 (s, 1H, OH). $^{13}C$ NMR ($CDCl_3$, 101 MHz) δ: 59.78 (HO-C), 48.53 ($N_3$-C), 31.56 (R-$CH_2$-R). FT-IR $\nu_{max}/cm^{-1}$: 3375.65 (O-H), 2934.10 (C-H), 2091.77 (N=N=N), 1057.62 (C-O). GC-MS (EI)- m/z of $C_3H_6NO^-$: 72.0455 (calc.), 72.0448 (expt.).

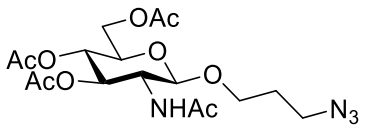
### 5.2.13. 3-bromopropyl-2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-β-D-glucopyranoside (**38**)




**38**

Acetimidate **36** (297 mg, 0.61 mmol) was dissolved in dry DCM (8 mL) with 4 Å molecular sieves under $N_2$. Commercially sourced 3-bromopropan-1-ol **37** (842 mg, 548 µL, 6.06 mmol, 10 eq.) was added and the solution was stirred for 1 hour at room temperature. The reaction mixture was cooled to -10 °C and TMSOTf (18 mg, 15 µL, 0.08 mmol, 0.14 eq.) was added. The reaction was allowed to stir for 25 minutes, filtered over celite, concentrated in vacuo, and purified by flash column chromatography (5:1 EtOAc:Hex) leaving 3-bromopropyl glycoside **38** as a colourless solid (67 mg, 0.14 mmol, 24%). $^1H$ NMR (400 MHz, $CDCl_3$) δ: 5.53 (d, *J* = 8.97 Hz, 1H, NH), 5.20 (t, *J* = 9.98 Hz, 1H, H3), 5.08 (t, *J* = 9.98 Hz, 1H, H4), 4.53 (d, *J* = 8.33 Hz, 1H, H1), 4.26 (dd, *J* = 4.61, 12.27 Hz, 1H, H6'), 4.14 (dd, *J* = 2.35, 12.27 Hz, 1H, H6), 3.99 – 3.91 (m, 1H, H2), 3.79 (dt, *J* = 5.74, 14.49 Hz, 2H, $OCH_2R$), 3.72 – 3.65 (m, 1H, H5), 3.58 – 3.52 (m, 2H, $RCH_2Br$), 2.17 (ddd, *J* = 4.90, 9.85, 19.28 Hz, 2H, $RCH_2R'$), 2.09 (s, 3H, $CH_3$), 2.04 (s, 3H, $CH_3$), 2.03 (s, 3H, $CH_3$), 1.97 (s, 3H, $CH_3$). $^{13}C$ NMR ($CDCl_3$, 101 MHz) δ: 161.54 (C=O), 101.42 (C1), 72.24 (C3), 71.70 (C5), 68.23 (C4), 67.07 ($OCH_2R$), 61.88 (C6), 54.25 (C2), 31.83 ($RCH_2Br$), 30.41 ($RCH_2R'$), 23.21 ($CH_3$), 20.53 ($CH_3$). FT-IR $\nu_{max}/cm^{-1}$: 3322.07 (C-H), 1736.37 (C=O), 1659.38 (C-N), 597.60 (C-Br). $[\alpha]_D$ -6.93 - - 8.31 (c 0.05 in DCM). HRMS (ESI): m/z calculated for $C_{17}H_{26}BrNNaO_9$ [(M+$Na^+$)]: 490.0683, found: 490.0683.

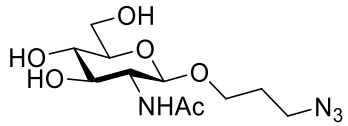
### 5.2.14. 3-azidopropyl-2-acetamido-3,4,6-tri-O-acetyl-2-deoxy-β-D-glucopyranoside (**27**)




**27**

3-bromopropyl glycoside **38** (67 mg, 0.14 mmol) was dissolved in DMF (2 mL). To this solution, sodium azide (47 mg, 0.72 mmol, 5 eq.) was added slowly with stirring. The resulting solution was then heated to 60 °C with stirring for 3 hr. The reaction mixture was then diluted with $Et_2O$ (10 mL) and washed with brine (10 mL). The aqueous layer was then extracted with $Et_2O$ (10 mL), the organic layers combined and dried over $MgSO_4$, and then concentrated in vacuo leaving 3-azidopropyl glycoside **27** as a colourless solid (27 mg, 0.06 mmol, 44%). $^1$H NMR (400 MHz, $CDCl_3$) δ: 5.49 (d, *J* = 8.9 Hz, 1H, NH), 5.21 (dd, *J* = 9.3, 10.7 Hz, 1H, H3), 5.07 (dd, *J* = 9.3, 9.9 Hz, 1H, H4), 4.59 (d, *J* = 8.3 Hz, 1H, H1), 4.24 (dd, *J* = 4.8, 12.3 Hz, 1H, H6'), 4.13 (dd, *J* = 2.5, 12.3 Hz, 1H, H6), 3.99 – 3.87 (m, 2H, $OCH_2R$, H2), 3.68 (ddd, *J* = 2.5, 4.7, 9.9 Hz, 1H, H5), 3.63 – 3.50 (m, 1H, $OCH_2R$), 3.43 – 3.29 (m, 2H, $RCH_2N_3$), 2.08 (s, 3H, $CH_3$), 2.02 (s, 3H, $CH_3$), 2.02 (s, 3H, $CH_3$), 1.95 (s, 3H, $CH_3$), 1.91 – 1.72 (m, 2H, $RCH_2R'$). $^{13}$C (400 MHz, $CDCl_3$) δ: 171.16 (C=O), 170.69 (C=O), 169.48 (C=O), 101.17 (C1), 72.46 (C3), 71.98 (C5), 68.53 (C4), 66.33 ($OCH_3R$), 62.17 (C6), 54.63 (C2), 48.11 ($RCH_3N_3$), 28.86 ($RCH_3R'$), 21.86 ($CH_3$), 20.78 ($CH_3$), 20.72 ($CH_3$). HRMS (ESI): m/z calculated for $C_{17}H_{26}N_4NaO_9$ [(M+Na$^+$)]: 453.1592, found: 453.1597.

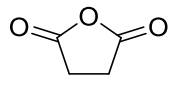
5.2.15. 3-azidopropyl-2-acetamido-2-deoxy-β-D-glucopyranoside (**8**)




**8**

3-azidopropyl glycoside **27** (24 mg, 0.06 mmol) was dissolved in MeOH (2 mL). To this solution, NaOMe (96 µL, 0.05 mmol, 0.86 eq) was added with stirring. The resulting solution was left at room temperature with stirring for 1 hr, following which the solution neutralised via addition of Amberlite IR-120 resin. The solution was then filtered and concentrated in vacuo leaving acceptor **8** as a red solid (14 mg, 0.05 mmol, 83%). $^1$H NMR (400 MHz, $D_2O$) δ: 4.50 (d, *J* = 8.3 Hz, 1H, H1), 4.02 – 3.87 (m, 2H, H3, $OCH_2R$), 3.78 – 3.64 (m, 4H, H2, H4, $RCH_2N_3$), 3.64 – 3.56 (m, 1H, H6'), 3.56 – 3.50 (m, 1H, H6), 3.48 – 3.31 (m, 2H, H5, $OCH_2R$), 2.04 (s, 3H, $CH_3$), 1.89 – 1.79 (m, 2H, $RCH_2R'$). $^{13}$C NMR (400 MHz, $D_2O$) δ: 170.89 (C=O), 101.21 (C1), 75.88 (C3), 73.78 (C5), 69.95 (C4), 67.12 ($OCH_3R$), 60.76 (C6), 55.61 (C2), 47.81 ($RCH_3N_3$), 28.14 ($RCH_3R'$), 22.17 ($CH_3$). IR $\nu_{max}$/cm$^{-1}$: 3257.46 (O-H), 2093.78 (N=N=N), 1650.58 (C=O). HRMS (ESI): m/z calculated for $C_{11}H_{20}N_4NaO_6$ [(M+Na$^+$)]: 327.1275, found: 327.1278.

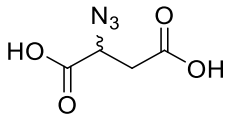
5.2.16. Dihydrofuran-2,5-dione (**40**)



**40**

Commercially sourced succinic acid **39** (2 g, 16.95 mmol) was dissolved in acetic anhydride (25 mL). The resulting solution was heated to 80 °C for 2 hr with stirring. After 2 hr, the solution was concentrated in vacuo leaving anhydride **40** as a white solid (1.646 g, 16.46 mmol, 97%). $^{1}$H NMR (400 MHz, DMSO-$d_6$) δ: 2.90 (s, 4H, $RCH_2CH_2R$). $^{13}$C NMR (400 MHz, DMSO-$d_6$) δ: 176.63 (C=O), 29.28 ($CH_2$). IR $\nu_{max}$/cm$^{-1}$: 1775.24 (C=O), 1056.13 (O=C-O-C=O).

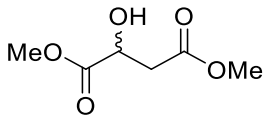
### 5.2.17. 2-azidosuccinic acid (**45**)




**45**

Commercially sourced 2-bromosuccinic acid **44** (1 g, 5.10 mmol) was dissolved in $H_2O$ (5 mL). To this solution, $NaN_3$ (0.660 g, 10.15 mmol, 2 eq.) was added slowly, and the resulting solution was heated to 60 °C for 3 d with stirring. After 3 d the resulting solution was lyophilised leaving a mixture of 2-bromosuccinic acid **44** and 2-azidosuccinic acid **45** as a light brown solid. $^{1}$H NMR (400 MHz, $D_2O$) δ: 4.37 (dd, *J* = 4.8, 7.7 Hz, 1H, $CHN_3$), 2.96 – 2.72 (m, 2H, $CH_2$). $^{13}$C NMR (400 MHz, $D_2O$) δ: 178.92 (C=O), 177.40 (C=O), 63.07 ($CHN_3$), 40.21 ($CH_2$). IR $\nu_{max}$/cm$^{-1}$: 3404.84 (O-H), 2111.92 (N=N=N), 1568.47 (C=O). HRMS (ESI): m/z calculated for $C_4H_4N_3O_4$ [(M-H)]: 158.0207, found: 158.0207.

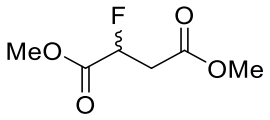
### 5.2.18. Dimethyl malate (**53**)




**53**

Commercially sourced DL-malic acid **52** (1 g, 7.46 mmol) and 2 heaped spatulas of Amberlite IR-120H resin were placed under $N_2$ in a round bottom flask. To this flask was added dry MeOH (20 mL) with stirring. The resulting solution was left under $N_2$ at room temperature with stirring for 48 hr. After 48 hr, the solution was filtered, concentrated in vacuo, resuspended in dioxane, and lyophilised, leaving dimethyl malate **53** as a red solid (894 mg, 5.52 mmol, 74%). $^{1}$H NMR (400 MHz, MeOD-$d_4$) δ: 4.55 – 4.46 (m, 1H, CHOH), 3.74 (s, 3H, $CH_3$), 3.68 (s, 3H, $CH_3$), 2.86 – 2.61 (m, 2H, $CH_2$). $^{13}$C NMR (400 MHz, MeOD-$d_4$) δ: 174.94 (C=O), 172.50 (C=O), 68.53 (C-OH), 52.69 ($CH_3$), 52.30 ($CH_3$), 39.86 ($CH_2$). IR $\nu_{max}$/cm$^{-1}$: 2957.44 (O-H), 1727.16 (C=O), 1161.53 (C-O). HRMS (ESI): m/z calculated for $C_6H_{10}NaO_5$ [(M+Na$^+$)]: 185.0420, found: 185.0424.

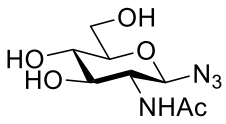
5.2.19. Dimethyl 2-fluorosuccinate (**54**)



**54**

Dimethyl malate **53** (100 mg, 0.62 mmol) was dissolved in dry DCM (9 mL) under $N_2$. The resulting solution was cooled to 0 °C. To this solution, DAST (122 µL, 0.93 mmol, 1.5 eq.) was added dropwise with stirring, and the solution was then allowed to warm to room temperature, where it was left for 1 hr with stirring. After 1 hr, the reaction mixture was quenched by addition of MeOH (245 µL). The resulting solution was concentrated in vacuo, and purified by flash column chromatography (9:1 $Et_2OAc$:Hex) leaving dimethyl 2-fluorosuccinate **54** as a red solid (49 mg, 0.30 mmol, 48%). $^1$H NMR (400 MHz, $(CD_3)_2CO$) δ: 5.31 (ddd, *J* = 4.2, 7.0, 47.0 Hz, 1H, CHF), 3.73 (s, 3H, $CH_3$), 3.65 (s, 3H, $CH_3$), 3.10 – 2.81 (m, 2H, $CH_2$). $^{19}$F NMR (376 MHz, $(CD_3)_2CO$) δ: -191.64 (ddd, *J* = 24.3, 26.0, 47.0 Hz, 1F, CHF). $^{13}$C NMR (400 MHz, $(CD_3)_2CO$) δ: 169.07 (d, *J* = 2.7 Hz, C=O), 168.70 (d, *J* = 23.2 Hz, C=O), 85.38 (d, *J* = 183.9 Hz, CHF), 51.88 ($CH_3$), 51.45 ($CH_3$), 36.81 (d, *J* = 22.8 Hz, $CH_2$). IR $\nu_{max}$/cm$^{-1}$: 1689.70 (C=O), 1264.77 (C-F). HRMS (ESI): m/z calculated for $C_6H_9FNaO_4$ [(M+Na$^+$)]: 187.0377, found: 187.0382.

5.2.20. 3-azido-2-acetamido-2-deoxy-β-D-glucopyranoside (**64**)



**64**

Commercially sourced *N*-acetyl-D-glucosamine **1** (50 mg, 0.23 mmol) was dissolved in $D_2O$ (1 mL). To this solution DMC **22** (115 mg, 0.68 mmol, 3 eq.), 2,6-lutidine (157 µL, 1.36 mmol, 6 eq.), and $NaN_3$ (184 mg, 2.83 mmol, 12.5 eq.) was added. The solution was then cooled to 4 °C and left for 48 hr with stirring. After 48 hr the reaction mixture was transferred to an NMR and formation of azido glycoside **64** was confirmed via 400 MHz NMR and HRMS. $^1$H NMR (400 MHz, $D_2O$) δ: 4.87 (d, *J* = 9.2 Hz, 1H, H1), 4.01 (dd, *J* = 2.2, 12.4 Hz, 1H, H3), 3.92 – 3.76 (m, 2H, H2, H4), 3.75 – 3.52 (m, 3H, H5, H6, H6'), 2.15 (s, 3H, $CH_3$). HRMS (ESI): m/z calculated for $C_8H_{14}N_4NaO_5$ [(M+Na$^+$)]: 269.0856, found: 269.0863.

## 5.3. Glycosynthase Assay Procedures

### 5.3.1. DspB Overexpression and Purification

Plasmids encoding for WT DspB and E248Q were commercially sourced from GenScript. Plasmids for encoding for the mutants D183A, Y278F, and Y278N from storage were sent for Eurofin Genomics LightRun sequencing which revealed that the plasmids had remained intact. 2 µL of each plasmid was added to 50 µL of BL21 DE3 chemically competent cells. These cells were then heated to 42 °C for 45 s, and then placed on ice for 5 mins. After 5 mins, 1 mL of Lysogeny Broth (LB) was added and the cells were incubated at 37 °C with 180 r.p.m shaking for 1 hour. 2 petri dishes with LB agar were prepared with 50 mg/mL kanamycin. After 1 hour, 50 µL of the cell culture was plated onto one of the LB agar plates, and the other 500 µL was centrifuged at 4000 xg for 2 mins. After centrifuging, 450 µL of the supernatant was decanted, and the cells were resuspended in the remaining 50 µL of media. This 50 µL was then plated onto the remaining LB agar plate and both plates were incubated at 37 °C for 18 hr. After 18 hr, individual colonies of cells were picked and inoculated into 10 mL of LB with 50 mg/mL kanamycin. These cultures were then incubated at 37 °C with 180 r.p.m. shaking for 18 hours. The resulting overnight cultures were then used to inoculate 800 mL of LB with 50 mg/mL kanamycin in a 1:100 dilution. These new cultures were incubated at 37 °C with 180 r.p.m. until the $OD_{600}$ of the cultures reached 0.6 A. Protein expression was then induced with IPTG at a final concentration of 0.2 M, and the cells were incubated for a further 4 hr. After 4 hr the cell cultures were centrifuged at 6000 xg for 20 mins at 4 °C, the supernatant decanted, and the cells harvested for further purification. To check that the expression had worked as intended, a 0.25 g portion of each mutant's cell pellet was taken and a BugBuster protocol was performed and the soluble fraction was then run in a 12% acrylamide SDS-PAGE gel.

The cell pellets were resuspended in 30 mL lysis buffer (20 mM Tris base, 500 mM NaCl, 5 mM imidazole, 30 µL Igepal CA-630, 12 µL benzonase, 100 µL protease inhibitor) and sonicated on ice for 30 s (x8 with 30 second intervals). The resulting suspension was centrifuged at 20,000 xg for 25 mins at 4 °C. The supernatant was decanted and loaded onto a 5 mL HisTrap column in binding buffer (20 mM Tris base, 500 mM NaCl, 5 mM imidazole, pH 7.5). The column was washed with 50 mL binding buffer and the protein was eluted by a gradient (0-50% over 5 C.V.) of elution buffer (20 mM Tris base, 500 mM NaCl, 100 mM imidazole, pH 7.5), followed by 5 C.V. of a 50:50 elution:binding buffer mix, and finally a step to 100% elution buffer. Fractions that had an absorbance at 280 nm were analysed by SDS-PAGE, and any fractions that contained high amounts of DspB were collated and dialysed into phosphate buffer saline (PBS) (100 mM phosphate, 200 mM NaCl, pH 5.9), flash frozen, and stored at -70°C.

### 5.3.2. WT DspB DTT Reduction

100 mM DTT was added to SDS sample buffer before addition a sample of to the collated purified fractions of the WT DspB purification. An SDS-PAGE gel was ran on the reduced and a non-reduced sample.

### 5.3.3. DMC **22** Inhibition Test Procedure

DMC **22** was dissolved in 100 mM pH 5.9 PBS at a concentration of 30 mM. 60 µL of this solution was added to a solution of WT DspB (100 nM and 2 µM respectively) and the reaction mixtures were incubated at 37 °C for 30 minutes. 20 µL aliquots were taken at timepoints 2, 5 ,10, 20, and 30 minutes and UV/Vis absorbance for each aliquot was measured at 405 nm.

### 5.3.4. TLC Assay General Procedure

1 µL aliquots of each sample was spotted onto the base of the TLC plate. TLCs were ran in ButOH:AcOH:$H_2O$ (2:1:1), stained with anisaldehyde, and developed by application of heat.

### 5.3.5. 2 eq. Oxazoline **5** Assay Procedure

Reactions were set up as follows:

| | **Stock Conc.** | **Reaction Conc.** | **Vol (µL)** |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 100 mM | 5 mM | 1.5 |
| **Oxazoline 5 in $dH_2O$** | 100 mM | 10 mM | 3 |
| **Enzyme** | 10 mg/mL | 1 mg/mL | 3 |
| **pH 5.9 PBS** | 200 mM | 100 mM | 15 |
| **$dH_2O$** | - | - | 7.5 |
| | | **Total Vol (µL)** | 30 |

Each reaction was repeated in triplicate. After the timepoints of 15 minutes, 1 hour, and 24 hours 1 µL aliquots were taken and analysed via TLC. After 24 hours the reaction mixtures were quenched by addition of 30 µL MeCN, frozen, and analysed by LC-MS.

No Enzyme Control:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 100 mM | 5 mM | 1.5 |
| **Oxazoline 5 in $dH_2O$** | 100 mM | 10 mM | 3 |
| **pH 5.9 PBS** | 200 mM | 100 mM | 15 |
| **$dH_2O$** | - | - | 9.5 |
| | | **Total Vol (µL)** | 30 |

5.3.6. 37 °C 2 eq. Oxazoline **5** Assay Procedure

Reactions were set up as described in section 5.3.5. Each reaction was incubated at 37 °C for 24 hours. At timepoints 15 mins, 30 mins, 45 mins, 1 hour, and 24 hours a 5 µL aliquot was taken and quenched by addition of 5 µL MeCN. These aliquots were then analysed by LC-MS.

5.3.7. 37 °C, pH 7, 2 eq. Oxazoline **5** Assay Procedure

Reactions were set up as follows:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 100 mM | 5 mM | 1.5 |
| **Oxazoline 5 in $dH_2O$** | 100 mM | 10 mM | 3 |
| **Enzyme** | 10 mg/mL | 1 mg/mL | 3 |
| **pH 7 PBS** | 200 mM | 100 mM | 15 |
| **$dH_2O$** | - | - | 7.5 |
| | | **Total Vol (µL)** | 30 |

Each reaction was repeated in triplicate. After the timepoints of 15 minutes, 1 hour, and 24 hours 1 µL 5 µL aliquots were taken and quenched by addition of 5 µL MeCN, frozen, and analysed by LC-MS.

5.3.8. ITAG **59** Copper (I)-catalysed Azide-Alkyne Cycloaddition (CuAAC) Click Procedure

Each reaction from section 5.3.5 was diluted into 1 mL $dH_2O$, flash frozen, and lyophilised overnight. The lyophilised samples was resuspended in 24 µL $dH_2O$, and to the suspension, 1 µL of ITAG compound **59** (250 mM) was added. 10 µL of $CuSO_4$ solution (500 mM $CuSO_4 \cdot 5H_2O$, 500 mM THPTA, $dH_2O$) was then added, and the suspension was vortexed. 5 µL 1M sodium ascorbate in $dH_2O$ was added and the suspension was incubated at 20°C for 2 hours with 300 rpm on a rotary wheel. The suspension was then diluted in 1 mL $dH_2O$, flash frozen, and lyophilised overnight. The

lyophilised samples was then resuspended in 30 µL MeCN and analysed by reverse-phased LC-MS. Each reaction was done in triplicate.

### 5.3.9. 5 mM *p*NP-GlcNAc **4** UV/Vis Assay Procedure

*p*NP-GlcNAc **4** was dissolved in 100 mM pH 5.9 PBS at a concentration of 10 mM. 60 µL of this solution was added to 60 µL of a solution of DspB (50 nM and 2 µM respectively) in a 96 well plate. The plate was then incubated at 37 °C and UV/Vis measurements were made at 405 nm by the CLARIOStar Plate Reader at timepoints 2 mins, 5 mins, 10 mins, 20 mins, 30 mins, 45 mins, and 1 hour. After 1 hour, each reaction was quenched by addition of 120 µL MeCN. The quenched reaction mixtures were then analysed by TLC and LC-MS analysis. Each reaction was done in triplicate.

### 5.3.10. pH 7.4, 5 mM *p*NP-GlcNAc **4** UV/Vis Assay Procedure

Reaction mixtures were prepared as described in section 5.3.9, replacing the pH 5.9 PBS with 100 mM pH 7.4 PBS. UV/Vis measurements were taken every minute on the CLARIOStar Plate Reader.

### 5.3.11. *p*NP-GlcNAc **4** and Acceptor **8** UV/Vis Assay Procedure

Reactions were set up as follows in a 96 well plate:

| | **Stock Conc.** | **Reaction Conc.** | **Vol (µL)** |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 10 mM | 5 mM | 60 |
| ***p*NP-GlcNAc 4 in $dH_2O$** | 100 mM | 25 mM | 30 |
| **Enzyme** | 25 µM | 1 µM | 4.8 |
| **pH 7 PBS** | 1000 mM | 100 mM | 12 |
| **$dH_2O$** | - | - | 4.8 |
| | | **Total Vol (µL)** | 120 |

Reactions were done in triplicate for each enzyme. The reactions were incubated at 37 °C in a 96 well plate, and UV/Vis absorbance at 405 nm was measured every minute for 1 hour by the CLARIOStar Plate Reader. After 1 hour, each reaction mixture was analysed by TLC. The reaction mixtures were then incubated at 37 °C for 24 hours. UV/Vis measurements at 405 nm were then performed.

### 5.3.12. WT DspB Carbamate **7** UV/Vis Assay Procedure

Crude carbamate **7** was dissolved in 100 mM pH 5.9 PBS at a concentration of 2 µM. 60 µL of this solution was added to a 96 well plate. To this, 60 µL of WT DspB (100 nM and 2 µM

respectively) was added, and the plate was incubated at 37 °C. UV/Vis measurements at 410 nm were taken every minute by the CLARIOStar plate reader. Each reaction was done in triplicate.

### 5.3.13. Carbamate **7** Substrate Depletion Assay Procedure

A serial dilution of 2000 nM WT DspB using 100 mM pH 5.9 PBS was performed to produce solutions of WT DspB at the concentrations of 1000 nM, 500 nM, 250 nM, and 100 nM. 60 µL of each WT DspB solution was added to a 96 well plate. Crude carbamate **7** was dissolved in 100 mM pH 5.9 PBS at a concentration of 1 µM. 60 µL of this solution was added to the WT DspB solutions in the 96 well plate, and the plate was incubated at 37 °C. UV/Vis measurements at 410 nm were taken every minute by the CLARIOStar Plate Reader. Each reaction was done in triplicate.

### 5.3.14. 2 mM Carbamate **7** UV/Vis Assay Procedure

Purified carbamate **7** was dissolved in 100 mM pH 5.9 PBS at a concentration of 4 µM. 60 µL of this solution was added to a 96 well plate. To this, 60 µL of DspB (100 nM and 2 µM respectively) was added, and the plate was incubated at 37 °C. UV/Vis measurements at 410 nm were taken every minute by the CLARIOStar plate reader. After 1 hour, each reaction mixture was analysed by TLC. Each reaction was done in triplicate.

### 5.3.15. 5 eq. Oxazoline **5** Assay Procedure

Reactions were set up as follows:

| | **Stock Conc.** | **Reaction Conc.** | **Vol (µL)** |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 100 mM | 5 mM | 1.5 |
| **Oxazoline 5 in $dH_2O$** | 250 mM | 25 mM | 3 |
| **Enzyme** | 10 mg/mL | 1 mg/mL | 3 |
| **pH 5.9 PBS** | 200 mM | 100 mM | 15 |
| **$dH_2O$** | - | - | 7.5 |
| | | **Total Vol (µL)** | 30 |

Each reaction was repeated in triplicate. After 24 hours the reaction mixtures were quenched by addition of 30 µL MeCN, frozen, and analysed by LC-MS.

No Enzyme Control:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| **Acceptor 8 in $dH_2O$** | 100 mM | 5 mM | 1.5 |
| **Oxazoline 5 in $dH_2O$** | 250 mM | 25 mM | 3 |
| **pH 5.9 PBS** | 200 mM | 100 mM | 15 |
| **$dH_2O$** | - | - | 10.5 |
| | | **Total Vol (µL)** | 30 |

5.3.16. Oxazoline **5** Inhibition Assay Procedure

*p*NP-GlcNAc **4** only reaction set up as follows:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| ***p*NP-GlcNAc 4 in $dH_2O$** | 10 mM | 5 mM | 60 |
| **Enzyme** | 25 µM | 1 µM | 4.8 |
| **pH 5.9 PBS** | 1000 mM | 100 mM | 12 |
| **$dH_2O$** | - | - | 43.2 |
| | | **Total Vol (µL)** | 120 |

Oxazoline **5** reaction mixture set up as follows:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| ***p*NP-GlcNAc 4 in $dH_2O$** | 10 mM | 5 mM | 60 |
| **Oxazoline 5 in $dH_2O$** | 250 mM | 25 mM | 12 |
| **Enzyme** | 25 µM | 1 µM | 4.8 |
| **pH 5.9 PBS** | 1000 mM | 100 mM | 12 |
| **$dH_2O$** | - | - | 31.2 |
| | | **Total Vol (µL)** | 120 |

Each reaction was done in triplicate. The reactions were incubated at 37 °C in a 96 well plate, and UV/Vis absorbance at 405 nm was measured every minute for 1 hour by the CLARIOStar Plate Reader.

No Enzyme Control:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| ***p*NP-GlcNAc 4 in $dH_2O$** | 10 mM | 5 mM | 60 |
| **Oxazoline 5 in $dH_2O$** | 250 mM | 25 mM | 12 |
| **pH 5.9 PBS** | 1000 mM | 100 mM | 12 |
| **$dH_2O$** | - | - | 36 |
| | | **Total Vol (µL)** | 120 |

## 5.4. Putative PNAGase Enzyme Procedures

### 5.4.1. Protein Overexpression and Purification

Plasmids encoding for SlHex, SaHex, and ShHex were commercially sourced from GenScript. 2 µL of each plasmid was added to 50 µL of BL21 DE3 chemically competent cells. These cells were then heated to 42 °C for 45 s, and then placed on ice for 5 mins. After 5 mins, 1 mL of Lysogeny Broth (LB) was added and the cells were incubated at 37 °C with 180 r.p.m shaking for 1 hour. 2 petri dishes with LB agar were prepared with 50 mg/mL kanamycin. After 1 hour, 50 µL of the cell culture was plated onto one of the LB agar plates, and the other 500 µL was centrifuged at 4000 xg for 2 mins. After centrifuging, 450 µL of the supernatant was decanted, and the cells were resuspended in the remaining 50 µL of media. This 50 µL was then plated onto the remaining LB agar plate and both plates were incubated at 37 °C for 18 hr. After 18 hr, individual colonies of cells were picked and inoculated into 10 mL of LB with 50 mg/mL kanamycin. These cultures were then incubated at 37 °C with 180 r.p.m. shaking for 18 hours. The resulting overnight cultures were then used to inoculate 800 mL of LB with 50 mg/mL kanamycin in a 1:100 dilution. These new cultures were incubated at 37 °C with 180 r.p.m. until the $OD_{600}$ of the cultures reached 0.6 A. Protein expression was then induced with IPTG at a final concentration of 0.2 M, and the cells were incubated for a further 4 hr. After 4 hr the cell cultures were centrifuged at 6000 xg for 20 mins at 4 °C, the supernatant decanted, and the cells harvested for further purification. To check that the expression had worked as intended, a 0.25 g portion of each mutant's cell pellet was taken and a BugBuster protocol was performed and the soluble fraction was then run in a 12% acrylamide SDS-PAGE gel.

The cell pellets were resuspended in 30 mL lysis buffer (20 mM Tris base, 500 mM NaCl, 5 mM imidazole, 30 µL Igepal CA-630, 12 µL benzonase, 100 µL protease inhibitor) and sonicated on ice for 30 s (x8 with 30 second intervals). The resulting suspension was centrifuged at 20,000 xg for 25 mins at 4 °C. The supernatant was decanted and loaded onto a 5 mL HisTrap column in binding buffer (20 mM Tris base, 500 mM NaCl, 5 mM imidazole, pH 7.5). The column was washed with 50

mL binding buffer and the protein was eluted by a gradient (0-50% over 5 C.V.) of elution buffer (20 mM Tris base, 500 mM NaCl, 100 mM imidazole, pH 7.5), followed by 5 C.V. of a 50:50 elution:binding buffer mix, and finally a step to 100% elution buffer. Fractions that had an absorbance at 280 nm were analysed by SDS-PAGE, and any fractions that contained high amounts of protein were collated and dialysed into phosphate buffer saline (100 mM phosphate, 200 mM NaCl, pH 5.9), flash frozen, and stored at -70°C.

#### 5.4.2. SlHex Activity Test Procedure

50 µL of the concentrated fractions containing SlHex was added to 50 µL of 2.5 mM *p*NP-GlcNAc **4.** The reaction mixture was incubated at 37 °C overnight. UV/Vis analysis of the reaction mixture was then performed.

#### 5.4.3. S200 SEC Purification of SlHex

The concentrated fractions containing SlHex were loaded onto a S200 SEC column and eluted using elution buffer (20 mM Tris base, 500 mM NaCl, 100 mM imidazole, pH 7.5). Fractions that had an absorbance at 280 nm were analysed by SDS-PAGE, and via the assay described in section 5.4.2.

#### 5.4.4. Protein Expression Trials

Expression trials were conducted by transforming the desired cell line with the desired plasmid as described in section 5.4.1. These cells were then cultured in 50 mL LB media with 50 mg/mL kanamycin at a range of temperatures with 180 r.p.m. shaking until the $OD_{600}$ of the cultures reached 0.6 A. Protein expression was then induced with IPTG at a final concentration of 0.2 mM for a range of induction times, and centrifuged at 6000 xg for 20 mins at 4 °C. The cells were lysed using BugBuster Lysis Buffer, and analysed by SDS-PAGE and Western Blot with a nitrocellulose membrane and HRP-conjugated anti-His antibody as the blotting reagent.

## 5.5. *In vitro* PNAG Modification procedures

#### 5.5.1. PNAG Purification Procedure

40 mL 0.5% xylose in TSB was inoculated with *S. carnosus* pTXicaADBC and incubated at 37 °C overnight. The overnight culture was then used to inoculate 1.2L 0.5% xylose in TSB, which was then incubated at 37 °C overnight with 50 r.p.m. The resulting biofilm was resuspended in the TSB media and the cells were then centrifuged at 5000 xg for 15 mins at 4 °C. The supernatant was decanted and the pellet was resuspended in 35 mL of 50 mM pH 7.5 PB, and sonicated on ice for 30 s (4 x at 50% amplitude of a 500 W system). The resulting suspension was centrifuged at 5000 xg for 15 mins at 4 °C. The supernatant was decanted and centrifuged again at 12,000 xg for 15

mins at 4 °C. The supernatant was decanted and TCA was added to create a 5% (w/v) solution, which was incubated on ice for 10 mins. The solution was then centrifuged again at 27,000 xg for 20 mins at 4 °C. The supernatant was decanted and dialysed using Spectra Por 12-14 kDa membrane into 50 mM pH 7.5 PB overnight at 4 °C. The dialysed supernatant was concentrated down to 6 mL using a CentriPrep 10 kDa MWCO spin concentrator at 3000 xg at 4 °C. The sample was then loaded onto a S200 SEC column and eluted using 50 mM pH 7.5 PBS. PNAG was detected via monitoring absorbance at 206 and 280 nm. Fractions were analysed using Dot Blot analysis using biotinylated succinylated WGA-aglutinin as a PNAG binding reagent, streptavidin-HRP as the blotting reagent, and 2% BSA solution as a blocking solution. Fractions containing PNAG were collated and lyophilised.

### 5.5.2. Generation of $N_3$-PNAG

DMC Method:

25 mg of PNAG was dissolved in 1 mL $D_2O$ and the method described in section 5.2.20 was repeated. After 48 hr the reaction mixture was transferred to an NMR tube and analysed using 700 MHz $^1$H NMR.

ADMP Method:

25 mg of PNAG was dissolved in 1 mL 4:1 $D_2O$:MeCN. To this solution TEA (74 µL, 0.57 mmol) was added and the solution was cooled to 4°C. ADMP **63** (97 mg, 0.34 mmol) was added and the solution was left at 4 °C with stirring. 250 µL aliquots were taken and lyophilised at the 3 hr, 24 hr, and 7 d timepoints. After 7 d the solution was lyophilised.

### 5.5.3. 488 nm Fluorescent Tag Click Procedure

Each timepoint taken from the ADMP $N_3$-PNAG reaction was dissolved in 500 µL of pH 7.4 PBS. 250 µL aliquots of the dilutions were then taken and either underwent CuAAC or SPAAC Click reactions.

CuAAC Method:

The reactions were set up as follows under red light:

| | Stock Conc. | Reaction Conc. | Vol (µL) |
|---|---|---|---|
| **Timepoint Aliquot** | - | - | 250 |
| **Na Ascorbate in $dH_2O$** | 15 mg/mL | 1.33 mg/mL | 25 |
| **$CuSO_4$ in $dH_2O$** | 200 mM | 4 mM | 5.62 |
| **AF 488 Alkyne** | 5 mM | 10 µM | 0.562 |
| | | **Total Vol (µL)** | 281.182 |

The solutions were wrapped in foil, and incubated at room temperature for 1 hour on a table top rotary wheel. After 1 hour, the solutions were frozen. Each solution was then analysed by TSDS-PAGE.

SPAAC Method:

2.5 µL of AF488-DBCO (25 mM in DMSO) was added to each timepoint aliquot under red light to give a reaction concentration of 25 µM. The solutions were wrapped in foil, and incubated at room temperature for 1 hour on a table top rotary wheel. After 1 hour, the solutions were frozen. Each solution was then analysed by TSDS-PAGE.

### 5.5.4. $N_3$-PNAG Purification Procedure

The lyophilised $N_3$-PNAG reaction mixture was redissolved in 5 mL of $dH_2O$. 2.5 mL of this solution was dialysed using 3.5 kDa MWCO dialysis tubing overnight into $dH_2O$. The solution was then lyophilised.

### 5.5.5. CycloOct-GFP SPAAC Click Procedure

$N_3$-PNAG (crude and purified) was dissolved at 4 mg/mL in pH 7.4 PBS. 100 µL of each $N_3$-PNAG solution was added to 100 µL of CycloOct-GFP (4 mg/mL) in darkness. The solutions were wrapped in foil and incubated at 37 °C for 1 hr. After 1 hr each reaction mixture was frozen. Samples were then analysed by SDS-PAGE analysis.

### 5.5.6. $N_3$-PNAG-GFP DspB Digest

50 µL of WT DspB (2 mg/mL) was added to of the 50 µL $N_3$-PNAG-GFP (2 mg/mL) reaction mixture. The solution was wrapped in foil and incubated at 37 °C for 1 hr. After 1 hr the reaction mixture was frozen. Samples were then analysed by SDS-PAGE analysis.

### 5.5.7. CycloOct-GFP Concentration Test

A serial dilution of CycloOct-GFP (4 mg/mL) was performed using pH 7.4 PBS to generate several concentrations of CycloOct-GFP (2, 1, 0.5, 0.25, 0.125, 0.06, 0.03, 0.02 mg/mL). 25 µL of each

solution was added to 25 µL $N_3$-PNAG (4 mg/mL), each reaction mixture was wrapped in foil, and incubated at 37 °C for 1 hr. After 1 hr the reaction mixtures were frozen. Samples were then analysed via native PAGE analysis.

#### 5.5.8. Generation of dPNAG

25 mg of PNAG was dissolved in 12.5 mL 5M NaOH in $dH_2O$ and incubated at 37 °C overnight. The solution was then placed on ice and neutralised by addition of 5M HCl. The resulting solution was then dialysed using 3.5 kDa MWCO dialysis tubing overnight at 4 °C into $dH_2O$. The solution was removed from the dialysis tubing and lyophilised. The resulting solid was redissolved in $D_2O$, transferred to an NMR tube and analysed by 700 MHz NMR.

#### 5.5.9. Generation of dsPNAG

25 mg of PNAG was dissolved in 12.5 mL 10% $NH_4OH$ in $H_2O$ and incubated at 37 °C overnight. The solution was then placed on ice and neutralised by addition of 1M HCl. The resulting solution was then dialysed using 3.5 kDa MWCO dialysis tubing overnight at 4 °C into $dH_2O$. The solution was removed from the dialysis tubing and lyophilised. The resulting solid was redissolved in $D_2O$, transferred to an NMR tube and analysed by 700 MHz NMR.

## 5.6. PNAG Detection Assay Procedures

#### 5.6.1. Biotinylation of DspB Y278F

Biotinylation of DspB Y278F was performed using the Pierce Antibody Biotinylation Kit for IP. Biotinylation was confirmed using SDS-PAGE and Western Blot analysis using a nitrocellulose membrane. 2% BSA solution was used as the blocking reagent, and Streptavidin-HRP in a 1:5000 dilution was used as a binding reagent.

#### 5.6.2. Ponceau Red Staining Procedure

Ponceau Red Staining Solution: Ponceau Red dye (0.1% w/v), 5% glacial AcOH, 95% $dH_2O$.

Post transfer the Western Blot Membrane was submerged in Ponceau Red Staining Solution for 5 mins. After 5 mins the membrane was washed using $dH_2O$ until the protein bands become visible. The blot was then imaged using G:BOX imager from Syngene.

# References


1. L. F. Mackenzie, Q. P. Wang, R. A. J. Warren and S. G. Withers, *J Am Chem Soc*, 1998, **120**, 5583-5584.
2. D. Lim, M. A. Brimble, R. Kowalczyk, A. J. Watson and A. J. Fairbanks, *Angew Chem Int Ed Engl*, 2014, **53**, 11907-11911.
3. Z. Tan, W. Yang, N. A. O'Brien, X. Pan, S. Ramadan, T. Marsh, N. Hammer, C. Cywes-Bentley, M. Vinacur, G. B. Pier, J. C. Gildersleeve and X. Huang, *Nat Commun*, 2024, **15**, 3420.
4. D. Shental-Bechor and Y. Levy, *Proc Natl Acad Sci U S A*, 2008, **105**, 8256-8261.
5. M. He, X. Zhou and X. Wang, *Signal Transduct Target Ther*, 2024, **9**, 194.
6. K. Wagh, B. H. Hahn and B. Korber, *Curr Opin HIV AIDS*, 2020, **15**, 267-274.
7. E. Fischer, *Ber. Dtsch. Chem. Ges.*, 1891, **24**, 1836-1845.
8. R. A. Laine, *Glycobiology*, 1994, **4**, 759-767.
9. K. Ramaprabha, S. Venkat Kumar, P. Saravanan, R. Rajeshkannan, M. Rajasimman, H. Kamyab and Y. Vasseghian, *Environ Res*, 2024, **240**, 117521.
10. H. J. Blumenthal and S. Roseman, *J Bacteriol*, 1957, **74**, 222-224.
11. M. D. Lenardon, C. A. Munro and N. A. Gow, *Curr Opin Microbiol*, 2010, **13**, 416-423.
12. G. L. Clark and A. F. Smith, *J Phys Chem*, 1936, **40**, 863-879.
13. P. J. Herring, *Pure Appl Chem*, 1979, **51**, 1901-1911.
14. G. P. Wagner, J. Lo, R. Laine and M. Almeder, *Experientia*, 1993, **49**, 317-319.
15. P. Karrer and A. Hofmann, *Helv Chim Acta*, 2004, **12**, 616-637.
16. M. R. Salton and J. G. Pavlik, *Biochim Biophys Acta*, 1960, **39**, 398-407.
17. W. Weidel, H. Frank and H. H. Martin, *J Gen Microbiol*, 1960, **22**, 158-166.
18. W. Vollmer, D. Blanot and M. A. de Pedro, *FEMS Microbiol Rev*, 2008, **32**, 149-167.
19. R. R. Yocum, J. R. Rasmussen and J. L. Strominger, *J Biol Chem*, 1980, **255**, 3977-3986.
20. E. Gordon, N. Mouz, E. Duee and O. Dideberg, *J Mol Biol*, 2000, **299**, 477-485.
21. R. K. Reed, K. Lilja and T. C. Laurent, *Acta Physiol Scand*, 1988, **134**, 405-411.
22. M. Averbeck, C. A. Gebhardt, S. Voigt, S. Beilharz, U. Anderegg, C. C. Termeer, J. P. Sleeman and J. C. Simon, *J Invest Dermatol*, 2007, **127**, 687-697.
23. D. A. Swann and E. L. Radin, *J Biol Chem*, 1972, **247**, 8069-8073.
24. M. W. Holmes, M. T. Bayliss and H. Muir, *Biochem J*, 1988, **250**, 435-441.
25. K. Meyer and J. W. Palmer, *J Biol Chem*, 1934, **107**, 629-634.
26. G. D. Holt and G. W. Hart, *J Biol Chem*, 1986, **261**, 8049-8057.
27. C. R. Torres and G. W. Hart, *J Biol Chem*, 1984, **259**, 3308-3317.
28. M. A. DeWinter, D. A. Wong, R. Fernandez, W. Kightlinger, A. H. Thames, M. P. DeLisa and M. C. Jewett, *ACS Chem Biol*, 2024, **19**, 1570-1582.
29. C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat Rev Nephrol*, 2019, **15**, 346-366.
30. E. Tsuda, G. Kawanishi, M. Ueda, S. Masuda and R. Sasaki, *Eur J Biochem*, 1990, **188**, 405-411.
31. T. Ueda, K. Tomita, Y. Notsu, T. Ito, M. Fumoto, T. Takakura, H. Nagatome, A. Takimoto, S. Mihara, H. Togame, K. Kawamoto, T. Iwasaki, K. Asakura, T. Oshima, K. Hanasaki, S. Nishimura and H. Kondo, *J Am Chem Soc*, 2009, **131**, 6237-6245.
32. K. Zheng, C. Bantog and R. Bayer, *MAbs*, 2011, **3**, 568-576.
33. L. Wells, K. Vosseller and G. W. Hart, *Science*, 2001, **291**, 2376-2378.
34. G. W. Hart, M. P. Housley and C. Slawson, *Nature*, 2007, **446**, 1017-1022.
35. J. A. Hanover, *FASEB J*, 2001, **15**, 1865-1876.
36. W. B. Dias and G. W. Hart, *Mol Biosyst*, 2007, **3**, 766-772.
37. P. Beltrao, P. Bork, N. J. Krogan and V. van Noort, *Mol Syst Biol*, 2013, **9**, 714.
38. C. Slawson and G. W. Hart, *Curr Opin Struct Biol*, 2003, **13**, 631-636.
39. D. E. Moormeier and K. W. Bayles, *Mol Microbiol*, 2017, **104**, 365-376.

40. F. Al-Wrafy, E. Brzozowska, S. Gorska and A. Gamian, *Postepy Hig Med Dosw (Online)*, 2017, **71**, 78-91.
41. L. L. Greiner, J. L. Edwards, J. Shao, C. Rabinak, D. Entz and M. A. Apicella, *Infect Immun*, 2005, **73**, 1964-1970.
42. X. Wang, J. F. Preston, 3rd and T. Romeo, *J Bacteriol*, 2004, **186**, 2724-2734.
43. C. Cywes-Bentley, D. Skurnik, T. Zaidi, D. Roux, R. B. Deoliveira, W. S. Garrett, X. Lu, J. O'Malley, K. Kinzel, T. Zaidi, A. Rey, C. Perrin, R. N. Fichorova, A. K. Kayatani, T. Maira-Litran, M. L. Gening, Y. E. Tsvetkov, N. E. Nifantiev, L. O. Bakaletz, S. I. Pelton, D. T. Golenbock and G. B. Pier, *Proc Natl Acad Sci U S A*, 2013, **110**, E2209-2218.
44. E. A. Izano, I. Sadovskaya, E. Vinogradov, M. H. Mulks, K. Velliyagounder, C. Ragunath, W. B. Kher, N. Ramasubbu, S. Jabbouri, M. B. Perry and J. B. Kaplan, *Microb Pathog*, 2007, **43**, 1-9.
45. K. Agladze, X. Wang and T. Romeo, *J Bacteriol*, 2005, **187**, 8237-8246.
46. C. Heilmann, O. Schweitzer, C. Gerke, N. Vanittanakom, D. Mack and F. Gotz, *Mol Microbiol*, 1996, **20**, 1083-1091.
47. D. McKenney, J. Hubner, E. Muller, Y. Wang, D. A. Goldmann and G. B. Pier, *Infect Immun*, 1998, **66**, 4711-4720.
48. J. B. Kaplan, K. Velliyagounder, C. Ragunath, H. Rohde, D. Mack, J. K. Knobloch and N. Ramasubbu, *J Bacteriol*, 2004, **186**, 8213-8220.
49. C. Vuong, J. M. Voyich, E. R. Fischer, K. R. Braughton, A. R. Whitney, F. R. DeLeo and M. Otto, *Cell Microbiol*, 2004, **6**, 269-275.
50. E. A. Izano, H. Wang, C. Ragunath, N. Ramasubbu and J. B. Kaplan, *J Dent Res*, 2007, **86**, 618-622.
51. A. Kropec, T. Maira-Litran, K. K. Jefferson, M. Grout, S. E. Cramton, F. Gotz, D. A. Goldmann and G. B. Pier, *Infect Immun*, 2005, **73**, 6868-6876.
52. L. K. Vestby, T. Gronseth, R. Simm and L. L. Nesse, *Antibiotics (Basel)*, 2020, **9**, 59.
53. G. D. Christensen, W. A. Simpson, A. L. Bisno and E. H. Beachey, *Infect Immun*, 1982, **37**, 318-326.
54. D. Mack, W. Fischer, A. Krokotsch, K. Leopold, R. Hartmann, H. Egge and R. Laufs, *J Bacteriol*, 1996, **178**, 175-183.
55. T. Maira-Litran, A. Kropec, D. A. Goldmann and G. B. Pier, *Infect Immun*, 2005, **73**, 6752-6762.
56. C. Vuong, S. Kocianova, J. M. Voyich, Y. Yao, E. R. Fischer, F. R. DeLeo and M. Otto, *J Biol Chem*, 2004, **279**, 54881-54886.
57. I. Sadovskaya, E. Vinogradov, S. Flahaut, G. Kogan and S. Jabbouri, *Infect Immun*, 2005, **73**, 3007-3017.
58. J. G. Joyce, C. Abeygunawardana, Q. Xu, J. C. Cook, R. Hepler, C. T. Przysiecki, K. M. Grimm, K. Roper, C. C. Ip, L. Cope, D. Montgomery, M. Chang, S. Campie, M. Brown, T. B. McNeely, J. Zorman, T. Maira-Litran, G. B. Pier, P. M. Keller, K. U. Jansen and G. E. Mark, *Carbohydr Res*, 2003, **338**, 903-922.
59. K. E. Atkin, S. J. MacDonald, A. S. Brentnall, J. R. Potts and G. H. Thomas, *FEBS Lett*, 2014, **588**, 1869-1872.
60. A. H. Choi, L. Slamti, F. Y. Avci, G. B. Pier and T. Maira-Litran, *J Bacteriol*, 2009, **191**, 5953-5963.
61. WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance, https://www.who.int/publications/i/item/9789240093461, (accessed 18th October 2024).
62. P. Yoong, C. Cywes-Bentley and G. B. Pier, *mBio*, 2012, **3**, e00217-00212.
63. A. R. Fullen, J. L. Gutierrez-Ferman, K. S. Yount, C. F. Love, H. G. Choi, M. A. Vargas, D. Raju, K. N. Corps, P. L. Howell, P. Dubey and R. Deora, *PLoS Pathog*, 2022, **18**, e1010764.

64. E. M. Atkinson and S. R. Long, *Mol Plant Microbe Interact*, 1992, **5**, 439-442.
65. P. L. DeAngelis, J. Papaconstantinou and P. H. Weigel, *J Biol Chem*, 1993, **268**, 19181-19184.
66. C. Gerke, A. Kraft, R. Sussmuth, O. Schweitzer and F. Gotz, *J Biol Chem*, 1998, **273**, 18586-18593.
67. Y. Wang, A. Andole Pannuri, D. Ni, H. Zhou, X. Cao, X. Lu, T. Romeo and Y. Huang, *J Biol Chem*, 2016, **291**, 10046-10057.
68. D. J. Little, J. Poloczek, J. C. Whitney, H. Robinson, M. Nitz and P. L. Howell, *J Biol Chem*, 2012, **287**, 31126-31137.
69. Y. Itoh, J. D. Rice, C. Goller, A. Pannuri, J. Taylor, J. Meisner, T. J. Beveridge, J. F. Preston, 3rd and T. Romeo, *J Bacteriol*, 2008, **190**, 3670-3680.
70. D. J. Little, R. Pfoh, F. Le Mauff, N. C. Bamford, C. Notte, P. Baker, M. Guragain, H. Robinson, G. B. Pier, M. Nitz, R. Deora, D. C. Sheppard and P. L. Howell, *PLoS Pathog*, 2018, **14**, e1006998.
71. R. Pfoh, A. S. Subramanian, J. Huang, D. J. Little, A. Forman, B. R. DiFrancesco, N. Balouchestani-Asli, E. N. Kitova, J. S. Klassen, R. Pomes, M. Nitz and P. L. Howell, *PLoS Pathog*, 2022, **18**, e1010750.
72. InterPro family IP002656: Acyltransferase 3 domain, https://www.ebi.ac.uk/interpro/entry/InterPro/IPR002656/, (accessed 19th October 2024).
73. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, *Nature*, 2021, **596**, 583-589.
74. M. Mirdita, K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat Methods*, 2022, **19**, 679-682.
75. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin, *Protein Sci*, 2021, **30**, 70-82.
76. J. Dinkelaar, H. Gold, H. S. Overkleeft, J. D. Codee and G. A. van der Marel, *J Org Chem*, 2009, **74**, 4208-4216.
77. L. G. Weaver, Y. Singh, J. T. Blanchfield and P. L. Burn, *Carbohydr Res*, 2013, **371**, 68-76.
78. M. L. Gening, Y. E. Tsvetkov, G. B. Pier and N. E. Nifantiev, *Carbohydr Res*, 2007, **342**, 567-575.
79. A. Fekete, A. Borbas, G. Gyemant, L. Kandra, E. Fazekas, N. Ramasubbu and S. Antus, *Carbohydr Res*, 2011, **346**, 1445-1453.
80. Q. Chen, F. Yang and Y. Du, *Carbohydr Res*, 2005, **340**, 2476-2482.
81. M. Fridman, D. Solomon, S. Yogev and T. Baasov, *Org Lett*, 2002, **4**, 281-283.
82. L. G. Melean, K. R. Love and P. H. Seeberger, *Carbohydr Res*, 2002, **337**, 1893-1916.
83. K. I. Kanno and K. Hatanaka, *Polym J*, 1998, **30**, 678-680.
84. M. L. Gening, T. Maira-Litrán, A. Kropec, D. Skurnik, M. Grout, Y. E. Tsvetkov, N. E. Nifantiev and G. B. Pier, *Infect Immun*, 2010, **78**, 764-772.
85. A. P. Breslawec, S. Wang, C. Li and M. B. Poulin, *J Biol Chem*, 2021, **296**, 100203.
86. A. P. Breslawec, S. Wang, K. N. Monahan, L. L. Barry and M. B. Poulin, *FEBS J*, 2022, DOI: 10.1111/febs.16624.
87. S. Wang, A. P. Breslawec, E. Alvarez, M. Tyrlik, C. Li and M. B. Poulin, *ACS Chem Biol*, 2019, **14**, 1998-2005.
88. S. Wang, A. P. Breslawec, C. Li and M. B. Poulin, *Chemistry*, 2020, **26**, 10719-10723.
89. F. Yang and Y. Du, *Carbohydr Res*, 2003, **338**, 495-502.
90. F. Yang, H. M. He and Y. G. Du, *Tetrahedron Lett*, 2002, **43**, 7561-7563.
91. D. J. Silva, H. Wang, N. M. Allanson, R. K. Jain and M. J. Sofia, *The Journal of Organic Chemistry*, 1999, **64**, 5926-5929.

92. S. Grann Hansen and T. Skrydstrup, *Eur J Org Chem*, 2007, **2007**, 3392-3401.
93. X. Huang, L. Huang, H. Wang and X. S. Ye, *Angew Chem Int Ed Engl*, 2004, **43**, 5221-5224.
94. J. Defaye, A. Gadelle and C. Pedersen, *Carbohydr Res*, 1989, **186**, 177-188.
95. C. Leung, A. Chibba, R. F. Gomez-Biagi and M. Nitz, *Carbohydr Res*, 2009, **344**, 570-575.
96. S. Aldrich, Safety Data Sheet: 70% Hydrogen Fluoride Pyridine, https://www.sigmaaldrich.com/GB/en/sds/aldrich/184225?userType=anonymous, (accessed 26th November 2024).
97. M. H. Caruthers, *Science*, 1985, **230**, 281-285.
98. R. B. Merrifield, *Science*, 1965, **150**, 178-185.
99. L. Kröck, D. Esposito, B. Castagner, C.-C. Wang, P. Bindschädler and P. H. Seeberger, *Chem Sci*, 2012, **3**, 1617.
100. O. J. Plante, E. R. Palmacci and P. H. Seeberger, *Science*, 2001, **291**, 1523-1527.
101. N. V. Ganesh, K. Fujikawa, Y. H. Tan, K. J. Stine and A. V. Demchenko, *Org Lett*, 2012, **14**, 3036-3039.
102. D. Budhadev, K. Saxby, J. Walton, G. Davies, P. C. Tyler, R. Schworer and M. A. Fascione, *Org Biomol Chem*, 2019, **17**, 1817-1821.
103. A. A. Joseph, A. Pardo-Vargas and P. H. Seeberger, *J Am Chem Soc*, 2020, **142**, 8561-8564.
104. C. Kelly-Quintos, A. Kropec, S. Briggs, C. L. Ordonez, D. A. Goldmann and G. B. Pier, *J Infect Dis*, 2005, **192**, 2012-2019.
105. D. Skurnik, A. Kropec, D. Roux, C. Theilacker, J. Huebner and G. B. Pier, *Clin Infect Dis*, 2012, **55**, 1188-1197.
106. D. Skurnik, M. Merighi, M. Grout, M. Gadjeva, T. Maira-Litran, M. Ericsson, D. A. Goldmann, S. S. Huang, R. Datta, J. C. Lee and G. B. Pier, *J Clin Invest*, 2010, **120**, 3220-3233.
107. D. McKenney, K. L. Pouliot, Y. Wang, V. Murthy, M. Ulrich, G. Doring, J. C. Lee, D. A. Goldmann and G. B. Pier, *Science*, 1999, **284**, 1523-1527.
108. C. Pozzi, K. Wilk, J. C. Lee, M. Gening, N. Nifantiev and G. B. Pier, *PLoS One*, 2012, **7**, e46648.
109. B. Henrissat and G. Davies, *Curr Opin Struct Biol*, 1997, **7**, 637-644.
110. B. Henrissat, *Biochem J*, 1991, **280 ( Pt 2)**, 309-316.
111. D. E. Koshland, *Biol Rev Camb Philos Soc*, 1953, **28**, 416-436.
112. L. P. McIntosh, G. Hand, P. E. Johnson, M. D. Joshi, M. Korner, L. A. Plesniak, L. Ziser, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1996, **35**, 9958-9966.
113. M. L. Sinnott, *Chem Rev*, 2002, **90**, 1171-1202.
114. L. F. Sobala, G. Speciale, S. Zhu, L. Raich, N. Sannikova, A. J. Thompson, Z. Hakki, D. Lu, S. Shamsi Kazem Abadi, A. R. Lewis, V. Rojas-Cervellera, G. Bernardo-Seisdedos, Y. Zhang, O. Millet, J. Jimenez-Barbero, A. J. Bennet, M. Sollogoub, C. Rovira, G. J. Davies and S. J. Williams, *ACS Cent Sci*, 2020, **6**, 760-770.
115. N. N. Thi, W. A. Offen, F. Shareck, G. J. Davies and N. Doucet, *Biochemistry*, 2014, **53**, 1789-1800.
116. J. B. Kaplan, C. Ragunath, N. Ramasubbu and D. H. Fine, *J Bacteriol*, 2003, **185**, 4693-4698.
117. J. B. Kaplan, C. Ragunath, K. Velliyagounder, D. H. Fine and N. Ramasubbu, *Antimicrob Agents Chemother*, 2004, **48**, 2633-2636.
118. A. Males, O. V. Moroz, E. Blagova, A. Munch, G. H. Hansen, A. H. Johansen, L. H. Ostergaard, D. R. Segura, A. Eddenden, A. V. Due, M. Gudmand, J. Salomon, S. R. Sorensen, J. P. L. Franco Cairo, M. Nitz, R. A. Pache, R. M. Vejborg, S. Bhosale, D. J. Vocadlo, G. J. Davies and K. S. Wilson, *Acta Crystallogr D Struct Biol*, 2025, **81**, 130-146.
119. S. G. Manuel, C. Ragunath, H. B. Sait, E. A. Izano, J. B. Kaplan and N. Ramasubbu, *FEBS J*, 2007, **274**, 5987-5999.
120. E. Fazekas, L. Kandra and G. Gyemant, *Carbohydr Res*, 2012, **363**, 7-13.
121. I. R. Greig, F. Zahariev and S. G. Withers, *J Am Chem Soc*, 2008, **130**, 17620-17628.
122. S. Knapp, D. Vocadlo, Z. N. Gao, B. Kirk, J. P. Lou and S. G. Withers, *J Am Chem Soc*, 1996, **118**, 6804-6805.

123. B. L. Mark, D. J. Vocadlo, S. Knapp, B. L. Triggs-Raine, S. G. Withers and M. N. James, *J Biol Chem*, 2001, **276**, 10330-10337.
124. D. J. Vocadlo and S. G. Withers, *Biochemistry*, 2005, **44**, 12809-12818.
125. Schrodinger LLC., *Journal*, 2015.
126. S. J. Williams, B. L. Mark, D. J. Vocadlo, M. N. James and S. G. Withers, *J Biol Chem*, 2002, **277**, 40055-40065.
127. I. Tews, A. Perrakis, A. Oppenheim, Z. Dauter, K. S. Wilson and C. E. Vorgias, *Nat Struct Biol*, 1996, **3**, 638-648.
128. D. J. Little, G. Li, C. Ing, B. R. DiFrancesco, N. C. Bamford, H. Robinson, M. Nitz, R. Pomes and P. L. Howell, *Proc Natl Acad Sci U S A*, 2014, **111**, 11013-11018.
129. H. T. T. Nguyen, T. H. Nguyen and M. Otto, *Comput Struct Biotec*, 2020, **18**, 3324-3334.
130. H. S. Joo and M. Otto, *Chem Biol*, 2012, **19**, 1503-1513.
131. K. L. Frank and R. Patel, *Infect Immun*, 2007, **75**, 4728-4742.
132. A. B. Higginson, MBiochem Master's Dissertation, University of York, 2022.
133. R. Patel, K. E. Piper, M. S. Rouse, J. R. Uhl, F. R. Cockerill, 3rd and J. M. Steckelberg, *J Clin Microbiol*, 2000, **38**, 4262-4263.
134. K. L. Frank, E. J. Reichert, K. E. Piper and R. Patel, *Antimicrob Agents Chemother*, 2007, **51**, 888-895.
135. K. M. Conlon, H. Humphreys and J. P. O'Gara, *J Bacteriol*, 2002, **184**, 4400-4408.
136. N. Cerca, J. L. Brooks and K. K. Jefferson, *J Bacteriol*, 2008, **190**, 6530-6533.
137. K. K. Jefferson, D. B. Pier, D. A. Goldmann and G. B. Pier, *J Bacteriol*, 2004, **186**, 2449-2456.
138. S. G. Withers, K. Rupitz, D. Trimbur and R. A. Warren, *Biochemistry*, 1992, **31**, 9979-9985.
139. B. v. Hofsten, *Biochim Biophys Acta*, 1961, **48**, 159-163.
140. Y. L. Zhang, J. Bommuswamy and M. L. Sinnott, *J Am Chem Soc*, 1994, **116**, 7557-7563.
141. N. S. Banait and W. P. Jencks, *J Am Chem Soc*, 2002, **113**, 7951-7958.
142. J. L. Copa-Patino, Y. Zhang, B. Padmaperuma, I. Marsden, P. Broda and M. L. Sinnott, *Biochem J*, 1993, **293 ( Pt 2)**, 591-594.
143. J. K. C. Knowles, P. Lentovaara, M. Murray and M. L. Sinnott, *J Chem Soc Chem Commun*, 1988, DOI: 10.1039/c39880001401, 1401.
144. H. D. Ly and S. G. Withers, *Annu Rev Biochem*, 1999, **68**, 487-522.
145. E. J. Hehre, C. F. Brewer and D. S. Genghof, *J Biol Chem*, 1979, **254**, 5942-5950.
146. M. R. Hayes and J. Pietruszka, *Molecules*, 2017, **22**.
147. Y. Ichikawa, Y. C. Lin, D. P. Dumas, G. J. Shen, E. Garcia-Junceda, M. A. Williams, R. Bayer, C. Ketcham and L. E. Walker, *J Am Chem Soc*, 1992, **114**, 9283-9298.
148. M. Umekawa, W. Huang, B. Li, K. Fujita, H. Ashida, L. X. Wang and K. Yamamoto, *J Biol Chem*, 2008, **283**, 4469-4479.
149. E. A. Martinez, H. Boer, A. Koivula, E. Samain, H. Driguez, S. Armand and S. Cottaz, *J Mol Catal B-Enzym*, 2012, **74**, 89-96.
150. J. Muschiol and A. S. Meyer, *Z Naturforsch C J Biosci*, 2019, **74**, 85-89.
151. K. Schmölzer, M. Weingarten, K. Baldenius and B. Nidetzky, *ACS Catalysis*, 2019, **9**, 5503-5514.
152. K. Schmolzer, M. Weingarten, K. Baldenius and B. Nidetzky, *Org Biomol Chem*, 2019, **17**, 5661-5665.
153. A. G. Santana, G. Vadlamani, B. L. Mark and S. G. Withers, *Chem Commun (Camb)*, 2016, **52**, 7943-7946.
154. G. Tegl, J. Hanson, H. M. Chen, D. H. Kwan, A. G. Santana and S. G. Withers, *Angew Chem Int Ed Engl*, 2019, **58**, 1632-1637.
155. K. Slámová, J. Krejzová, P. Marhol, L. Kalachova, N. Kulik, H. Pelantová, J. Cvačka and V. Křen, *Adv Synth Catal*, 2015, **357**, 1941-1950.
156. I. Tews, A. C. T. vanScheltinga, A. Perrakis, K. S. Wilson and B. W. Dijkstra, *J Am Chem Soc*, 1997, **119**, 7954-7959.

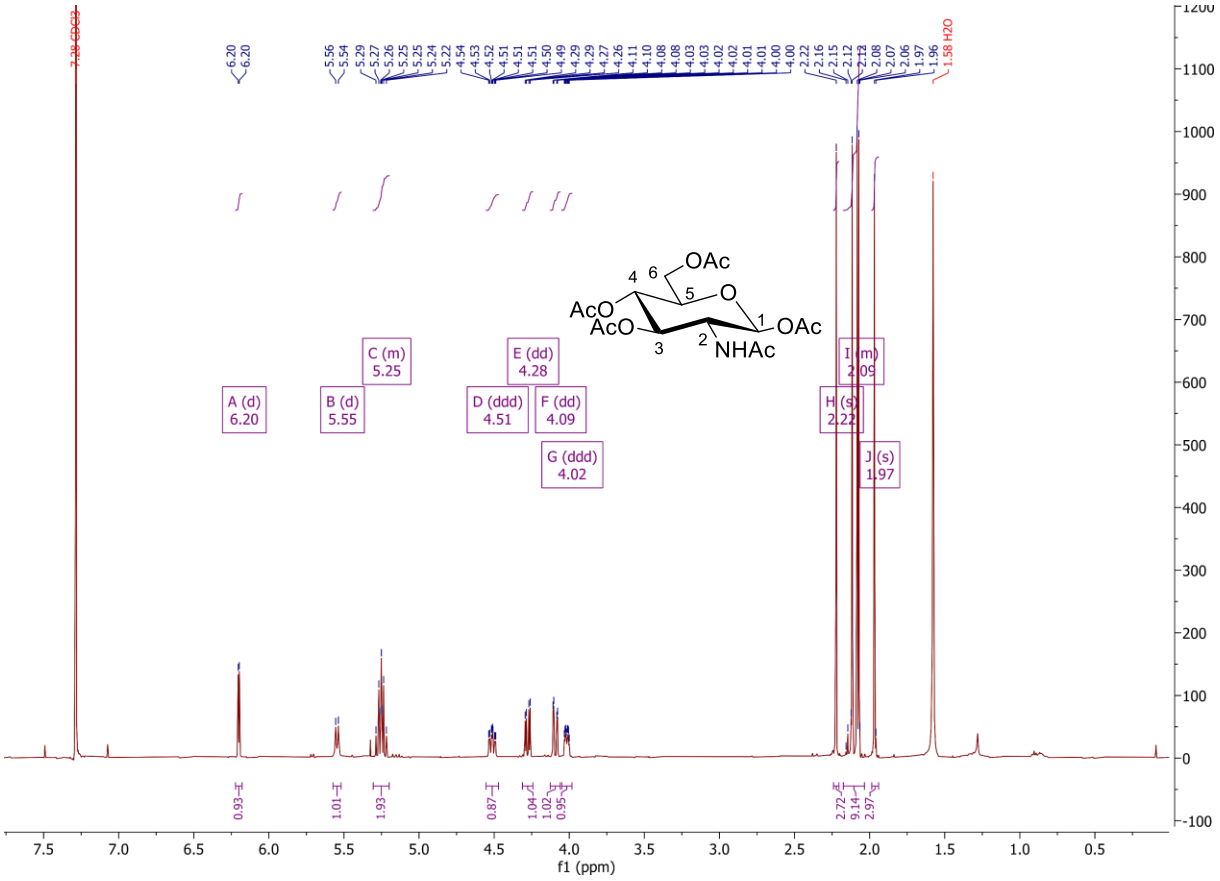
157. Christoph Mayer, David L. Zechel, Stephen P. Reid, R. Antony J. Warren and S. G. Withers, *FEBS Letters*, 2000, **466**, 40-44.
158. C. Malet and A. Planas, *FEBS Lett*, 1998, **440**, 208-212.
159. C. Malet and A. Planas, *FEBS Letters*, 1998, **440**, 208-212.
160. J. H. Wei, X. Lv, Y. Lu, G. Z. Yang, L. F. Fu, L. Yang, J. J. Wang, J. H. Gao, S. H. Cheng, Q. Duan, C. Jin and X. B. Li, *Eur J Org Chem*, 2013, **2013**, 2414-2419.
161. F. W. Ballardie, B. Capon, W. M. Dearie and R. L. Foster, *Carbohydr Res*, 1976, **49**, 79-92.
162. B. Cobucci-Ponzano, F. Conte, E. Bedini, M. M. Corsaro, M. Parrilli, G. Sulzenbacher, A. Lipski, F. Dal Piaz, L. Lepore, M. Rossi and M. Moracci, *Chemistry & Biology*, 2009, **16**, 1097-1108.
163. B. Cobucci-Ponzano, C. Zorzetti, A. Strazzulli, S. Carillo, E. Bedini, M. M. Corsaro, D. A. Comfort, R. M. Kelly, M. Rossi and M. Moracci, *Glycobiology*, 2011, **21**, 448-456.
164. H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew Chem Int Ed Engl*, 2001, **40**, 2004-2021.
165. N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem Biol*, 2006, **1**, 644-648.
166. N. J. Pedowitz and M. R. Pratt, *RSC Chem Biol*, 2021, **2**, 306-321.
167. F. Liu, H. M. Chen, Z. Armstrong and S. G. Withers, *ACS Cent Sci*, 2022, **8**, 656-662.
168. K. Yamamoto and B. G. Davis, *Angew Chem Int Ed Engl*, 2012, **51**, 7449-7453.
169. D. Teze, J. Zhao, M. Wiemann, Z. G. A. Kazi, R. Lupo, B. Zeuner, M. Vuillemin, M. E. Ronne, G. Carlstrom, J. O. Duus, Y. H. Sanejouand, M. J. O'Donohue, E. Nordberg Karlsson, R. Faure, H. Stalbrand and B. Svensson, *Chemistry*, 2021, **27**, 10323-10334.
170. X. Chen, L. Xu, L. Jin, B. Sun, G. Gu, L. Lu and M. Xiao, *Appl Environ Microbiol*, 2016, **82**, 5642-5652.
171. A. Chibba, S. Dasgupta, N. Yakandawala, S. Madhyastha and M. Nitz, *J Carbohydr Chem*, 2011, **30**, 549-558.
172. J. Sakamoto, T. Watanabe, Y. Ariga and S. Kobayashi, *Chemistry Letters*, 2001, **30**, 1180-1181.
173. M. Fujita, S. Shoda, K. Haneda, T. Inazu, K. Takegawa and K. Yamamoto, *Biochim Biophys Acta*, 2001, **1528**, 9-14.
174. H. Ochiai, W. Huang and L. X. Wang, *Carbohydr Res*, 2009, **344**, 592-598.
175. I. Sittel and M. C. Galan, *Org Biomol Chem*, 2017, **15**, 3575-3579.
176. S. J. Richards, T. Keenan, J. B. Vendeville, D. E. Wheatley, H. Chidwick, D. Budhadev, C. E. Council, C. S. Webster, H. Ledru, A. N. Baker, M. Walker, M. C. Galan, B. Linclau, M. A. Fascione and M. I. Gibson, *Chem Sci*, 2021, **12**, 905-910.
177. C. S. Jones, A. C. Anderson and A. J. Clarke, *Proc Natl Acad Sci U S A*, 2021, **118**, e2103602118.
178. T. Keenan, H. S. Chidwick, M. Best, E. K. P. Flack, N. D. J. Yates, N. E. Hatton, M. E. Warnes and M. A. Fascione, *Chem Commun (Camb)*, 2024, **60**, 1428-1431.
179. InterPro Family IPR000182: GNAT domain, https://www.ebi.ac.uk/interpro/entry/InterPro/IPR000182/, (accessed 5th May 2025).
180. R. Sadamoto, T. Matsubayashi, M. Shimizu, T. Ueda, S. Koshida, T. Koda and S. Nishimura, *Chemistry*, 2008, **14**, 10192-10195.
181. S. M. Andersen, M. Heuckendorff and H. H. Jensen, *Org Lett*, 2015, **17**, 944-947.
182. R. G. G. Leenders, R. Ruytenbeek, E. W. P. Damen and H. W. Scheeren, *Synthesis-Stuttgart*, 1996, **11**, 1309-&.
183. R. R. Schmidt, *Angew Chem Int Ed Engl*, 1986, **25**, 212-235.
184. R. U. Lemieux, K. B. Hendriks, R. V. Stick and K. James, *J Am Chem Soc*, 1975, **97**, 4056-4062.
185. A. Chibba, PhD Thesis, University of Toronto, 2014.
186. A. Paiotta, G. D'Orazio, R. Palorini, F. Ricciardiello, L. Zoia, G. Votta, L. De Gioia, F. Chiaradonna and B. La Ferla, *Eur J Org Chem*, 2018, **2018**, 1946-1952.

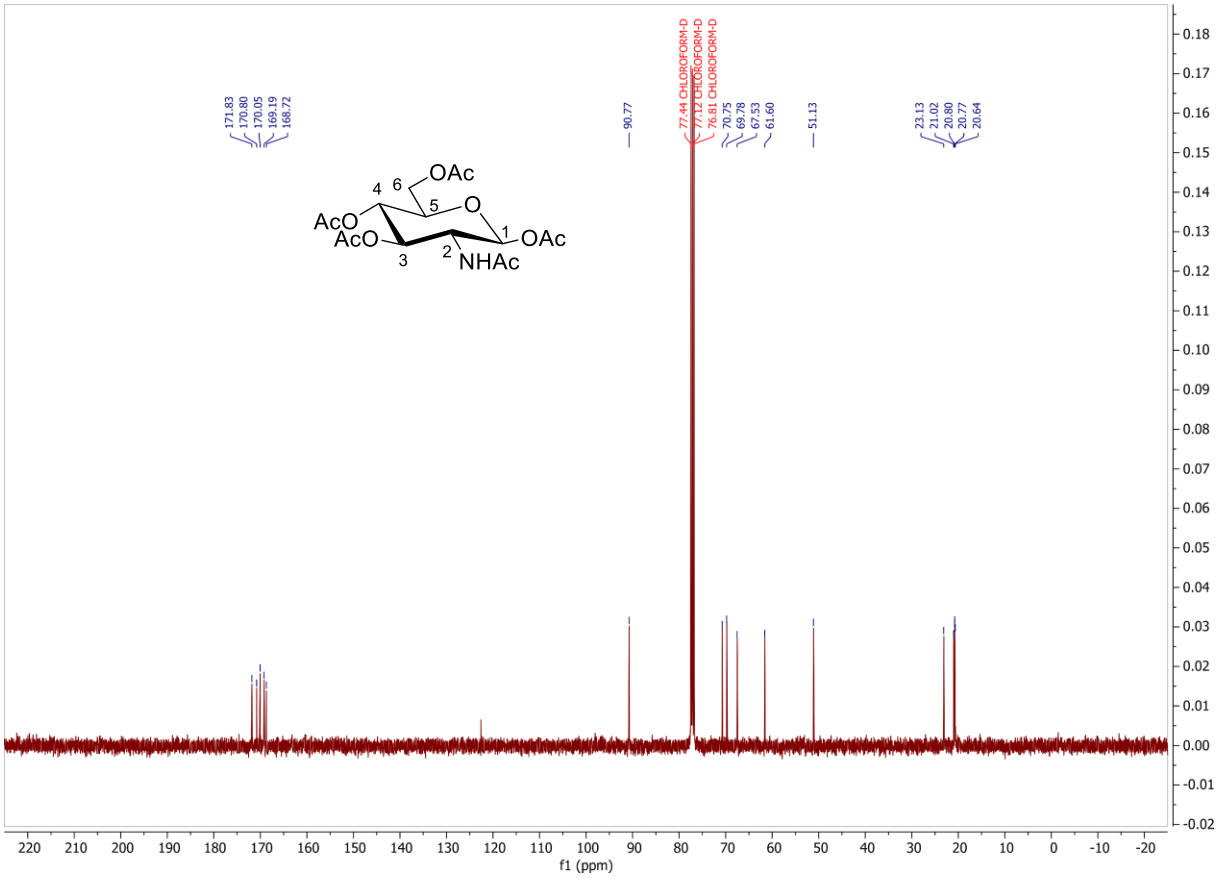
187. M. Noguchi, T. Tanaka, H. Gyakushi, A. Kobayashi and S. Shoda, *J Org Chem*, 2009, **74**, 2210-2212.
188. M. Noguchi, T. Fujieda, W. C. Huang, M. Ishihara, A. Kobayashi and S.-i. Shoda, *Helv Chim Acta*, 2012, **95**, 1928-1936.
189. H. Christensen, M. S. Christiansen, J. Petersen and H. H. Jensen, *Org Biomol Chem*, 2008, **6**, 3276-3283.
190. M. R. Rasmussen, M. H. Marqvorsen, S. K. Kristensen and H. H. Jensen, *J Org Chem*, 2014, **79**, 11011-11019.
191. R. P. McGeary, K. Wright and I. Toth, *J Org Chem*, 2001, **66**, 5102-5105.
192. C. P. Zheng, T. Bavaro, S. Tengattini, A. G. Mascherpa, M. Sollogoub, Y. M. Zhang and M. Terreni, *Eur J Org Chem*, 2019, **2019**, 3622-3631.
193. H. Liu, T. Hansen, S. Y. Zhou, G. E. Wen, X. X. Liu, Q. J. Zhang, J. D. C. Codee, R. R. Schmidt and J. S. Sun, *Org Lett*, 2019, **21**, 8713-8717.
194. I. Choudhury Mukherjee, N. Minoura and H. Uzawa, *Carbohydr Res*, 2003, **338**, 1265-1270.
195. H. Shimizu, Y. Ito, Y. Matsuzaki, H. Iijima and T. Ogawa, *Biosci Biotechnol Biochem*, 1996, **60**, 73-76.
196. M. E. Jung and P. Koch, *Org Lett*, 2011, **13**, 3710-3713.
197. Y. Y. Lin, S. H. Chan, Y. P. Juang, H. M. Hsiao, J. H. Guh and P. H. Liang, *Eur J Med Chem*, 2018, **143**, 1942-1958.
198. J. G. Badiang and J. Aube, *J Org Chem*, 1996, **61**, 2484-2487.
199. K. Lau, V. Thon, H. Yu, L. Ding, Y. Chen, M. M. Muthana, D. Wong, R. Huang and X. Chen, *Chem Commun (Camb)*, 2010, **46**, 6066-6068.
200. C.-W. Chang, S.-S. Chang, C.-S. Chao and K.-K. T. Mong, *Tetrahedron Lett*, 2009, **50**, 4536-4540.
201. S. R. Lu, Y. H. Lai, J. H. Chen, C. Y. Liu and K. K. Mong, *Angew Chem Int Ed Engl*, 2011, **50**, 7315-7320.
202. F. Manoni, C. Cornaggia, J. Murray, S. Tallon and S. J. Connon, *Chem Commun (Camb)*, 2012, **48**, 6502-6504.
203. E. Eskil, M. J. Hansson, K. H. J. Ehinger and S. Moss, *WO Pat.,* 2015155231, 2015.
204. K. Freudenberg, G. Piazolo and C. Knoevenagel, *Justus Liebigs Annalen der Chemie*, 1939, **537**, 197-204.
205. B. Neises and W. Steglich, *Angew Chem Int Ed Engl*, 1978, **17**, 522-524.
206. B. Li, M. Berliner, R. Buzon, C. K. Chiu, S. T. Colgan, T. Kaneko, N. Keene, W. Kissel, T. Le, K. R. Leeman, B. Marquez, R. Morris, L. Newell, S. Wunderwald, M. Witt, J. Weaver, Z. Zhang and Z. Zhang, *J Org Chem*, 2006, **71**, 9045-9050.
207. W. Li, A. Santra, H. Yu, T. J. Slack, M. M. Muthana, D. Shi, Y. Liu and X. Chen, *J Org Chem*, 2019, **84**, 6697-6708.
208. J. M. Risley, J. P. Kastanis and A. M. Young, *J Fluorine Chem*, 2011, **132**, 269-275.
209. S. M. Smith, *Methods Mol Biol*, 2017, **1485**, 389-400.
210. S.-H. Lin and G. Guidotti, in *Methods in Enzymology*, eds. R. R. Burgess and M. P. Deutscher, Academic Press, 2009, vol. 463, pp. 619-629.
211. E. C. Minnihan, D. D. Young, P. G. Schultz and J. Stubbe, *J Am Chem Soc*, 2011, **133**, 15942-15945.
212. A. J. Walklett, E. K. P. Flack, H. S. Chidwick, N. E. Hatton, T. Keenan, D. Budhadev, J. Walton, G. H. Thomas and M. A. Fascione, *Angew Chem Int Ed Engl*, 2024, **63**, e202318523.
213. T. Tanaka, H. Nagai, M. Noguchi, A. Kobayashi and S. Shoda, *Chem Commun (Camb)*, 2009, DOI: 10.1039/b905761g, 3378-3379.
214. T. Monsen, E. Lovgren, M. Widerstrom and L. Wallinder, *J Clin Microbiol*, 2009, **47**, 2496-2501.
215. M. Monsigny, A. C. Roche, C. Sene, R. Maget-Dana and F. Delmotte, *Eur J Biochem*, 1980, **104**, 147-153.

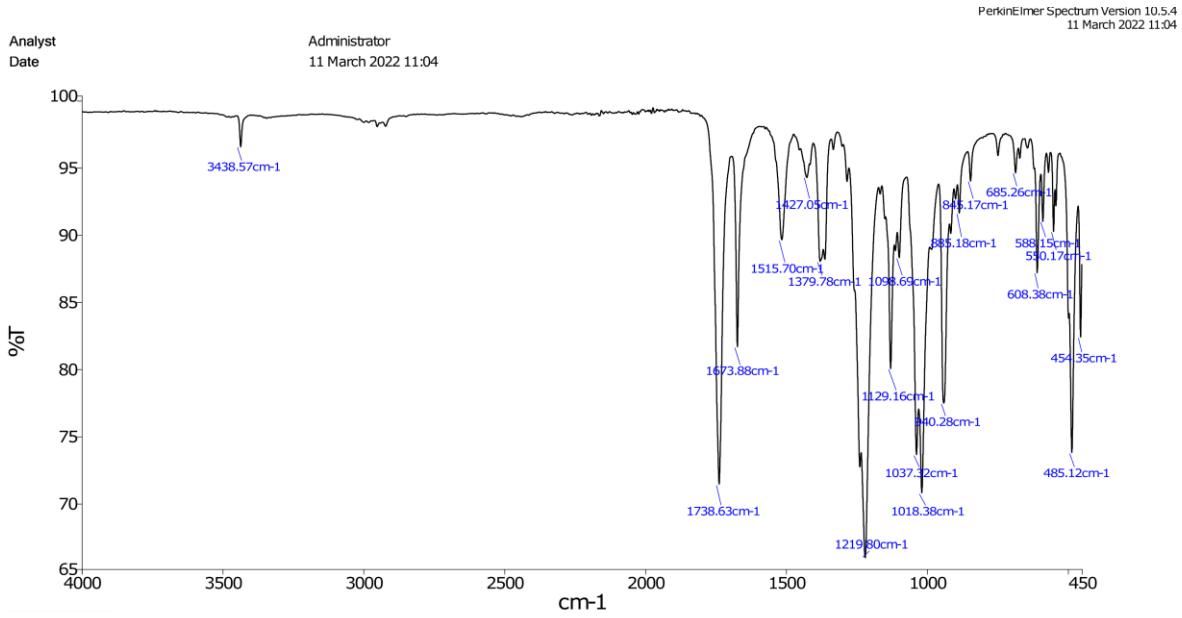
216. R. L. Brabham, PhD, University of York, 2019.
217. G. Meng, T. Guo, T. Ma, J. Zhang, Y. Shen, K. B. Sharpless and J. Dong, *Nature*, 2019, **574**, 86-89.
218. R. Kamimura, H. Kanematsu, A. Ogawa, T. Kogo, H. Miura, R. Kawai, N. Hirai, T. Kato, M. Yoshitake and D. M. Barry, *Materials (Basel)*, 2022, **15**.
219. Y. Yang, D. W. Jung, D. G. Bai, G. S. Yoo and J. K. Choi, *Electrophoresis*, 2001, **22**, 855-859.
220. R. G. E. Krause and J. P. D. Goldring, *Anal Biochem*, 2019, **566**, 107-115.
221. H. Towbin, T. Staehelin and J. Gordon, *Proceedings of the National Academy of Sciences*, 1979, **76**, 4350-4354.
222. J. Renart, J. Reiser and G. R. Stark, *Proceedings of the National Academy of Sciences*, 1979, **76**, 3116-3120.
223. A. Eddenden, E. N. Kitova, J. S. Klassen and M. Nitz, *ACS Chem Biol*, 2020, **15**, 1204-1211.
224. C. D. Spicer and B. G. Davis, *Nat Commun*, 2014, **5**, 4740.
225. J. M. Becker and M. Wilchek, *Biochim Biophys Acta*, 1972, **264**, 165-170.
226. S. Sokalingam, G. Raghunathan, N. Soundrarajan and S. G. Lee, *PLoS One*, 2012, **7**, e40410.

# Appendix

Appendix 1. Compound **2**:

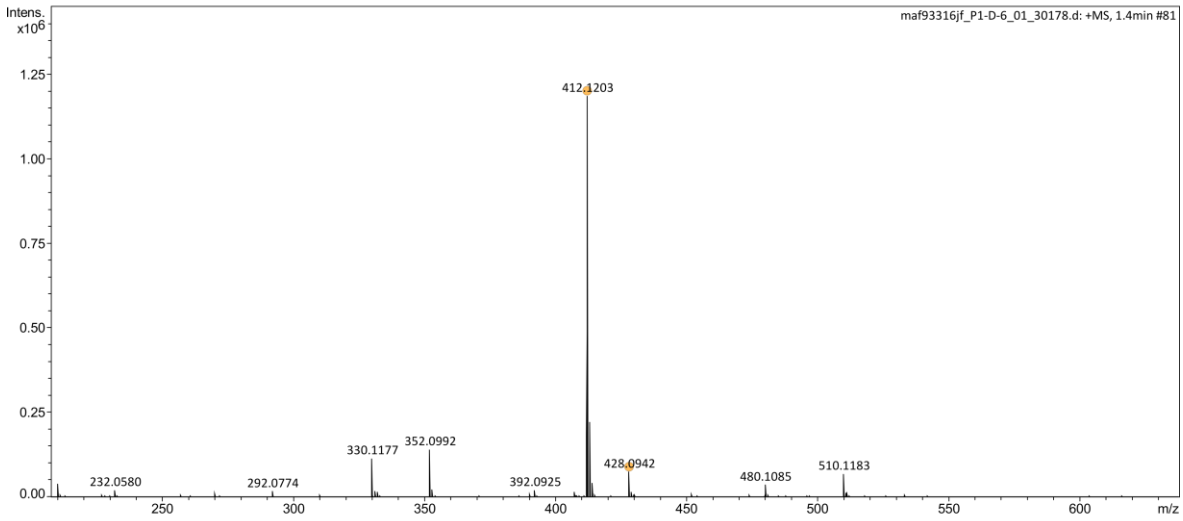



171.83
170.80
170.05
169.19
168.72
90.77
77.44 CHLOROFORM-D
77.12 CHLOROFORM-D
76.81 CHLOROFORM-D
70.75
69.78
67.53
61.60
51.13
23.13
21.02
20.80
20.77
20.64
OAc
AcO
AcO
OAc
NHAc
f1 (ppm)


PerkinElmer Spectrum Version 10.5.4
11 March 2022 11:04
Analyst
Administrator
Date
11 March 2022 11:04
%T
cm-1
3438.57cm-1
1738.63cm-1
1673.88cm-1
1515.70cm-1
1427.05cm-1
1379.78cm-1
1219.80cm-1
1098.69cm-1
1037.12cm-1
1018.38cm-1
1129.16cm-1
940.28cm-1
885.18cm-1
843.17cm-1
685.26cm-1
608.38cm-1
588.15cm-1
550.17cm-1
485.12cm-1
454.35cm-1

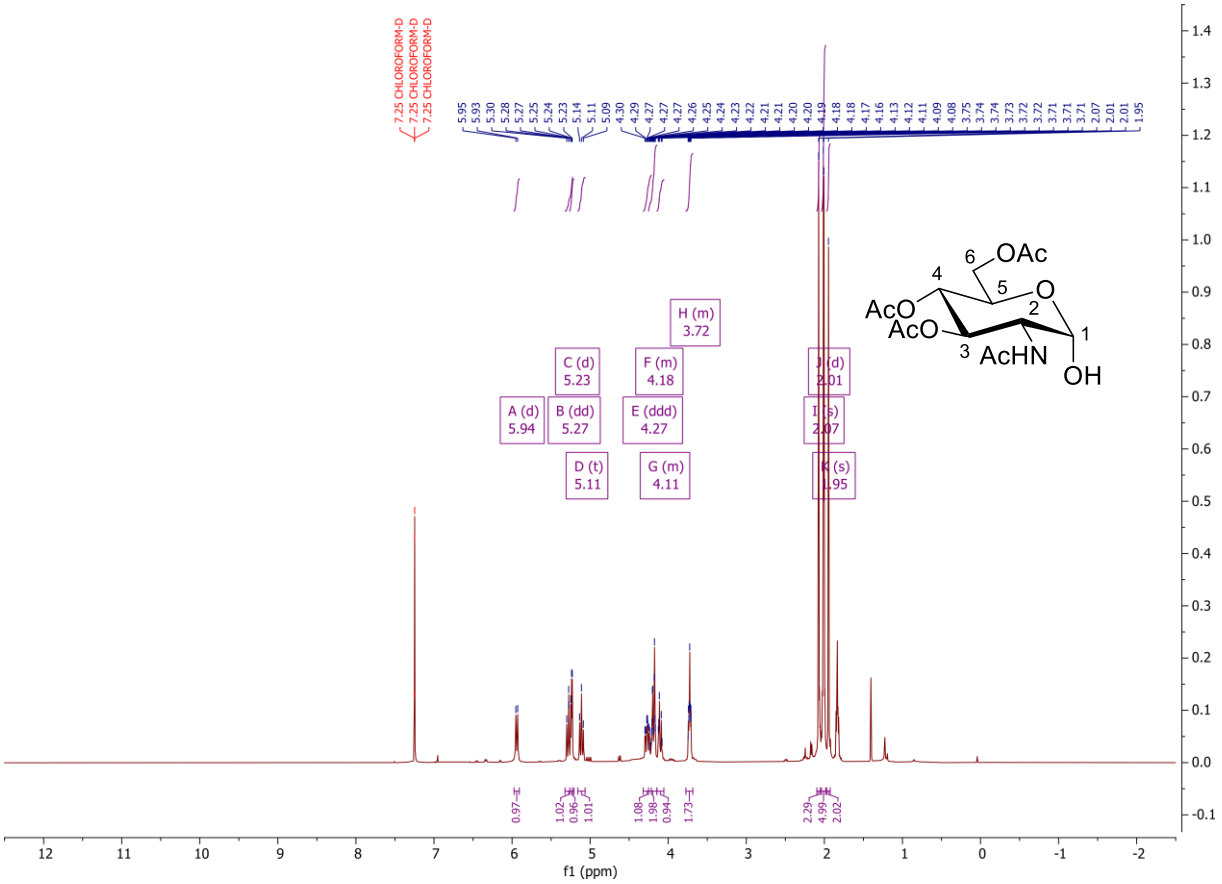
**Analysis Information**

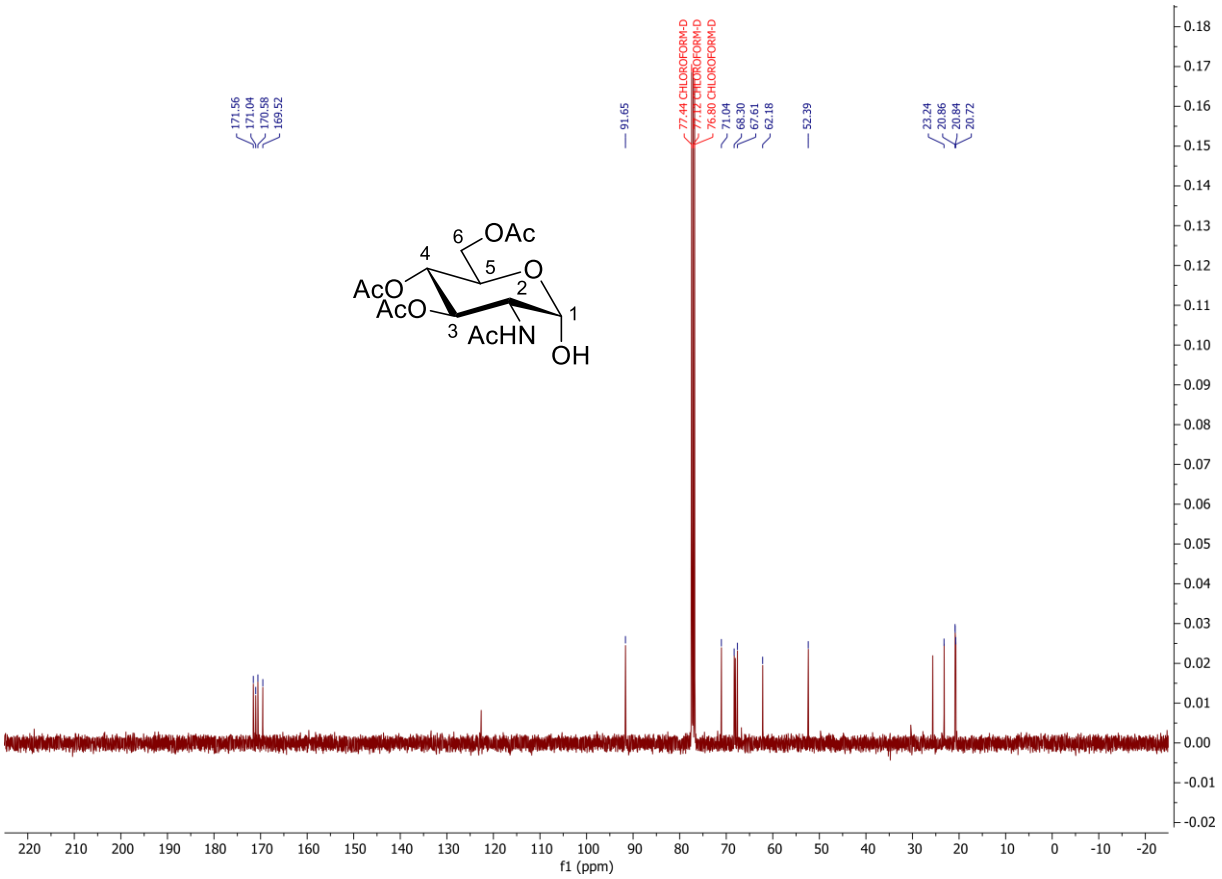
| | | | |
|---|---|---|---|
| Analysis Filename | maf93316jf_P1-D-6_01_30178.d | Acquisition Date | 08/11/2021 11:14:14 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf93316jf | ESI | Positive |

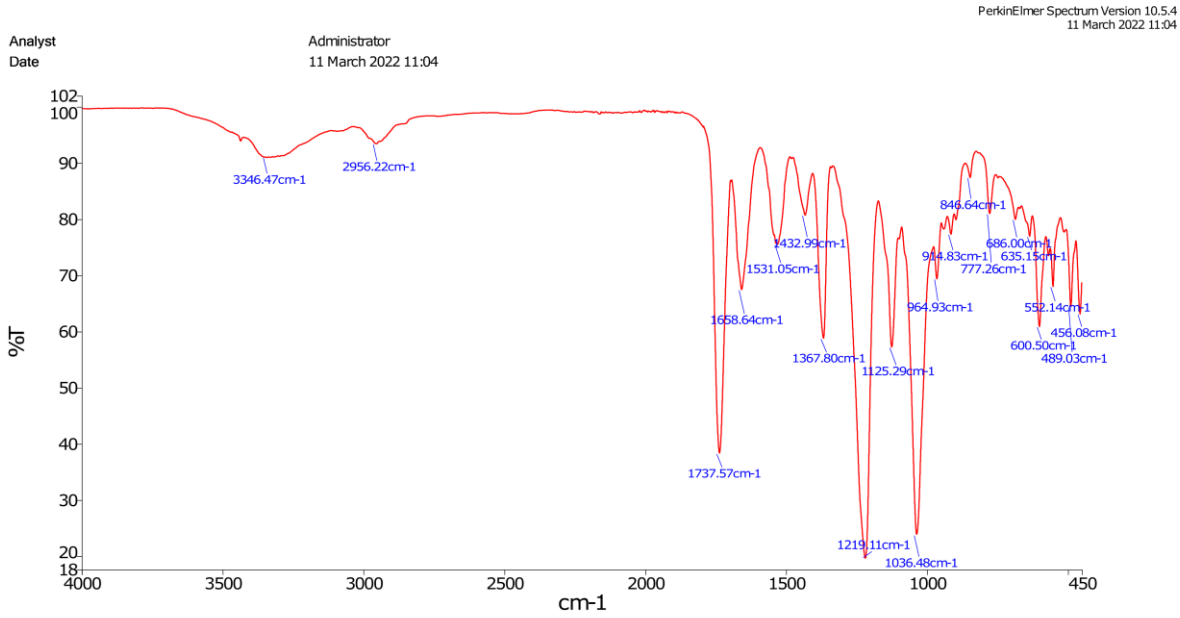



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 412.1203 | 1 | C16H23NNaO10 | 412.1214 | 2.8 | 1.2 | 2.2 | 2.5 |
| 428.0942 | 1 | C16H23KNO10 | 428.0954 | 2.7 | 1.1 | 55.1 | 1.5 |

Appendix 2. Compound **16**:

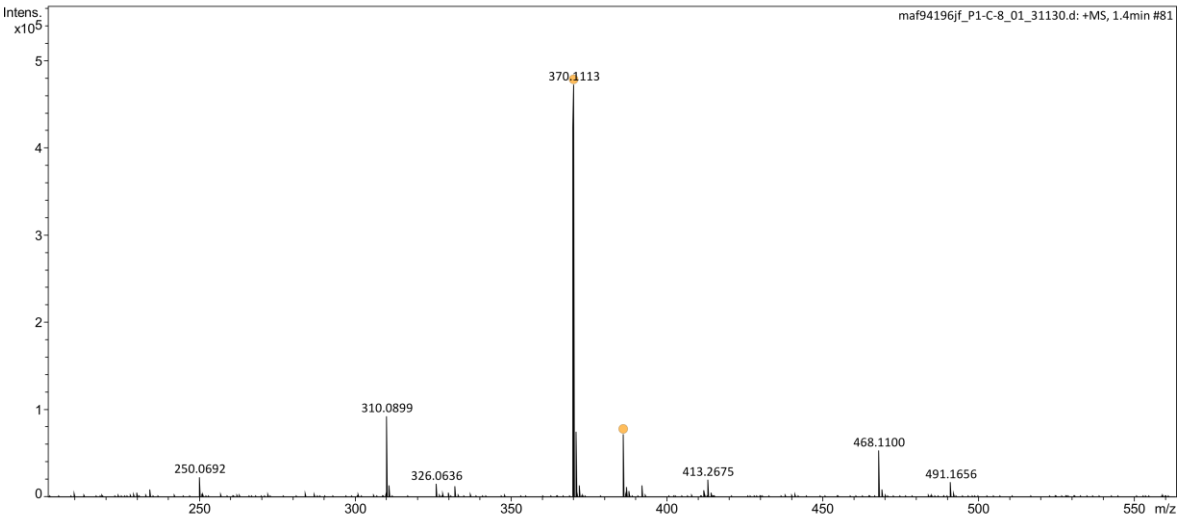




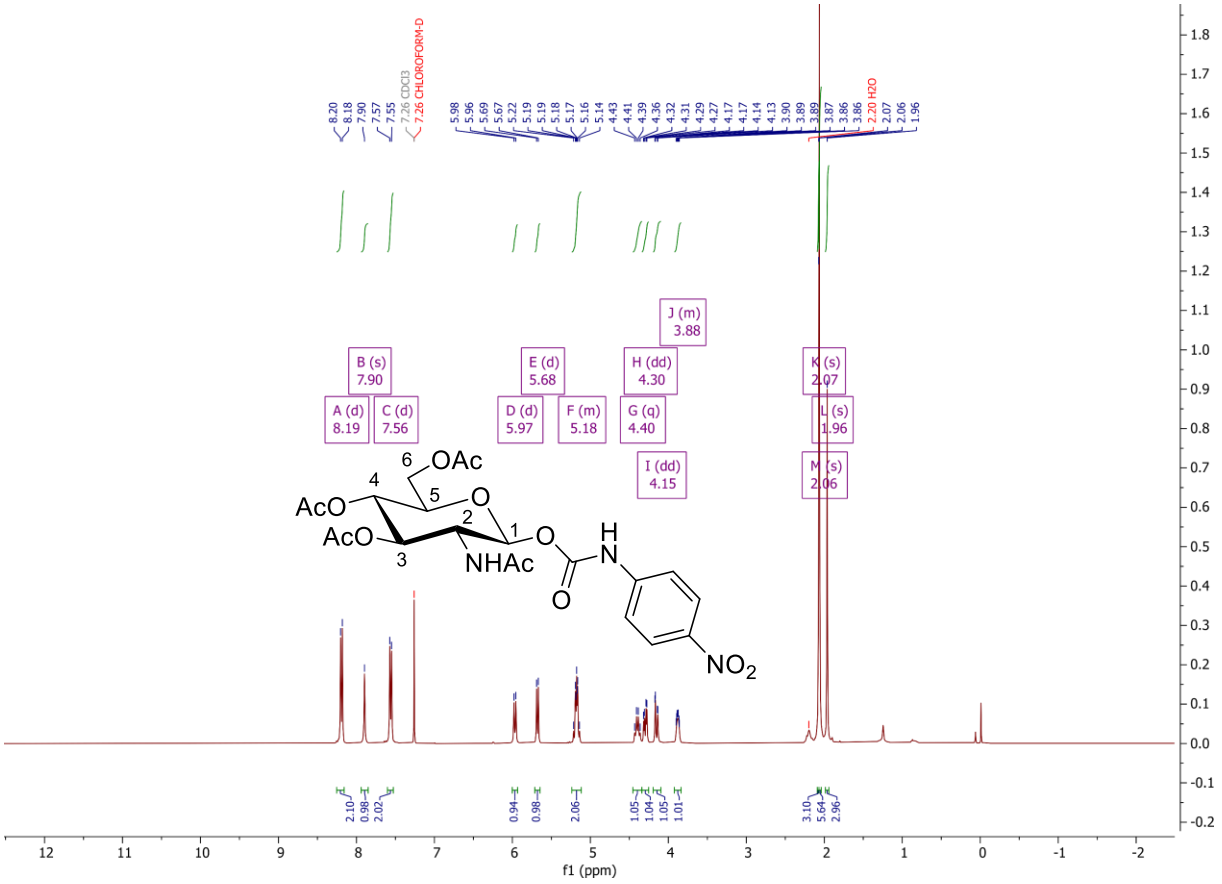

**Analysis Information**

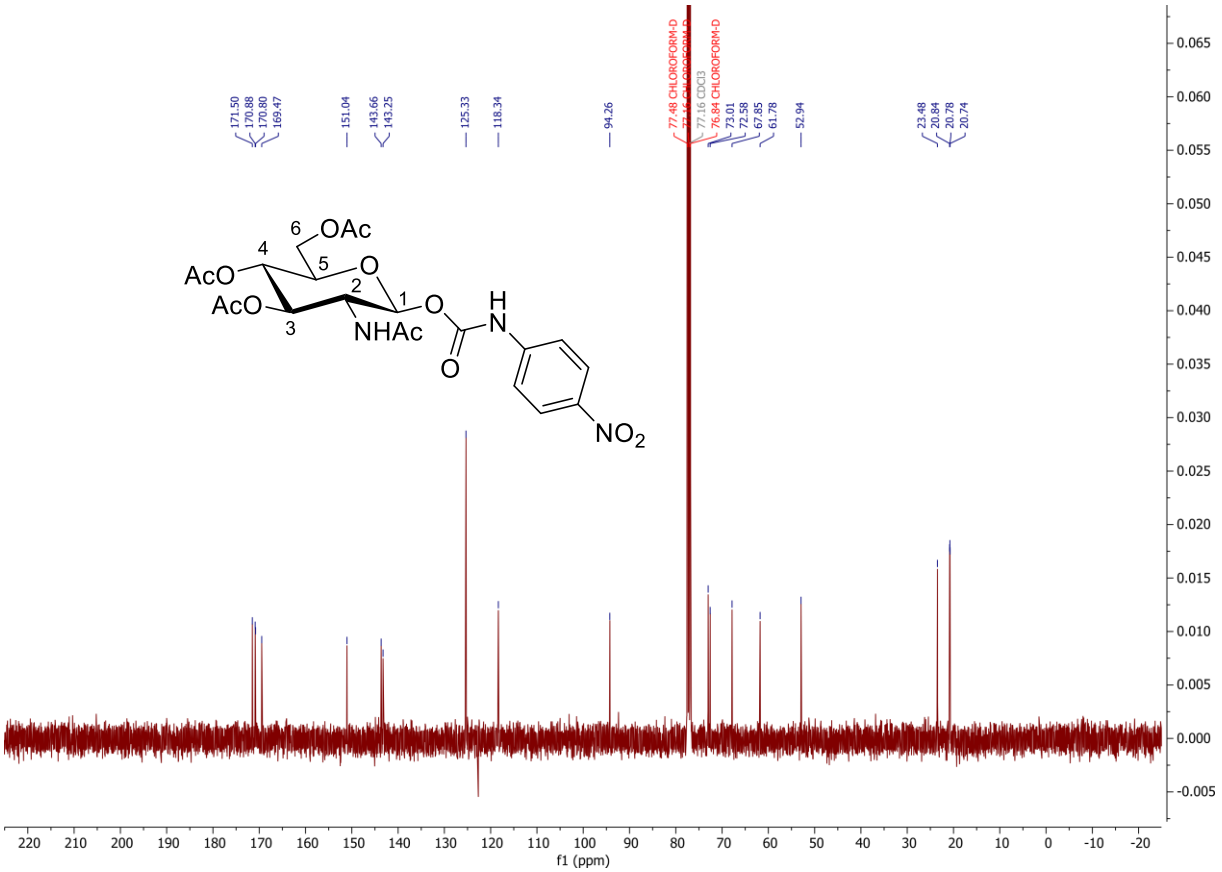
| | | | |
|---|---|---|---|
| Analysis Filename | maf94196jf_P1-C-8_01_31130.d | Acquisition Date | 13/12/2021 09:31:43 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf94196jf | ESI | Positive |

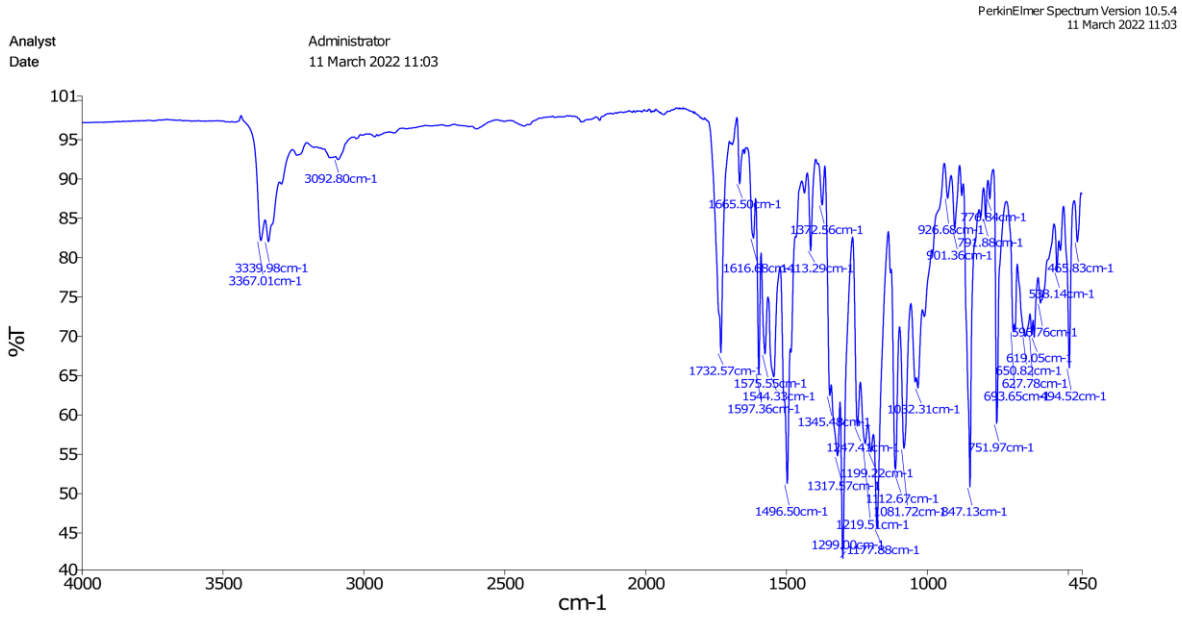



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 370.1113 | 1 | C14H21NNaO9 | 370.1109 | -1.1 | -0.4 | 1.7 | -1.1 |
| 386.0853 | 1 | C14H21KNO9 | 386.0848 | -1.2 | -0.5 | 3.5 | -2.3 |

Appendix 3. Compound **17:**

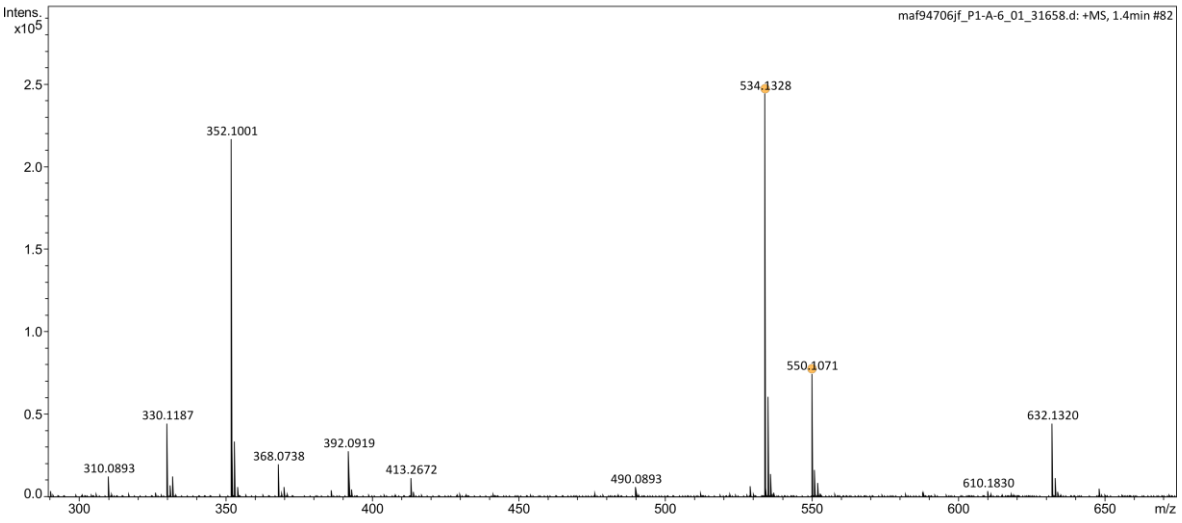




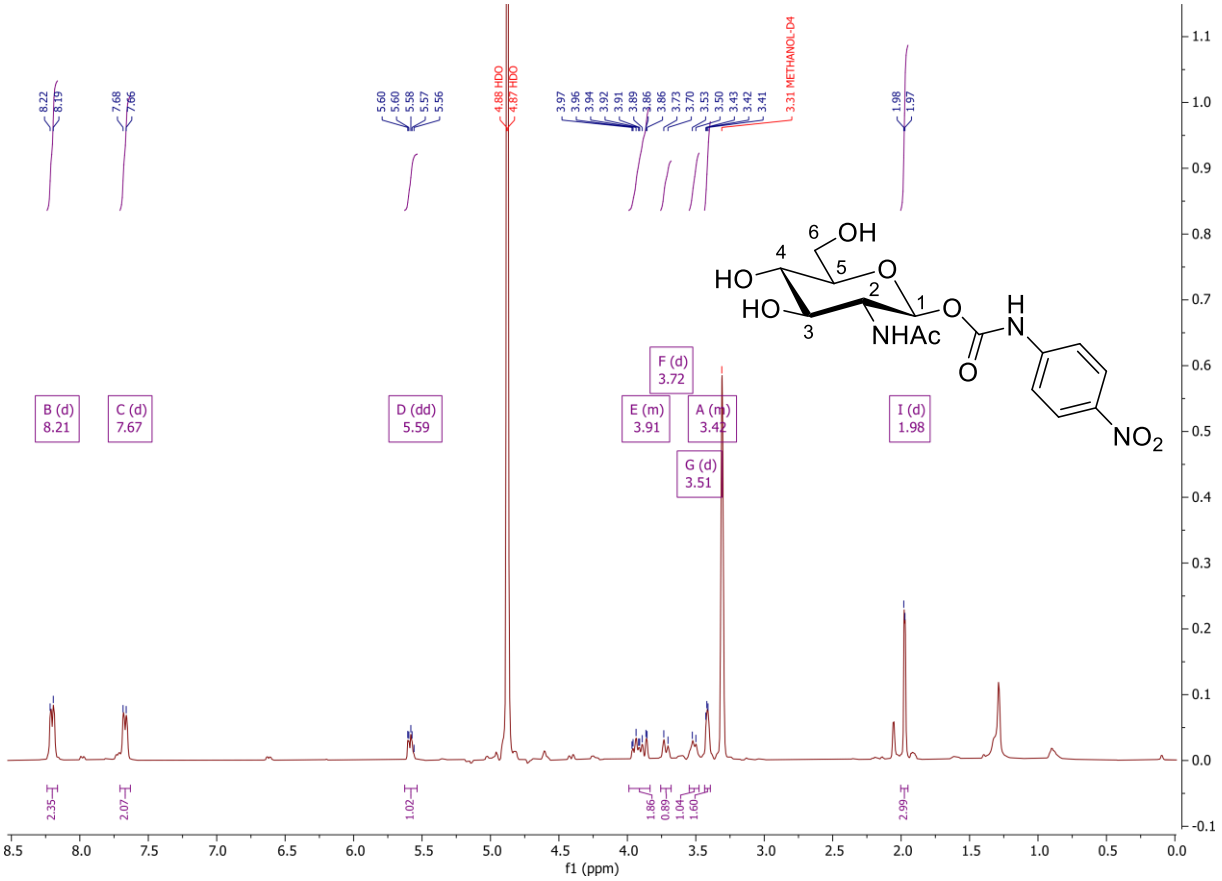

**Analysis Information**

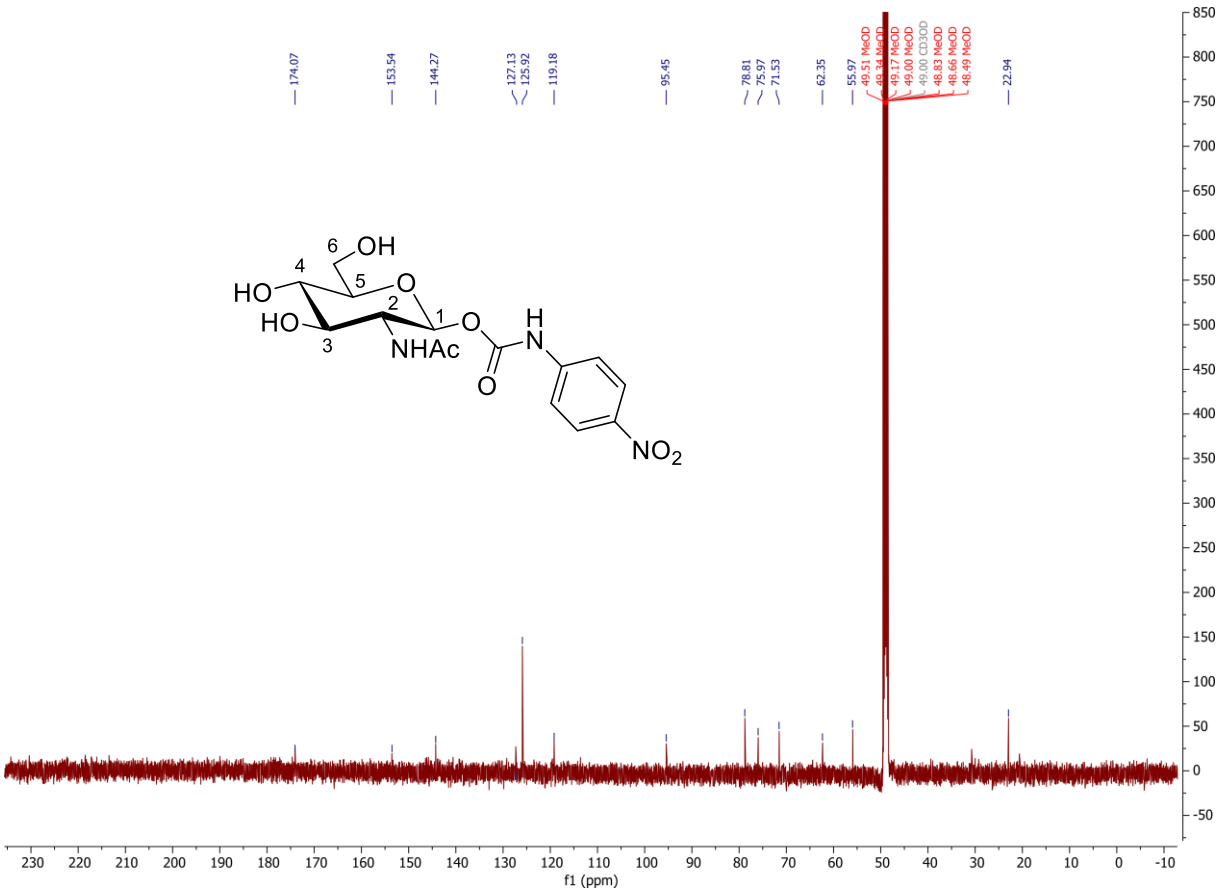
| | | | |
|---|---|---|---|
| Analysis Filename | maf94706jf_P1-A-6_01_31658.d | Acquisition Date | 21/01/2022 12:06:53 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf94706jf | ESI | Positive |

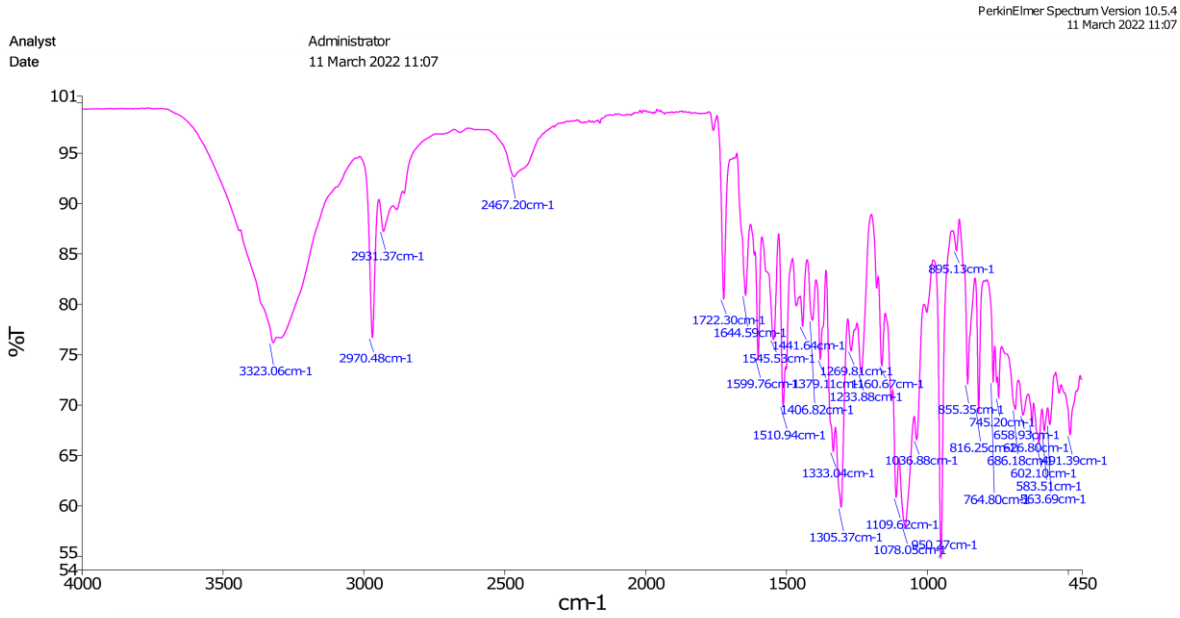



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 534.1328 | 1 | C21H25N3NaO12 | 534.1330 | 0.4 | 0.2 | 2.8 | -1.0 |
| 550.1071 | 1 | C21H25KN3O12 | 550.1070 | -0.2 | -0.1 | 13.6 | -0.8 |

Appendix 4. Compound **7**:

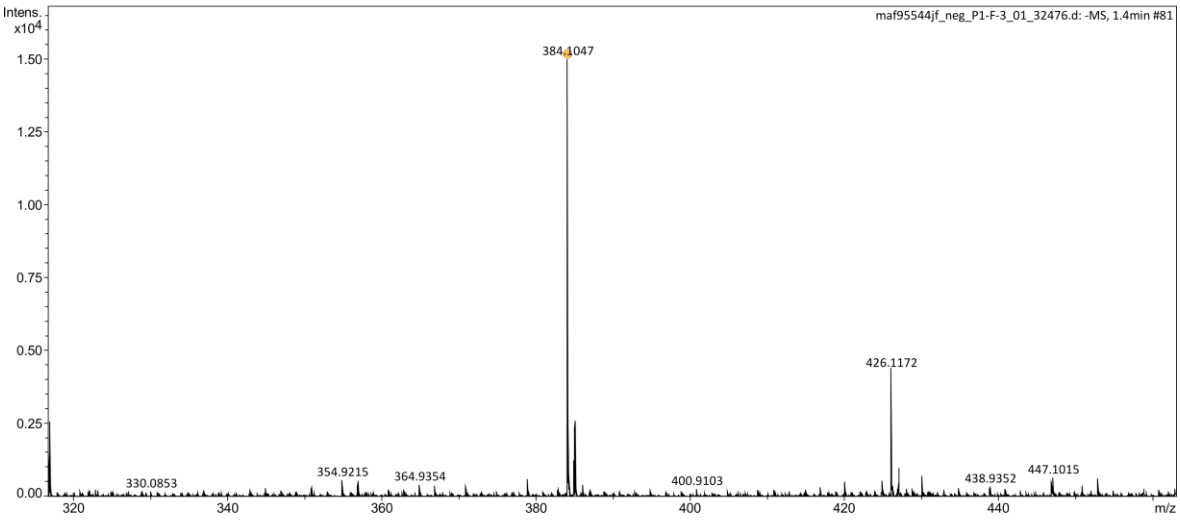




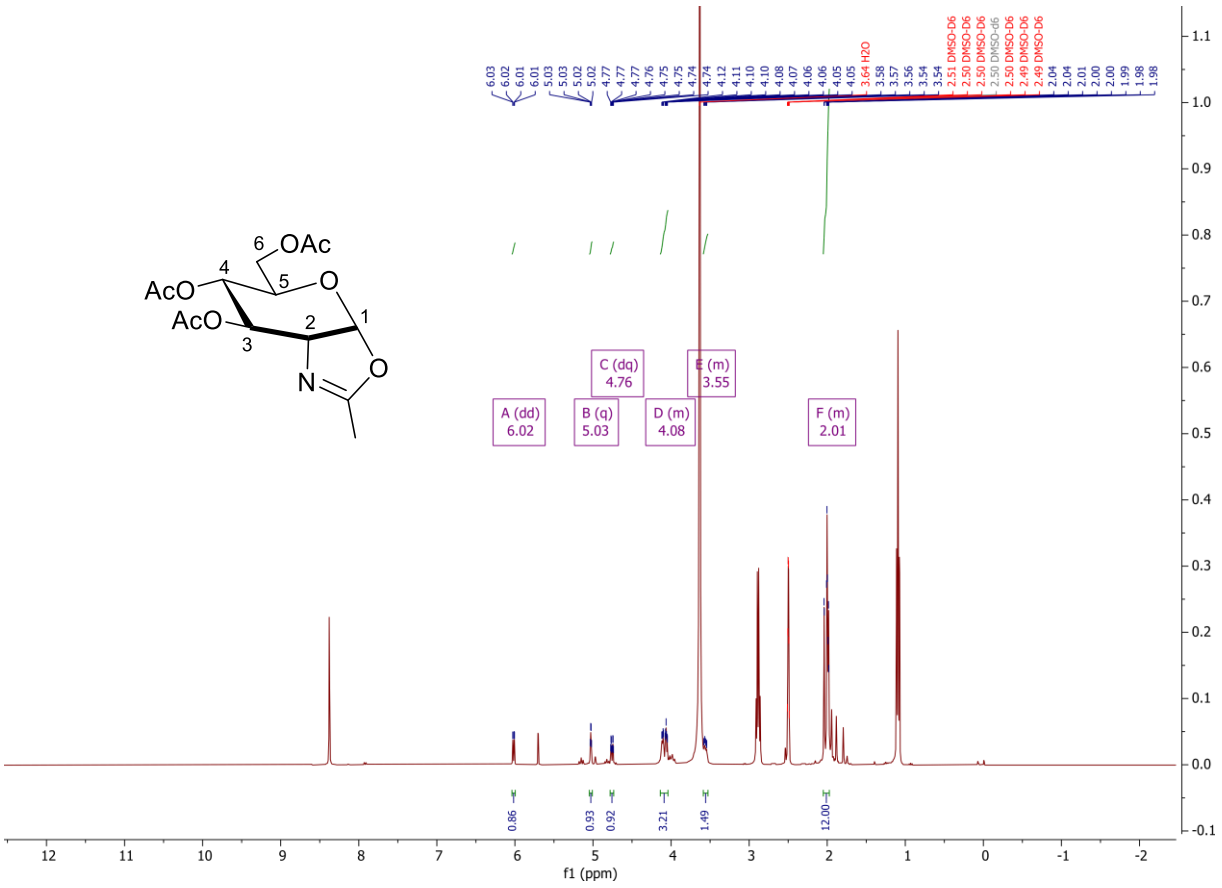

**Analysis Information**

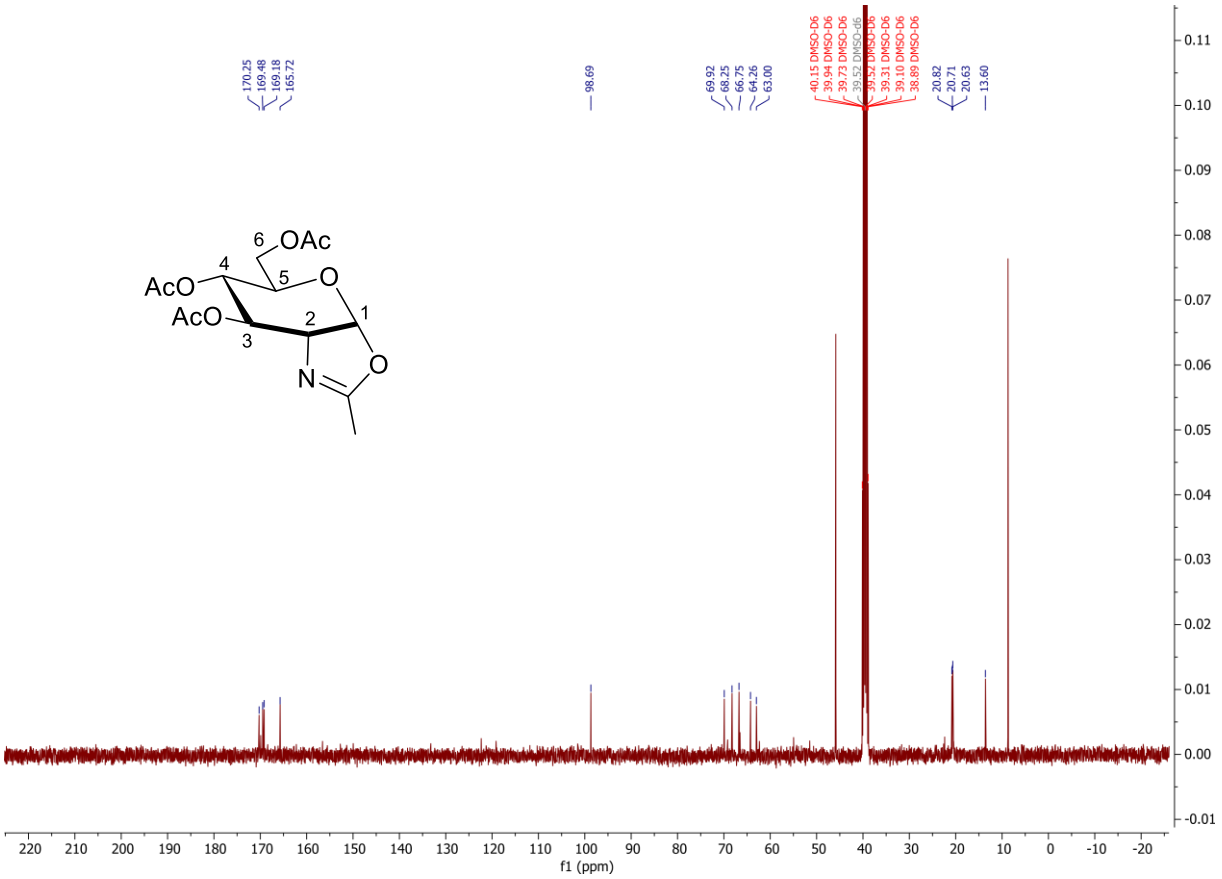
| | | | |
|---|---|---|---|
| Analysis Filename | maf95544jf_neg_P1-F-3_01_32476.d | Acquisition Date | 24/02/2022 16:17:32 |
| Method | ESI_low mass neg_2c1s.m | Instrument | compact |
| Submission Name | maf95544jf_neg | ESI | Negative |




| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 384.1047 | 1 | C15H18N3O9 | 384.1049 | 0.5 | 0.2 | 5.0 | 0.7 |

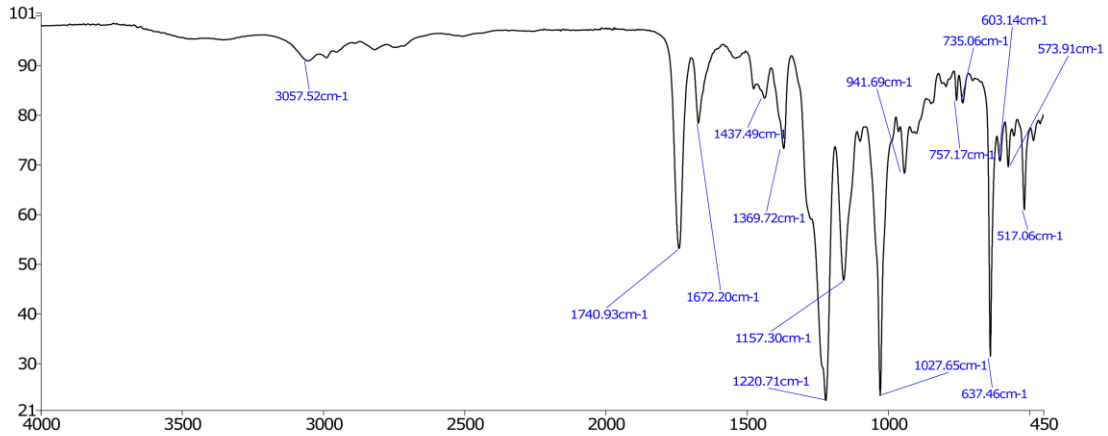
Appendix 5. Compound **21**:



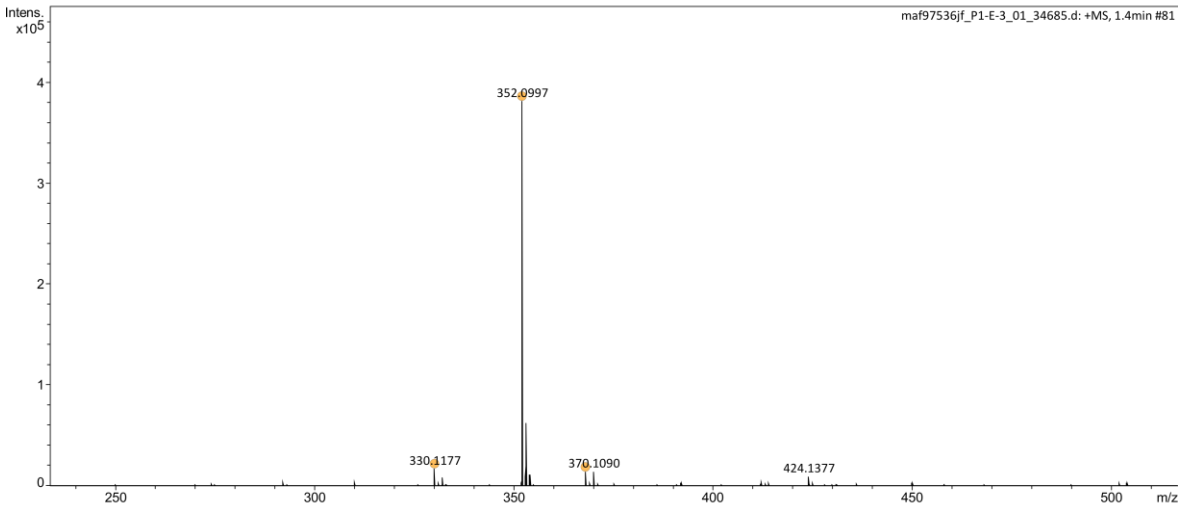
PerkinElmer Spectrum Version 10.5.4
21 July 2022 15:51

Analyst Administrator
Date 21 July 2022 15:51



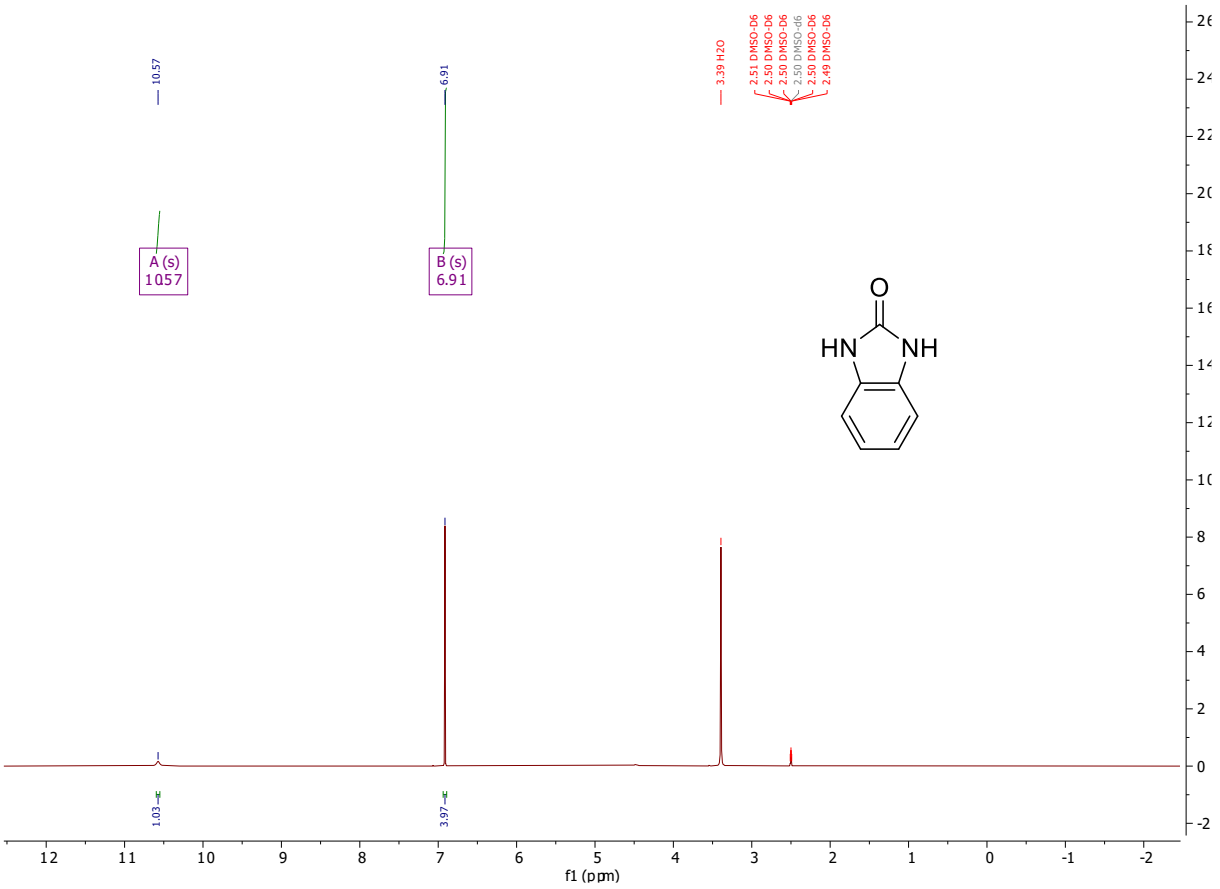

**Analysis Information**

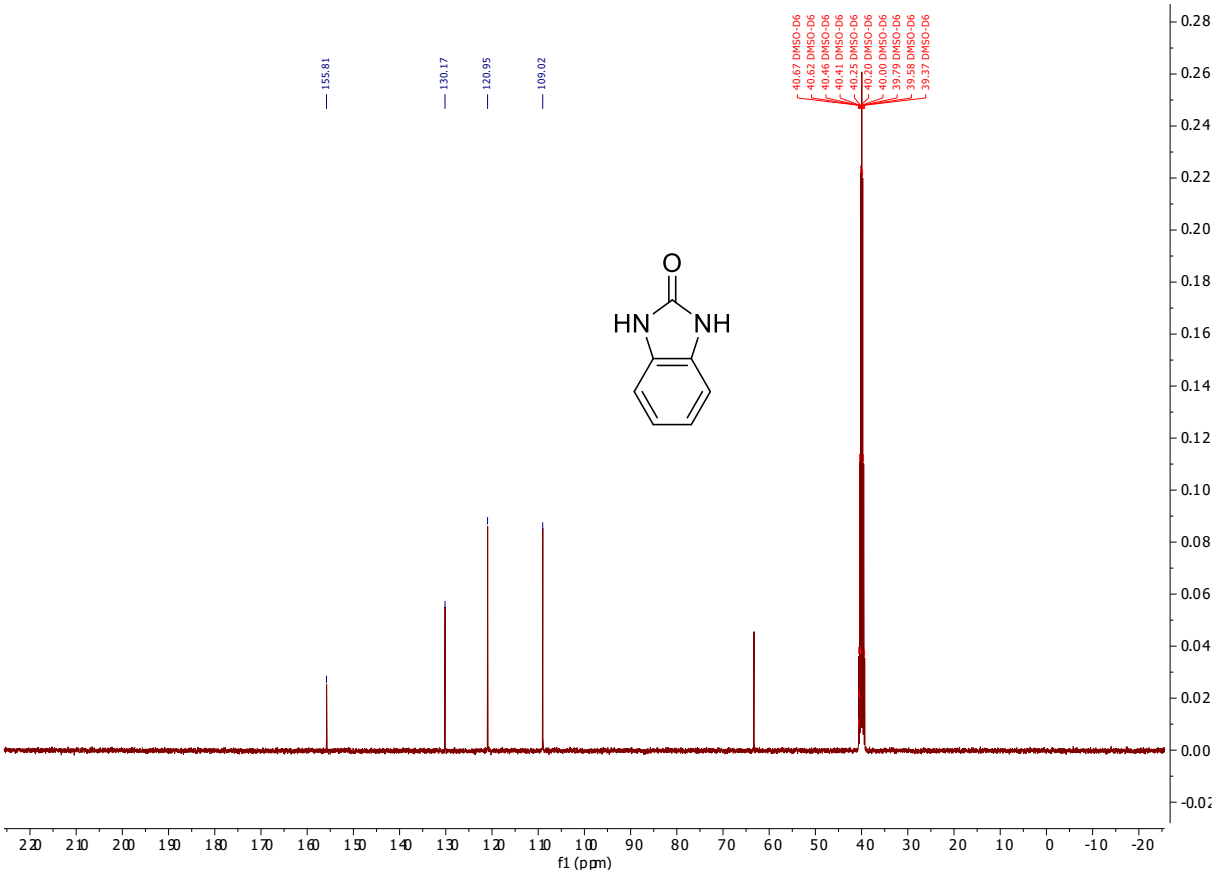
| | | | |
|---|---|---|---|
| Analysis Filename | maf97536jf_P1-E-3_01_34685.d | Acquisition Date | 13/06/2022 11:07:27 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf97536jf | ESI | Positive |

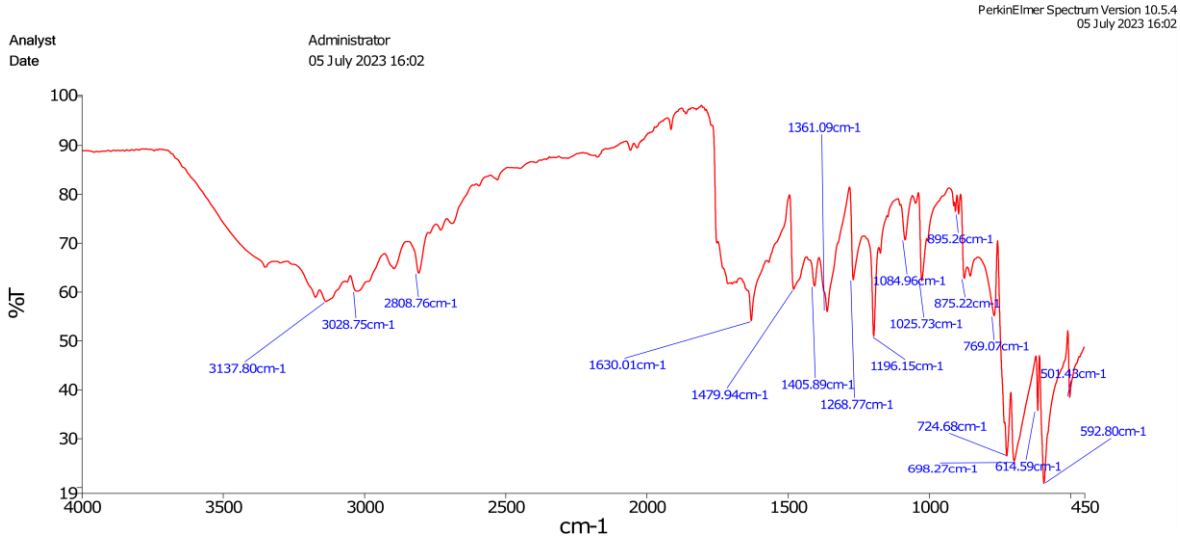



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 330.1177 | 1 | C14H20NO8 | 330.1183 | 2.0 | 0.6 | 16.6 | 2.0 |
| 352.0997 | 1 | C14H19NNaO8 | 352.1003 | 1.6 | 0.6 | 2.2 | 1.6 |
| 368.0732 | 1 | C14H19KNO8 | 368.0742 | 2.9 | 1.1 | 52.0 | 1.1 |

Appendix 6. Compound **26**:

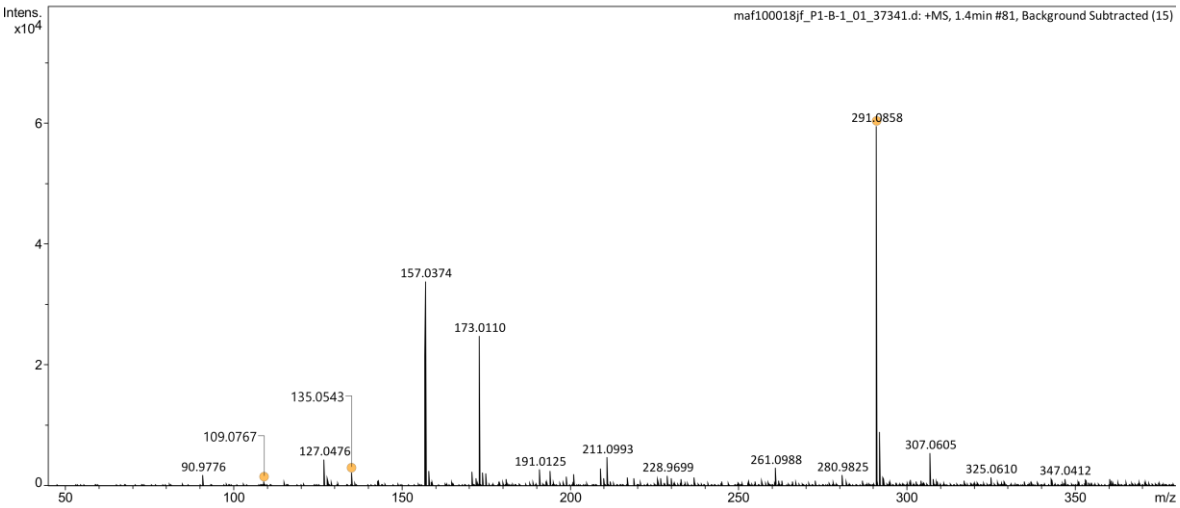




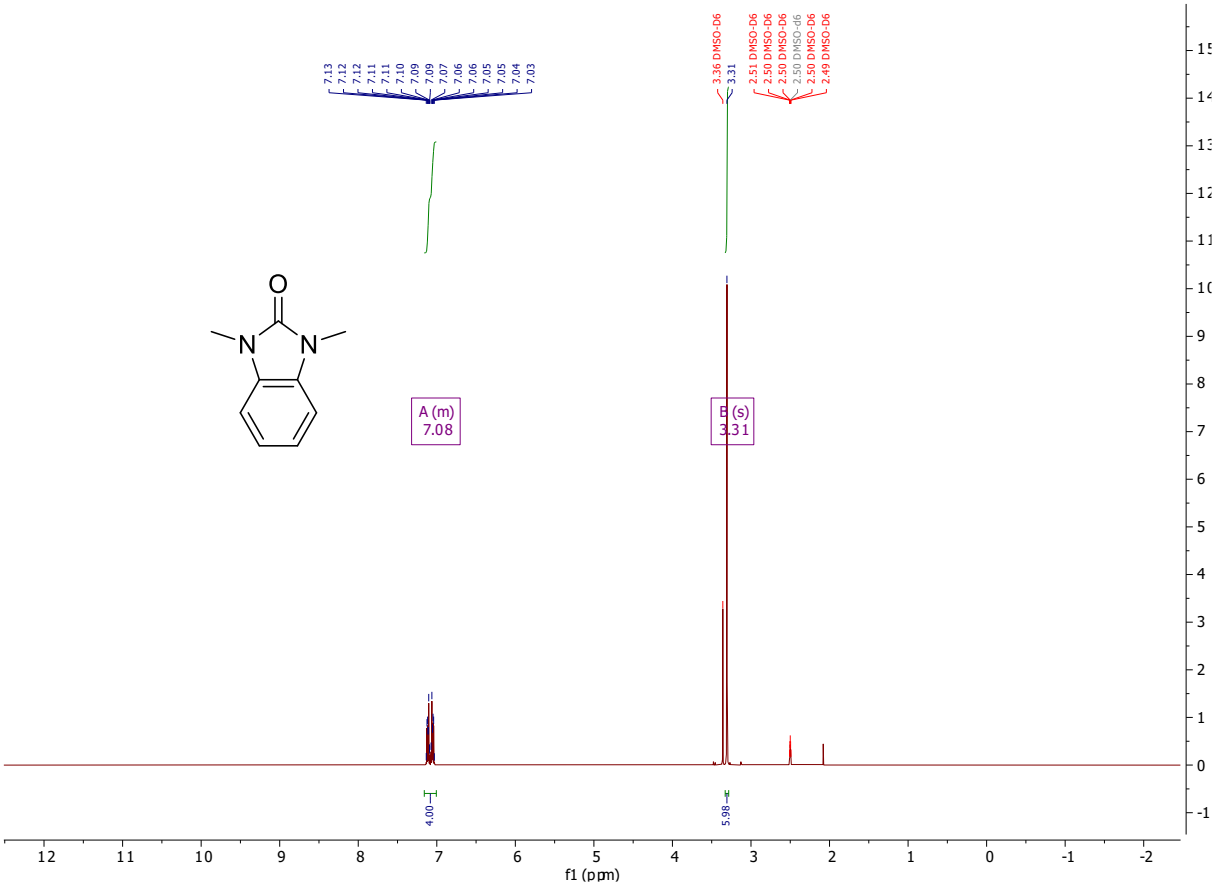

**Analysis Information**

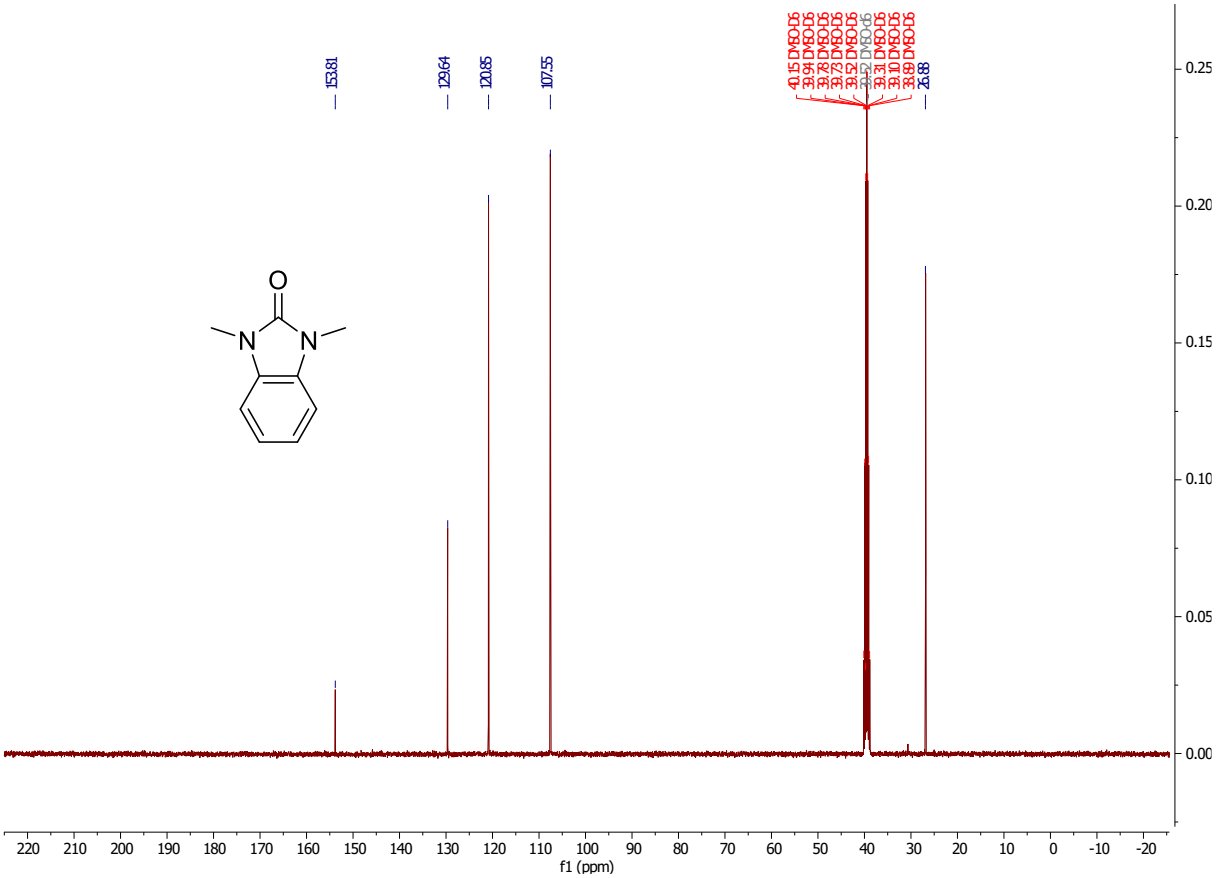
| | | | |
|---|---|---|---|
| Analysis Filename | maf100018jf_P1-B-1_01_37341.d | Acquisition Date | 24/10/2022 09:21:01 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf100018jf | ESI | Positive |

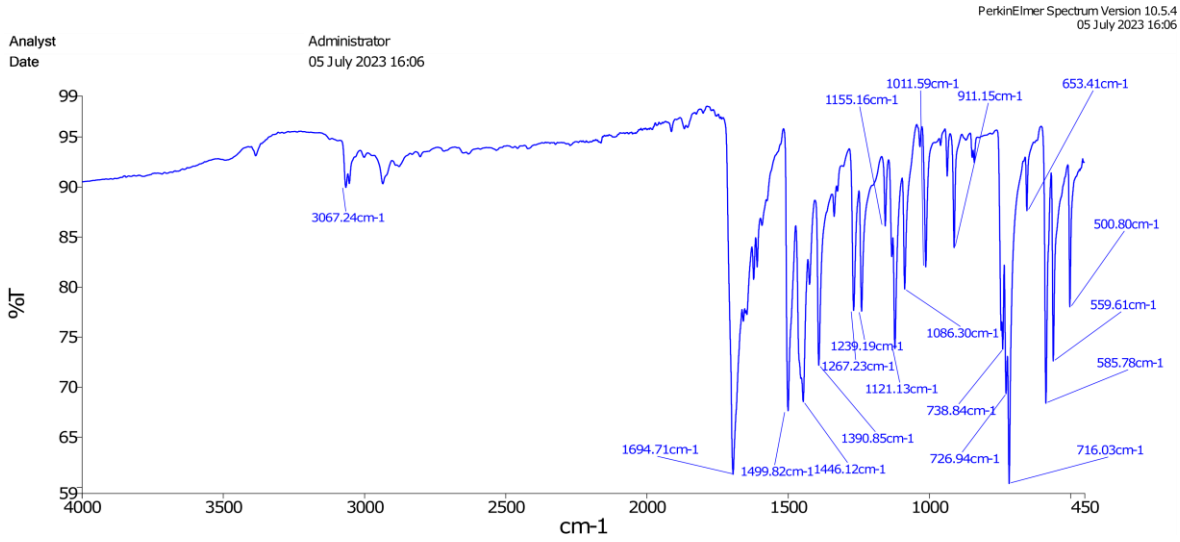



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 109.0767 | 1 | C6H9N2 | 109.0760 | -6.6 | -0.7 | n.a. | -3.7 |
| 135.0543 | 1 | C7H7N2O | 135.0553 | 7.7 | 1.0 | n.a. | 4.3 |
| 291.0858 | 1 | C14H12N4NaO2 | 291.0852 | -2.0 | -0.6 | 11.3 | -2.3 |

Appendix 7. Compound **24**:

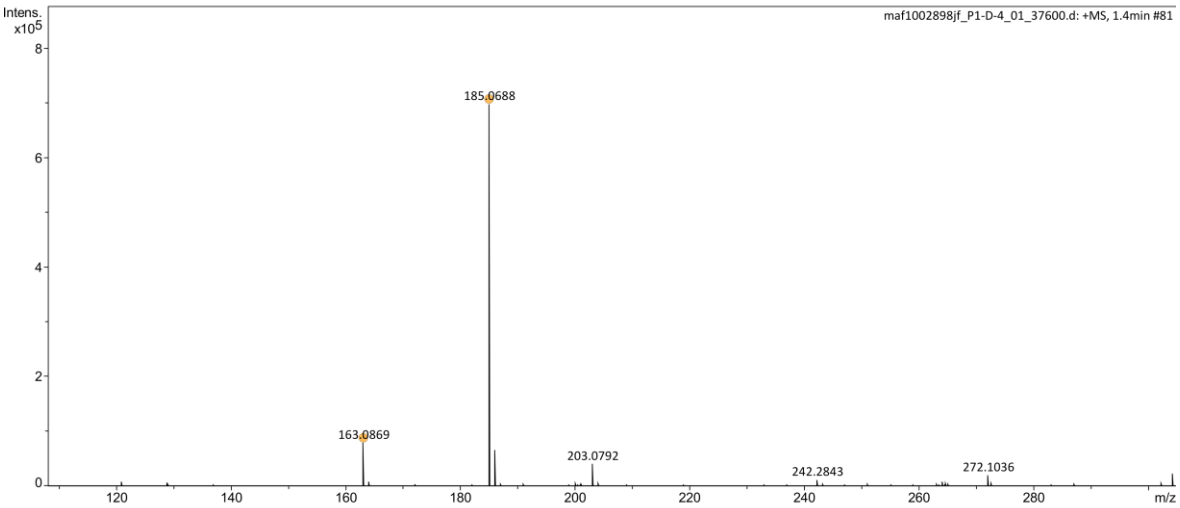




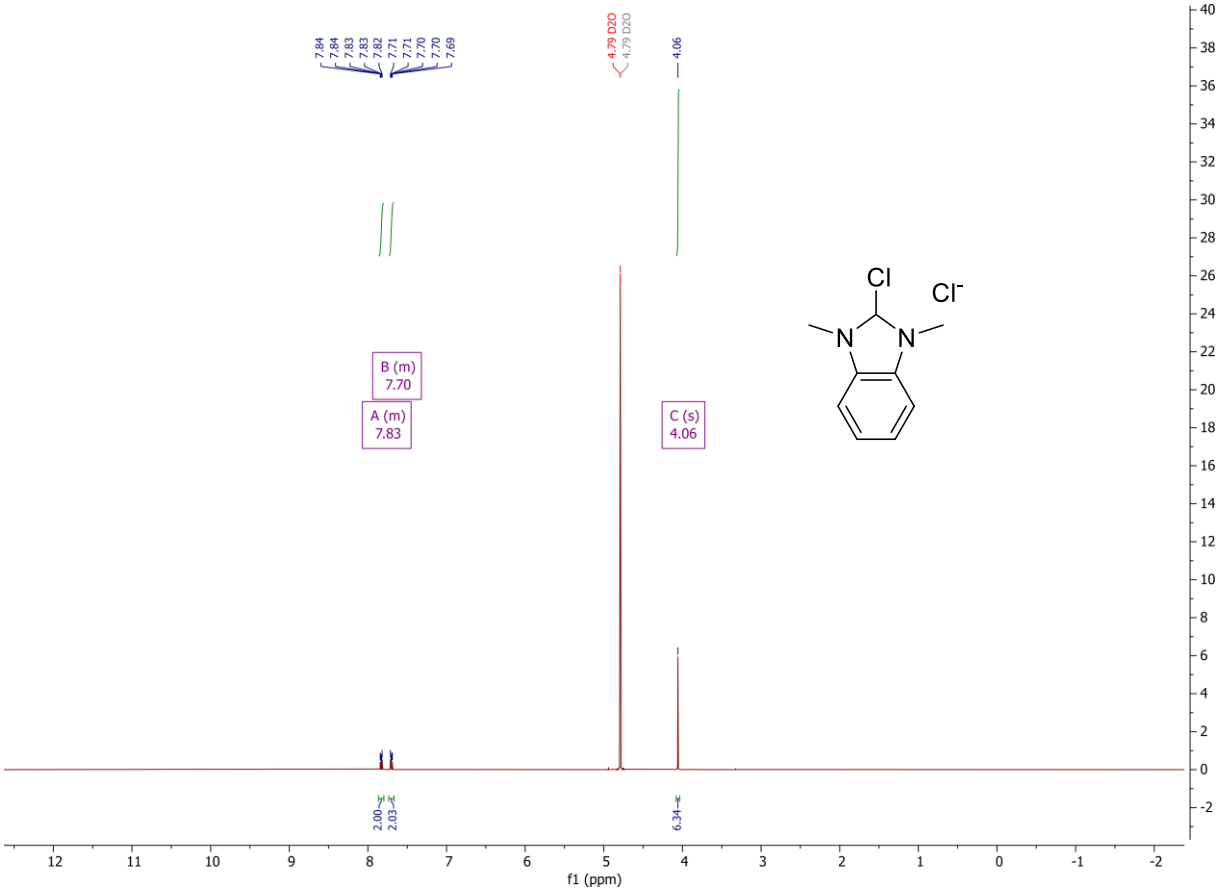

**Analysis Information**

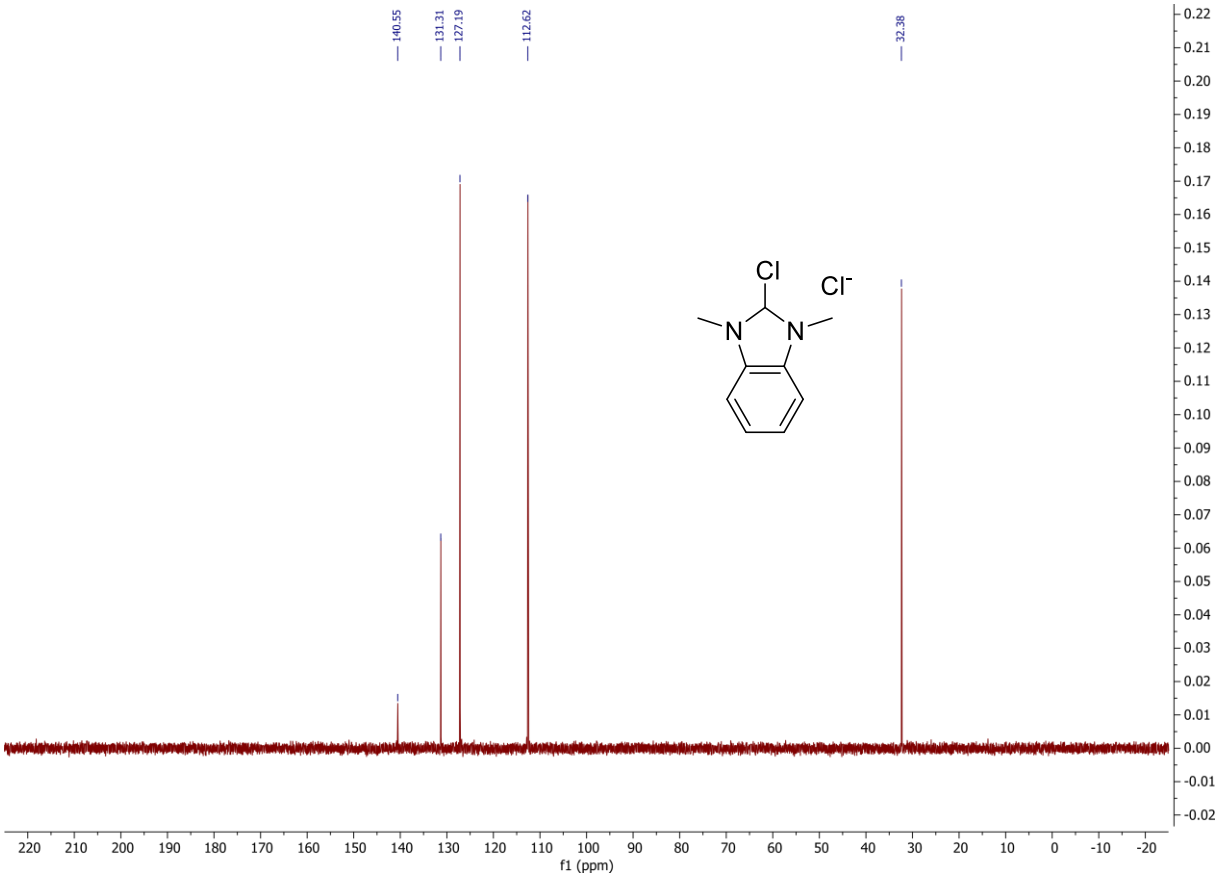
| | | | |
|---|---|---|---|
| Analysis Filename | maf1002898jf_P1-D-4_01_37600.d | Acquisition Date | 03/11/2022 14:05:47 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf1002898jf | ESI | Positive |

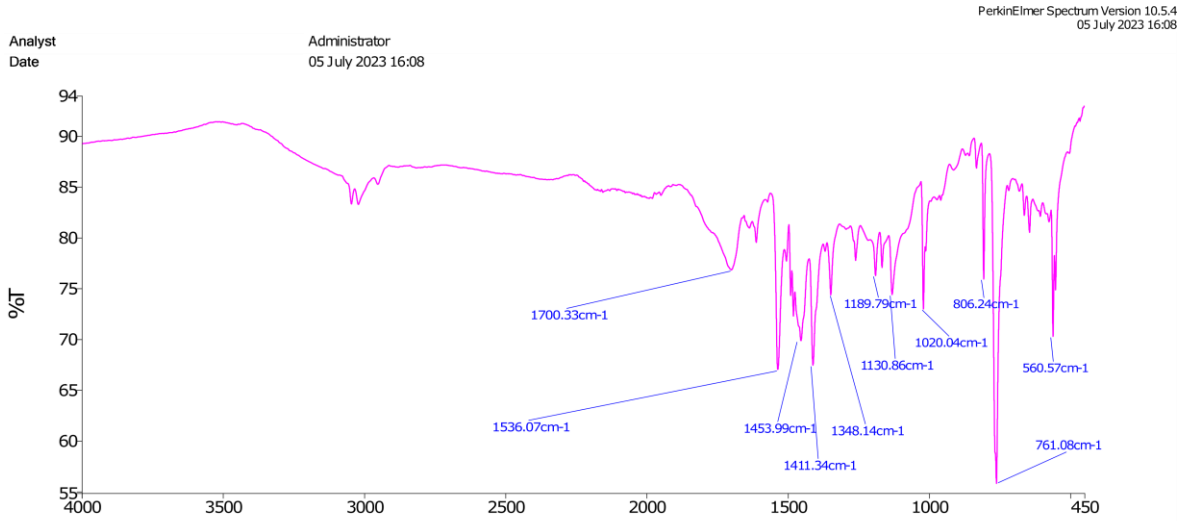



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 163.0869 | 1 | C9H11N2O | 163.0866 | -1.8 | -0.3 | 4.9 | -1.0 |
| 185.0688 | 1 | C9H10N2NaO | 185.0685 | -1.5 | -0.3 | 6.7 | -0.8 |

Appendix 8. Compound **23:**

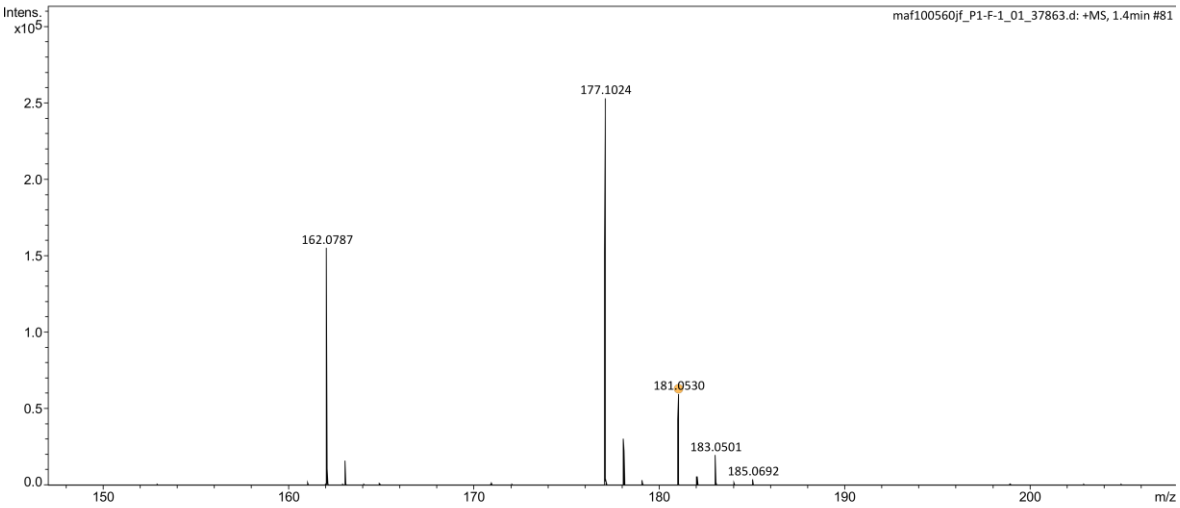




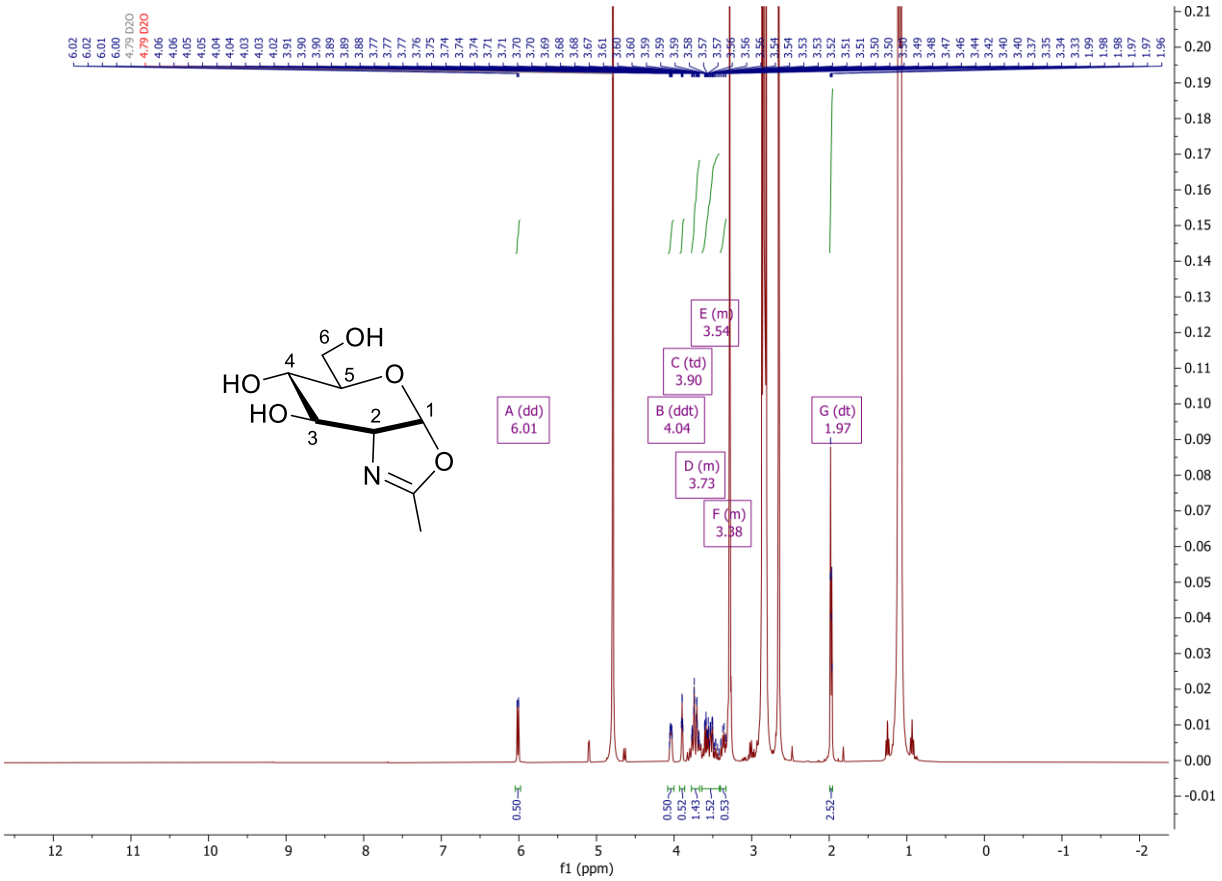

**Analysis Information**

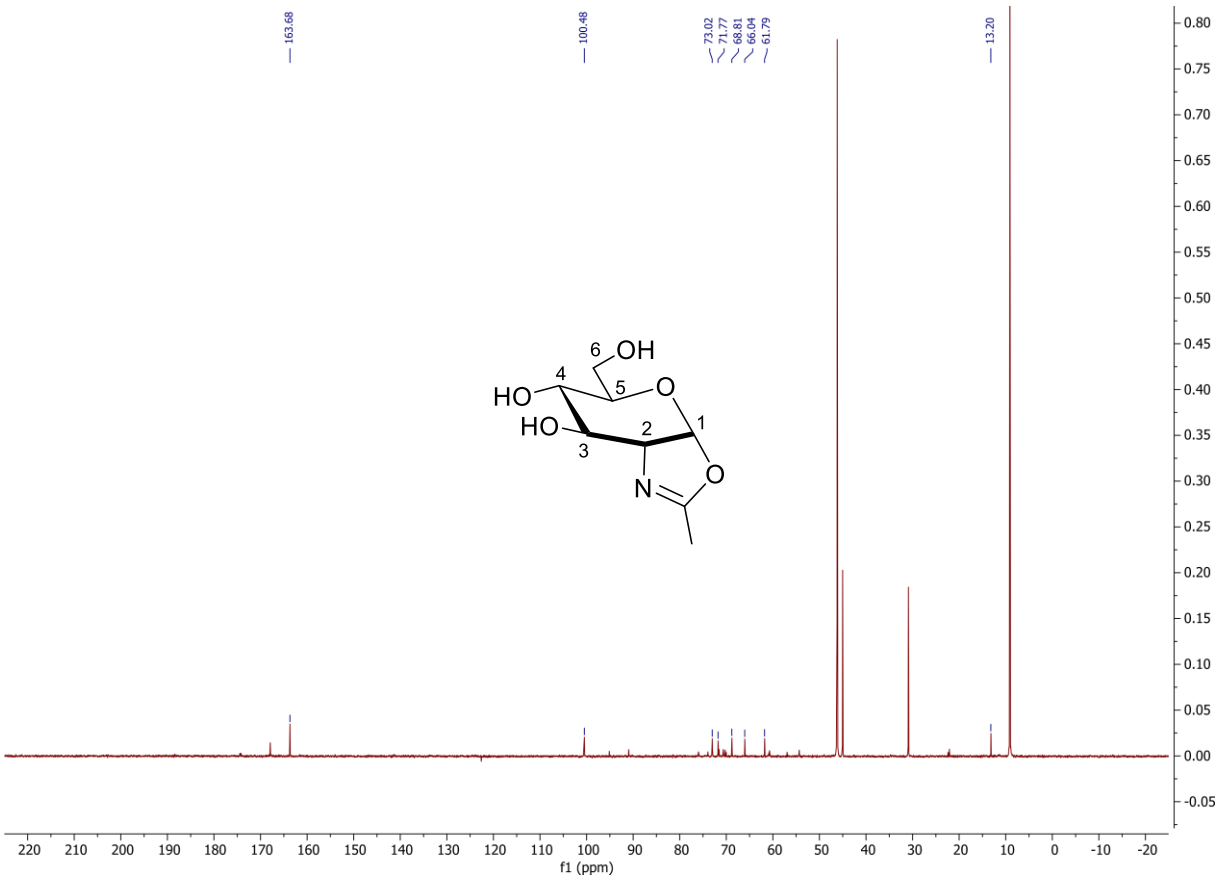
| | | | |
|---|---|---|---|
| Analysis Filename | maf100560jf_P1-F-1_01_37863.d | Acquisition Date | 14/11/2022 11:34:06 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf100560jf | ESI | Positive |

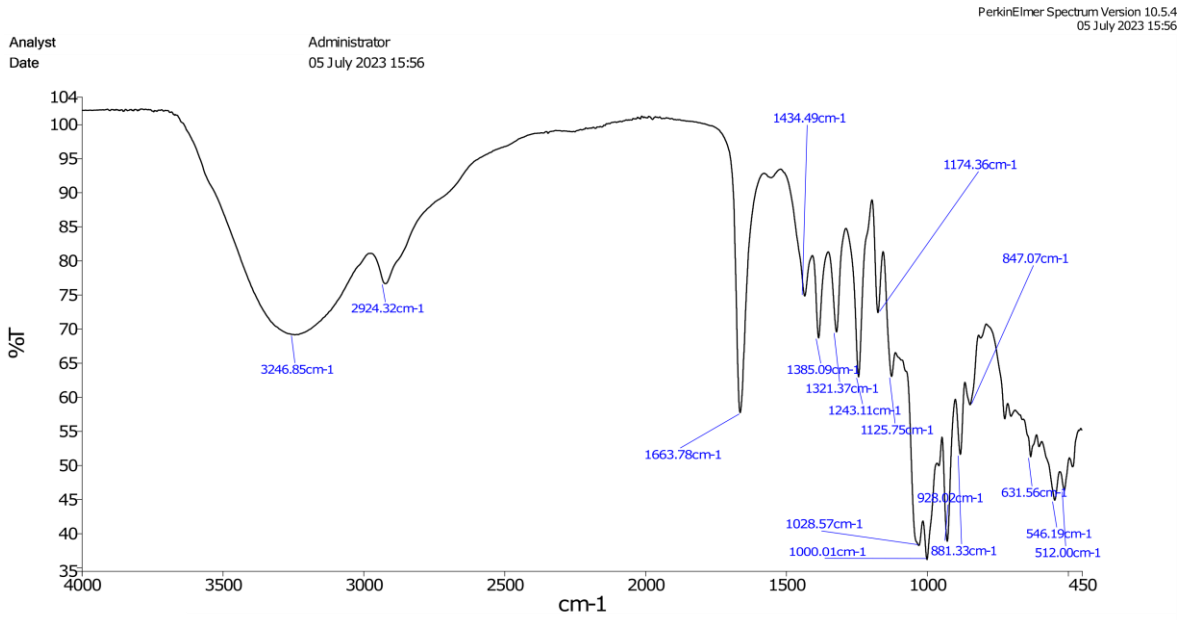



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 181.0530 | 1 | C9H10ClN2 | 181.0527 | -1.6 | -0.3 | 6.3 | -2.8 |

Appendix 9. Compound **5:**

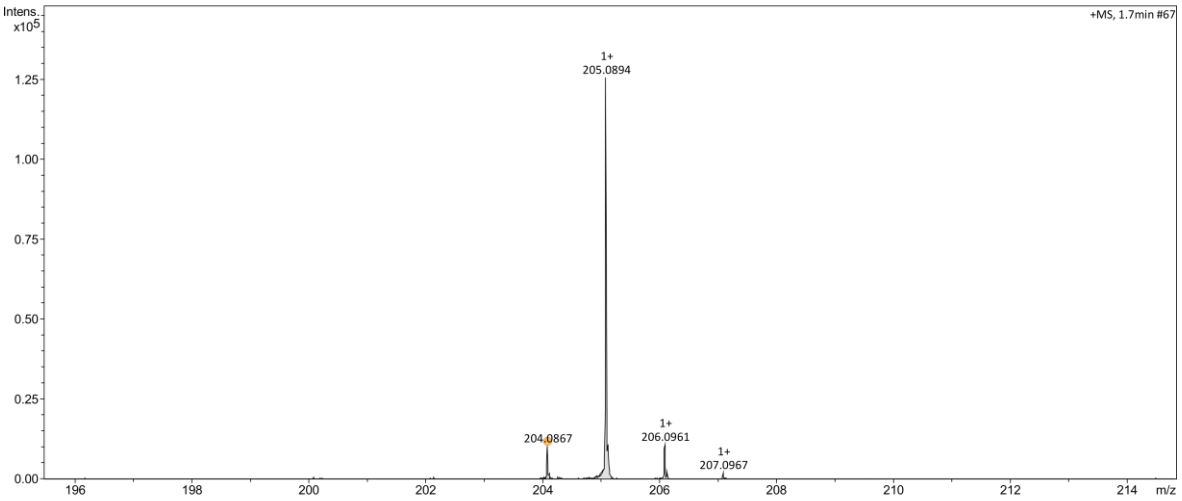




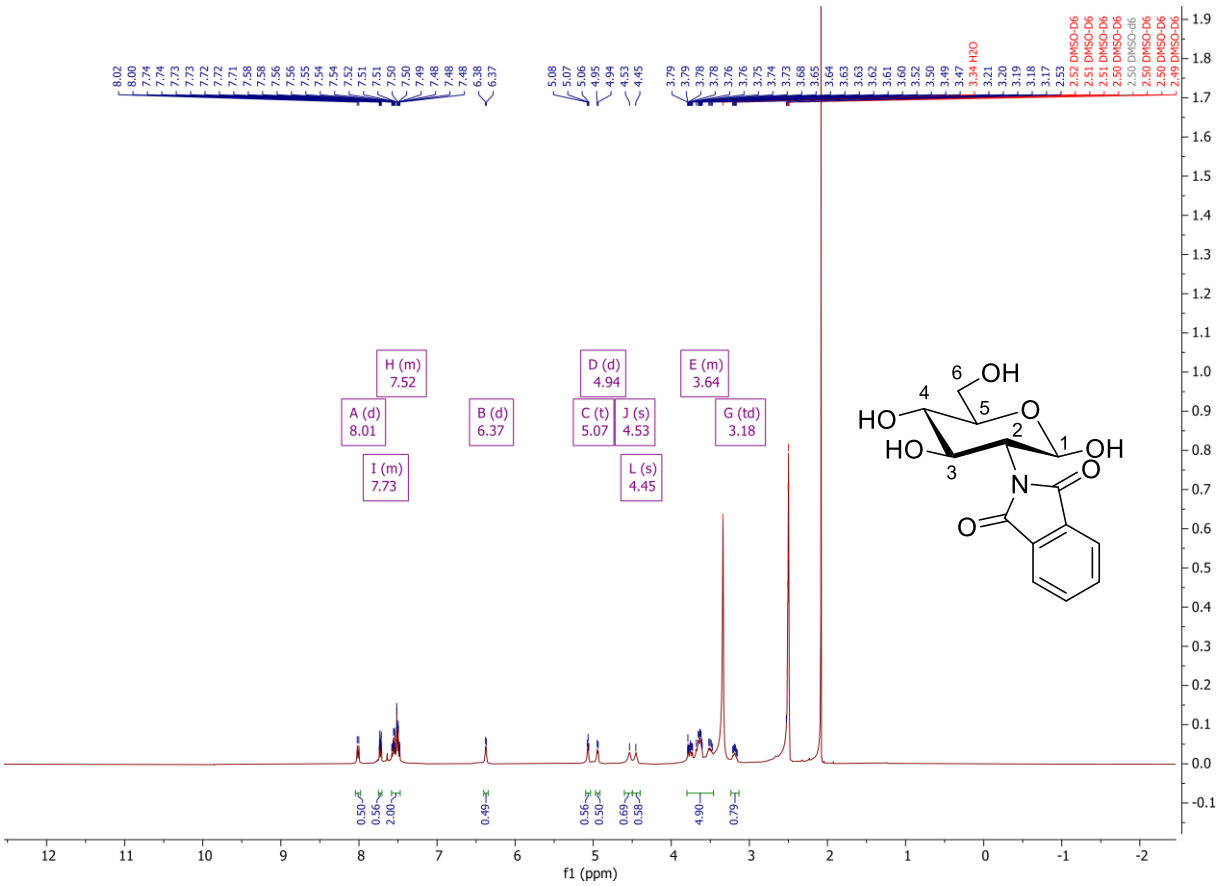

**Analysis Information**

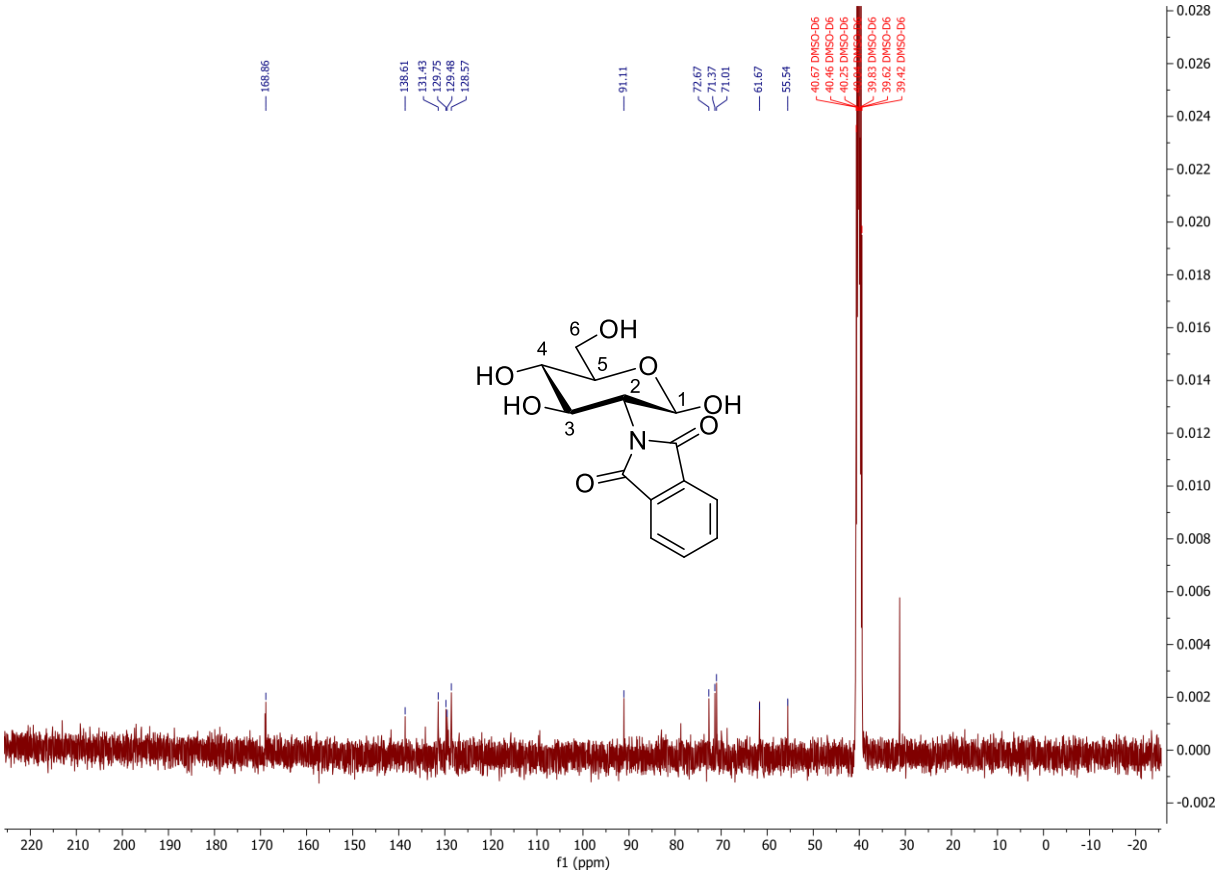
| | | | |
|---|---|---|---|
| Analysis Filename | maf97363jf.d | Acquisition Date | 01/06/2022 08:07:39 |
| Method | APCI_LOW MASS_oTOF34 METHODS.m | Instrument | micrOTOF |
| Submission Name | jwg71051rm_3 | APCI | Positive |

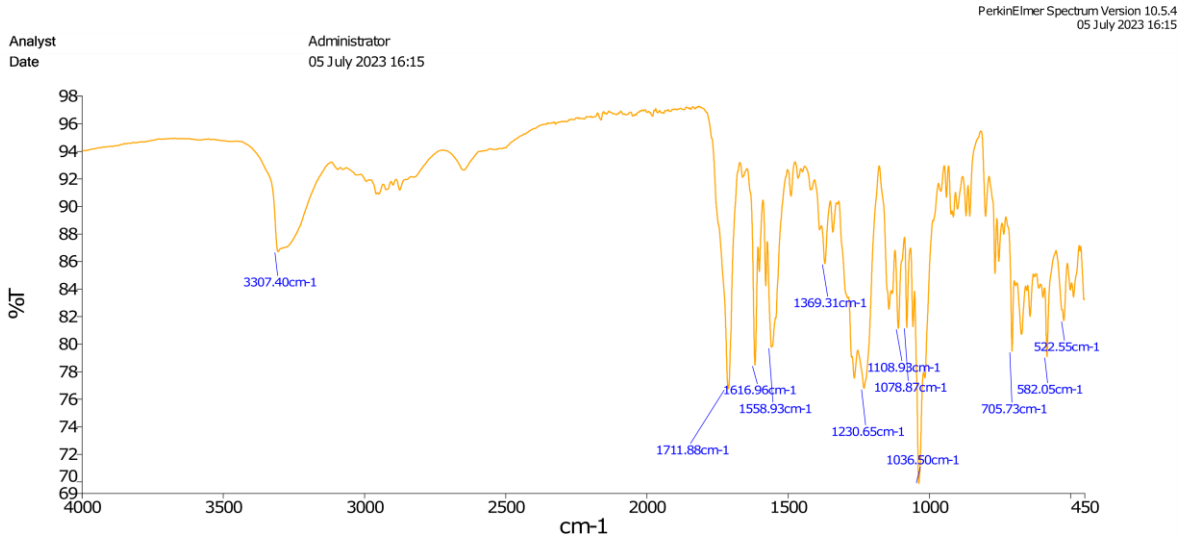



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 204.086675 | 1 | C8H14NO5 | 204.086649 | 0.1 | 0.0 | 745.7 | 0.0 |

Appendix 10. Compound **28:**

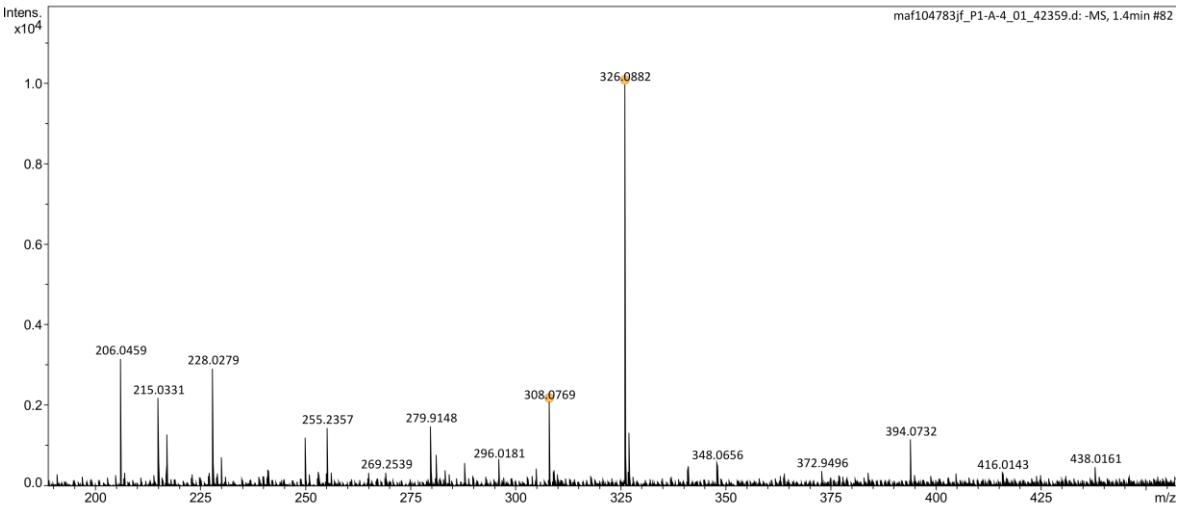




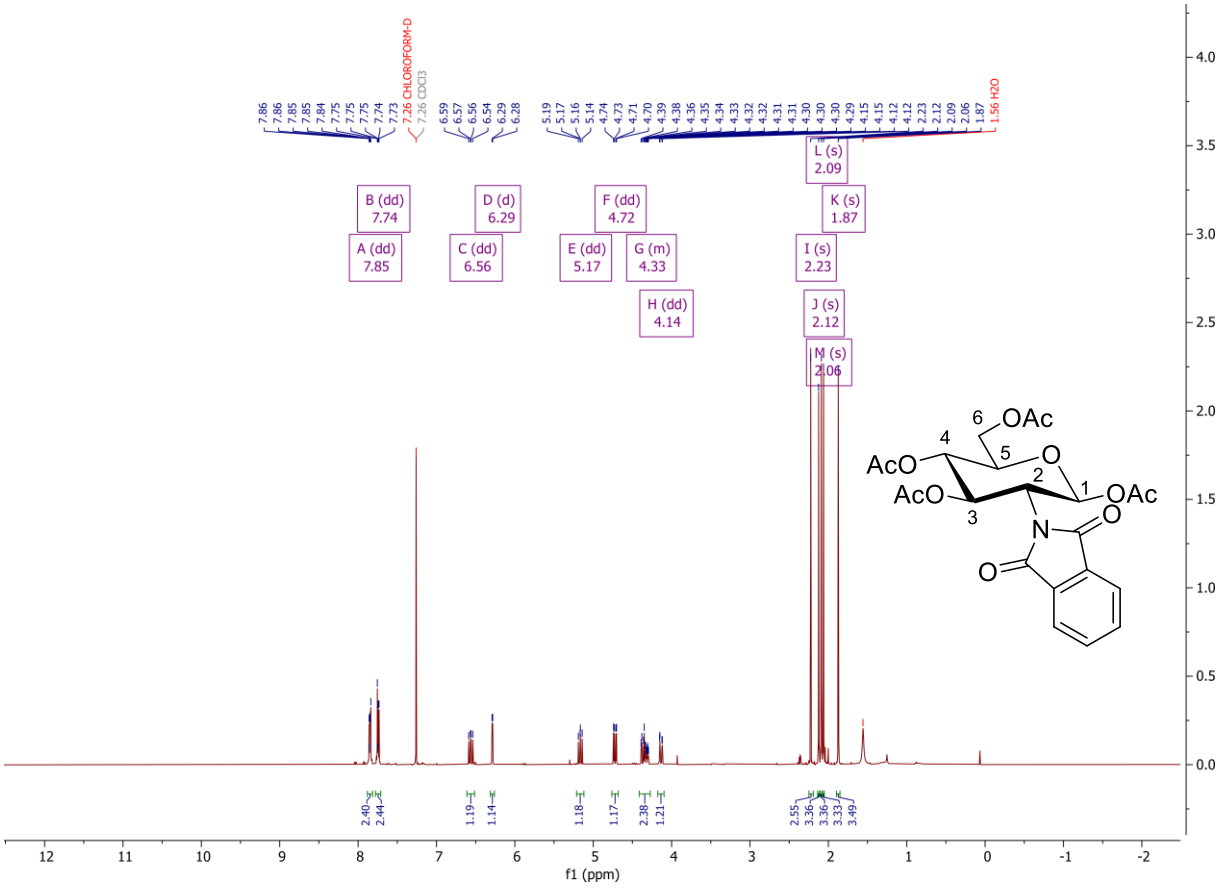

**Analysis Information**

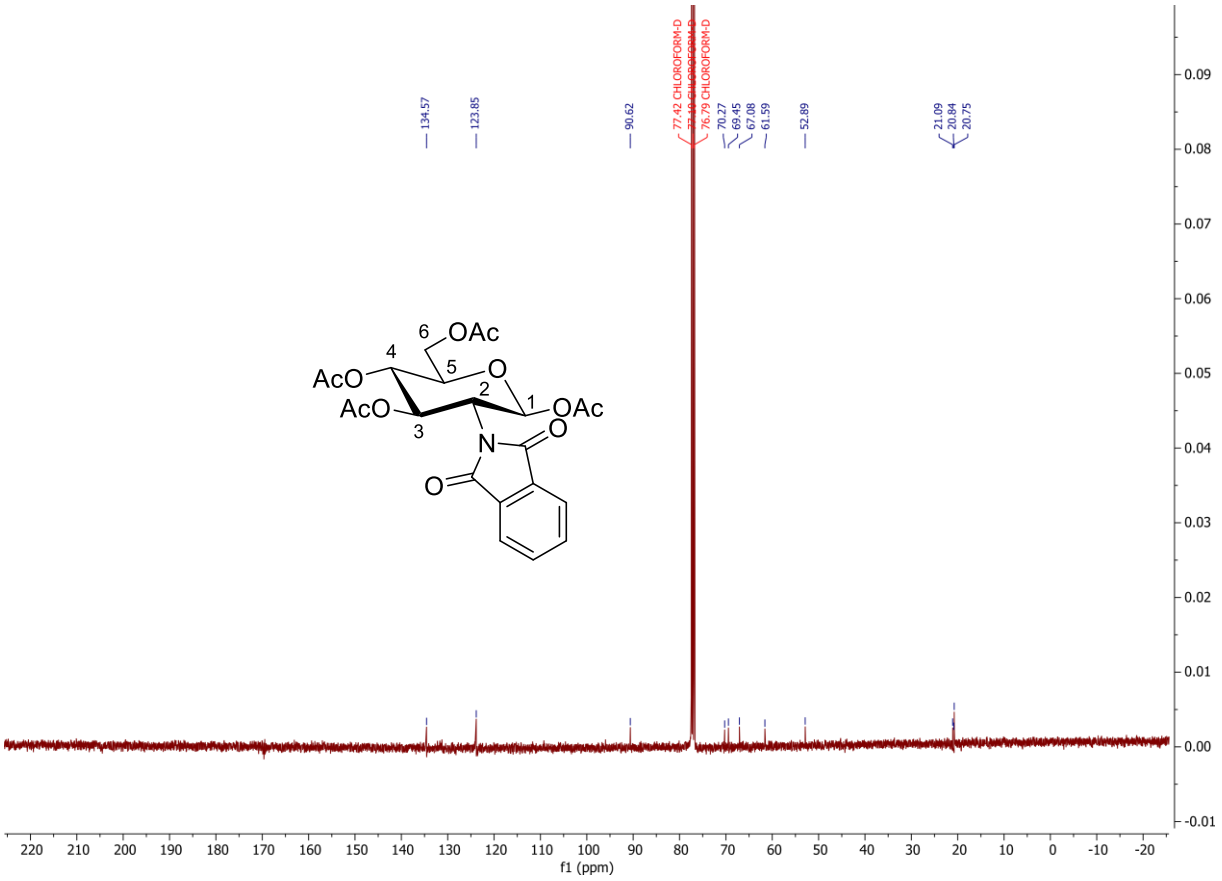
| | | | |
|---|---|---|---|
| Analysis Filename | maf104783jf_P1-A-4_01_42359.d | Acquisition Date | 13/06/2023 15:24:49 |
| Method | ESI_low mass neg_2c1s.m | Instrument | compact |
| Submission Name | maf104783jf | ESI | Negative |

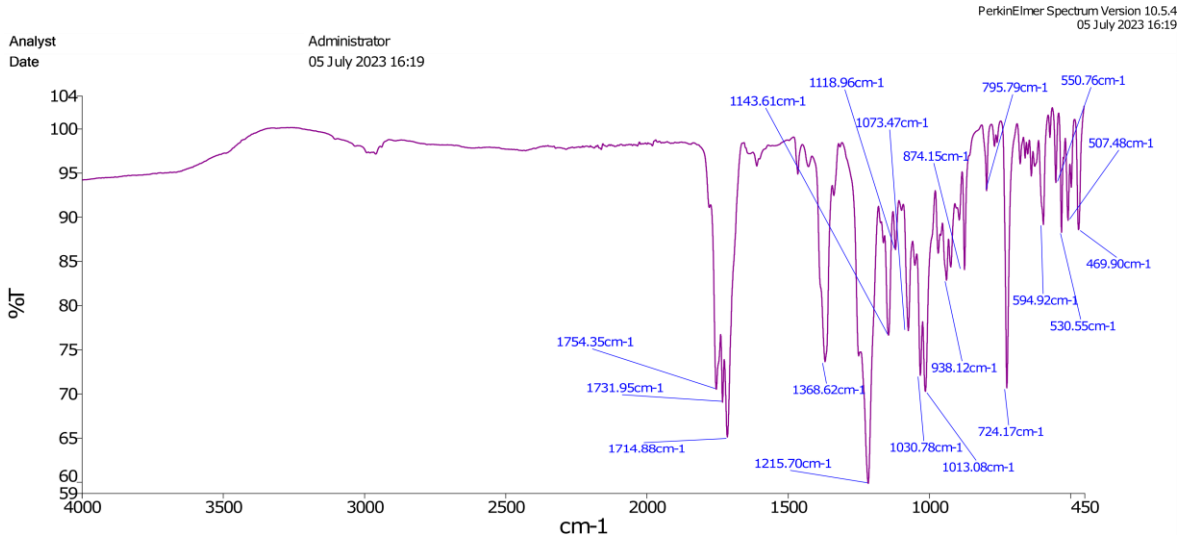



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 308.0769 | 1 | C14H14NO7 | 308.0776 | 2.2 | 0.7 | n.a. | 1.2 |
| 326.0882 | 1 | C14H16NO8 | 326.0881 | -0.1 | -0.0 | 22.7 | 1.9 |

Appendix 11. Compound **29**:

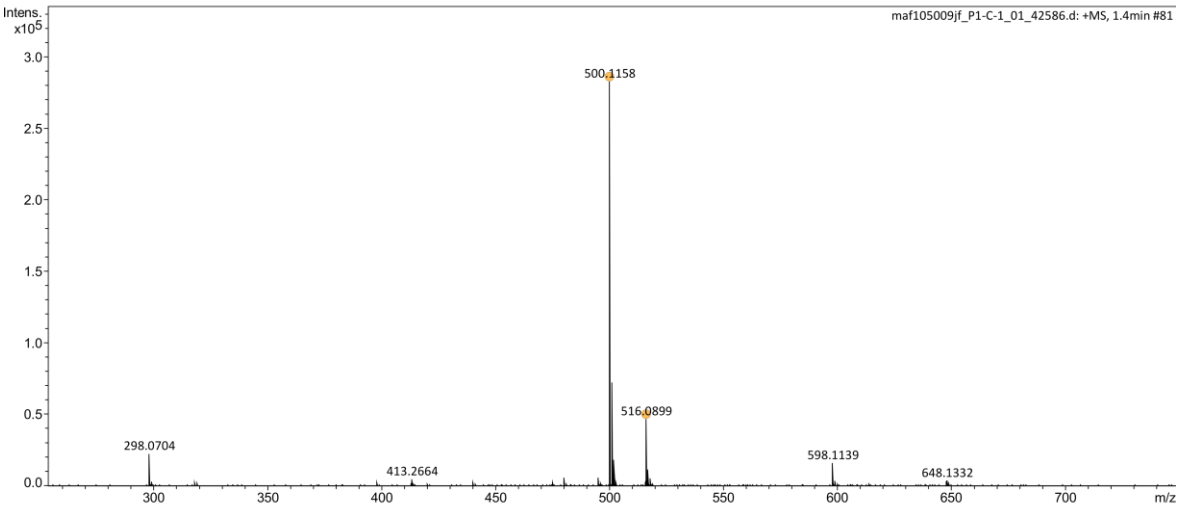




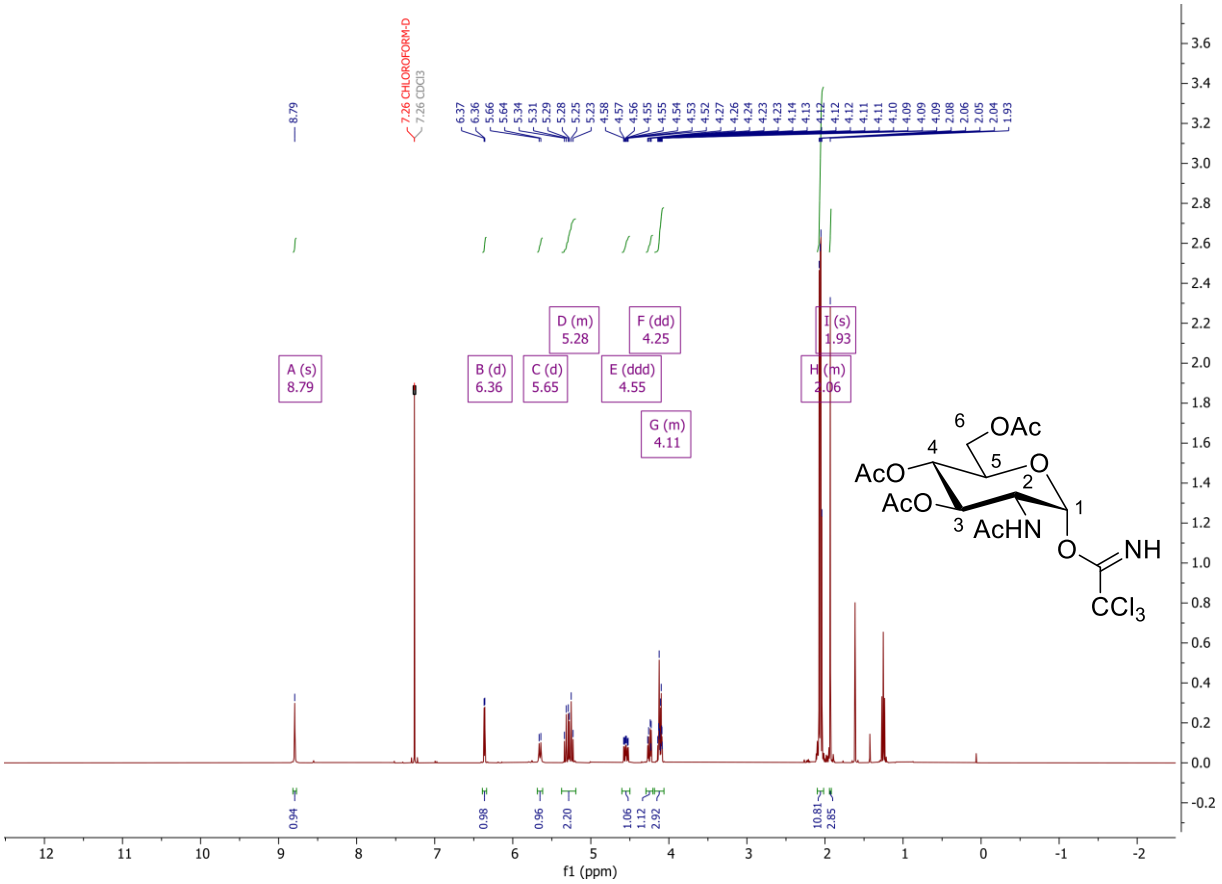

**Analysis Information**

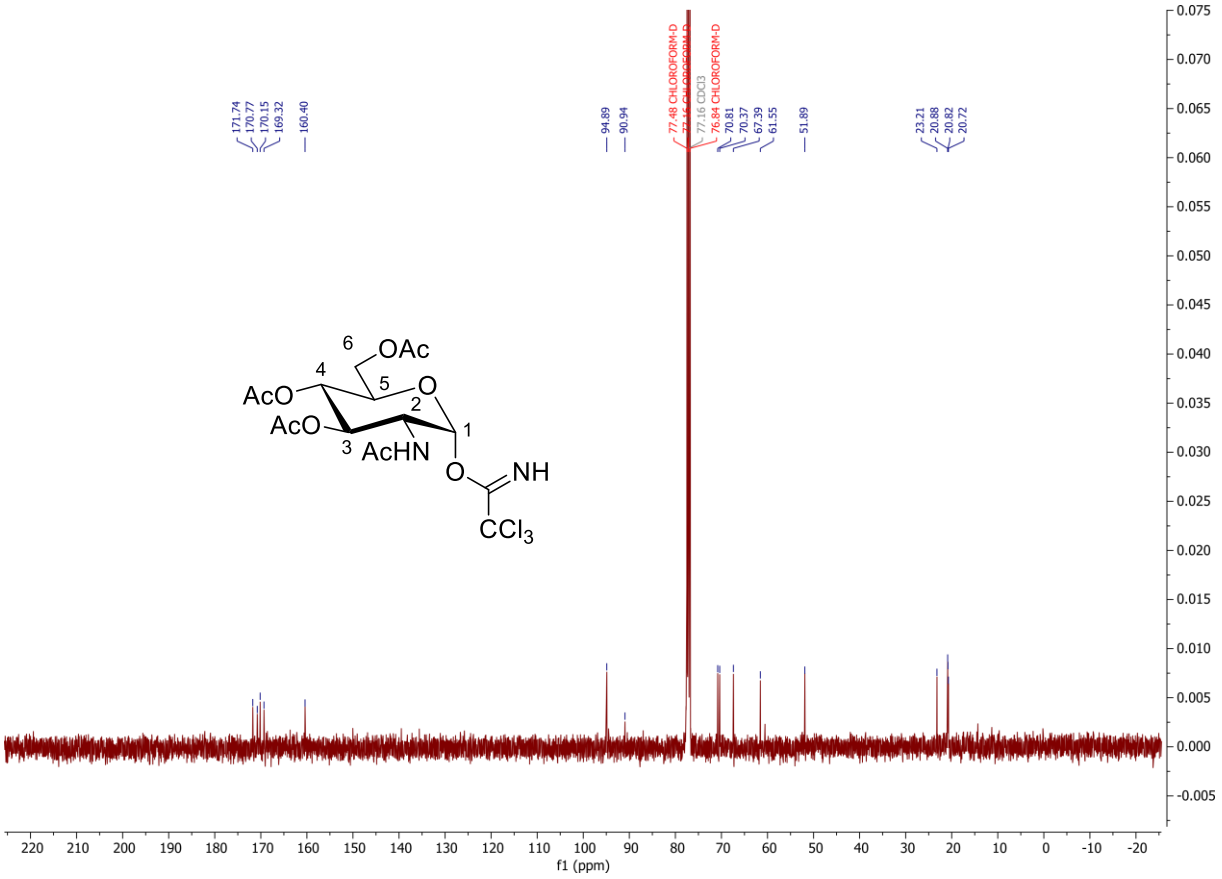
| | | | |
|---|---|---|---|
| Analysis Filename | maf105009jf_P1-C-1_01_42586.d | Acquisition Date | 30/06/2023 06:56:55 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf105009jf | ESI | Positive |

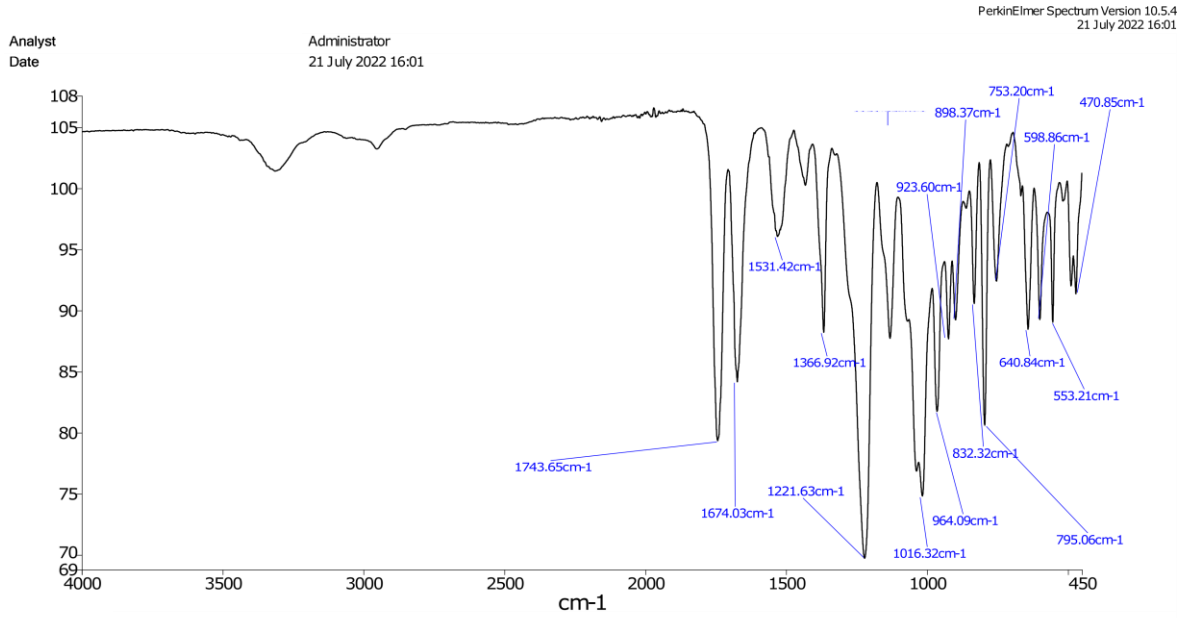



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 500.1158 | 1 | C22H23NNaO11 | 500.1163 | 1.0 | 0.5 | 7.9 | -0.4 |
| 516.0899 | 1 | C22H23KNO11 | 516.0903 | 0.8 | 0.4 | 5.8 | 2.4 |

Appendix 12. Compound **36:**

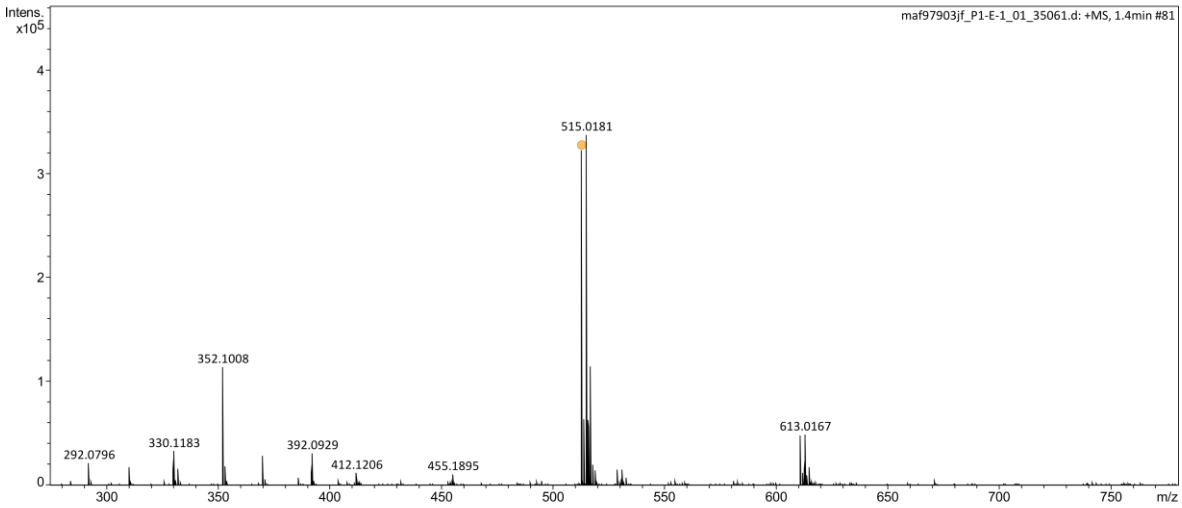




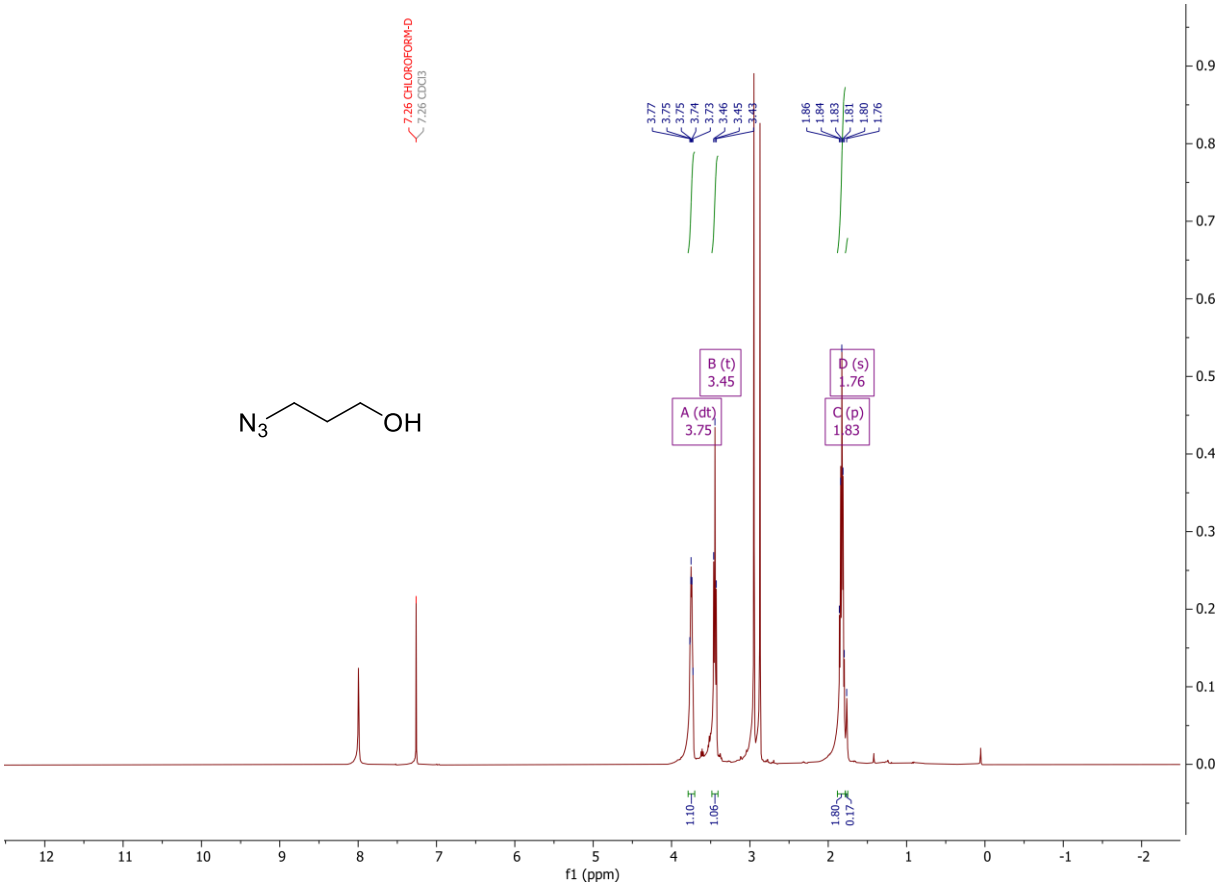

**Analysis Information**

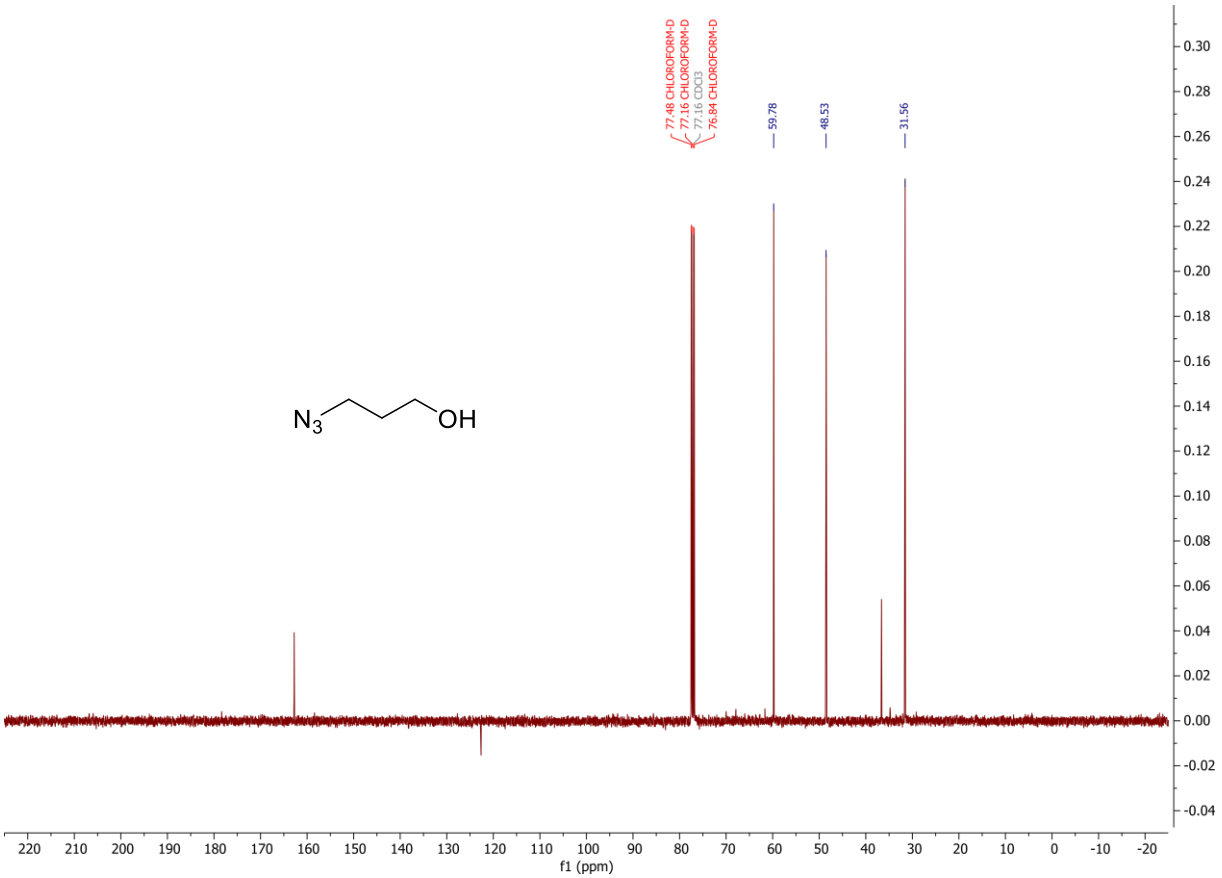
| | | | |
|---|---|---|---|
| Analysis Filename | maf97903jf_P1-E-1_01_35061.d | Acquisition Date | 01/07/2022 11:04:21 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf97903jf | ESI | Positive |

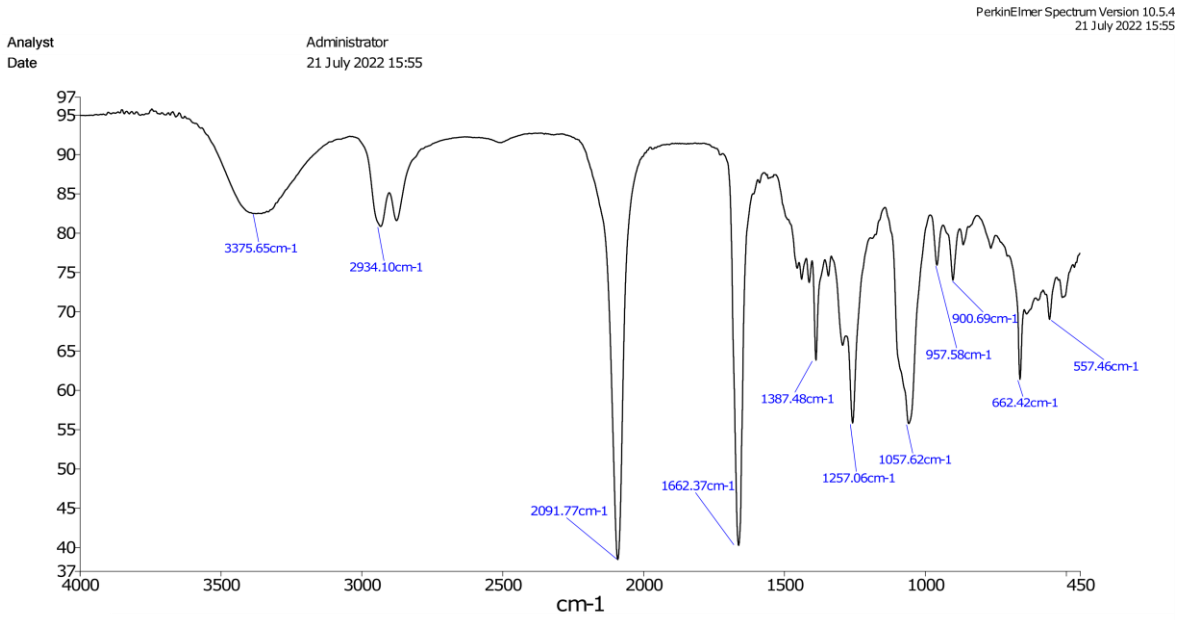


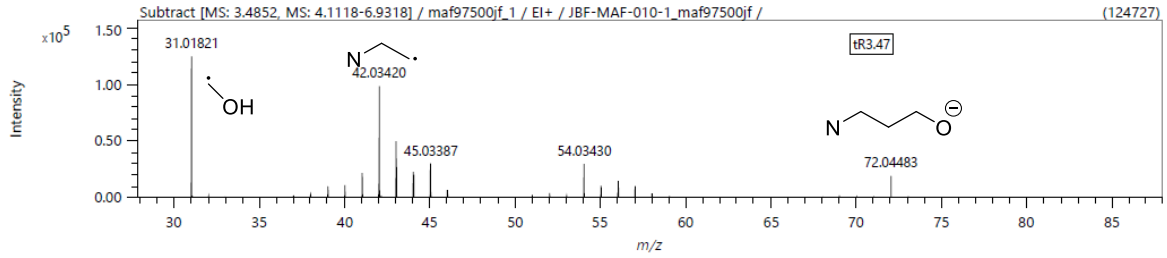

| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 513.0207 | 1 | C16H21Cl3N2NaO9 | 513.0205 | -0.4 | -0.2 | 16.8 | -0.4 |

Appendix 13. Compound **34:**

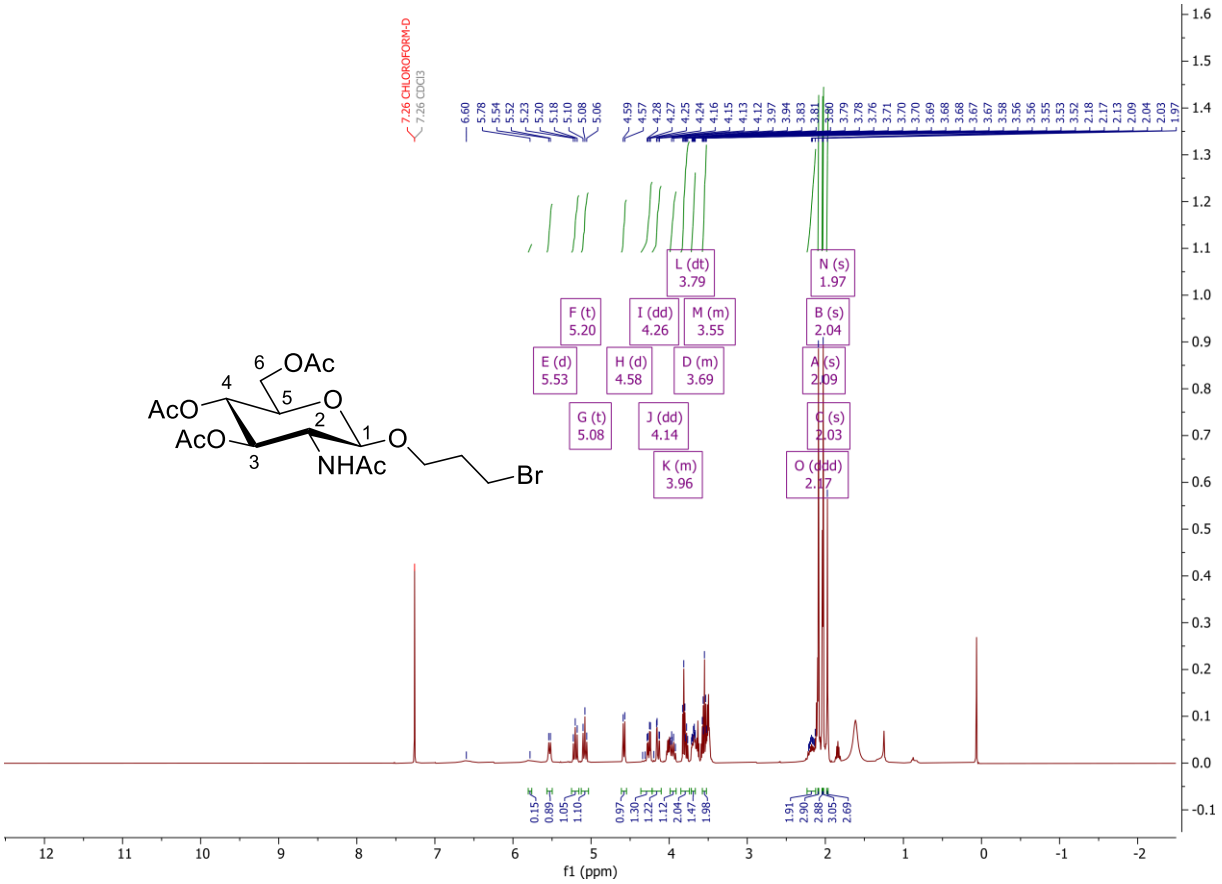


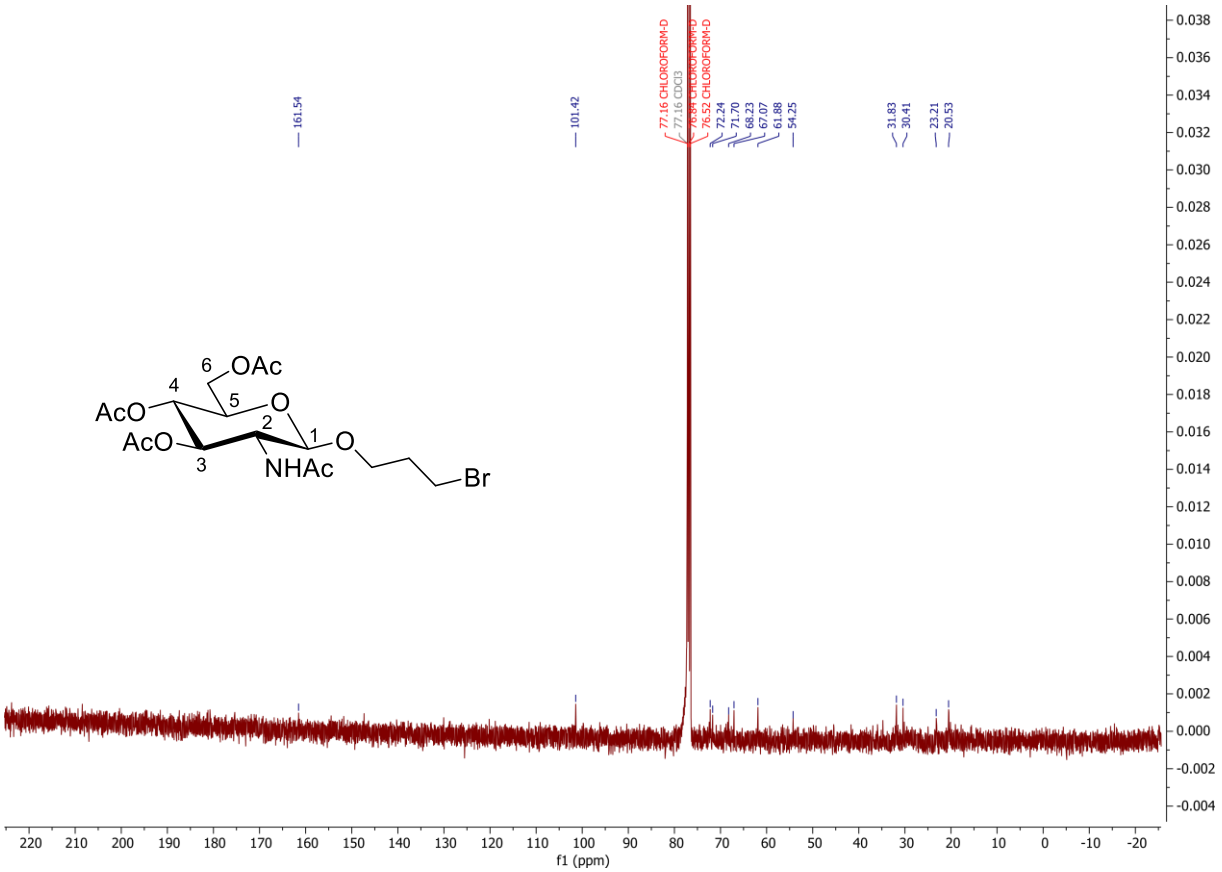



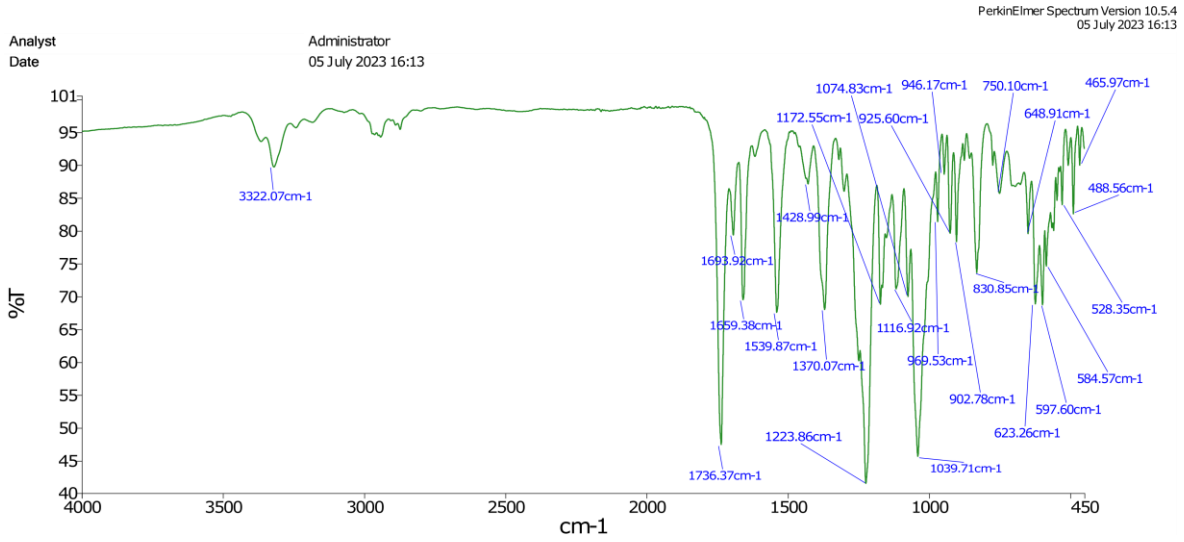

PerkinElmer Spectrum Version 10.5.4
21 July 2022 15:55
Analyst
Administrator
Date
21 July 2022 15:55
3375.65cm-1
2934.10cm-1
2091.77cm-1
1662.37cm-1
1387.48cm-1
1257.06cm-1
1057.62cm-1
957.58cm-1
900.69cm-1
662.42cm-1
557.46cm-1
cm-1


Subtract [MS: 3.4852, MS: 4.1118-6.9318] / maf97500jf_1 / EI+ / JBF-MAF-010-1_maf97500jf /
(124727)
Intensity
31.01821
42.03420
45.03387
54.03430
72.04483
tR3.47
m/z

Appendix 14. Compound **38:**

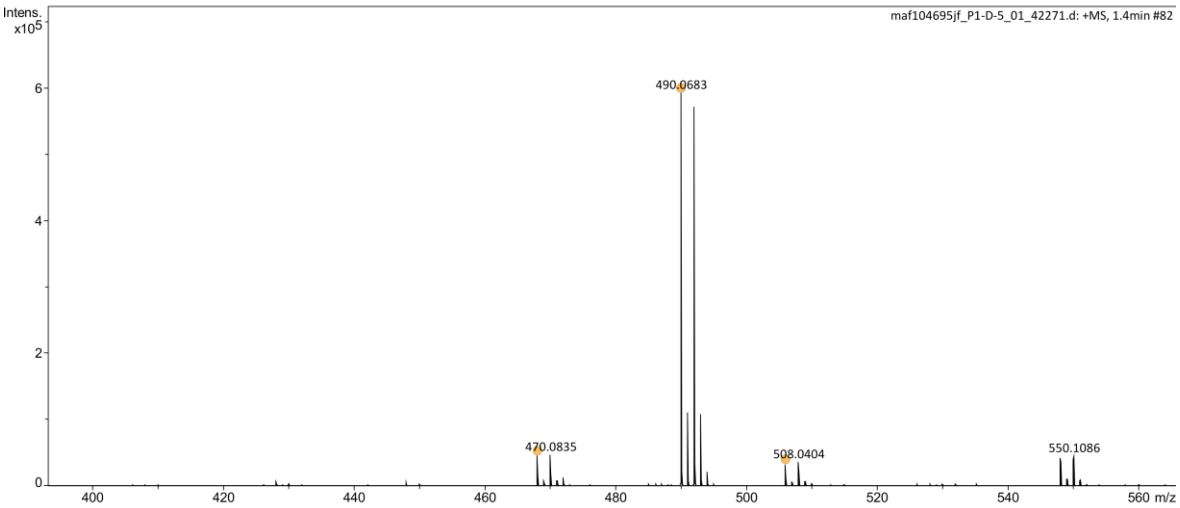




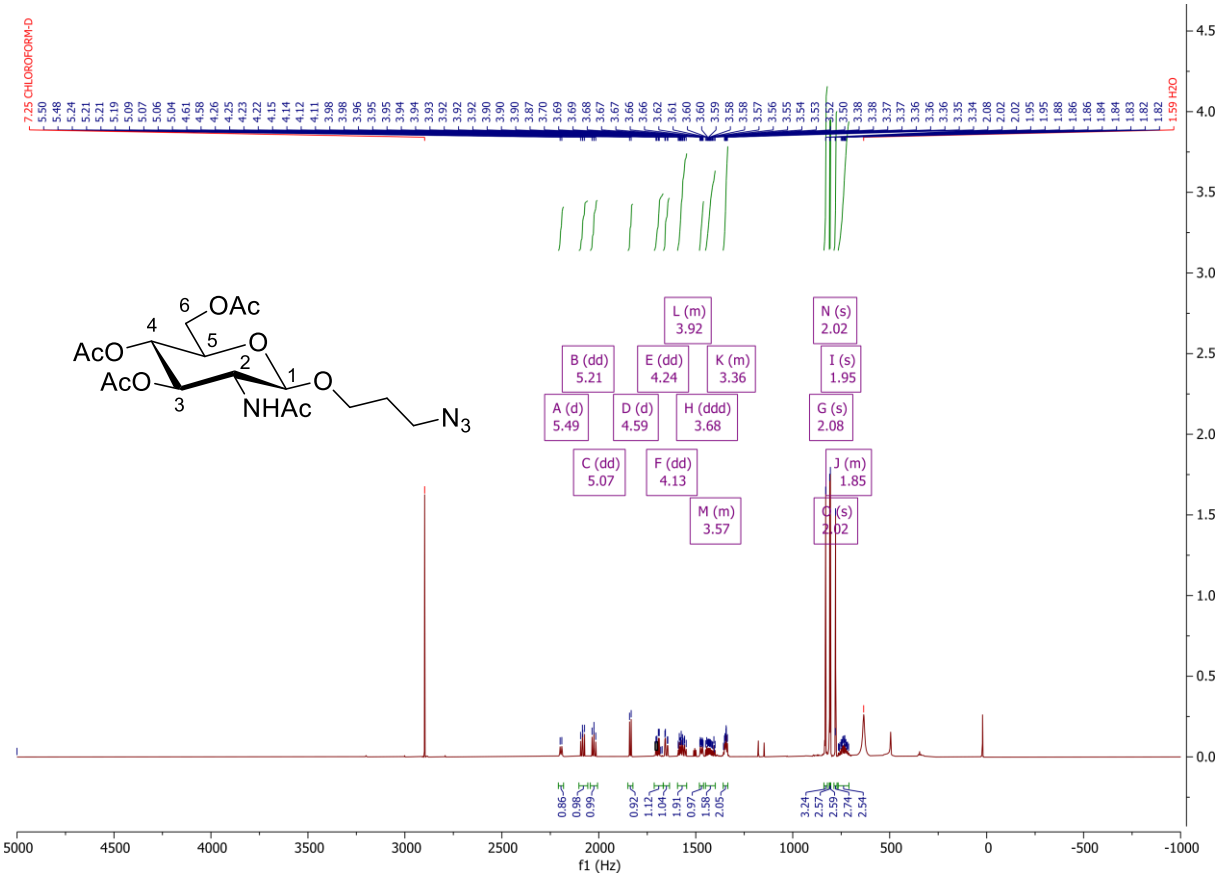

**Analysis Information**

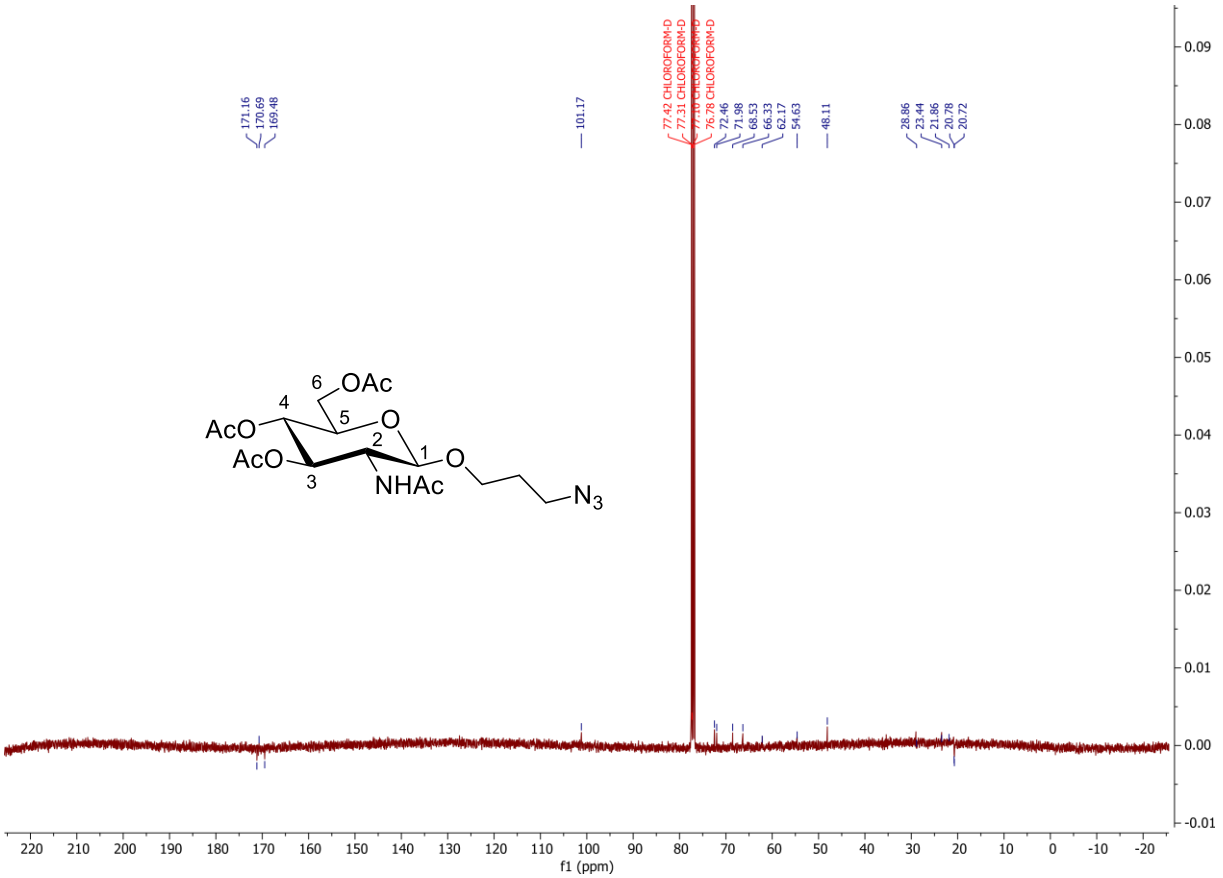
| | | | |
|---|---|---|---|
| Analysis Filename | maf104695jf_P1-D-5_01_42271.d | Acquisition Date | 08/06/2023 06:52:37 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf104695jf | ESI | Positive |




| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 468.0882 | 1 | C17H27BrNO9 | 468.0864 | -4.0 | -1.9 | 21.0 | -0.8 |
| 490.0683 | 1 | C17H26BrNNaO9 | 490.0683 | 0.0 | 0.0 | 16.5 | 0.1 |
| 506.0420 | 1 | C17H26BrKNO9 | 506.0423 | 0.4 | 0.2 | 10.6 | -1.6 |

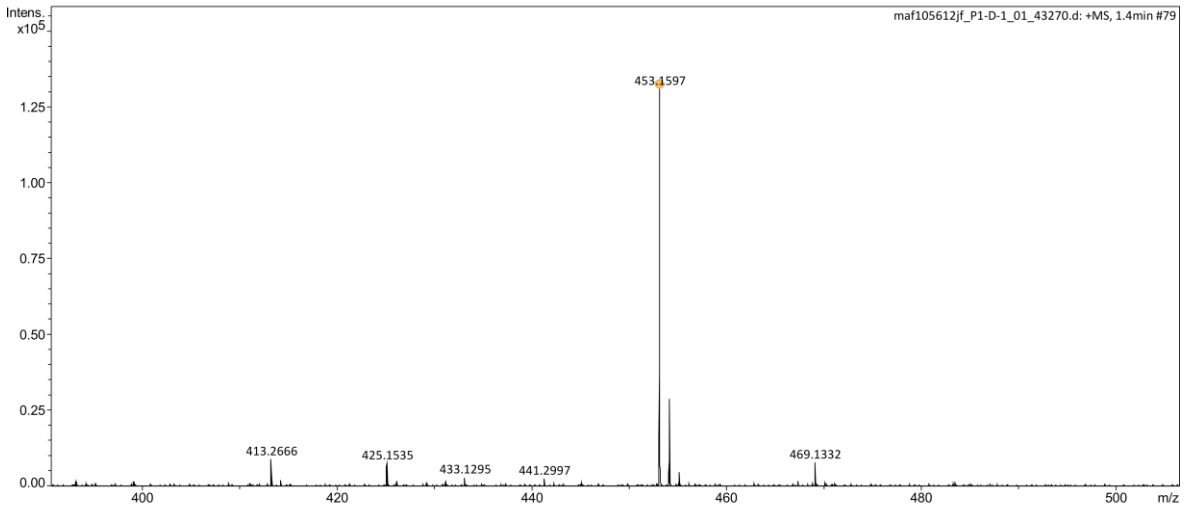
Appendix 15. Compound **27:**



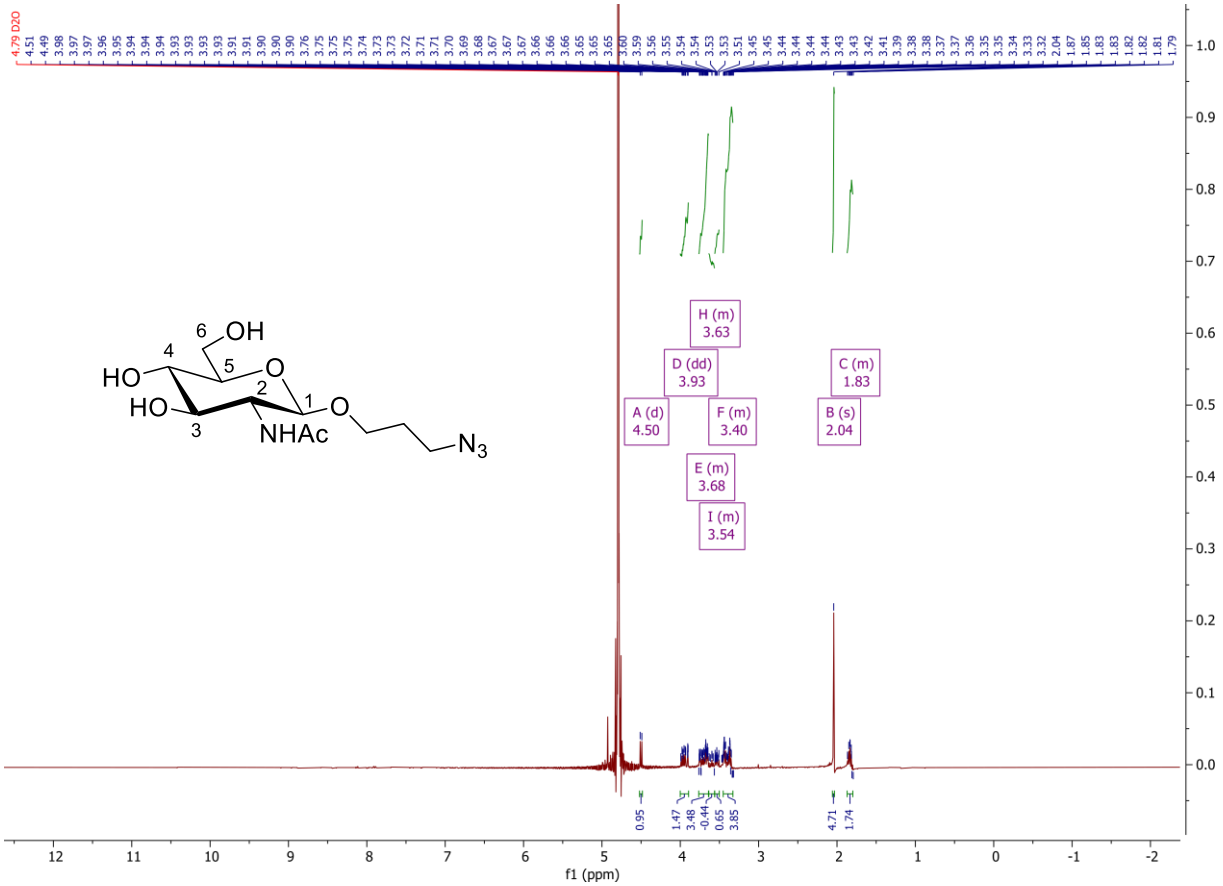
**Analysis Information**

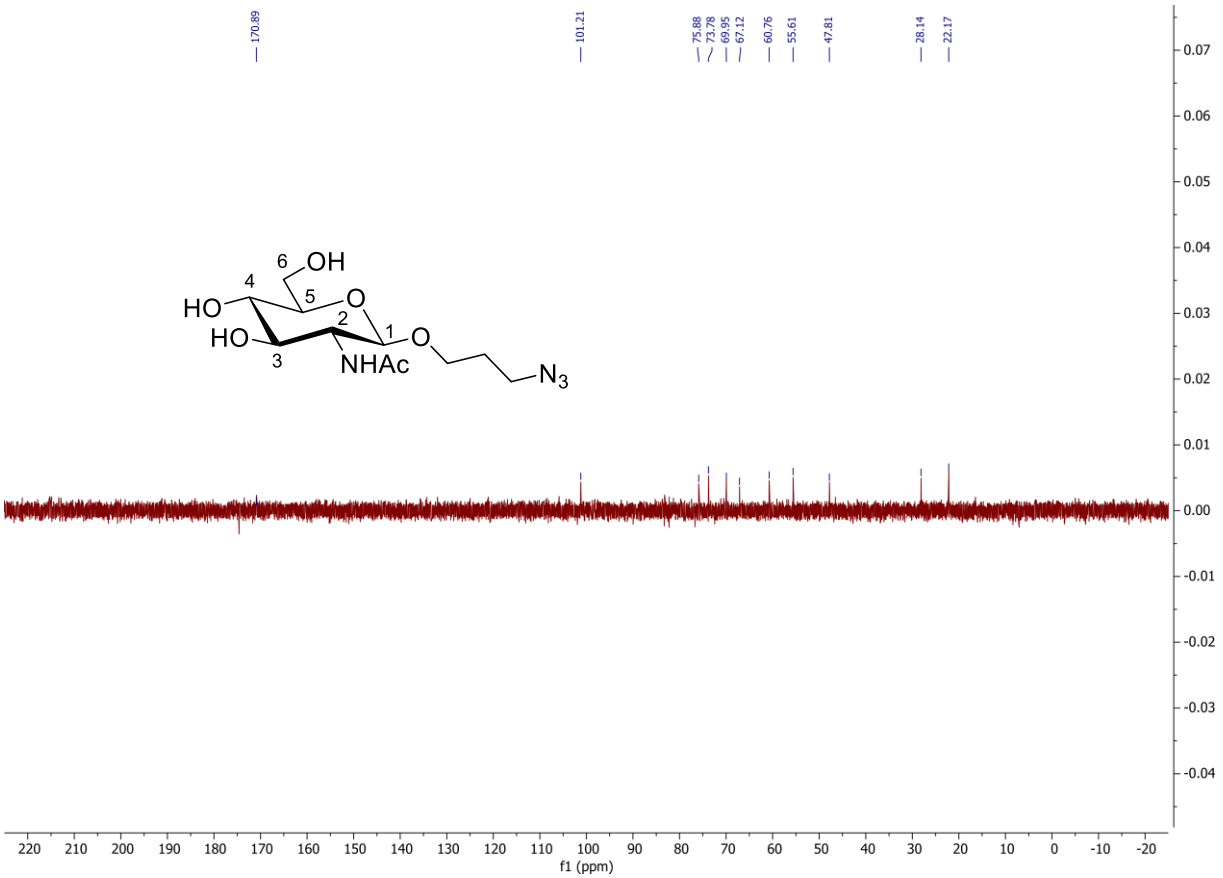
| | | | |
|---|---|---|---|
| Analysis Filename | maf105612jf_P1-D-1_01_43270.d | Acquisition Date | 10/08/2023 10:38:13 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf105612jf | ESI | Positive |

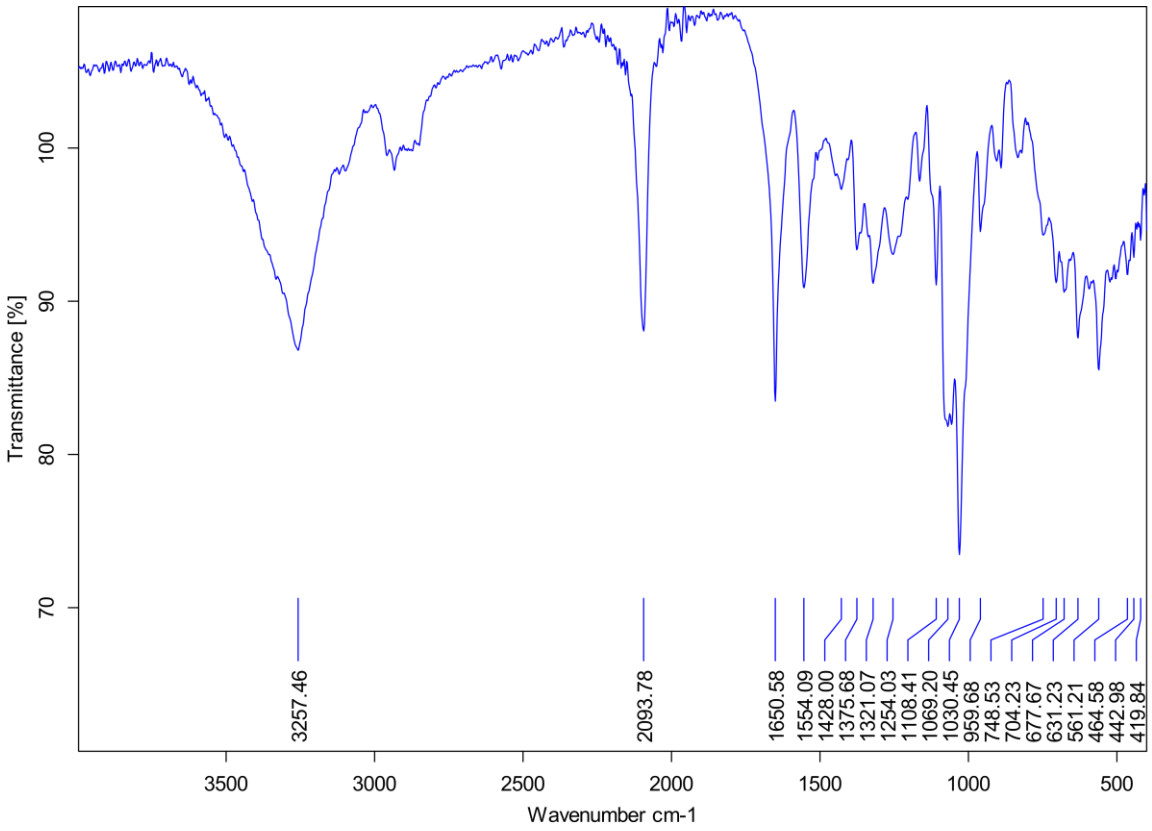



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 453.1597 | 1 | C17H26N4NaO9 | 453.1592 | -1.1 | -0.5 | 9.2 | -1.0 |

Appendix 16. Compound **8:**

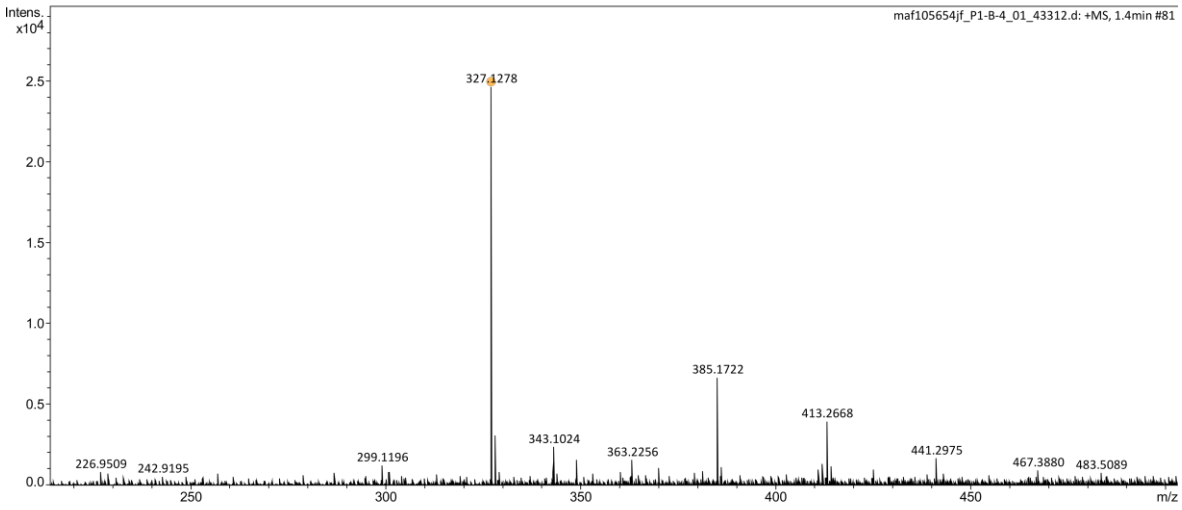




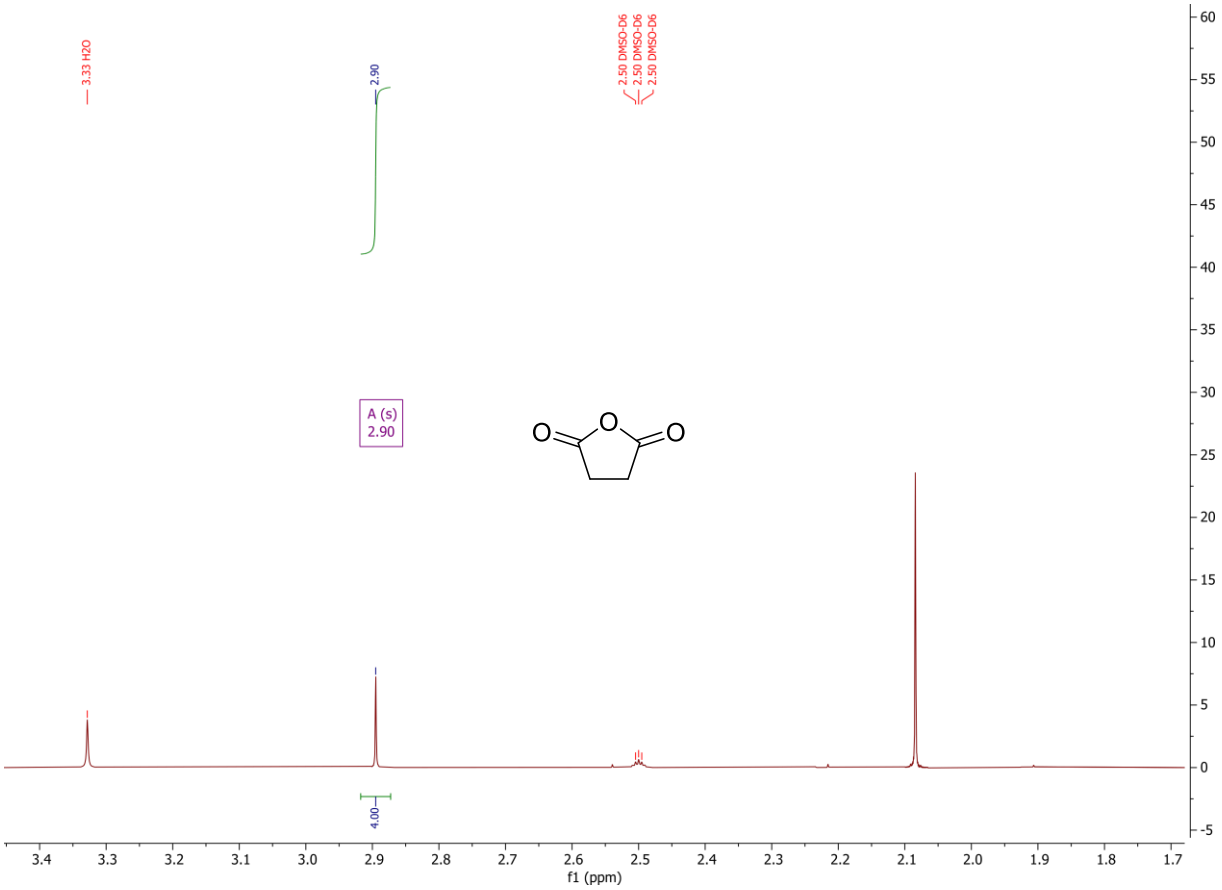

**Analysis Information**

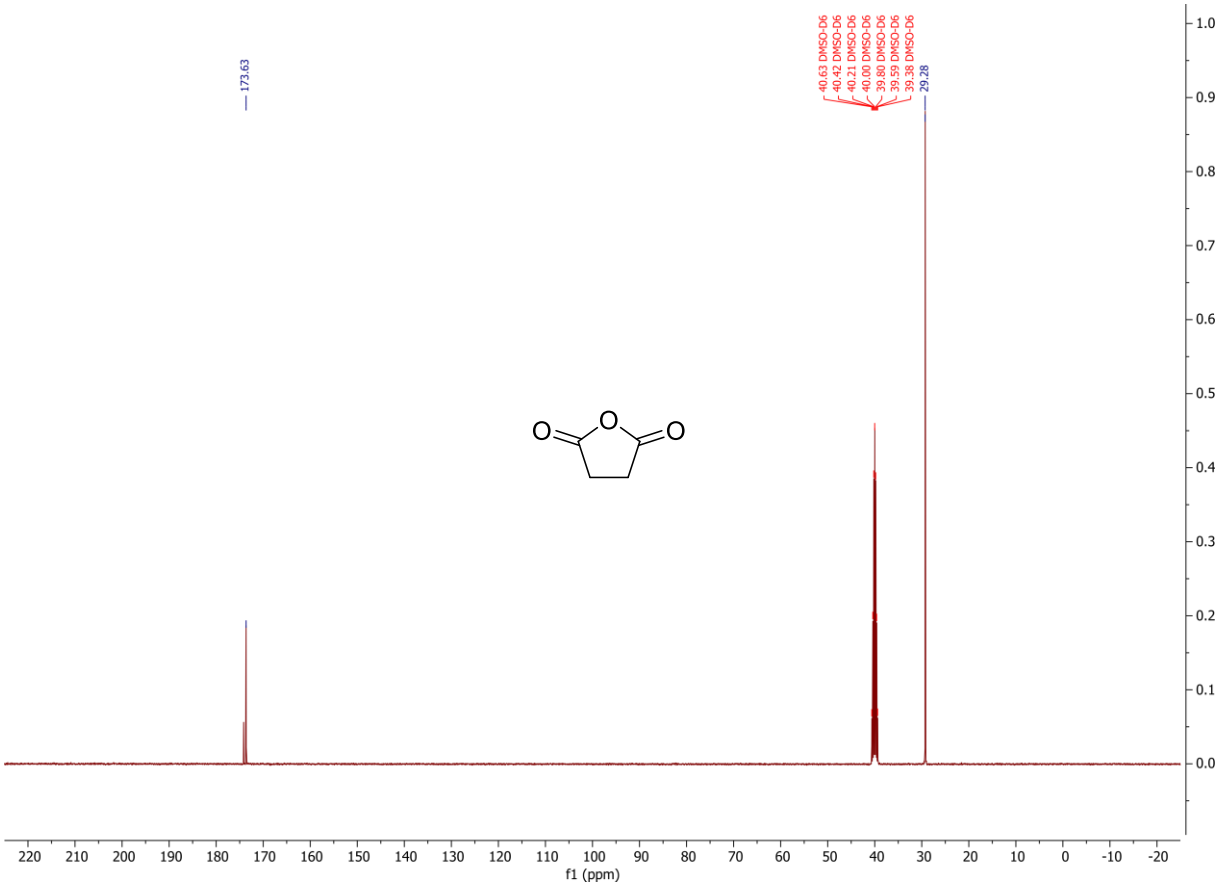
| | | | |
|---|---|---|---|
| Analysis Filename | maf105654jf_P1-B-4_01_43312.d | Acquisition Date | 14/08/2023 11:20:54 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf105654jf | ESI | Positive |

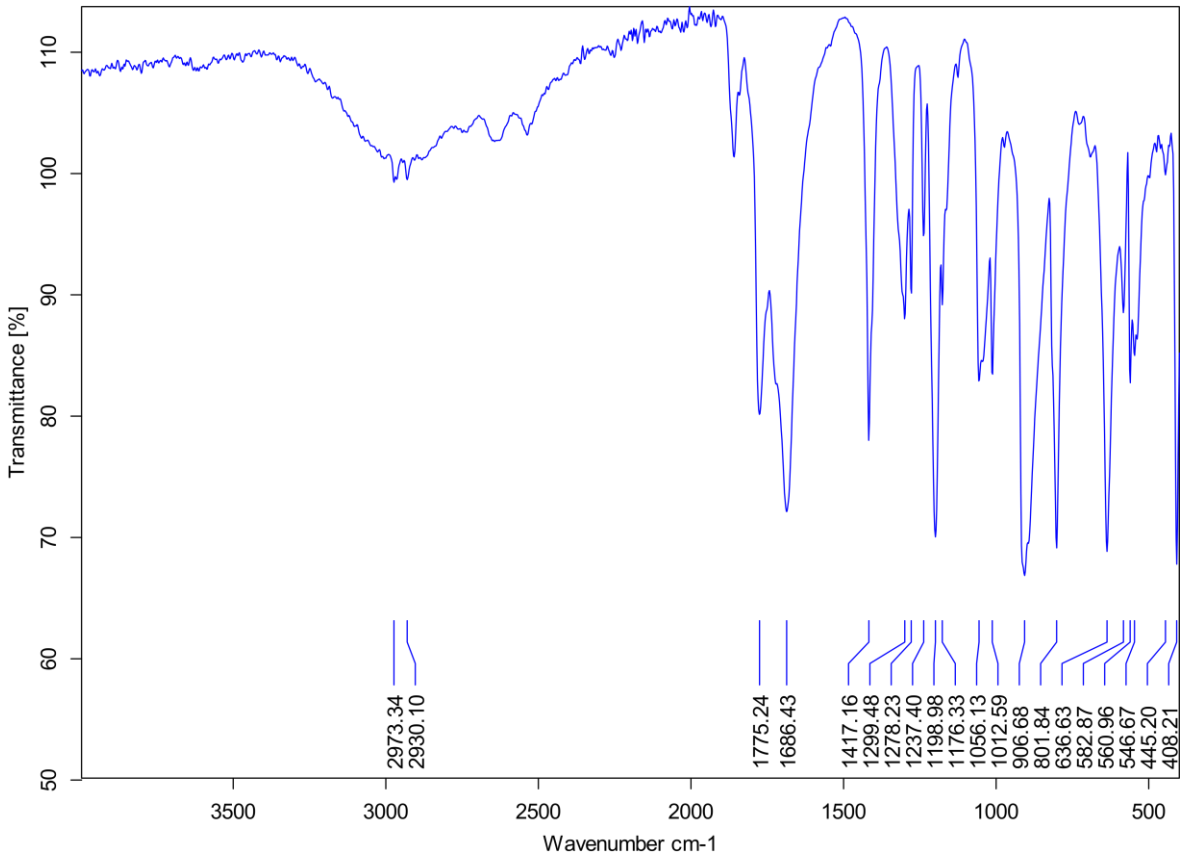



| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 327.1278 | 1 | C11H20N4NaO6 | 327.1275 | -0.9 | -0.3 | 10.8 | -3.1 |

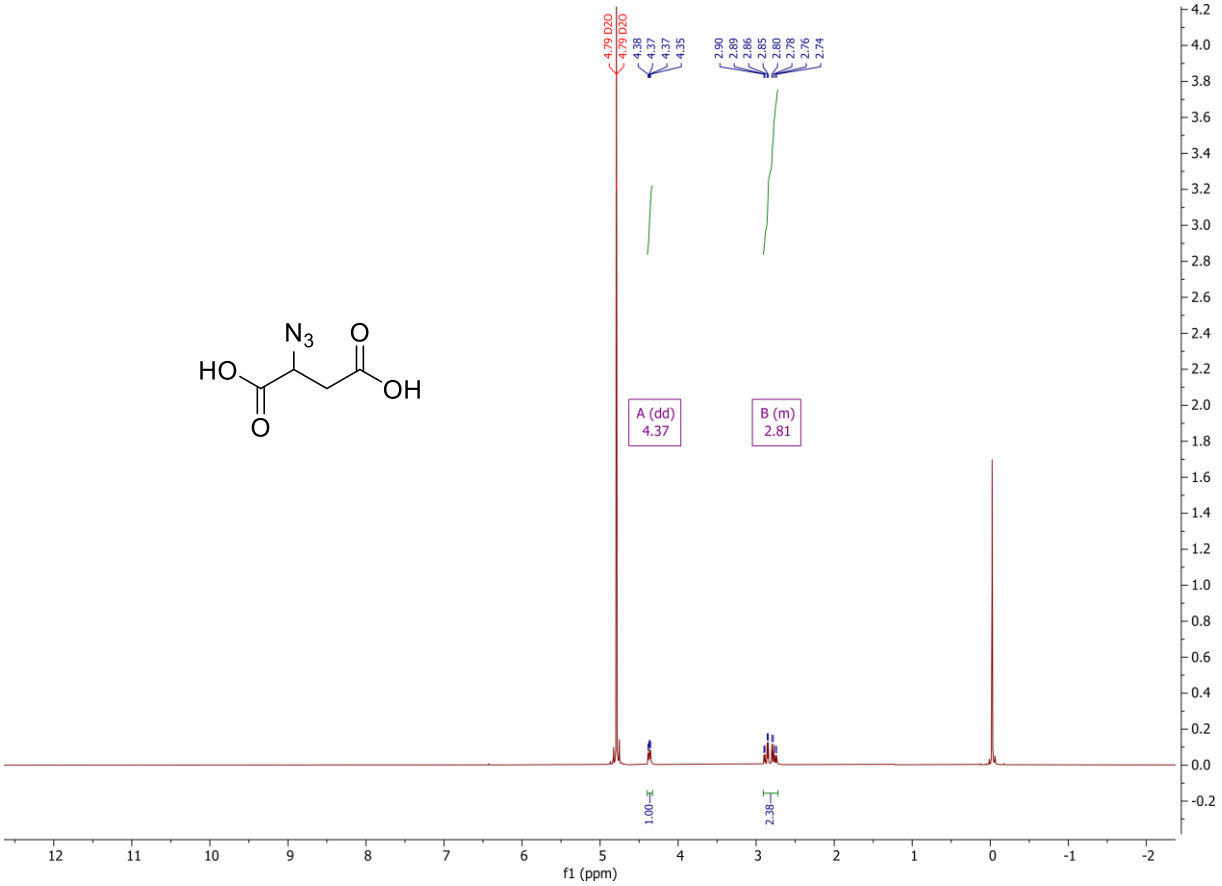
Appendix 17. Compound **40:**

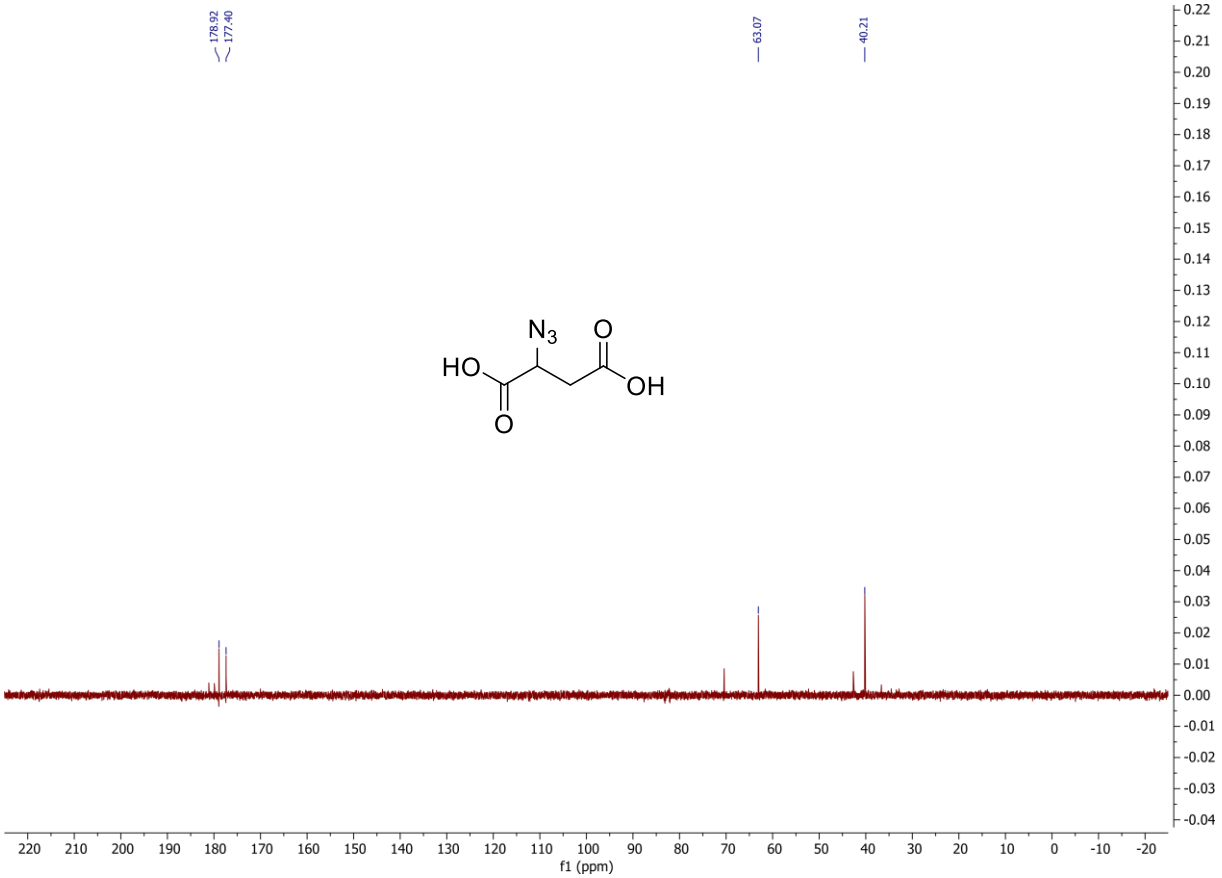



Transmittance [%]
110
100
90
80
70
60
50
3500
3000
2500
2000
1500
1000
500
Wavenumber cm-1
2973.34
2930.10
1775.24
1686.43
1417.16
1299.48
1278.23
1237.40
1198.98
1176.33
1056.13
1012.59
906.68
801.84
636.63
582.87
560.96
546.67
445.20
408.21

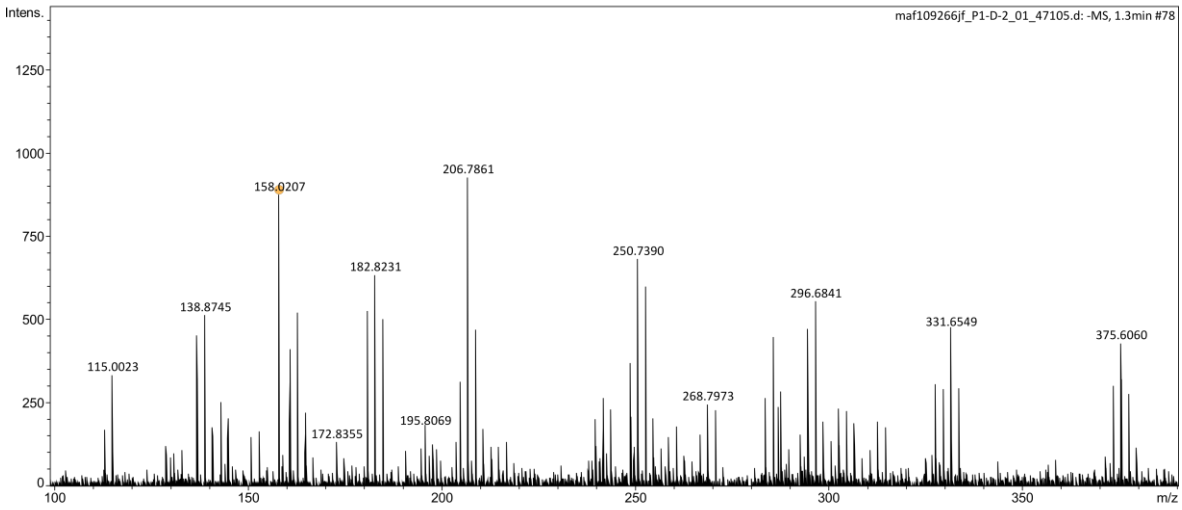
Appendix 18. Compound **45:**



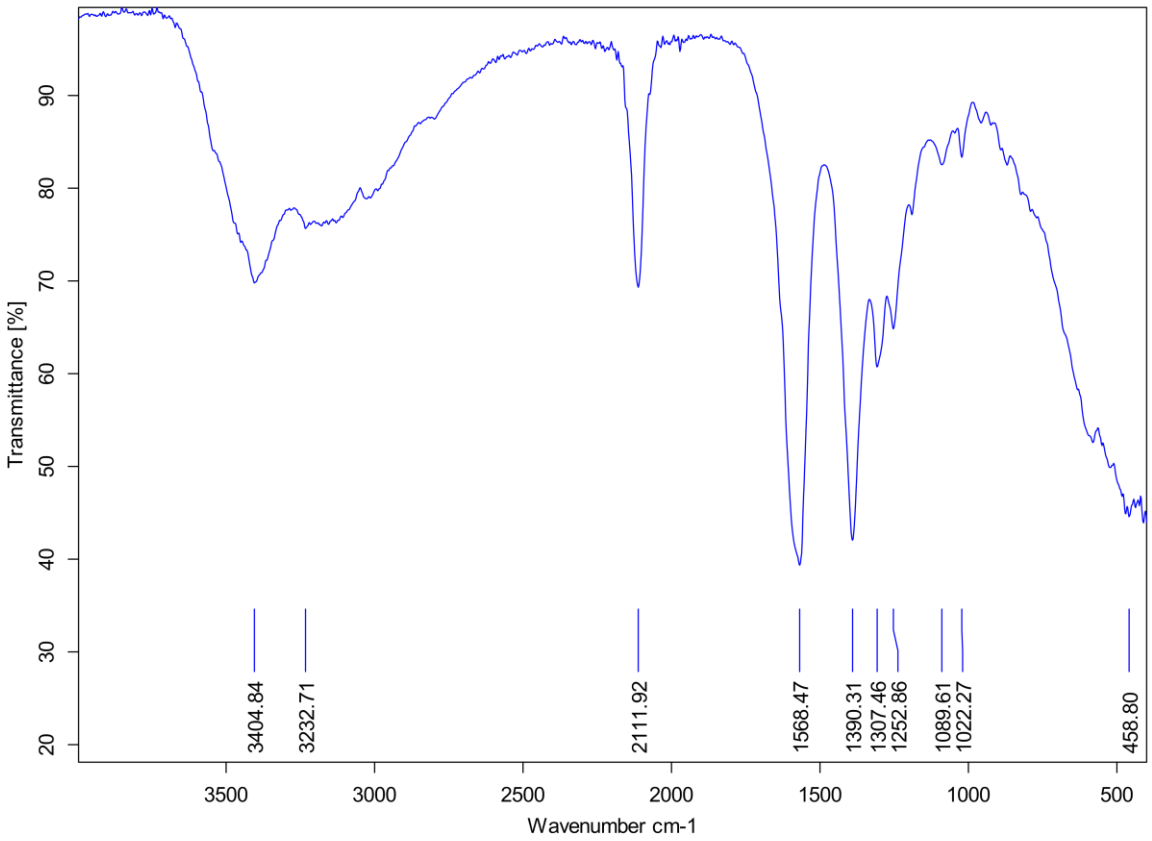



**Analysis Information**

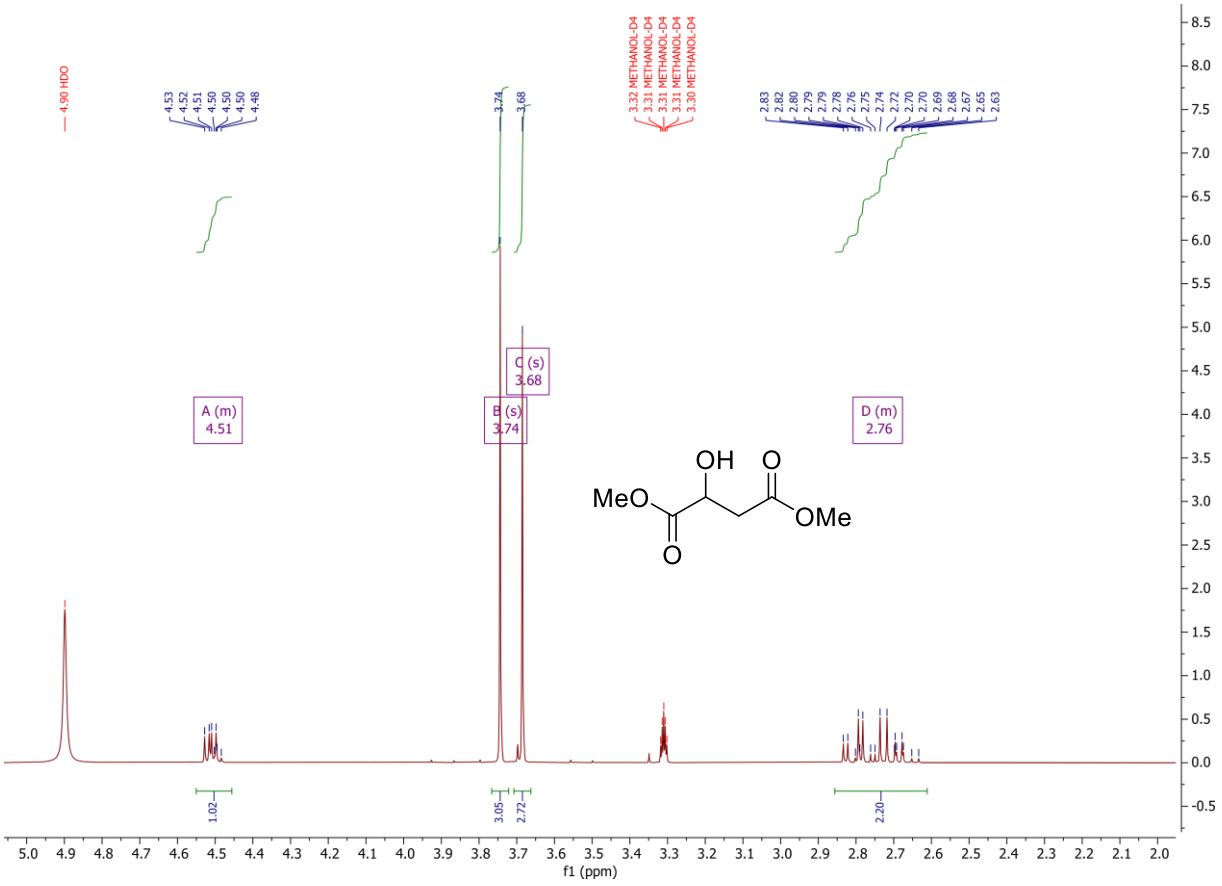
| | | | |
|---|---|---|---|
| Analysis Filename | maf109266jf_P1-D-2_01_47105.d | Acquisition Date | 19/02/2024 08:26:11 |
| Method | ESI_low mass neg_2c1s.m | Instrument | compact |
| Submission Name | maf109266jf | ESI | Negative |

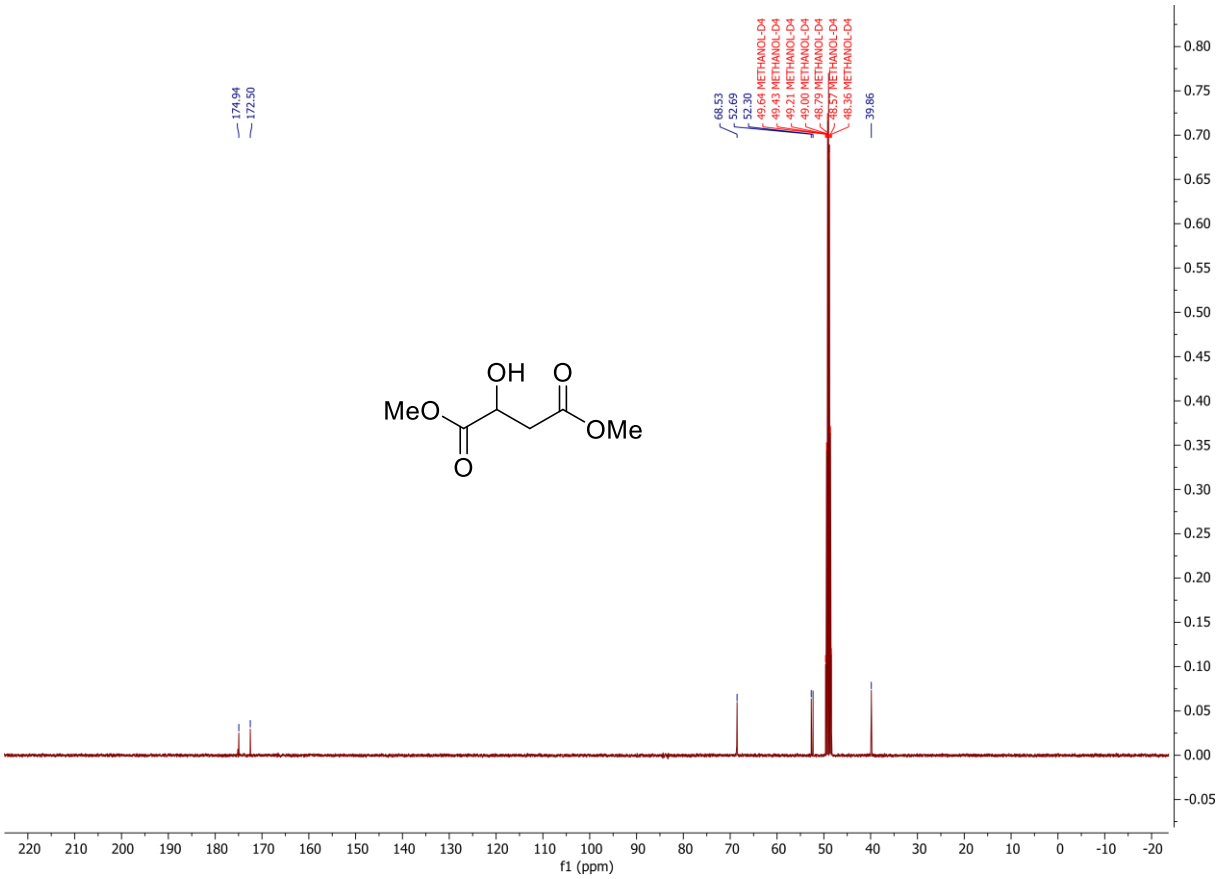


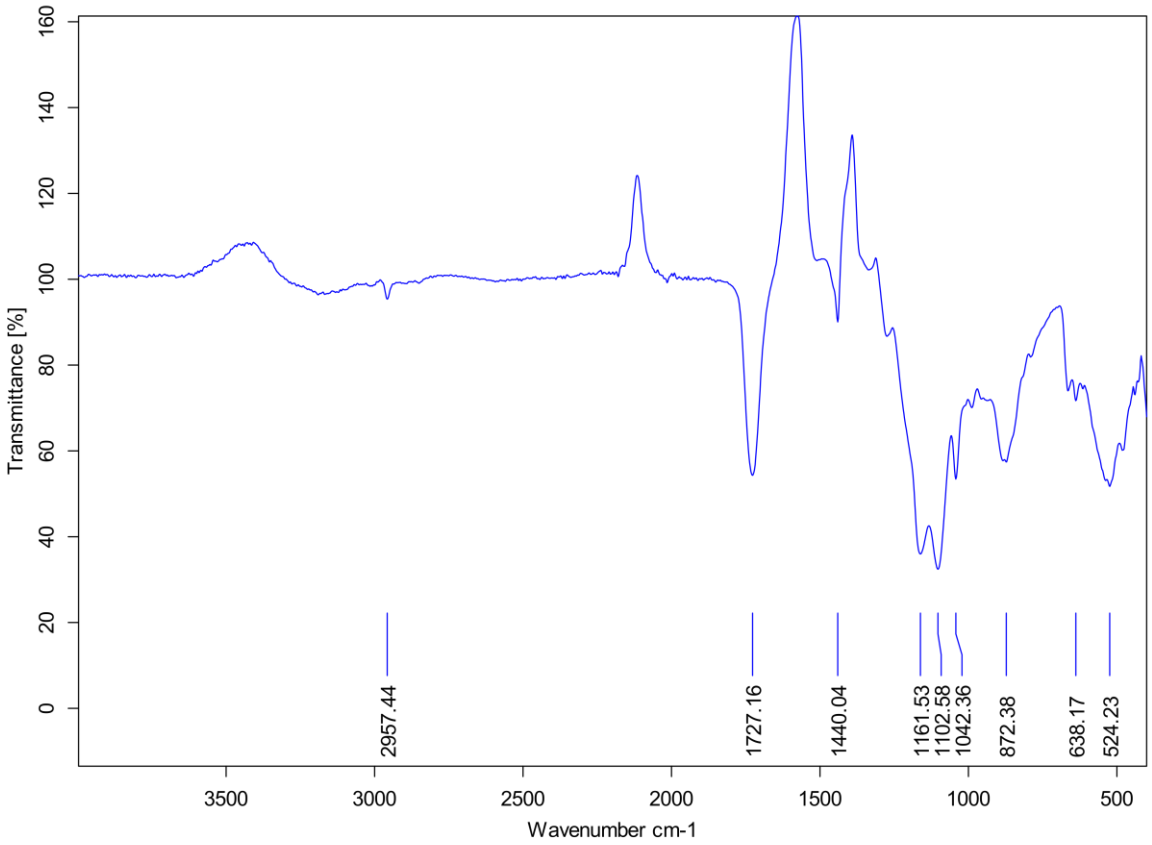

| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 158.0207 | 1 | C4H4N3O4 | 158.0207 | 0.5 | 0.1 | n.a. | 0.3 |



Appendix 19. Compound **53:**

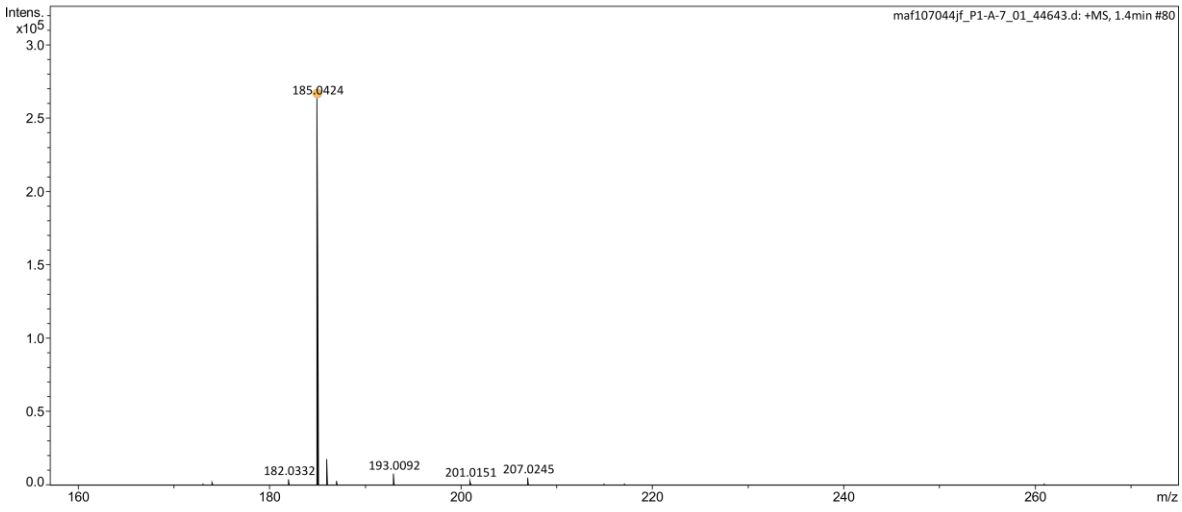




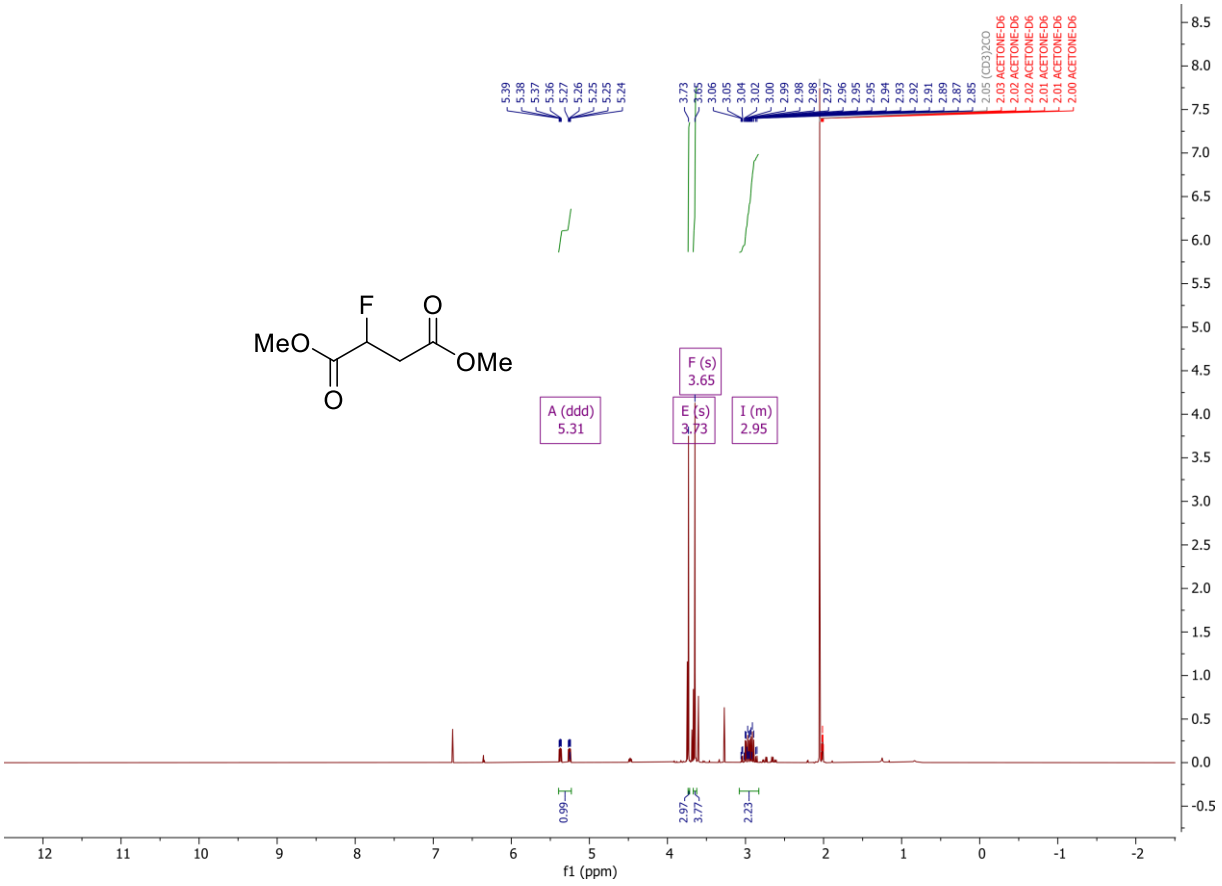

**Analysis Information**

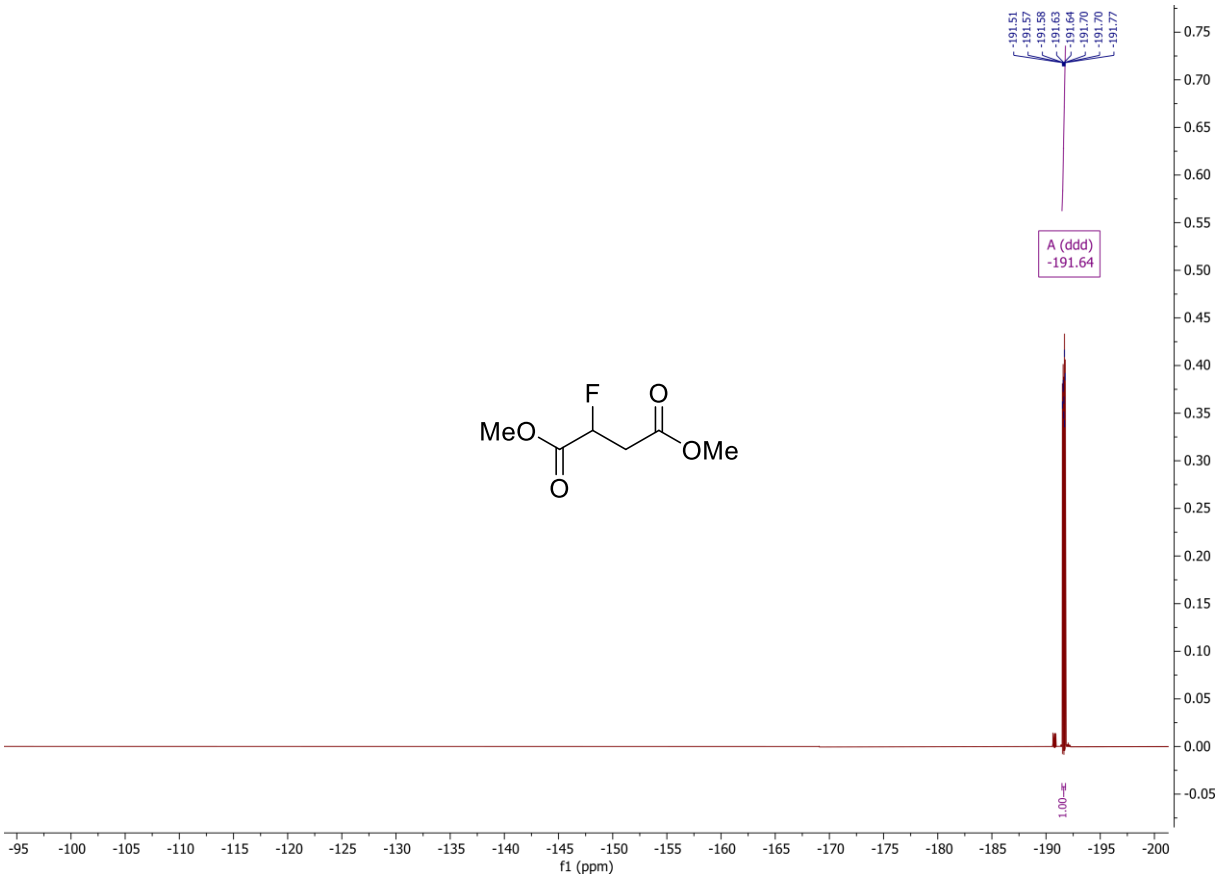
| | | | |
|---|---|---|---|
| Analysis Filename | maf107044jf_P1-A-7_01_44643.d | Acquisition Date | 03/11/2023 10:02:26 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf107044jf | ESI | Positive |

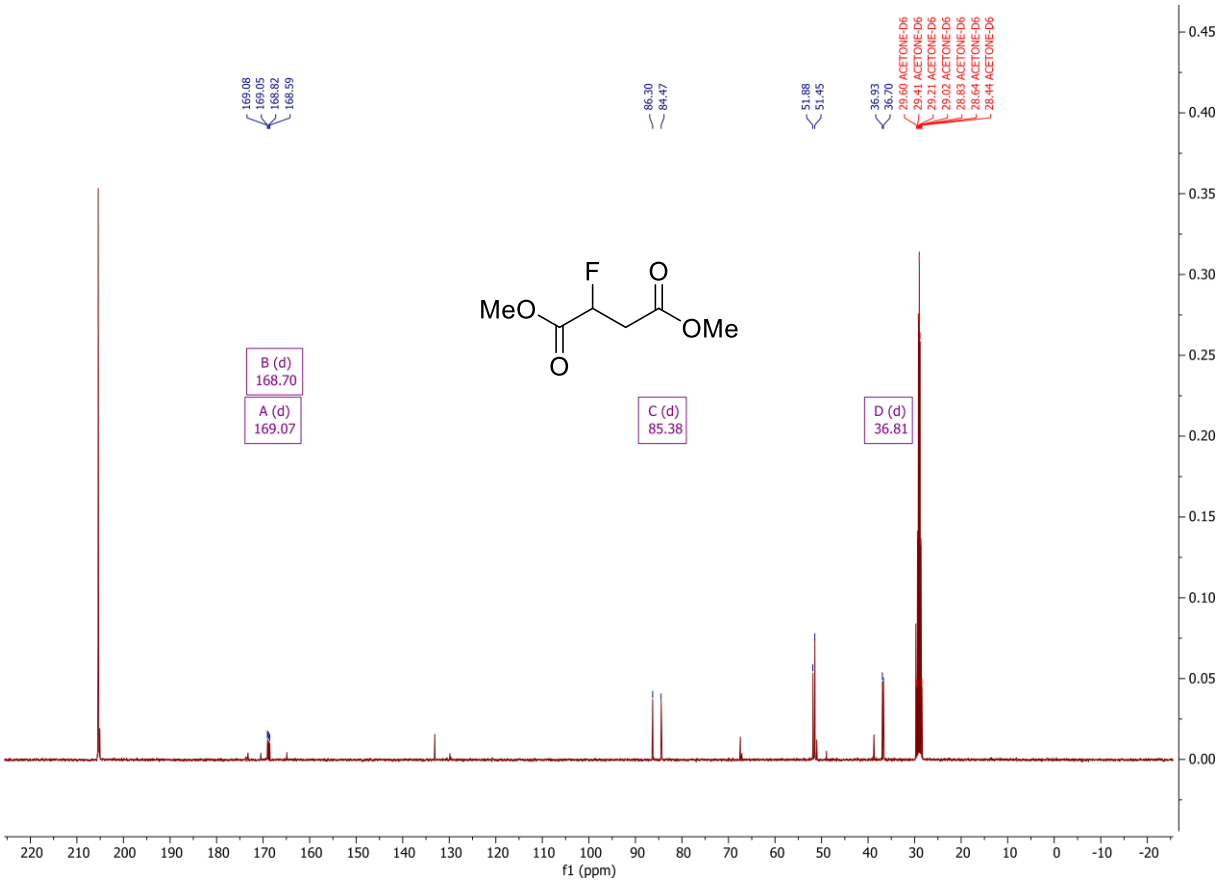


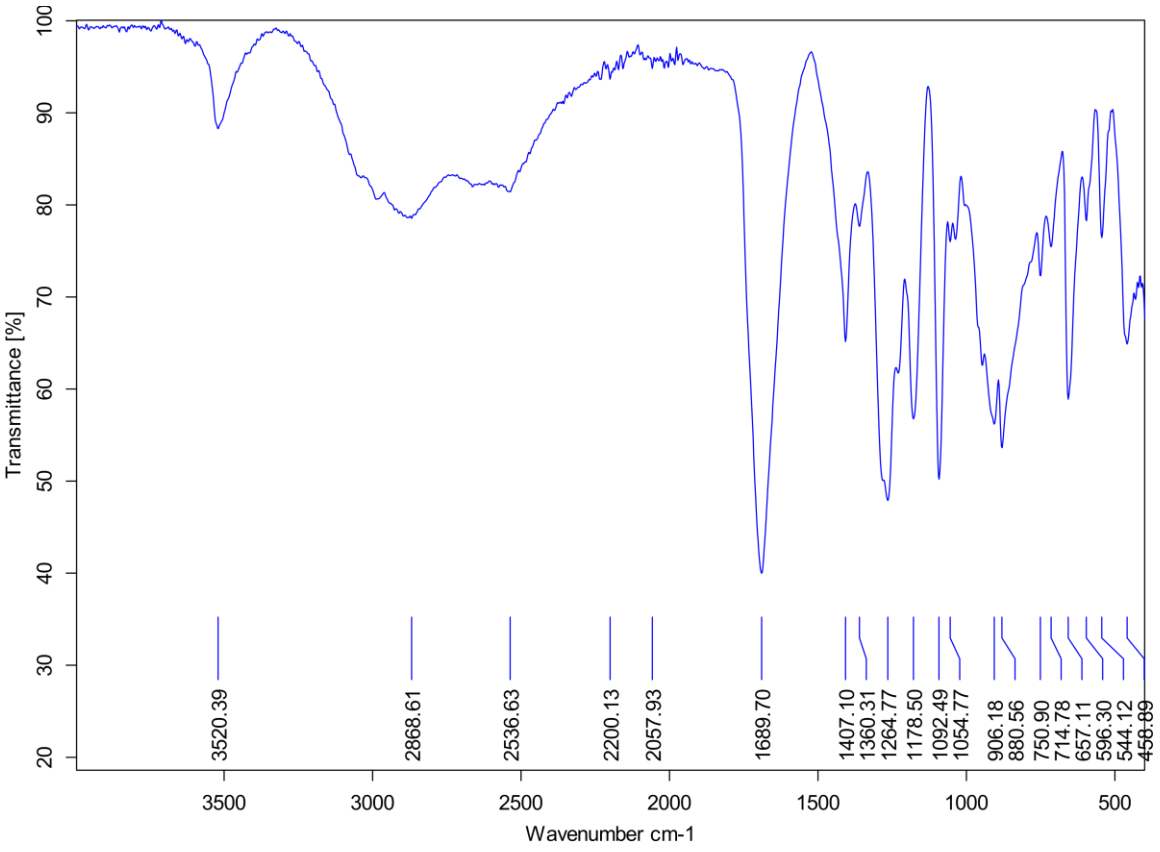

| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 185.0424 | 1 | C6H10NaO5 | 185.0420 | -1.7 | -0.3 | 0.3 | -2.4 |

Appendix 20. Compound **54:**



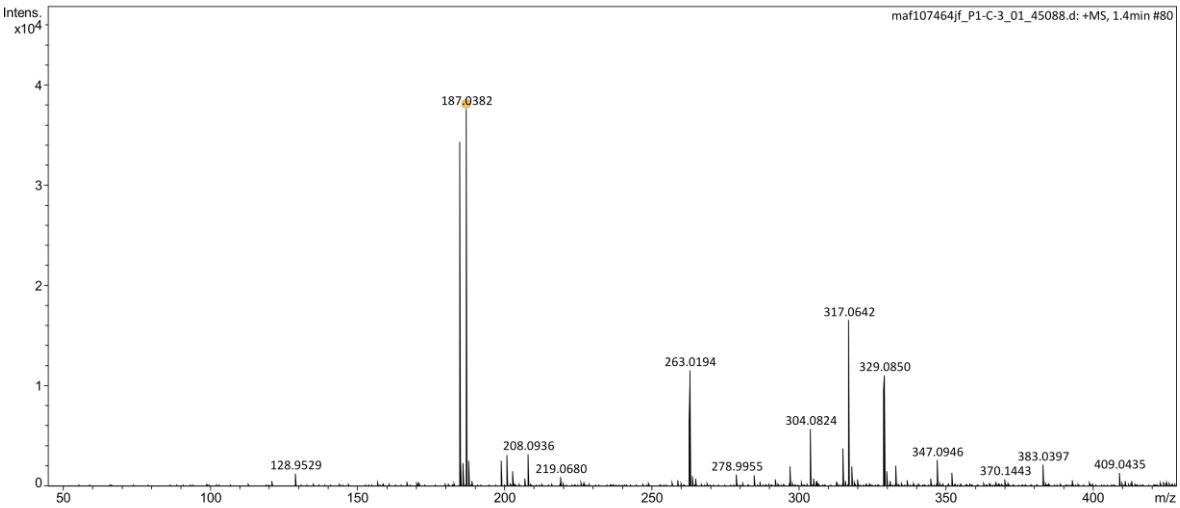




169.08
169.05
168.82
168.59
86.30
84.47
51.88
51.45
36.93
36.70
29.60 ACETONE-D6
29.41 ACETONE-D6
29.21 ACETONE-D6
29.02 ACETONE-D6
28.83 ACETONE-D6
28.64 ACETONE-D6
28.44 ACETONE-D6
F
O
MeO
OMe
O
B (d)
168.70
A (d)
169.07
C (d)
85.38
D (d)
36.81
f1 (ppm)


Transmittance [%]
3520.39
2868.61
2536.63
2200.13
2057.93
1689.70
1407.10
1360.31
1264.77
1178.50
1092.49
1054.77
906.18
880.56
750.90
714.78
657.11
596.30
544.12
458.89
Wavenumber cm-1

**Analysis Information**

| | | | |
|---|---|---|---|
| Analysis Filename | maf107464jf_P1-C-3_01_45088.d | Acquisition Date | 22/11/2023 06:59:08 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf107464jf | ESI | Positive |




| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 187.0382 | 1 | C6H9FNaO4 | 187.0377 | -2.4 | -0.4 | 2.9 | -4.3 |

## Appendix 21. WT DspB Sequence

DNA:

ATGAATTGTTGCGTAAAAGGAAACTCAATATATCCACAGAAGACGTCCACCAAACAGACCGGTTTGATGCT
GGACATCGCCAGACATTTTTATTCGCCGGAAGTCATAAAGTCGTTCATTGACACCATTAGCCTGTCTGGTGG
TAATTTTCTGCATCTGCACTTCAGCGATCATGAGAACTACGCGATTGAGAGCCATCTGTTGAATCAGCGTGC
AGAAAACGCTGTGCAAGGTAAGGACGGCATTTACATTAATCCGTATACCGGCAAACCGTTTCTGTCTTACC
GCCAACTGGATGATATTAAAGCGTATGCTAAGGCGAAGGGCATCGAGCTGATCCCGGAACTGGACTCCCC
GAACCACATGACCGCAATTTTTAAGCTGGTTCAAAAGGACCGCGGTGTTAAATATCTGCAGGGCTTGAAGA
GTCGTCAGGTTGATGATGAGATCGACATCACGAACGCGGATAGCATCACCTTCATGCAGAGCCTGATGAG
CGAAGTGATCGACATCTTCGGCGACACTTCTCAACATTTCCACATCGGCGGTGATGAGTTTGGTTATAGCGT
TGAAAGCAATCACGAGTTCATCACCTATGCGAATAAACTGTCATACTTCCTTGAGAAAAAAGGCTTGAAGA
CCCGTATGTGGAATGATGGCCTGATCAAGAACACCTTCGAGCAGATTAACCCGAATATTGAAATCACGTAC
TGGAGCTACGACGGTGACACCCAGGATAAAAACGAAGCGGCGGAACGTCGTGATATGCGCGTCAGCTTAC
CGGAGCTGCTGGCAAAAGGCTTTACCGTTTTGAACTACAACAGCTACTACCTTTATATCGTGCCGAAAGCTA
GCCCGACCTTCTCCCAAGATGCCGCGTTTGCAGCTAAGGACGTTATCAAGAACTGGGATCTGGGTGTGTGG
GACGGCCGTAATACGAAAAACCGCGTGCAAAACACCCATGAAATCGCGGGTGCAGCGTTGTCTATTTGGG
GTGAAGATGCTAAGGCCCTGAAGGACGAGACAATTCAGAAAAACACTAAATCCCTCTTGGAAGCGGTGAT
TCATAAAACCAACGGCGACGAGCACCACCACCACCACCAC

Amino Acid:

MNCCVKGNSIYPQKTSTKQTGLMLDIARHFYSPEVIKSFIDTISLSGGNFLHLHFSDHENYAIESHLLNQRAENAV
QGKDGIYINPYTGKPFLSYRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDRGVKYLQGLKSRQVDDEIDIT
NADSITFMQSLMSEVIDIFGDTSQHFHIGGDEFGYSVESNHEFITYANKLSYFLEKKGLKTRMWNDGLIKNTFEQ
INPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLAKGFTVLNYNSYYLYIVPKASPTFSQDAAFAAKDVIKNW
DLGVWDGRNTKNRVQNTHEIAGAALSIWGEDAKALKDETIQKNTKSLLEAVIHKTNGDEHHHHHH

Appendix 22. DspB D183A Sequences

DNA:

ATGAATTGTTGCGTAAAAGGCAATTCCATATATCCGCAAAAAACAAGTACCAAGCAGACCGGATTAATGCT
GGACATCGCCCGACATTTTTATTCACCCGAGGTGATTAAATCCTTTATTGATACCATCAGCCTTTCCGGCGG
TAATTTTCTGCACCTGCATTTTTCCGACCATGAAAACTATGCGATAGAAAGCCATTTACTTAATCAACGTGC
GGAAAATGCCGTGCAGGGCAAAGACGGTATTTATATTAATCCTTATACCGGAAAGCCATTCTTGAGTTATC
GGCAACTTGACGATATCAAAGCCTATGCTAAGGCAAAAGGCATTGAGTTGATTCCCGAACTTGACAGCCCG
AATCACATGACGGCGATCTTTAAACTGGTGCAAAAAGACAGAGGGGTCAAGTACCTTCAAGGATTAAAAT
CACGCCAGGTAGATGATGAAATTGATATTACTAATGCTGACAGTATTACTTTTATGCAATCTTTAATGAGTG
AGGTTATTGATATTTTTGGCGACACGAGTCAGCATTTTCATATTGGTGGCGCCGAATTTGGTTATTCTGTGG
AAAGTAATCATGAGTTTATTACGTATGCCAATAAACTATCCTACTTTTTAGAGAAAAAAGGGTTGAAAACCC

GAATGTGGAATGACGGATTAATTAAAAATACTTTTGAGCAAATCAACCCGAATATTGAAATTACTTATTGG
AGCTATGATGGCGATACGCAGGACAAAAATGAAGCTGCCGAGCGCCGTGATATGCGGGTCAGTTTGCCGG
AGTTGCTGGCGAAAGGCTTTACTGTCCTGAACTATAATTCCTATTATCTTTACATTGTTCCGAAAGCTTCACC
AACCTTCTCGCAAGATGCCGCCTTTGCCGCCAAAGATGTTATAAAAAATTGGGATCTTGGTGTTTGGGATG
GACGAAACACCAAAAACCGCGTACAAAATACTCATGAAATAGCCGGCGCAGCATTATCGATCTGGGGAGA
AGATGCAAAAGCGCTGAAAGACGAAACAATTCAGAAAAACACGAAAAGTTTATTGGAAGCGGTGATTCAT
AAGACGAATGGGGATGAGCACCACCACCACCACTGA

Amino Acid:

MNCCVKGNSIYPQKTSTKQTGLMLDIARHFYSPEVIKSFIDTISLSGGNFLHLHFSDHENYAIESHLLNQRAENAV
QGKDGIYINPYTGKPFLSYRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDRGVKYLQGLKSRQVDDEIDIT
NADSITFMQSLMSEVIDIFGDTSQHFHIGGAEFGYSVESNHEFITYANKLSYFLEKKGLKTRMWNDGLIKNTFEQ
INPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLAKGFTVLNYNSYYLYIVPKASPTFSQDAAFAAKDVIKNW
DLGVWDGRNTKNRVQNTHEIAGAALSIWGEDAKALKDETIQKNTKSLLEAVIHKTNGDEHHHHHH

Appendix 23. DspB Y278N Sequences

DNA:

ATGAATTGTTGCGTAAAAGGCAATTCCATATATCCGCAAAAAACAAGTACCAAGCAGACCGGATTAATGCT
GGACATCGCCCGACATTTTTATTCACCCGAGGTGATTAAATCCTTTATTGATACCATCAGCCTTTCCGGCGG
TAATTTTCTGCACCTGCATTTTTCCGACCATGAAAACTATGCGATAGAAAGCCATTTACTTAATCAACGTGC
GGAAAATGCCGTGCAGGGCAAAGACGGTATTTATATTAATCCTTATACCGGAAAGCCATTCTTGAGTTATC
GGCAACTTGACGATATCAAAGCCTATGCTAAGGCAAAAGGCATTGAGTTGATTCCCGAACTTGACAGCCCG
AATCACATGACGGCGATCTTTAAACTGGTGCAAAAAGACAGAGGGGTCAAGTACCTTCAAGGATTAAAAT
CACGCCAGGTAGATGATGAAATTGATATTACTAATGCTGACAGTATTACTTTTATGCAATCTTTAATGAGTG
AGGTTATTGATATTTTTGGCGACACGAGTCAGCATTTTCATATTGGTGGCGACGAATTTGGTTATTCTGTGG
AAAGTAATCATGAGTTTATTACGTATGCCAATAAACTATCCTACTTTTTAGAGAAAAAAGGGTTGAAAACCC
GAATGTGGAATGACGGATTAATTAAAAATACTTTTGAGCAAATCAACCCGAATATTGAAATTACTTATTGG
AGCTATGATGGCGATACGCAGGACAAAAATGAAGCTGCCGAGCGCCGTGATATGCGGGTCAGTTTGCCGG
AGTTGCTGGCGAAAGGCTTTACTGTCCTGAACTATAATTCCTATTATCTTAACATTGTTCCGAAAGCTTCACC
AACCTTCTCGCAAGATGCCGCCTTTGCCGCCAAAGATGTTATAAAAAATTGGGATCTTGGTGTTTGGGATG
GACGAAACACCAAAAACCGCGTACAAAATACTCATGAAATAGCCGGCGCAGCATTATCGATCTGGGGAGA
AGATGCAAAAGCGCTGAAAGACGAAACAATTCAGAAAAACACGAAAAGTTTATTGGAAGCGGTGATTCAT
AAGACGAATGGGGATGAGCACCACCACCACCACTGA

Amino Acid:

MNCCVKGNSIYPQKTSTKQTGLMLDIARHFYSPEVIKSFIDTISLSGGNFLHLHFSDHENYAIESHLLNQRAENAV
QGKDGIYINPYTGKPFLSYRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDRGVKYLQGLKSRQVDDEIDIT
NADSITFMQSLMSEVIDIFGDTSQHFHIGGDEFGYSVESNHEFITYANKLSYFLEKKGLKTRMWNDGLIKNTFEQ
INPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLAKGFTVLNYNSYYLNIVPKASPTFSQDAAFAAKDVIKN
WDLGVWDGRNTKNRVQNTHEIAGAALSIWGEDAKALKDETIQKNTKSLLEAVIHKTNGDEHHHHHH

Appendix 24. DspB Y278F Sequences

DNA:

ATGAATTGTTGCGTAAAAGGCAATTCCATATATCCGCAAAAAACAAGTACCAAGCAGACCGGATTAATGCT
GGACATCGCCCGACATTTTTATTCACCCGAGGTGATTAAATCCTTTATTGATACCATCAGCCTTTCCGGCGG
TAATTTTCTGCACCTGCATTTTTCCGACCATGAAAACTATGCGATAGAAAGCCATTTACTTAATCAACGTGC
GGAAAATGCCGTGCAGGGCAAAGACGGTATTTATATTAATCCTTATACCGGAAAGCCATTCTTGAGTTATC
GGCAACTTGACGATATCAAAGCCTATGCTAAGGCAAAAGGCATTGAGTTGATTCCCGAACTTGACAGCCCG
AATCACATGACGGCGATCTTTAAACTGGTGCAAAAAGACAGAGGGGTCAAGTACCTTCAAGGATTAAAAT
CACGCCAGGTAGATGATGAAATTGATATTACTAATGCTGACAGTATTACTTTTATGCAATCTTTAATGAGTG
AGGTTATTGATATTTTTGGCGACACGAGTCAGCATTTTCATATTGGTGGCGaCGAATTTGGTTATTCTGTGG
AAAGTAATCATGAGTTTATTACGTATGCCAATAAACTATCCTACTTTTTAGAGAAAAAAGGGTTGAAAACCC
GAATGTGGAATGACGGATTAATTAAAAATACTTTTGAGCAAATCAACCCGAATATTGAAATTACTTATTGG
AGCTATGATGGCGATACGCAGGACAAAAATGAAGCTGCCGAGCGCCGTGATATGCGGGTCAGTTTGCCGG
AGTTGCTGGCGAAAGGCTTTACTGTCCTGAACTATAATTCCTATTATCTTTTCATTGTTCCGAAAGCTTCACC
AACCTTCTCGCAAGATGCCGCCTTTGCCGCCAAAGATGTTATAAAAAATTGGGATCTTGGTGTTTGGGATG
GACGAAACACCAAAAACCGCGTACAAAATACTCATGAAATAGCCGGCGCAGCATTATCGATCTGGGGAGA
AGATGCAAAAGCGCTGAAAGACGAAACAATTCAGAAAAACACGAAAAGTTTATTGGAAGCGGTGATTCAT
AAGACGAATGGGGATGAGCACCACCACCACCACTGA

Amino Acid:

MNCCVKGNSIYPQKTSTKQTGLMLDIARHFYSPEVIKSFIDTISLSGGNFLHLHFSDHENYAIESHLLNQRAENAV
QGKDGIYINPYTGKPFLSYRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDRGVKYLQGLKSRQVDDEIDIT
NADSITFMQSLMSEVIDIFGDTSQHFHIGGDEFGYSVESNHEFITYANKLSYFLEKKGLKTRMWNDGLIKNTFEQ
INPNIEITYWSYDGDTQDKNEAAERRDMRVSLPELLAKGFTVLNYNSYYLFIVPKASPTFSQDAAFAAKDVIKNW
DLGVWDGRNTKNRVQNTHEIAGAALSIWGEDAKALKDETIQKNTKSLLEAVIHKTNGDEHHHHHH

Appendix 25. DspB E248Q Sequences

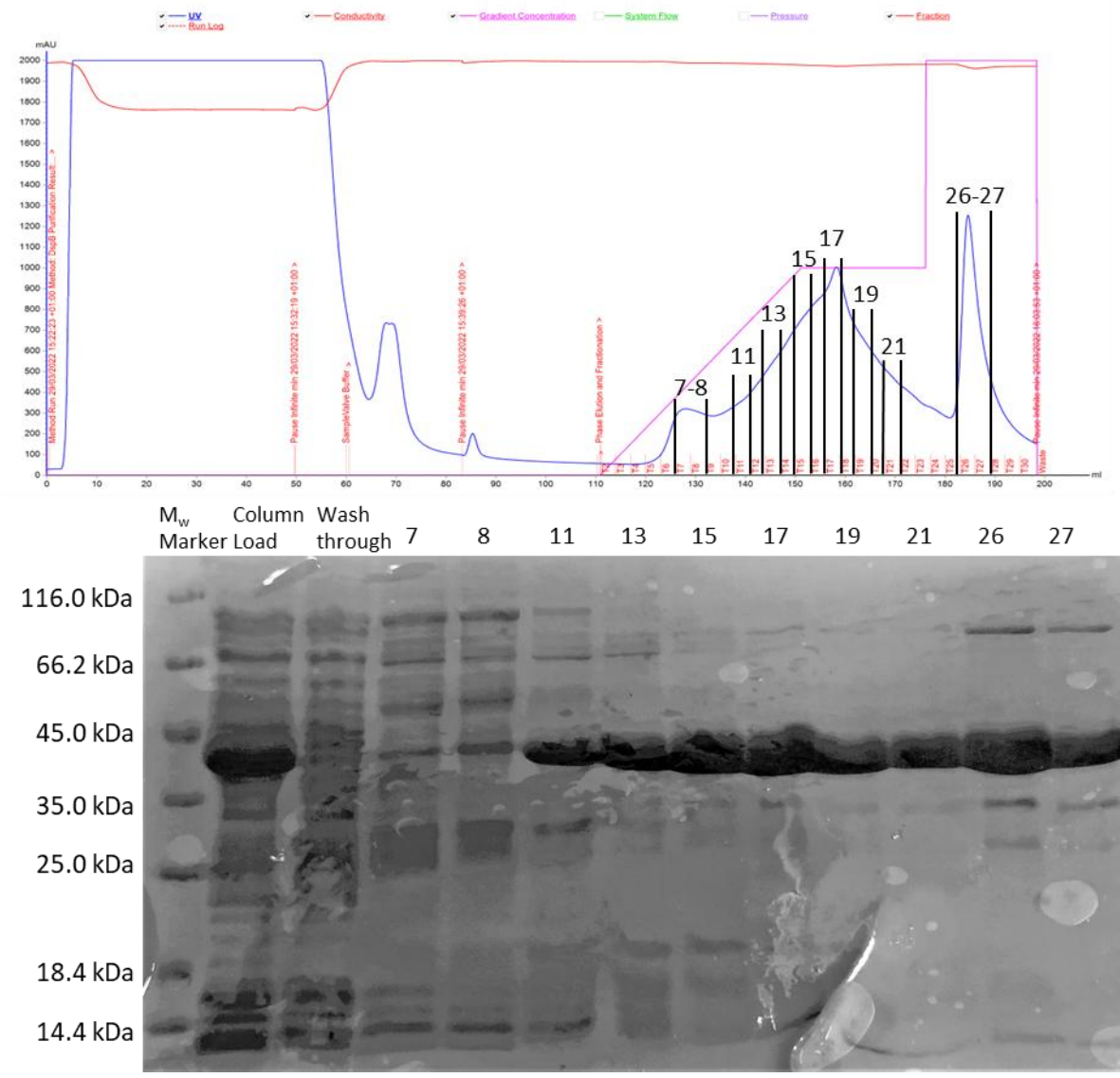
DNA:

CATATGAATTGTTGCGTAAAAGGAAACTCAATATATCCACAGAAGACCAGCACCAAACAGACCGGTTTAAT
GCTGGATATCGCCCGTCATTTCTATAGCCCGGAGGTCATCAAGTCGTTCATTGACACTATCAGCCTGAGCG
GTGGAAACTTTCTTCACCTCCACTTCAGCGACCACGAGAACTATGCGATTGAGTCACATCTTTTGAATCAAC
GTGCTGAAAACGCTGTGCAAGGTAAAGACGGCATCTACATCAACCCGTATACCGGCAAACCGTTTCTGTCC
TATCGTCAACTGGACGATATAAAGGCGTATGCGAAAGCGAAGGGTATCGAGCTGATTCCGGAGCTGGATA
GCCCGAACCACATGACGGCTATCTTCAAACTGGTTCAGAAGGACCGCGGTGTAAAATACCTGCAGGGCCT
GAAGAGCAGACAGGTTGATGACGAAATTGACATCACAAATGCGGATTCTATCACCTTCATGCAATCCTTAA
TGTCCGAAGTCATCGATATCTTCGGCGACACTTCGCAACATTTTCATATCGGTGGTGACGAGTTCGGTTATA
GTGTTGAAAGCAATCACGAGTTTATCACCTACGCTAATAAATTGTCTTACTTCTTGGAAAAGAAAGGCTTGA
AAACCCGTATGTGGAATGATGGCCTGATCAAGAACACTTTCGAGCAGATTAACCCGAATATTGAAATCACC
TACTGGTCTTACGATGGCGATACCCAGGATAAGAACCAGGCAGCGGAGCGCCGTGATATGCGTGTGAGCC
TACCGGAGCTGCTGGCAAAGGGCTTTACCGTGCTCAACTACAATAGCTACTACCTGTATATTGTTCCGAAA
GCCTCCCCGACCTTTAGCCAAGACGCCGCGTTTGCGGCGAAAGACGTGATTAAGAACTGGGATCTGGGTG
TTTGGGATGGCCGTAATACGAAGAACCGCGTGCAAAACACCCATGAAATTGCGGGCGCAGCTCTGTCCATT
TGGGGTGAAGACGCGAAGGCCCTGAAGGACGAAACCATTCAGAAAAACACGAAAAGCTTGCTGGAGGCA
GTTATTCACAAAACGAACGGTGACGAACTCGAGCACCACCACCACCACCACTGA

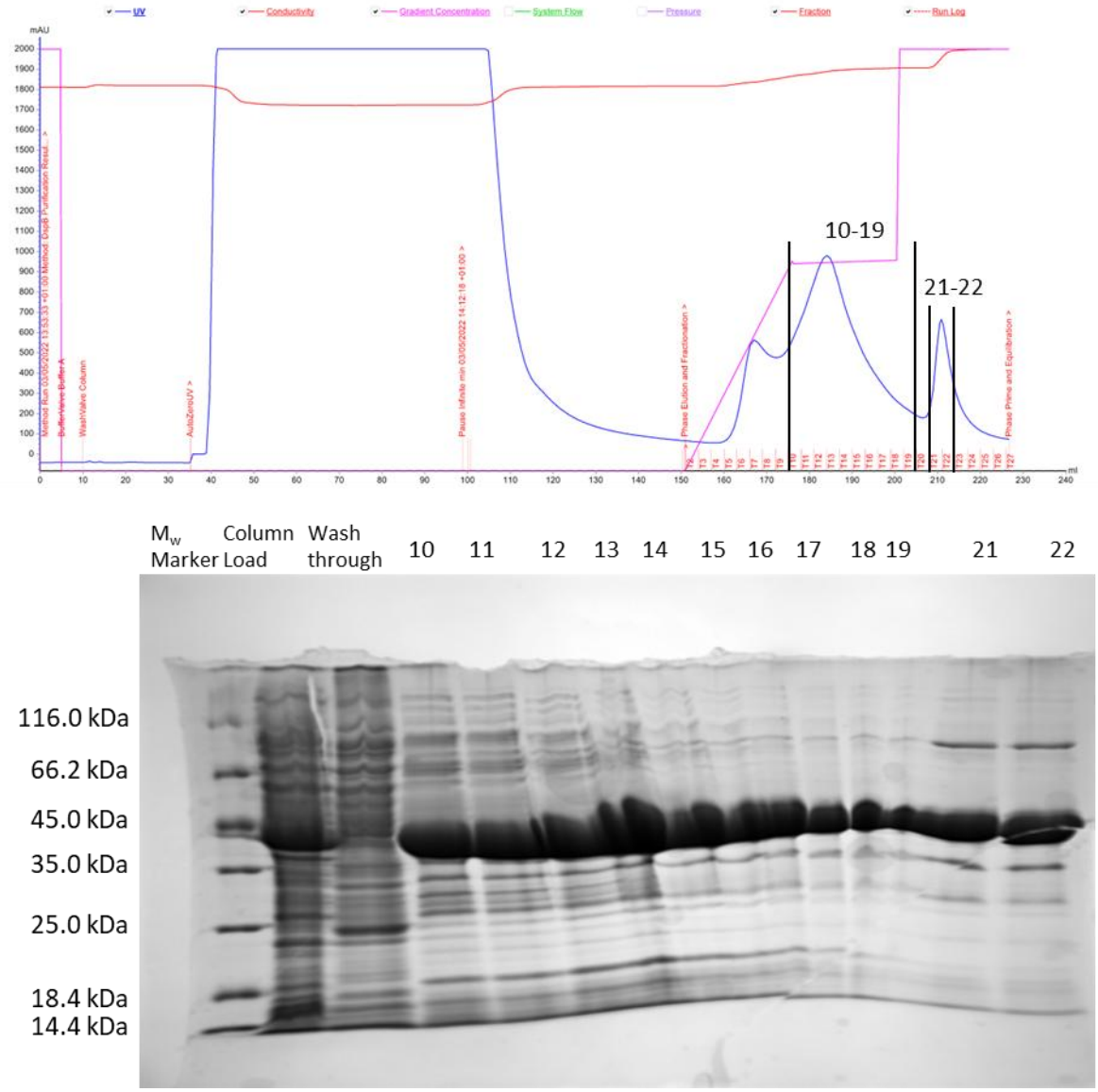
Amino Acid:

MNCCVKGNSIYPQKTSTKQTGLMLDIARHFYSPEVIKSFIDTISLSGGNFLHLHFSDHENYAIESHLLNQRAENAV
QGKDGIYINPYTGKPFLSYRQLDDIKAYAKAKGIELIPELDSPNHMTAIFKLVQKDRGVKYLQGLKSRQVDDEIDIT
NADSITFMQSLMSEVIDIFGDTSQHFHIGGDEFGYSVESNHEFITYANKLSYFLEKKGLKTRMWNDGLIKNTFEQ
INPNIEITYWSYDGDTQDKNQAAERRDMRVSLPELLAKGFTVLNYNSYYLYIVPKASPTFSQDAAFAAKDVIKN
WDLGVWDGRNTKNRVQNTHEIAGAALSIWGEDAKALKDETIQKNTKSLLEAVIHKTNGDEHHHHHH

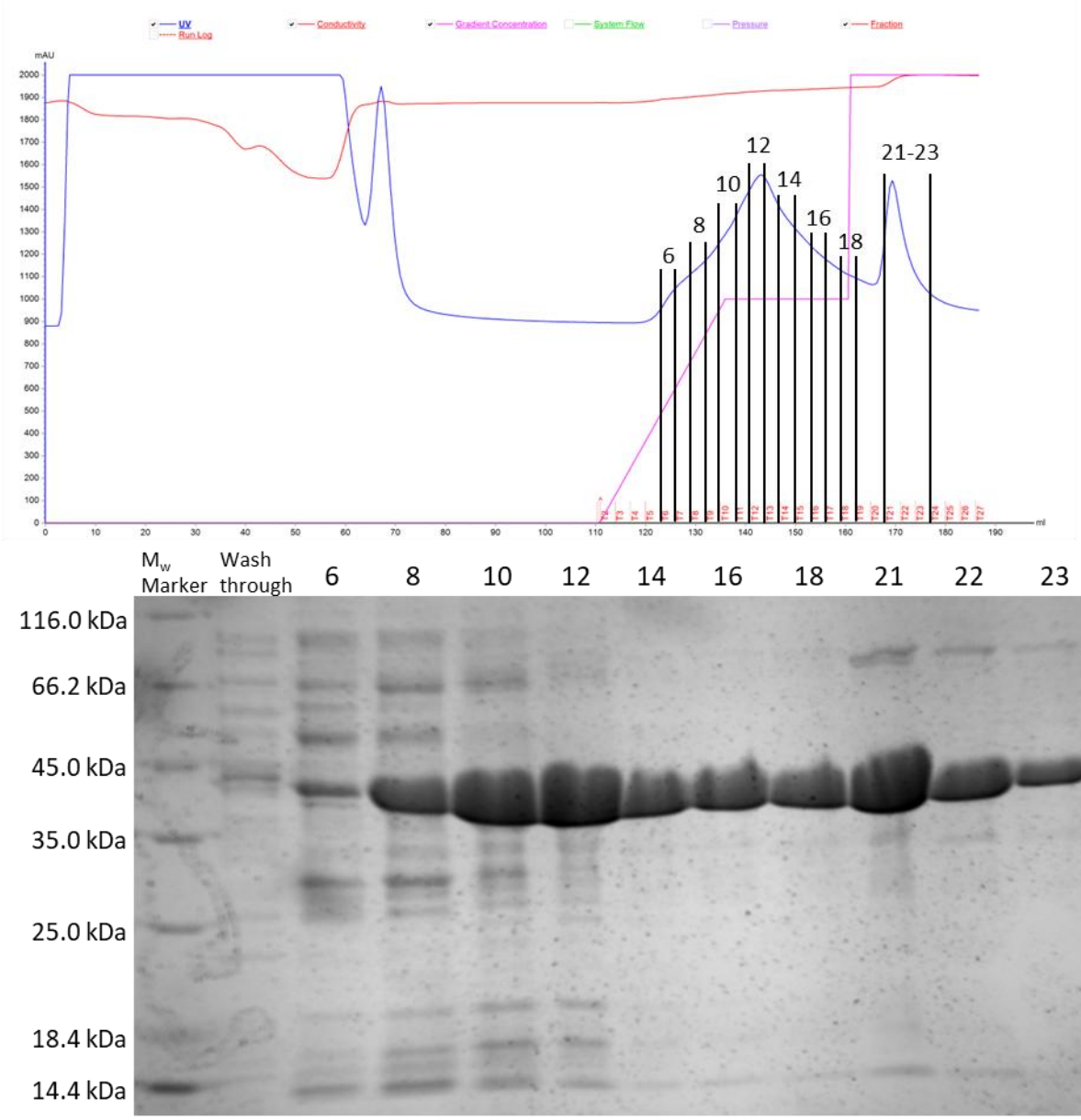
Appendix 26. DspB D183A Purification Chromatogram and SDS PAGE Gel



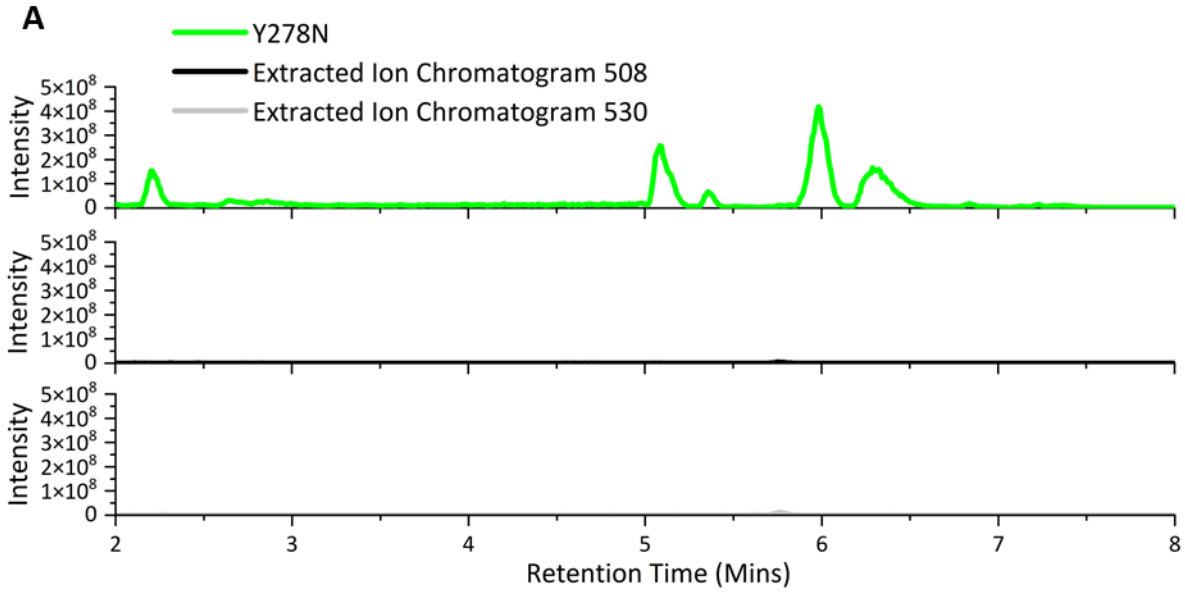
Appendix 27. DspB Y278F Purification Chromatogram and SDS PAGE Gel

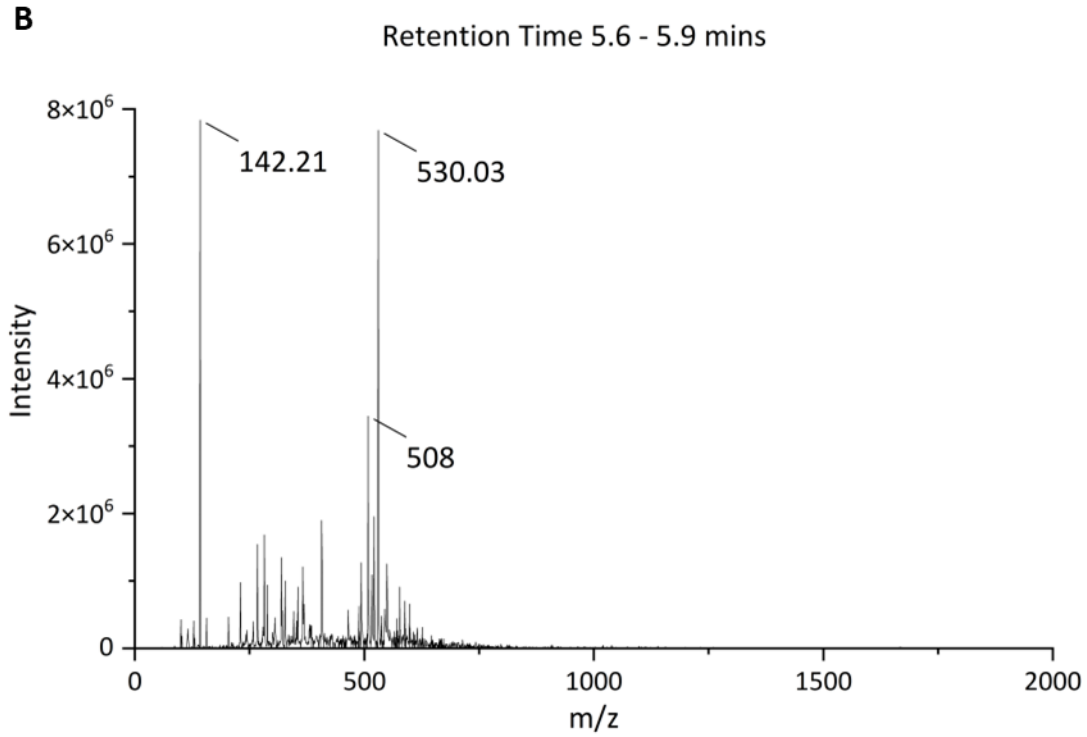


Appendix 28. DspB E248Q Purification Chromatogram

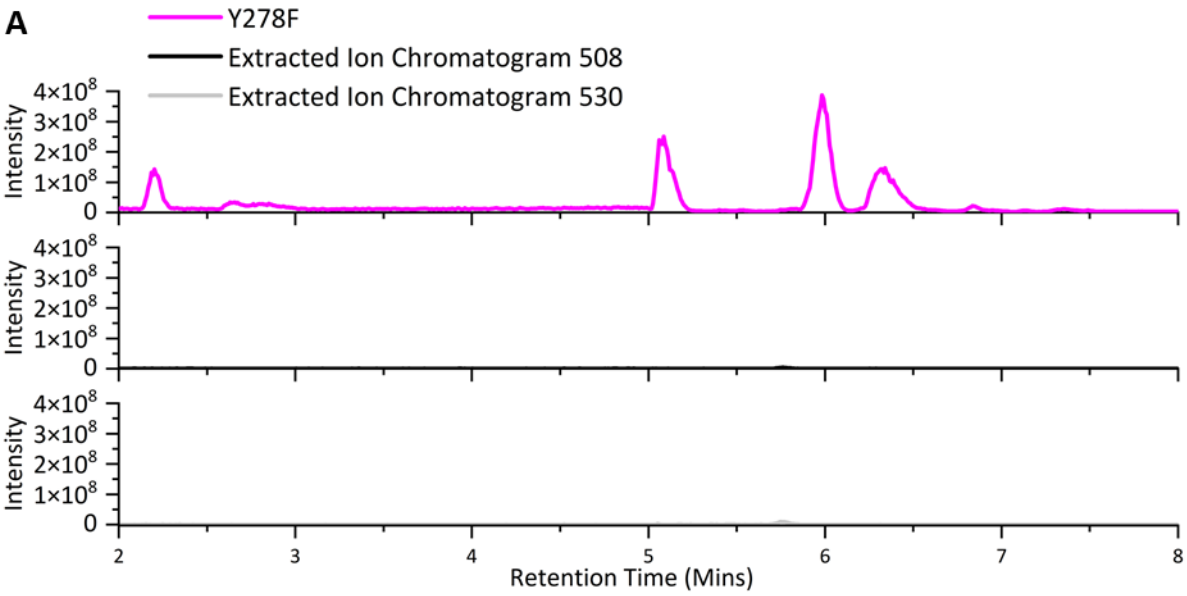


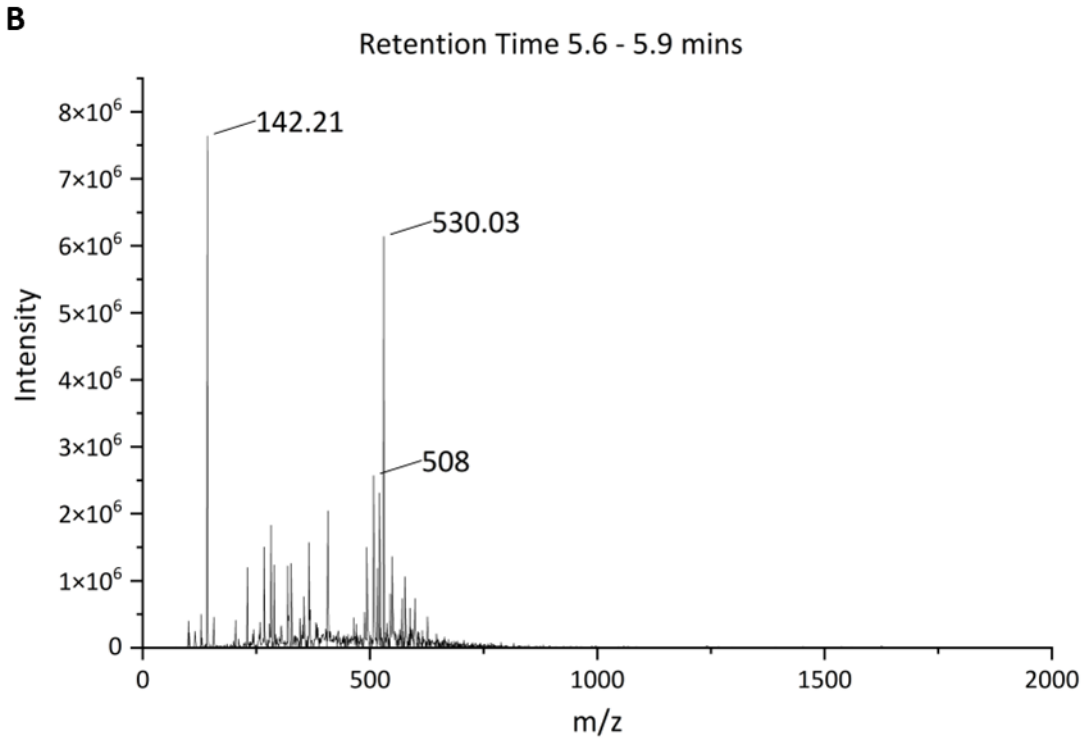
Appendix 29. Y278N Oxazoline **5** and Acceptor **8** TLC Assay LC-MS Chromatogram



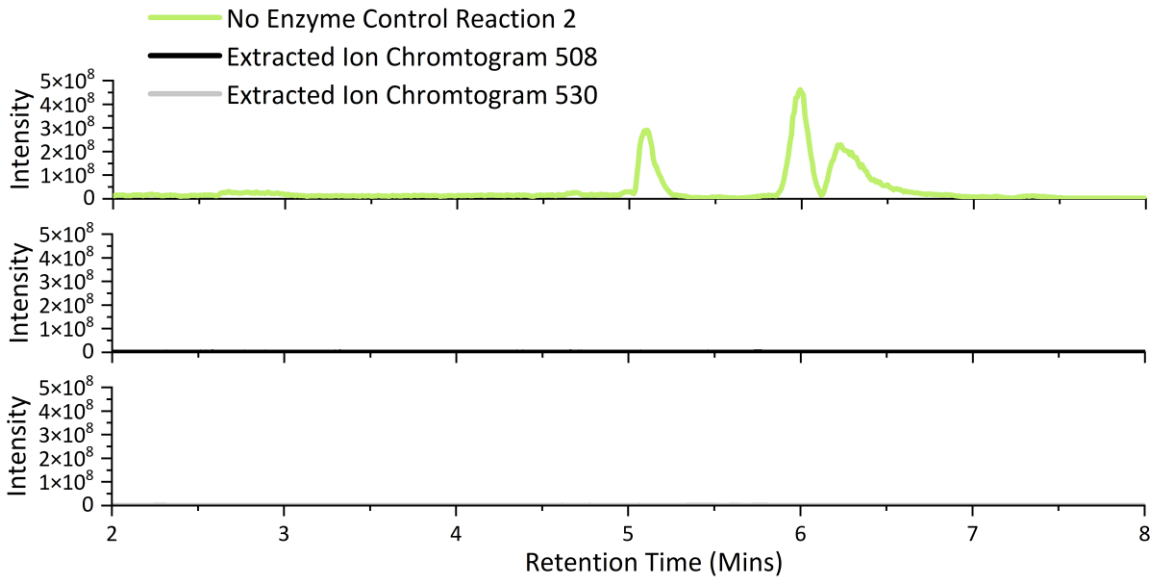



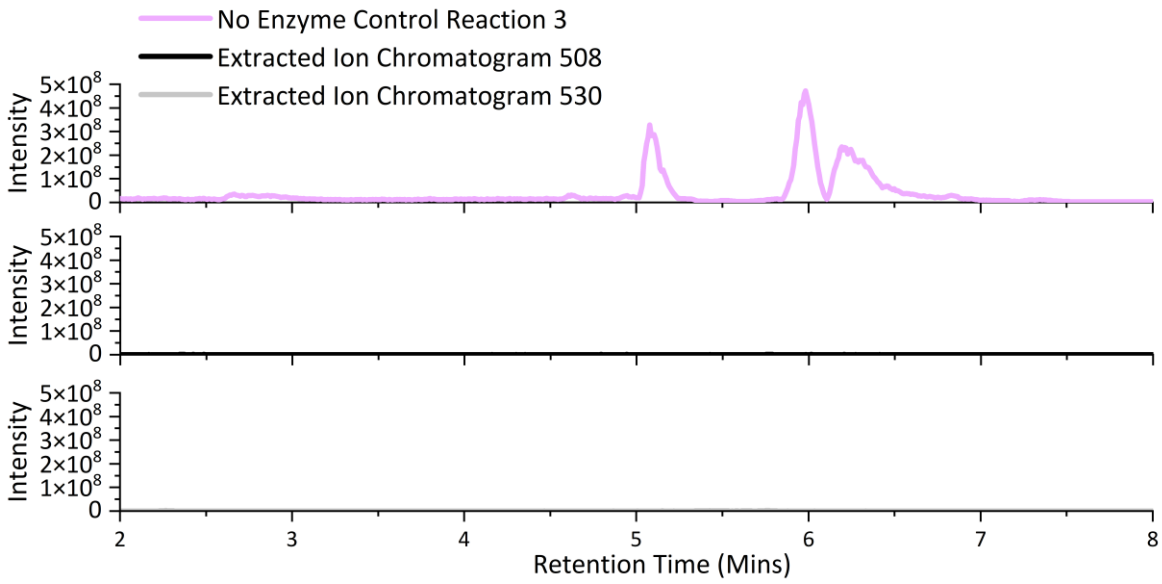
Appendix 30. Y278F Oxazoline **5** and Acceptor **8** TLC Assay LC-MS Chromatogram:

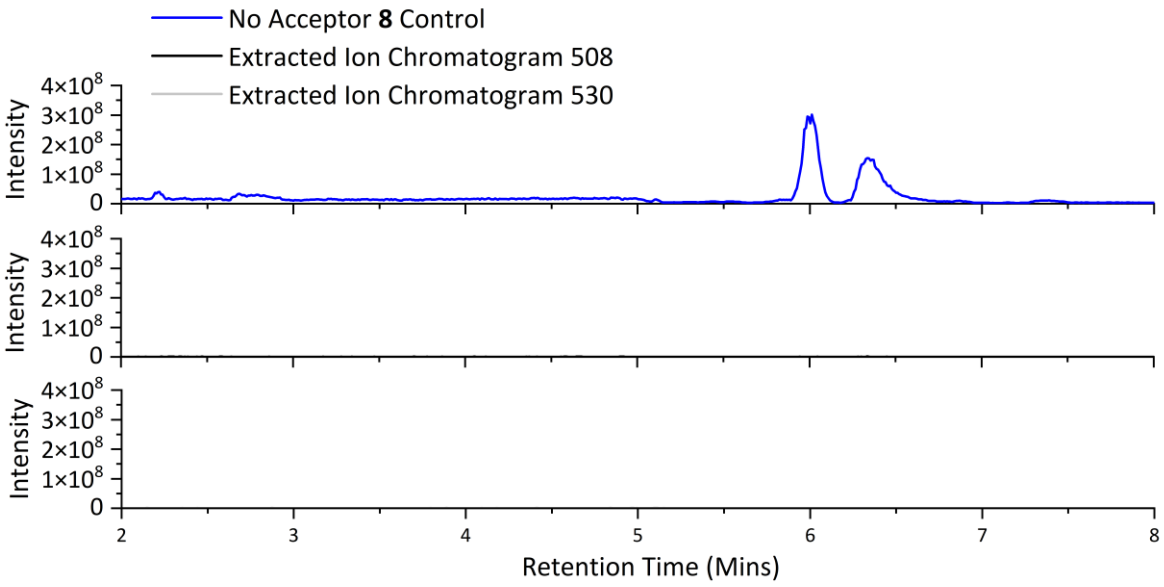


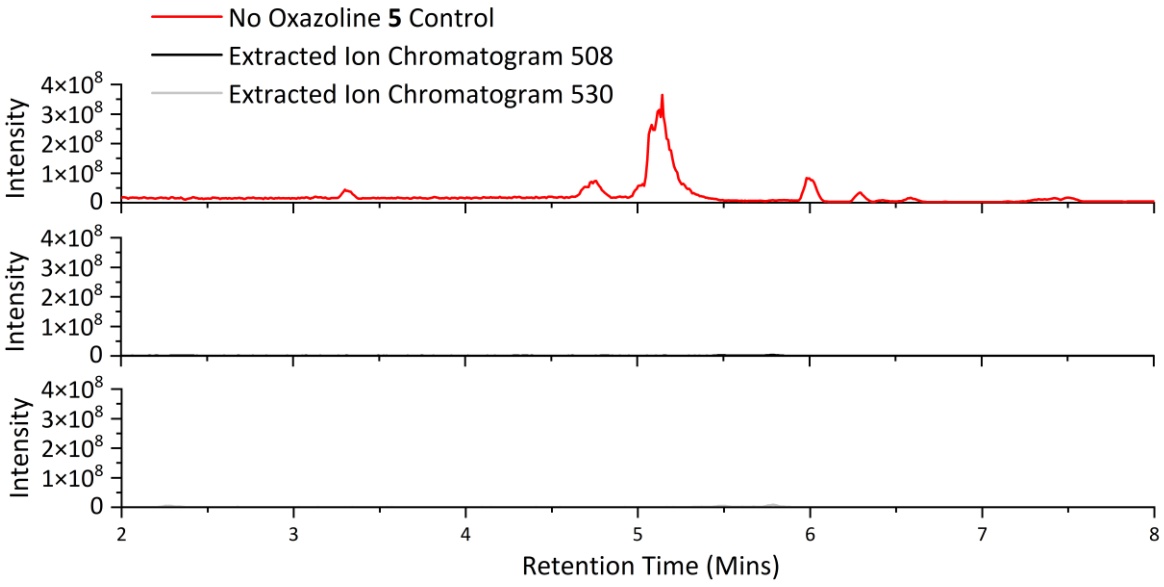



Appendix 31. TLC Assay LC-MS Chromatograms Control Reactions:

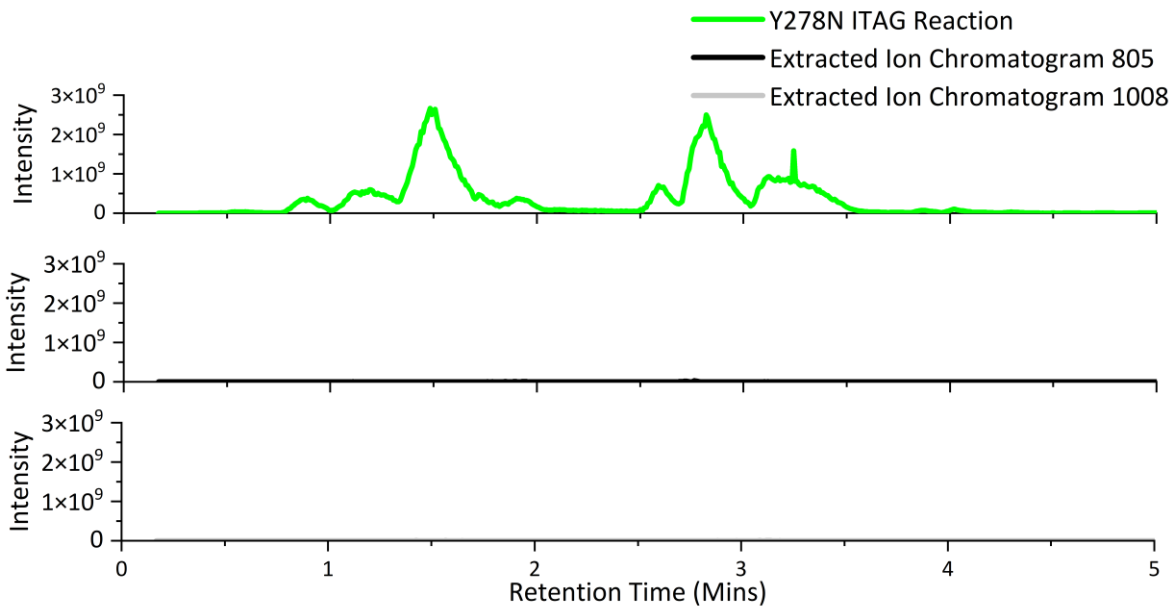


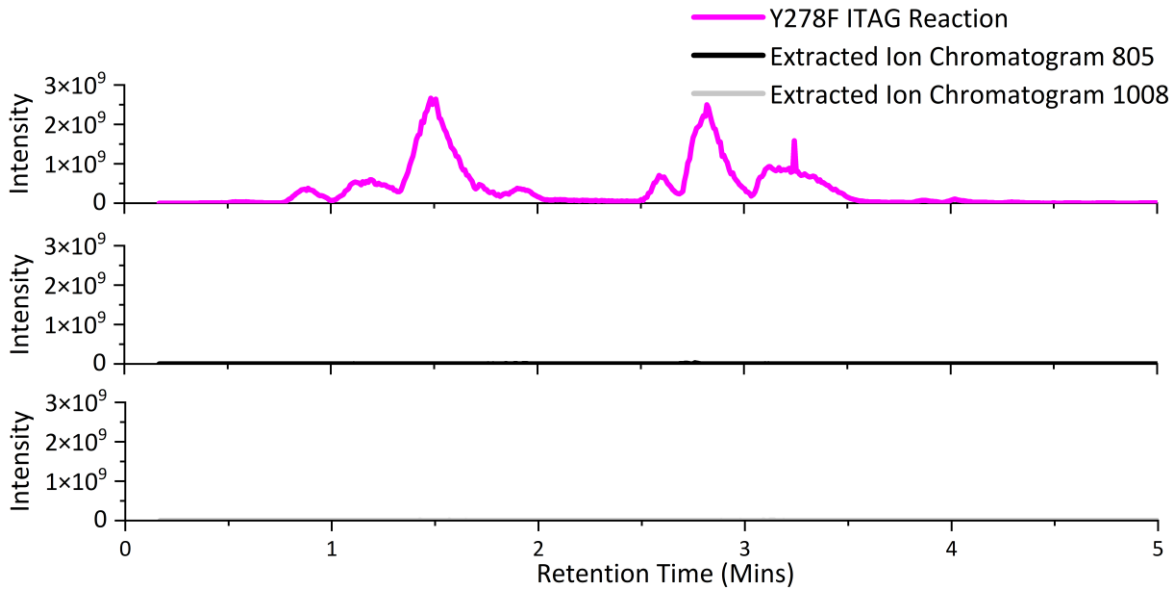

No Acceptor 8 Control
Extracted Ion Chromatogram 508
Extracted Ion Chromatogram 530
Intensity
4×10^8
3×10^8
2×10^8
1×10^8
0
2
3
4
5
6
7
8
Retention Time (Mins)


No Oxazoline 5 Control
Extracted Ion Chromatogram 508
Extracted Ion Chromatogram 530
Intensity
4×10^8
3×10^8
2×10^8
1×10^8
0
2
3
4
5
6
7
8
Retention Time (Mins)

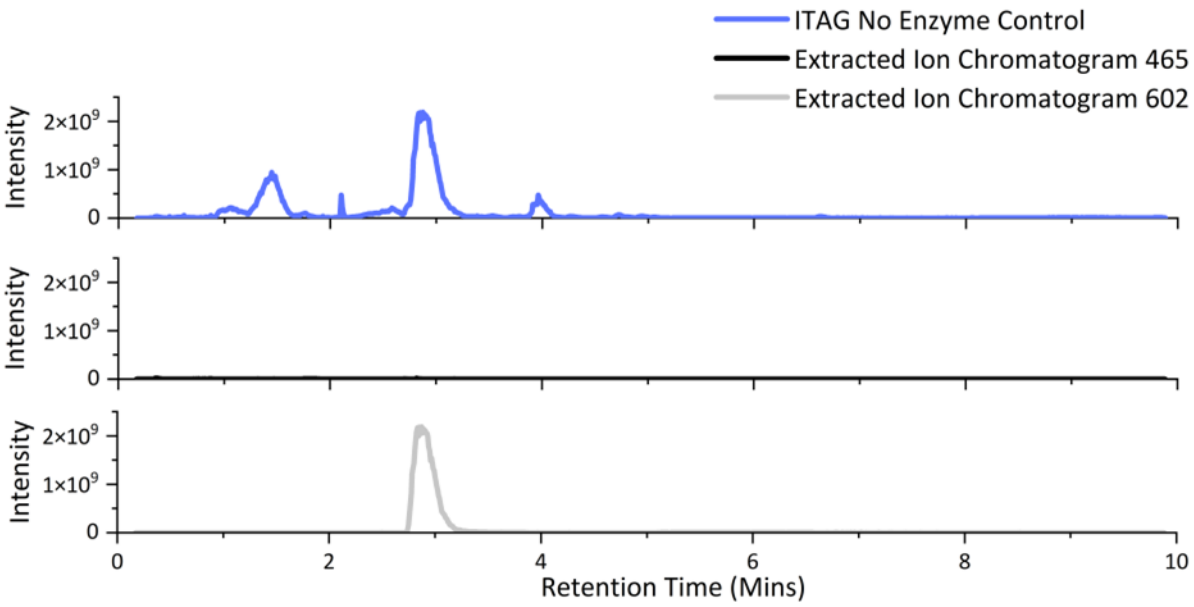
Appendix 32. Y278N ITAG Assay LC-MS Chromatogram:

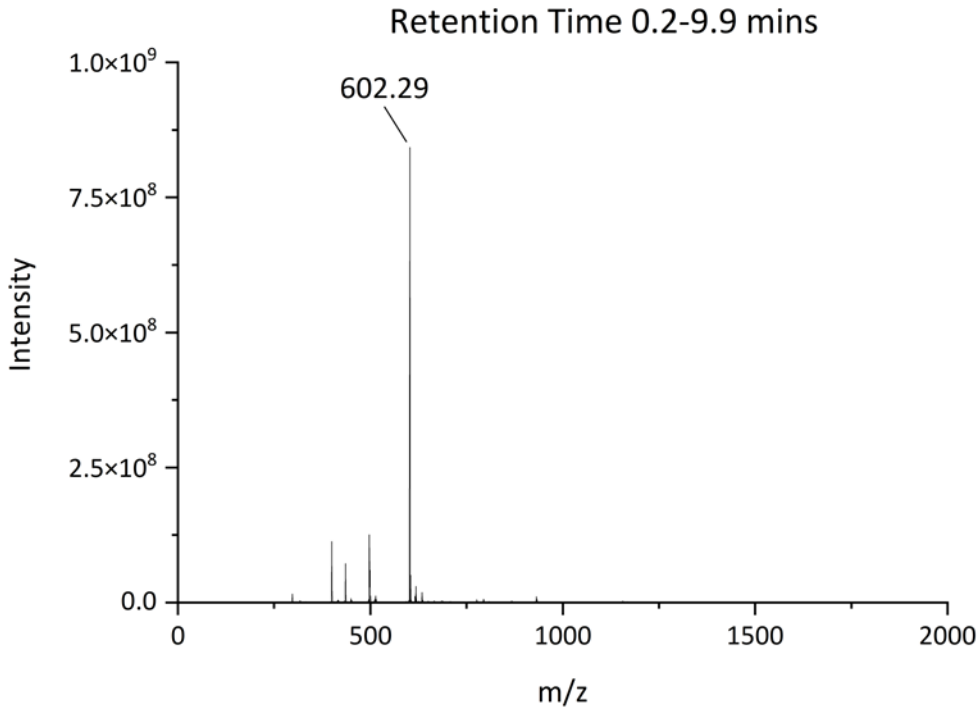



Appendix 33. Y278F ITAG Assay LC-MS Chromatogram:

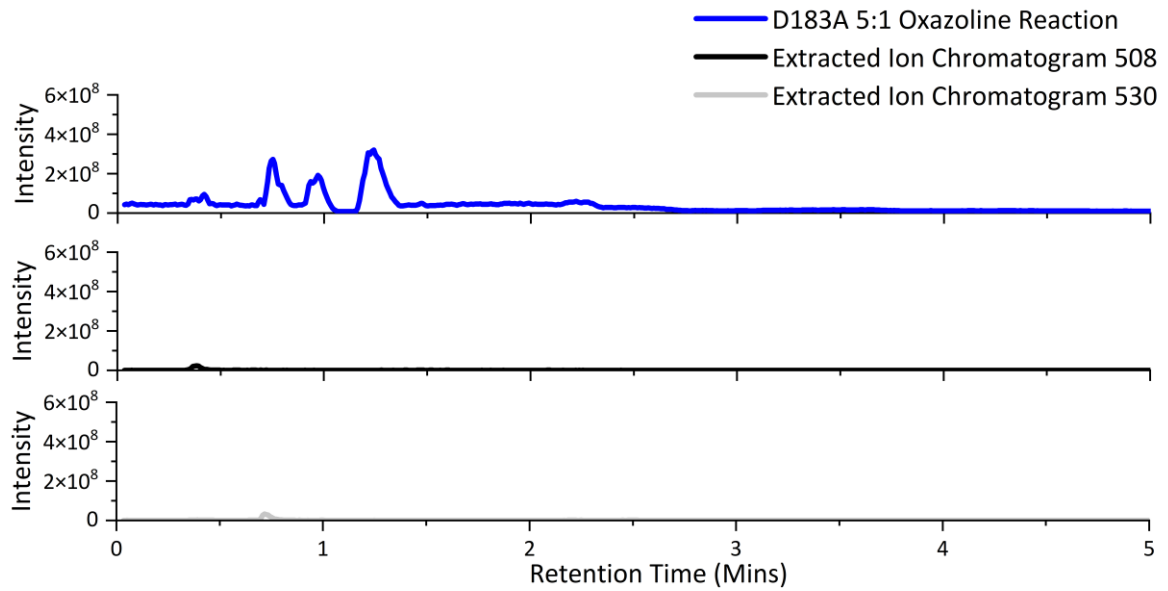


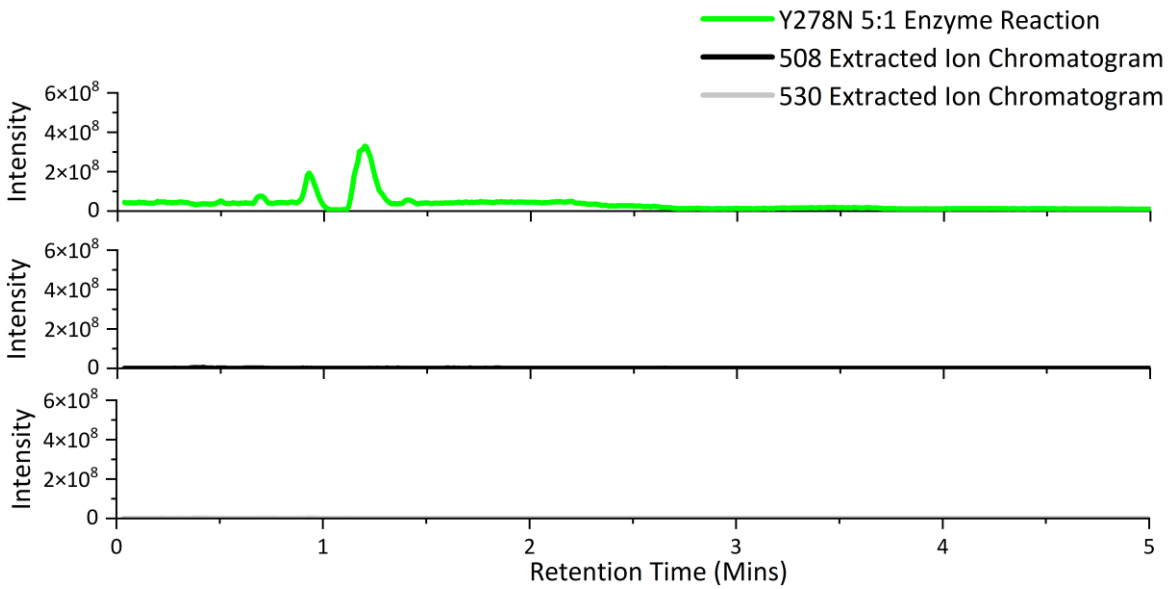
Appendix 34. ITAG Reaction No Enzyme Control

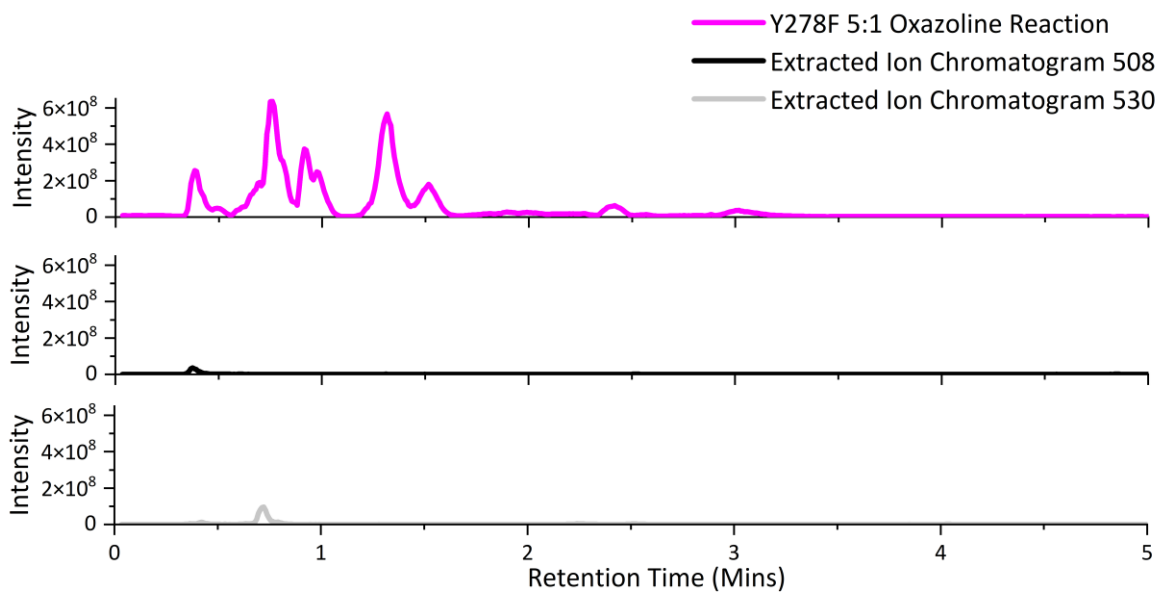


Appendix 35. 5:1 Oxazoline **5** Extracted Ion Chromatogram Analysis






Appendix 36. SlHex Sequences

DNA:

CATATGCAAAGTGATATTGAAAAAGGAATATCAATAGACATCGCCCGTACCCACTACACCAAGGACTCCAT
TAAAAAAATCATCGGCGAGCTGAGCAGAGTGAATGGCCGCTATCTGCAGCTTCACCTCGCGGATAACGAC
AACTATTCCATCTACAGCAATGTTCTGGGTCAAACCAGCACCCACAGTAACCATTATTACCTGACGAAGGCG
GAGCTGCGTGAATTGGTTCAATACGCAAATAAACACCACGTTCAATTGATCCCGGAATTGGACTTCCCGGC
ACACTCCAAGGCTATGCTGACGCTGCTGCATAAACACCATCCGAGCCAGTACCGTCAGGTGGTGTCTAGCT
ACGACAACACCATGCTGGACTTCCAGCAAAACCAAGCGGCGCTGGATGTTTCCCGCCAGTTGATCAACGAA
GTAGCTGATATCTTCTATCAGACCCCGTATAAGGACAATTTGAAGATGGTCATTGGCGGTGATGAAGTTCC
GGGTGGTGGCGCACACCAACGTGATTTCGTGTCTTACATGAATCAGCTGGCGGACACGGTCCAAGCCAAG
CACTACACGCCGAAAATGTGGAATGATAGCCTAACCCACGAAGGTCTGAAAAACCTGAACCACTCTATTAT
TATCATGTATTGGCATCAGCCGTCAAAGCAGTCCCCGAGCCCAACGGATTTTTTCACCAATCATTTTATGGT
CGAGAACTTTAACCGCAGCGTTTACTATGTGTTTCCTCGCGCTCAGCAGAGCACCCATTCGCTGGCGAAGC
AAAAAGCCGACATTGCGGACACTCGTCTTACCGATTTCAACACTGCGAATATGCGTAAAGATCCGCATTTCA
ACTCGTACATTAACGGCGAGTGCCTGACCTTTTGGGGTGAGTTTGCGAGCGATCTGAAGCAGATTAACTTG
ATCGAGTACGTGTATAAATTCATCCGTATTTACTTCAACAGCCTCGAGCACCACCACCACCACCAC

Amino Acid:

HMQSDIEKGISIDIARTHYTKDSIKKIIGELSRVNGRYLQLHLADNDNYSIYSNVLGQTSTHSNHYYLTKAELRELV
QYANKHHVQLIPELDFPAHSKAMLTLLHKHHPSQYRQVVSSYDNTMLDFQQNQAALDVSRQLINEVADIFYQT
PYKDNLKMVIGGDEVPGGGAHQRDFVSYMNQLADTVQAKHYTPKMWNDSLTHEGLKNLNHSIIIMYWHQP

SKQSPSPTDFFTNHFMVENFNRSVYYVFPRAQQSTHSLAKQKADIADTRLTDFNTANMRKDPHFNSYINGECL
TFWGEFASDLKQINLIEYVYKFIRIYFNSLEHHHHHH

Appendix 37. ShHex Sequences

DNA:

CATATGGCTGAAATAGAGAAAGGAGTAACACTAGATATCGCGCGTAAGCACTACAGCGCGGAAAGCATCA
AGAAAATCATCAAGAAGATCAGCGACTACAACGGCGACTACCTTCAGCTGCACTTCTCCGACAACGAGAAC
TACTCTATTTATTCCAAAATTCTGAAGCAGGATTCTGAAAAAAGCAACATGTACTTCCTGACCAAGAAAGAG
CTGAAATCTATAATCCAGTACGCCAATGACCGCGATGTGCAGGTTATTCCGGAACTGGATCTGCCGAGCCA
CAGCAAAGCGATTGTCACCCTGTTACGTAAACATGATTTTAAGCGCTATGAAAAGGTGGTGTCGAGCTATG
ATGAGTCCACCATTGATTTCCACGACAATCGTCATGCAGTTAAATTCTCCCAAGAGATGATTAAGGAGGTGT
CCAGCTTGTTTTATCAGAACAAATACAAAAACAACCAAAAGATTGCTCTGGGTGGTGACGAGGTTCCGGGT
AGCGGCAGCAACCAAGAGAGCTTTATCAAGTACATTAACACCATTTCTAACTACGCGAACAGCCAAAACTA
CGAGACGAAGATCTGGAATGACTCAATTACGAAGAACGGCCTGAAGAAGTTGAATCAAAATATCACGATC
ATGTATTGGAAACAAAAATATAACAAAGATCCGAACCCAAAAGACTTCTTCAAGTTGAACCATAATGTTGA
AAATTTTAACAGCAATGCACTGTATATCTTTCCGAGAAAACAGAATAACGAGGAAATTATCCAGAGCCAGA
ACTTCATCAAACAAACCAAGATGAACGACTTCAACACTAAGTCTCCGAATCAGAAAACCAATTACAATACCC
GTCTGAGCGGCAAATCCCTCTCGTTCTGGGGTGAATTTAGTCTGAAACTGTCGCAAAAAAAGTTGCTGGAA
TATATCTTCATGTTTATGGATACCTTTTTTAAGTCCCTCGAGCACCACCACCACCACCAC

Amino Acid:

HMAEIEKGVTLDIARKHYSAESIKKIIKKISDYNGDYLQLHFSDNENYSIYSKILKQDSEKSNMYFLTKKELKSIIQYA
NDRDVQVIPELDLPSHSKAIVTLLRKHDFKRYEKVVSSYDESTIDFHDNRHAVKFSQEMIKEVSSLFYQNKYKNN
QKIALGGDEVPGSGSNQESFIKYINTISNYANSQNYETKIWNDSITKNGLKKLNQNITIMYWKQKYNKDPNPKD
FFKLNHNVENFNSNALYIFPRKQNNEEIIQSQNFIKQTKMNDFNTKSPNQKTNYNTRLSGKSLSFWGEFSLKLSQ
KKLLEYIFMFMDTFFKSLEHHHHHH

Appendix 38. SaHex Sequences

DNA:

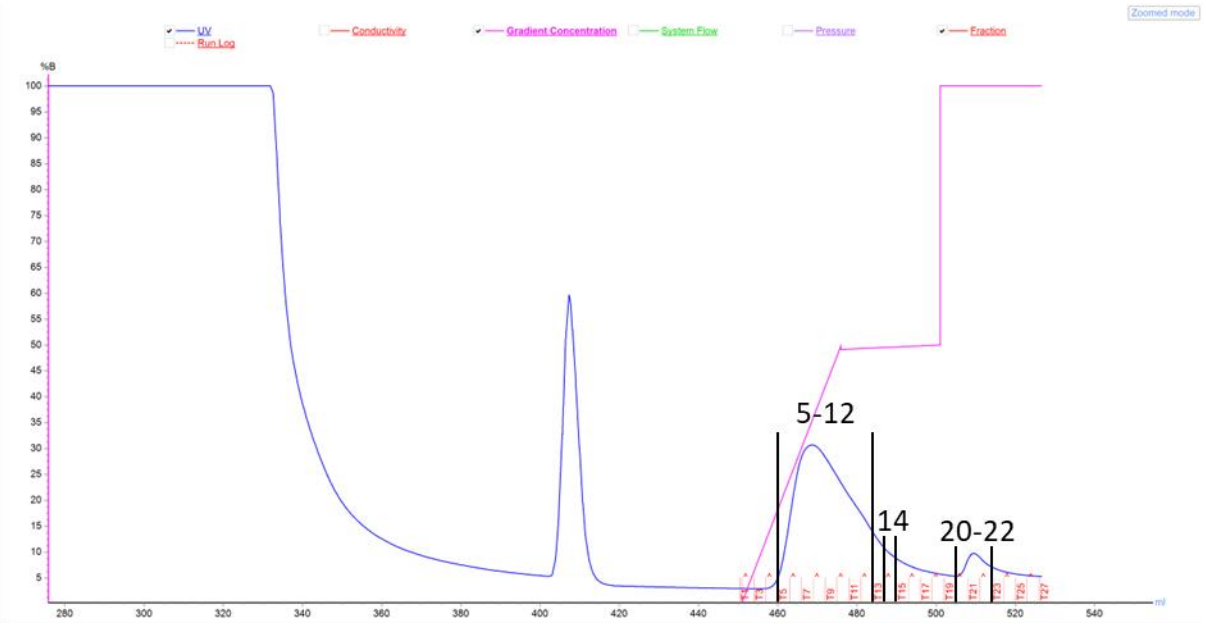
CATATGCAAGATTTTCAGAAAGGAATAAATGTAGACATCGCCCGTAAAGATTACTCCTTGAAGAGCTTGAA
GAAAATCGTTGACACCATCCACGAAAACAACGGCGACTACTTACAACTGCATTTTTCCGACAACGAGAACT
ACGCGATTGAGAGCCAATTTTTCAAGCACGAGAACATCGCGAGCCAAAACTACCTCAGTCAGCAGGAGCT
GAAGAACCTGATCCATTATTCCAACAAACTGAACATTATGGTCGTGCCGGAATTTGACCTGCCTTCCCATAG
CAAGGCGTGGCTGCTGCTCTTGAAGAATGAGAACAGCAACCTGCACGAGAACATTGTCTCGGACTACTCTG
ACGAAACTATTGATTTCTTCTCTAACCAAAAAGCTCTGGAAATCAGCAAACGTCAGATCAAAGAGATCCTGT

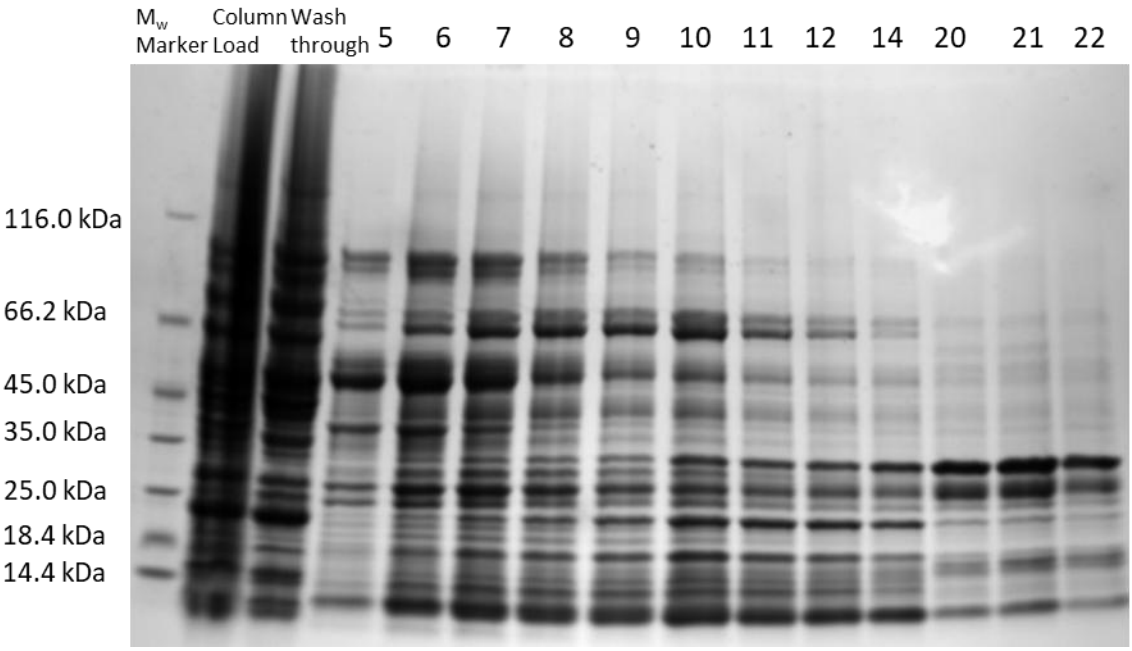
ACCTGTTTCACCAGCCGAATTTCCAGAAAGAACAGCGTATTGTTCTGGGTGGTGACGAGGTGCCAGGTGG
CAAATCATACCAAAACGACTTCATCAACTTCATGAATGAGATCGGCGAATATGCGTATCAGAATGGTTATG
AACCGCAGATTTGGAATGATTCCATTACGAAAAACGGCCTGAAGCTGTTGAAGAACTACTTTAGCGTTATTT
ATTGGAAACAGAGCAATAACGAAAATAATGAGCCGGGCATCACCGTTGAAGATTTCCTGGATTACAACTTT
AAGGTGTATAACTACAATTTTTATAGCTTGTACTTCCTGCCGTCGAAGAACTACTCTCCGACCGATATTGAG
GAACAAACCAGCTATATTAGCTGGGCATATAACCATAATTCGTTCTACTATTTGAAGAATCCGTATTACGAA
GTTGATTCTCTGAACATCCAAGGTAGCGCACTTAGCTTTTGGGGTGAGCACGCTACCGGTATGCGCGAAGA
GGAAGTGCTGAATCAGGAGCTGCCGCTGATCCGCACGTACTTGAATAAACTCGAG

Amino Acid:

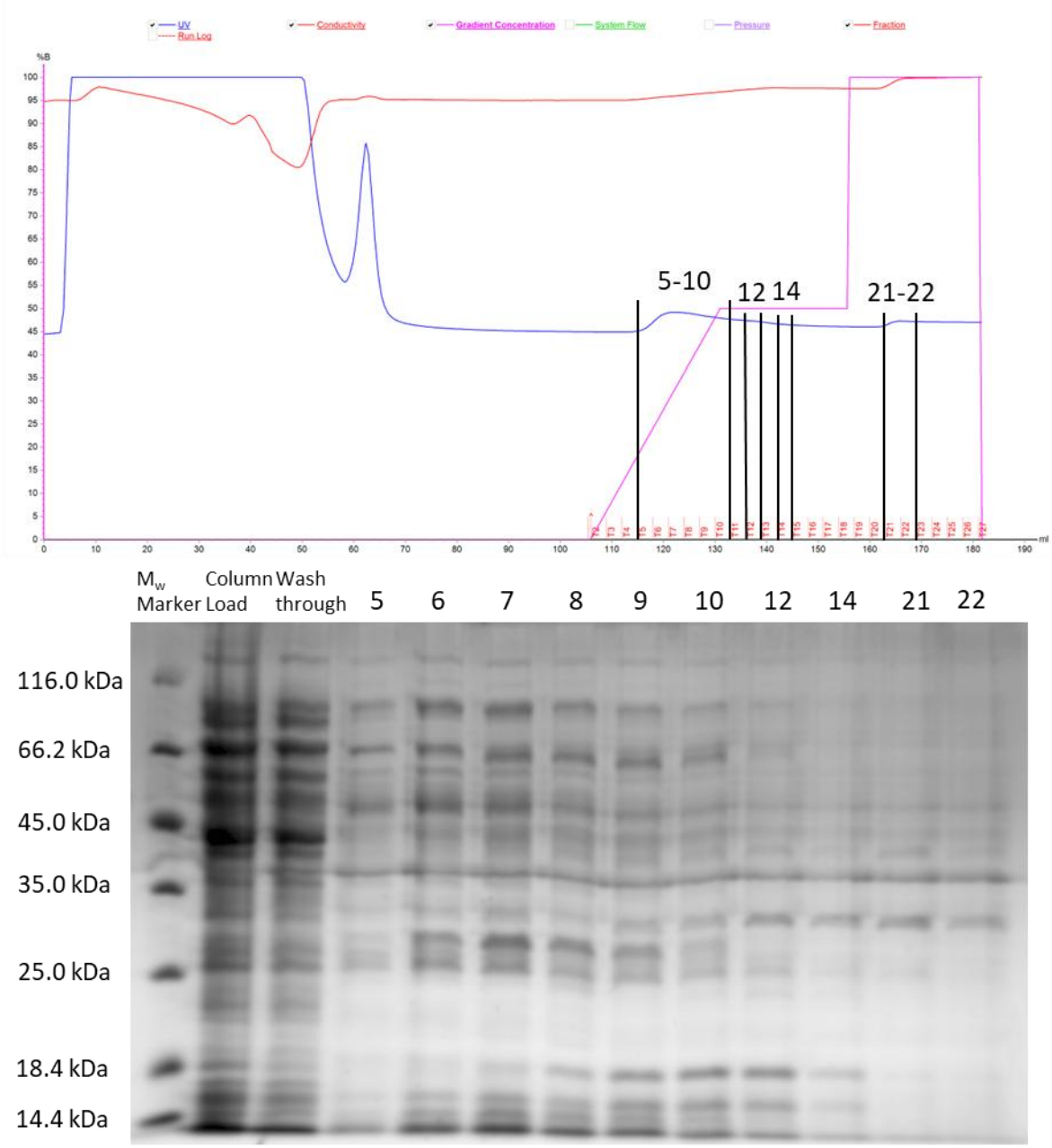
HMQDFQKGINVDIARKDYSLKSLKKIVDTIHENNGDYLQLHFSDNENYAIESQFFKHENIASQNYLSQQELKNLI
HYSNKLNIMVVPEFDLPSHSKAWLLLLKNENSNLHENIVSDYSDETIDFFSNQKALEISKRQIKEILYLFHQPNFQK
EQRIVLGGDEVPGGKSYQNDFINFMNEIGEYAYQNGYEPQIWNDSITKNGLKLLKNYFSVIYWKQSNNENNEP
GITVEDFLDYNFKVYNYNFYSLYFLPSKNYSPTDIEEQTSYISWAYNHNSFYYLKNPYYEVDSLNIQGSALSFWGE
HATGMREEEVLNQELPLIRTYLNKLE

Appendix 39. SlHex 8L Purification Attempt

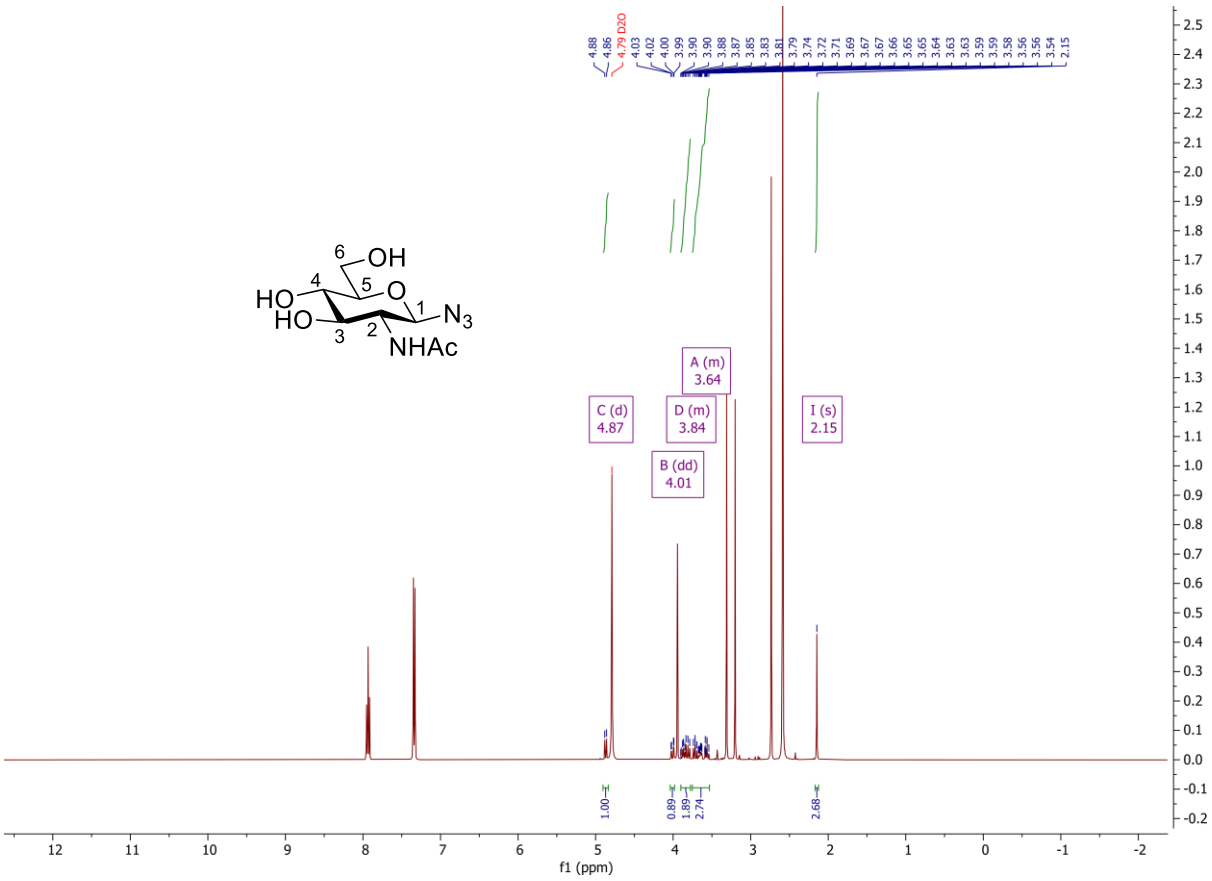


Appendix 40. ShHex Purification Attempt

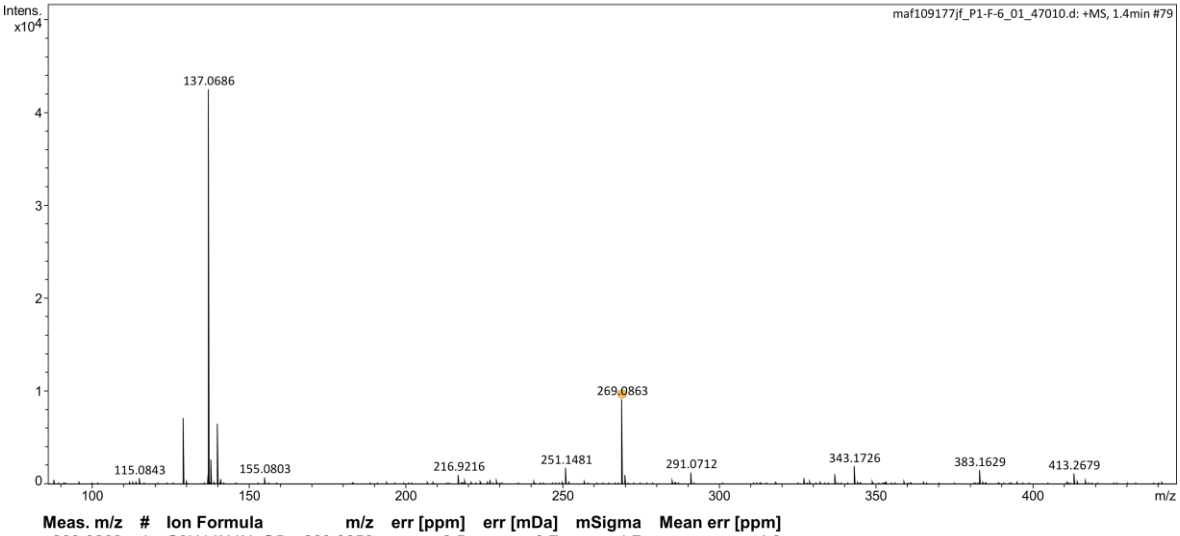


Appendix 41. Compound **64**:




**Analysis Information**

| | | | |
|---|---|---|---|
| Analysis Filename | maf109177jf_P1-F-6_01_47010.d | Acquisition Date | 14/02/2024 08:26:16 |
| Method | ESI_low mass_2c1s.m | Instrument | compact |
| Submission Name | maf109177jf | ESI | Positive |




| Meas. m/z | # | Ion Formula | m/z | err [ppm] | err [mDa] | mSigma | Mean err [ppm] |
|---|---|---|---|---|---|---|---|
| 269.0863 | 1 | C8H14N4NaO5 | 269.0856 | -2.5 | -0.7 | 4.7 | 4.3 |