University of Reading

Integrated workflows for Carbohydrate Analysis - Artificial Glycoconjugates as Tools for Glycan Analysis

Submitted to the University of Reading in partial fulfilment of the requirements for the degree of Doctor of Philosophy

School of Pharmacy

Oliver G A Hancox

January 2023

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged. This work has not been, and is not, being submitted for candidature for any other degree.

OLIVER G A HANCOX

Acknowledgements

I would like to express my deepest gratitude to Dr Sarah Allman and Professor Helen Osborn for their continuous hard work, guidance and support over the course of this process and without whom this journey would not have been possible.

I would also like to thank Dr Nicholas Michael of the University of Reading chemical analysis facility for his skill and guidance in the development of analytical methodologies in LCMS and for HRMS analysis. As well as to Dr Chris Smith of the University of Reading Chemistry department his expertise in 'click' chemistry and collaboration in providing three azido modified aromatic compounds for this project. I am greatly appreciative of all the other students at the University of Reading over the course of this study and especially to Ryan Coones for his enthusiastic discussion and constant supply of caffeine.

I also wish to thank my family for their unending support, encouragement and continuous help. Finally, to Sophie, for her constant support and encouragement throughout my studies, this would not have been possible without you.

Abstract

Glycosylation refers to the post-translational modification of biological molecules to introduce ornate saccharide structures. Glycans have been linked to varying functions within biological systems ranging from cell stabilisation and adhesion, to host-pathogen interactions and cell messaging. Glycosylation can, therefore, have a significant effect upon the activity or viability of a glycosylated biomolecule. The recent growth in the development of protein based biotherapeutics has also resulted in a rapid increase in the development of tools to understand glycan structure and function. Analysis of glycan structures provides significant analytical challenges, attributed to their high hydrophilicity and lack of optically active regions.

To overcome these challenges, derivatisation workflows are commonly employed to modify glycan structure in order to incorporate moieties which facilitate analysis. The work described in this thesis contributes to this area by focusing on the development and application of a series of novel multifunctional labelling strategies that enable modification of the reducing end of glycans using reductive amination. These strategies used derivatives of the commercially available labels 2-aminobenzamide and procainamide. Three multifunctional labels containing additional functionality in the form of a terminal alkyne were prepared. These labels were then compared with commercially available derivatives demonstrating they exhibited comparable Hydrophilic interaction liquid chromatography separation, over similar reaction times, to the commercial derivatives. These workflows were then applied for the multifunctional derivatisation of carbohydrates released from biological sources, and showed comparable separation could be achieved, and sialylation remained intact. This demonstrated that multifunctional labels can be successfully applied to glycan profiling workflows while maintaining further downstream capability. The downstream capability was then investigated though Cu(I) catalysed triazole formation between the multifunctionally labelled carbohydrate and azide conjugation partners. This was achieved though conjugation of labelled heptasaccharide to a library of azides ranging in non-polar surface area aimed at increasing hydrophobicity to impact ionisation efficiency in ESI-Mass spectrometry applications. This demonstrated that the analytical sensitivity for detection of low abundance carbohydrates was possible following reversed phase separation and detection by tandem mass spectrometry. This work therefore paves the way for the structural elucidation of undefined low abundance glycan species. Moreover the use of picomlar scale Cu(I) catalysed triazole formation on biologically derived carbohydrate provides a route towards the generation of neo-glycoconjugates enabling the further illumination of carbohydrate interactions on biological systems.

Abbreviations

2 -AA	2-Anthranilic acid
2-AB	2-Aminobenzamide
ABBE	Butyl-4-amino-benzoate
ABEE	Ethyl-4-amino-benzoate
ABME	Methyl-4-aminobenzoate
ACN	Acetonitrile
ANTS	8-aminonapthalene-1,3,6-trisulfonic acid
APS	Ammonium persulfate
APTS	8-aminopyrene-1,3,6-trisulfonic acid
Asn	Asparagine
ATR-IR	Attenuated total reflectance infrared spectroscopy
BACH	Biotinamidohexanoic acid hydrazide
BHK	Baby hamster kidney
BNAH	Biotinyl-L-3-(2-naphthyl)-alanine hydrazide
Boc	tert-Butyloxycarbonyl
BPC	Base peak chromatogram
CE	Capillary electrophoresis
Cer	Ceramide
CHCA	α-Cyano-4-hydroxycinnamic acid
CHO	Chinese hamster ovary
CID	collision induced dissociation
Cl-CCA	4-Chloro-α-cyanocinnamic acid
CS	Chondroitin sulfate
CuAAc	Copper catalysed azide-alkyne cycloaddition
CV	Column volume
Da	Dalton
DCC	N,N'-Dicyclocarbodiimide
DCM	Dichloromethane
DCU	Dicyclohexylurea
DHB	Dihydroxybenzoic acid
DI-MS	Direct injection mass spectrometry
DMF	Dimethylformamide

DMSO	Dimethyl sulfoxide
Dol-P-P	Dolichol-diphosphate
DS	Dermatan sulfate
ECM	Extra cellular matrix
EIC	Extracted ion chromatogram
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
ESI-MS	Electrospray ionisation mass spectrometry
FA	Formic acid
FLD	Fluorescence detection
Fuc	Fucose
FUT8	α1-6 fucosyltransferase
GAGs	Glycosaminoglycans
Gal	Galactose
GalNAc	N-Acetylgalactosamine
GHP	Glucose homopolymer
Glc	Glucose
GlcNAc	N-Acetylglucosamine
Gly	Glycine
GnT	Glucose N-acetyltransferase
GPI	Glycosylphosphatidylinositol
GU	Glucose units
HA	Hyaluronic acid
Hep	Heparin
HILIC	Hydrophilic interaction liquid chromatography
HPAEC	High peformance anion exchange chromatography
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HS	Heparan sulfate
ID	Internal diameter
IEEDA	Inverse electron-demand Diels-Alder
IgG	Immunoglobin G
KS	Keratan sulfate
LC	Liquid chromatography

LCMS	Liquid chromatography mass sprectometry
LOD	Limit of detection
LOQ	Limit of quantitation
Lys	Lysine
M/Z	Mass/charge ratio
mAbs	Monoclonal antibody
MALDI	Matrix assisted laser desorption ionisation
MALDI-ToF	Matrix assisted laser desorption ionisation - Time of Flight
Man	Mannose
MMC	Mouse myeloma cells
MS	Mass spectrometry
MWCO	Molecular weight cut off
Neu5Ac	N-Acetylneuraminic acid
Neu5Gc	N-Glycolylneuraminic acid
NHS	N-Hydroxysuccinimide
NMP	1-(2-Naphthyl)-3-methyl-5-pyrazolone
NMR	Nuclear magnetic resonance
NP	Normal phase
OST	oligosaccharyl transferase
PG	Proteoglycans
PGC	Porous graphitized carbon
PMP	1-Phenyl-3-methyl-5-pyrazolone
PNGase	Peptide N-glycosidase
PPM	Parts per million
ProA	Procainamide
QTOF	Quadrupole Time of Flight
RF-MS	RapiFluor-MS
RNase B	Bovine pancreatic ribonuclease B
RP	Reversed phase
Rt	Room temperature
RT	Retention time
sDHB	Super-DHB
SDS	Sodium dodecyl sulphate
Ser	Serine

SPAAC	Strain promoted azide-alkyne cycloaddition
SPE	Solid phase extraction
TEA	Triethylamine
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
THAP	2,4,6-Trihydroxyacetophenone
THF	Tetrahydrofuran
Thr	Thionine
TIC	Total ion chromatogram
TLC	Thin layer chromatography
TOF	Time of flight
UDP	Uridine diphosphate
UDP-GalNAc	Uridine diphosphate N-actylgalactosesamine
UDP-GlcNAc	Uridine diphosphate N-actylglucosamine
UDP-Neu5Ac	Uridine diphosphate N-acetylneuraminic acid
UV	Ultraviolet
UV-Vis	Ultraviolet visible
WAX	Weak anion exchange
Xyl	Xylose

Table of Contents

Table of Figures

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

Table of Tables

Table of Schemes

CHAPTER 1

CHAPTER 2

CHAPTER 3

Chapter 1- Introduction

This thesis broadly aims to develop glycan derivatisation tools to facilitate the analysis of glycans at low abundance though the use of multifunctional labelling techniques. Hence this introduction will focus on the structure and analysis of glycans within existing glycan analysis workflows.

1.1 Glycan structure

Glycans are oligosaccharide structures covalently attached to biomolecules such as proteins, lipids and other small molecules in a process called glycosylation. In the case of protein and peptide glycosylation, this occurs as a post translational modification in which carbohydrates are linked to nascent protein molecules during translation [1-3]. Glycosylation has been shown to aid a number of significant biological processes, ranging from improved protein stability and cell signalling to cell- cell interaction events [2, 4]. Glycans are formed as a polymerised structure of monosaccharide subunits synthesized in a non-template driven sequence (**Figure 1.1**). This differs from the biological synthesis routes of the genome and proteome, in that the synthesis does not follow a template [5]. Instead, the process relies upon a cascade of glycosyltransferase and glycosidase enzymes responsible for the attachment, trimming and elongation of glycan structures [5]. Although lacking a formal template, the synthesis route is tightly controlled at multiple levels in the endoplasmic reticulum (ER) and Golgi apparatus [1, 6].

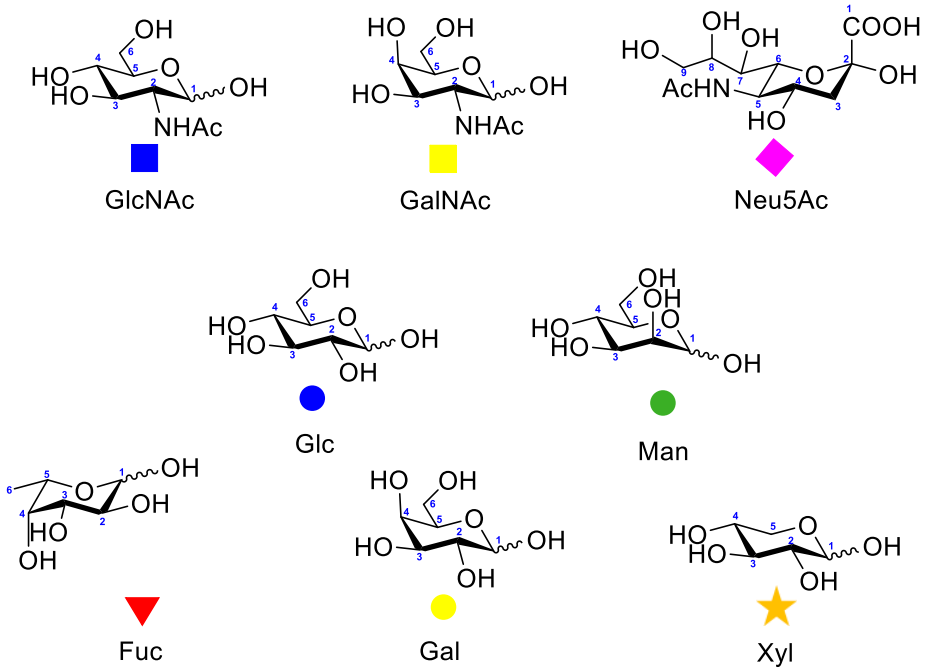


Figure 1.1 - Monosaccharide subunits commonly found in eukaryotic glycan structures accompanied by their respective graphical symbols according to SNFG [7].

Despite multiple levels of control associated with glycan synthesis, the population of glycans produced display high levels of heterogenicity and can differ in structure, length, and linkage. Moreover, carbohydrates frequently exist as positional or structural isomers of each other, as in the case of the varying glycosylation patterns found on immunoglobulin G (IgG)[8, 9].

The complex and heterogeneous structural variations present between oligosaccharide structures pose a significant challenge for traditional synthetic routes. Large reaction volumes are often required to yield relatively small quantities of complete glycan [10]. This is largely due to the requirement for specific hydroxyl modification, in the presence of many others, and is further complicated by challenges in forming specific isomers during glycosidic bond formation [11]. These challenges have resulted in hybrid chemo-enzymatic and enzymatic routes, based upon biologically derived enzymes to form oligosaccharides under *in vitro* conditions. The alternative to synthesis is the enzymatic cleavage of glycans from biological sources, which by the action of PNGase F sees the cleavage of intact *N*-linked glycans from the protein backbone. These released glycans can then be extracted from the deglycosylated protein fragments by solid phase extraction (SPE).

The position of attachment of glycans to proteins can be subdivided into two main categories. *N*-Linked glycans are attached to asparagine residues (Asn) via an amide bond to an *N*-acetylglucosamine (GlcNAc) sugar. *O*-Linked glycans are affixed to proteins by the oxygen of either serine (Ser) or threonine (Thr) residues [12].

1.1.1 *N*-Glycosylation

Although often referred to as a posttranslational modification, the addition of glycans to protein molecules takes place earlier in protein synthesis, occurring during the translation process and is more accurately described as a co-translational modification [13]. *N*-Linked glycans are found extensively throughout eukaryotic life. They are identified by their covalent linkage to the protein residue via the nitrogen found on asparagine residues within the protein's primary structure [14]. In many cases, the presence of an *N*-linked glycan can be predicted based upon the amino acid sequence surrounding a particular asparagine residue [15, 16]. This sequence presents as Asn-X-Ser/Thr with X pertaining to any amino acid except for proline as shown in **Figure 1.2**. Although glycans are almost exclusively found at these sites, not all sites in the finished protein will bear glycan modification. Glycosylation will not occur at sequences containing proline, due to the presence of a ring constrained amine. This forms a region where the peptide chain orientation is not conducive to complementary enzyme active site fitment this is due to the cyclic structure of proline.

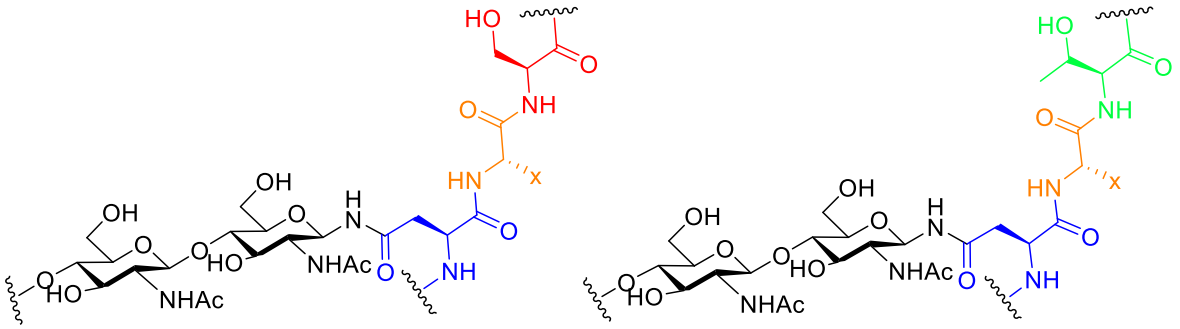


Figure 1.2 - *N*-Linked GlcNAc structures attached to Asn residues in a Asn-X-Ser (Left) or Asn-X Thr (Right) where X is any amino acid with the exception of Proline.

The biosynthetic pathway for *N*-linked glycosylation and attachment is a complex but well researched area of glycomic analysis [1]. The synthesis of *N*-glycans begins with the formation of a lipid linked carrier shown in **Figure 1.3**. The formation of dolichol phosphate begins on the cytoplasmic side of the endoplasmic reticulum, where a single phosphate molecule is affixed to dolichol. A GlcNAc-1- phosphotransferase enzyme then sees the conversion and transfer of uridinediphosphate (UDP) UDP-GlcNAc to Dolichol-diphosphate (Dol-P-P) Dol-P-P-GlcNAc via the intermediate GlcNAc-1-P. The dolichol linked precursor is then elongated to form a $GlcNAc_2Man_5$ oligosaccharide. This glycolipid is then flipped into the lumen of the Endoplasmic reticulum (ER) by an enzyme belonging to the family known as flippases [17]. Following its entry to the ER, the glycolipid carrier is further elongated by a family of glycosyltransferase enzymes. This results in the formation of a mature lipid-

linked precursor consisting of $GlcNAc_2Man_9Glc_3$. The completed saccharyl intermediate is then transferred from the dolichol lipid carrier to nascent protein molecules by the action of oligosaccharyl transferase (OST). OST exhibits a high degree of specificity towards the completed oligosaccharide and therefore when incomplete oligosaccharides are generated, their transfer to proteins is reduced. This manifests as hypoglycosylation and the formation of mature proteins with empty *N*-glycosylation sites [18]. During this process, OST cleaves the GlcNAC-P bond and in so doing, releases Dol-P-P before the resulting GlcNAc containing oligosaccharide is attached to the asparagine residue of the glycoprotein.

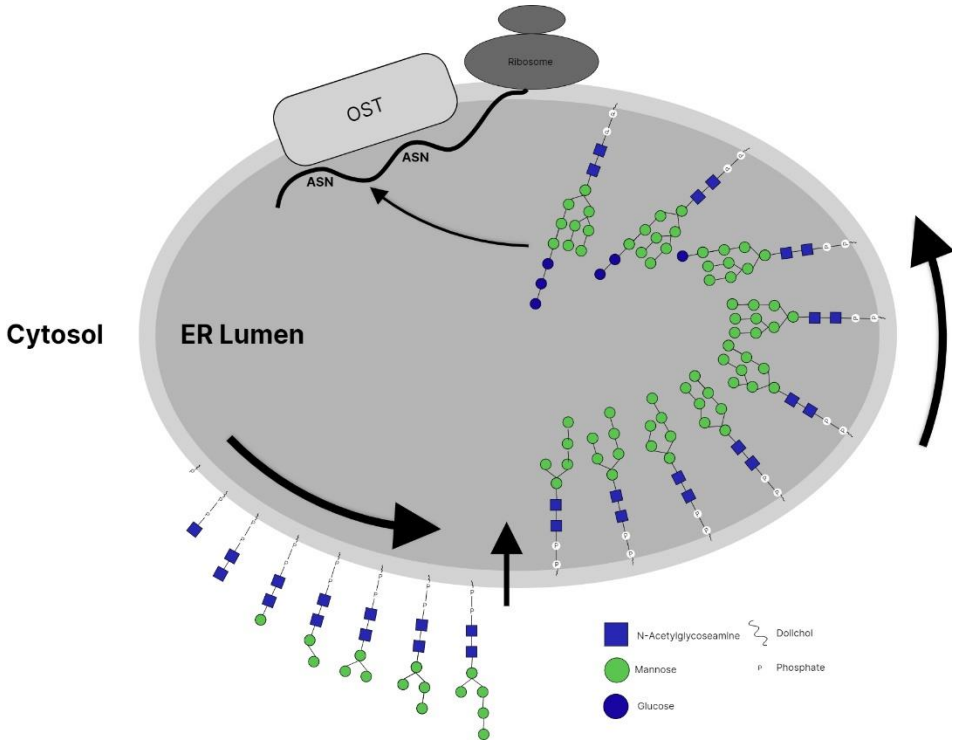


Figure 1.3 - *N*-Linked glycosyation takes place at both the cytosolic and luminal faces of the Endoplasmic reticulum (ER). This begins with the synthesis of a dolichol linked oligo saccharide (DLO) on the cytosolic face of the ER initiated by a cascade of membrane anchored glycotransferases. This culminates in the generation of Man5-GlcNAc2-PP-dolichol structure which is flipped to the luminal face to undergo further modification resulting in a final DLO Glc_3-Man_9-$GlcNAc_2$-PP-dolichol. The final DLO is then transferred to As residues by oligosaccharyl transferase OST complexes (Adapted from figure obtained under creative commons license CC-BY) [19].

The protein bound *N*-linked glycan then undergoes a process of sequential trimming. During this stage, α-glucosidase I and II remove three Glc residues at the terminal end of the glycan. It is also common for a single mannose unit to be cleaved during this stage, resulting in a size reduction and the formation of Asn-$GlcNAc_2Man_{8\text{-}9}$ [20].

The glycan containing protein is then transferred to the Golgi with most oligosaccharides leaving the ER in an Asn-$GlcNAc_2Man_8$ configuration. Upon entering the Golgi, further monosaccharide subunits may be added by glycosyltransferase enzymes, or removed by the action of α-mannosidase I and II. Removal of α1-2 Man residues in the cis region of the Golgi yields glycans of Asn-$GlcNAc_2Man_5$, a key intermediate for further glycan processing and the generation of completed glycan structures as shown in **Figure 1.4**.

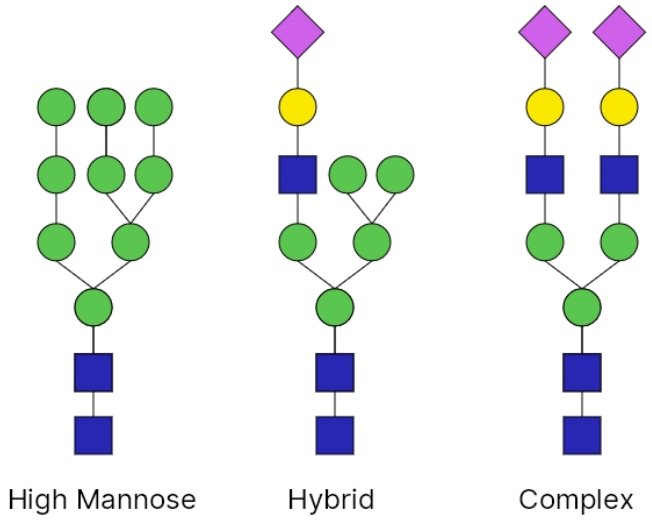


Figure 1.4 - Examples of three *N*-linked glycan structures present in eukaryotic organisms. High mannose (Right) contain unsubstituted terminal mannose subunits. Hybrid (Centre) contain both unsubstituted terminal mannose as well as substituted subunits of N-actylglucoseamine. Complex (Left) exhibit N-actylglucoseamine at both α-3 and α-6 mannose sites resulting in bi-, tri-, or tetra antennary structures.

During this process, some Man_{5-9} glycans escape further processing and go on to form a variety of high mannose glycans on the final mature protein. Others remain and undergo further processing [21]. At this point, transfer of a GlcNAc to the α-1,3 arm of $GlcNAc_2Man_5$ by the action of glycosyltransferase Glucose *N*-acetyltransferase 1 (GnT-1), can trigger one of two actions for the finished glycan. The glycan can be further modified with a Galactose (Gal), triggering the addition of a sialic acid residue, such as Neu5Ac **2**. This forms a hybrid type glycan containing one or more antennae [22]. Alternatively, the $GlcNAc_3Man_5$ oligosaccharide can be further reduced in size with the removal of two mannose residues from the α1,6 arm of the glycan structure, followed by the addition of a GlcNAc residue to the same region by the action of GnT-2 enzyme. This intermediate oligosaccharide can then be decorated further with GlcNAc and Gal residues, leading to the addition of sialic acid. These structures can be further manipulated by the addition of GlcNAc residues at both α1,3 and 1,6 arms of the glycan. The resulting glycoform forms a complex glycan with tri- or tetra-antennary structures. Following completion of modifications to the main structure of both complex

and hybrid glycans, a process called capping can see the addition of a variety of monosaccharide subunits to the Asn-linked GlcNAc. In most vertebrate *N*-glycans, this modification sees the addition of an α1,6 fucose by α1-6 fucosyltransferase (FUT8) forming fucosylated glycan species. Invertebrates exhibit heterogeneity in linkage exhibiting both α1,3 and 1,6 linkages. While in plants, it is common for xylose or mannose additions to be present.

The sequential biosynthesis of glycans described above requires that glycosyltransferase and glycosidase concentrations differ depending on their requirement within different regions of the Golgi. This is achieved in the Golgi by a process of cisternal maturation [23, 24]. Rather than glycan migration within the Golgi, glycans/glycoproteins remain in a single cisternae, because they are constructed moving from cis to trans Golgi. During this period, glycosyltransferase and glycosidase are transported back to earlier cisternae of the Golgi via a retrograde transport system [25]. This results in a higher concentration of sialyl transferase and galactotransferase enzymes in the trans-golgi, where a higher proportion of mature glycans are found. Meanwhile, the cis-golgi contains higher concentrations of mannosidase and glycose *N*-actyltransferase so that less structurally developed glycans encounter these enzymes first, thus avoiding random and incorrect glycan synthesis.

The biosynthesis of *N*-linked glycan structures results in a heterogeneous population of structurally diverse oligosaccharides. The competitive nature of enzyme action within the Golgi leads to carbohydrate elongation and capping. This results in the transformation of a small pool of potential glycan structures into an extensive repertoire of mature complex and hybrid glycans. These structures share a $GlcNAc_2Man_3$ core, but display variations in monosaccharide composition and chain length, as well as GlcNAc branching and capping.

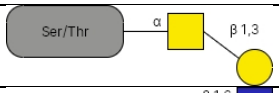
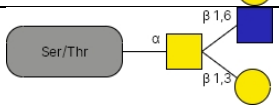
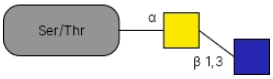
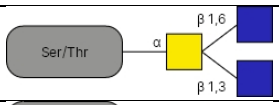
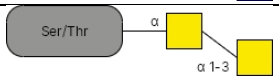
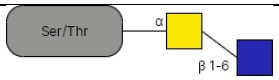
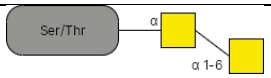
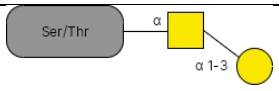
1.1.2 *O*-Linked glycosylation

Glycan attachment is also achieved at residues other than asparagine by the action of *O*-linked glycosylation. In these cases, a glycosidic bond is formed between a monosaccharide and the hydroxyl group of either Ser or Thr (**Figure 1.5**). Unlike *N*-linked glycosylation, research has been unable to isolate a consensus sequence that leads to *O*-linked glycosylation sites, however computational prediction is possible through statistical machine learning [16, 26]. *O*-Linked glycosylation, although harder to predict, has been much more widely observed than *N*-linked glycosylation. These structures are often simpler and have been observed in both prokaryotic and eukaryotic life [27, 28].

Figure 1.5 - *O*-linked glycosylation to Serine residues (left) and Threonine residues (Right)

By far the most common *O*-linked addition is that of *N*-acetylgalactosamine (GalNAc) [29]. These structures are categorised as mucin type glycans. The glycosylation of *O*-GalNAc glycans differs from *N*-linked glycans in that a lipid linked carrier is not required. Instead, the co-translational process begins in the ER/Cis-Golgi. GalNAc is transferred to a peptide donor from UDP-GalNAc by *N*-acetylgalactosaminetransferase, a family of enzymes found primarily in the trans-Golgi region. At this stage, the monosaccharide structure is added to the protein, forming a core for the addition of subsequent carbohydrate subunits. The addition of *O*-linked oligosaccharides can give rise to the formation of eight possible core structures outlined in **Table 1.1**. Core structures one through to four represent the most commonly occurring *O*-linked modifications. Structures five through to eight are less widely observed [30]. Mucin type glycosylation frequently contains highly heterogeneous populations of these core structures, with many of the structures displaying antigenic properties.

Table 1.1 - *O*-linked glycosylation core structures with Cores 1-4 being the most widely observed within eukaryotic life.

O-Glycan Core	Structure	
Core 1		Galβ1-3GalNAc
Core 2		GlcNAcβ1-6(Galβ1-3)GalNAc
Core 3		GlcNAcβ1-3GalNAc
Core 4		GlcNAcβ1-6(GlcNAcβ1-3)GalNAc
Core 5		GalNAcα1-3GalNAc
Core 6		GlcNAcβ1-6GalNAc
Core 7		GalNAcα1-6GalNAc
Core 8		Galα1-3GalNAc

The effect of *O*-linked glycosylation has been shown to significantly influence the structure and activity of proteins bearing their modification [31]. Mucins represent a branch of glycoproteins that exhibit a high degree of sialyation along much of the protein's surface. The presence of this addition has been shown to contribute to a net negative charge over the protein's surface, affecting the overall conformation of the protein. This has the effect of increasing protein rigidity, while increasing the hydrophilic properties of the protein. These factors prevent globular protein formation and lead to the proteins forming a stiff linear structure. The addition of the *O*-GalNAc glycan in mucins is also essential for mucin function. The mucin serves to both hydrate and protect the epithelial lining of both the gastrointestinal tract and respiratory tract by trapping water, forming a hydrated lubricant [32]. This also acts to form a protective layer against proteolysis on exposed epithelial cell walls, as well as a first line of immune response, due to trapping bacteria before they enter more exposed regions of the respiratory tract [33, 34].

1.1.3 Glycosaminoglycans

Glycosaminoglycans (GAGs) represent a family of glycoforms knows as mucopolysaccharides. GAGs are formed of repeating disaccharide subunits carrying a negative charge [35]. Initially their role was thought to be limited to cell-hydration and scaffolding, however significant evidence suggests that GAGs play an important role in cell signalling for the regulation of cell proliferation and growth, as well as cell adhesion[36].

GAGs display a vast array of structural variation and therefore are a result of differing synthesis routes. However, they can be separated into four main categories based upon their disaccharide core; these include heparin/heparan sulfate, chondroitin sulfate/dermatan sulfate, keratan sulfate, and hyaluronic acid [37, 38].

Although their synthesis is application dependant, heparin (Hep)/heparan sulfate (HS) [37], chondroitin sulfate (CS)/dermatan sulfate (DS) and keratan sulfate(KS) GAGs, all exhibit similar synthetic routes beginning with the generation of a uridine diphosphate activated carbohydrate subunit. Depending on the GAG being synthesised, this could be one of five UDP activated subunits **Figure 1.6**.

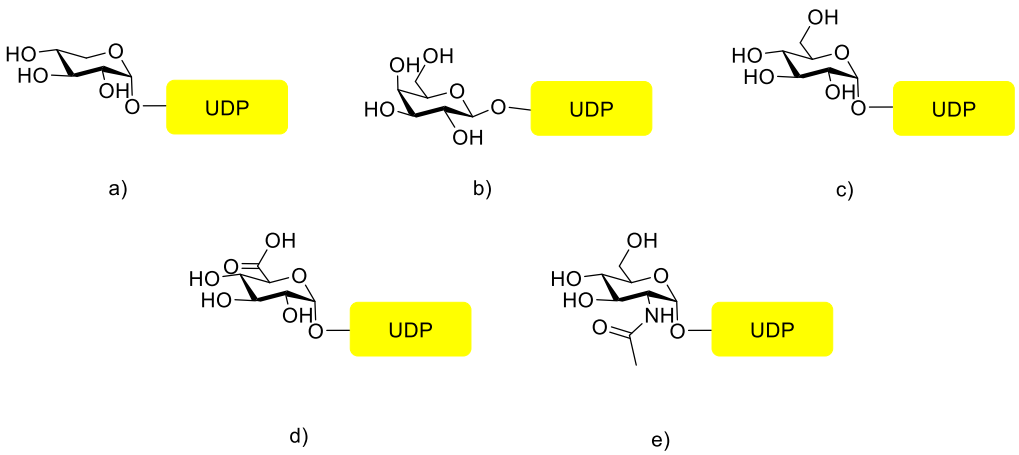


Figure 1.6 - UDP activated carbohydrate subunits used in the synthesis of glycosaminoglycan structures a) UDP-Xylose, b) UDP-Galactose, c) UDP-Glucose, d) UDP-Glucuronic acid, e) UDP-GlcNAc.

All GAG structures, with the exception of hyaluronic acid (HA), require additional modification in the Golgi. Primarily, this modification is the sulfation of hydroxyl groups and is carried out by a sulfate donor, typically in the form of 3-phosphoadenosine-5-phosphosulfate. The availability of this compound significantly affects the rate of synthesis of sulfated GAGs, and serves to control and regulate their production.

1.1.4 Hyaluronic acid GAGs

The biosynthesis of GAG structures is then continued within the Golgi with the exception of HA glycans, which are recruited directly by the cell membrane without undergoing further modification or sulfation. HA represents the simplest structure of all the GAGs produced by living systems and consists of sequentially bound repeating units of glucuronic acid and *N*-acetylglucosamine sub-units [39]. These subunits are recruited from the cytoplasm by the plasma membrane and HA synthesis is carried out on the inner face of the cell membrane by family of enzymes, known as hyaluronan synthase isoenzyme (HAS1, HAS2, HAS3), which recruit UDP-*N*-acetylneuraminic acid (UDP-Neu5Ac) and UDP-glucuronic acid (UDP-GlcUA) [38]. While many glycoforms are smaller in size, HA has been shown to exhibit masses ranging from 3000 kDa in umbilical HA, to up to 7000 kDa in joint synovial fluids [40]. These large structures are stiffened by a network of hydrogen bonds, which sees the production of a polar and non-polar face, causing a twisting helical configuration that can trap up to 1000x the molecule's weight in water. The completed biomolecule serves to act as a form of fluid retention and lubrication. Predominantly found in the extra cellular matrix (ECM), HA also serves to

increase the elastoviscosity in connective tissue, as well as joint synovial fluid. In this function, HA serves to bind water and provide lubrication, while limiting evaporation from surfaces such as the eye, where HA provides both lubrication and forms a key component of vitreous fluid [41].

1.1.5 Proteoglycans

Following modification, both heparin/heparan sulfate and chondroitin/dermatan sulfate groups of GAG are covalently anchored to protein molecules, known as proteoglycans (PG). This process sees the linkage of modified GAG glycoforms to serine residues present on the protein. Glycoforms are attached via a common tetrasaccharide linker, formed of glucuronic acid– galactose–galactose–xylose, previously affixed to Ser-Gly sites along the proteins primary structure [42].

These GAG molecules are made up of repeating disaccharide units of *N*-acetylglucosamine and hexuronic acid residues in HS and Hep GAG structures while CS and DS structures exhibit *N*-acetylgalactosamine, followed by either iduronic acid in DS and glucuronic acid in CS [43, 44]. While HA GAG structures can be upwards of 10,000 disaccharide units long, these GAG structures have been isolated at lengths ranging from 10-200 disaccharide units. However, this is still significantly larger than the smaller structures exhibited by both *N*- and *O*-linked glycosylation [45]. The structure and degree of sulfation exhibited by these molecules has a significant impact upon the activity and function of the biomolecules. While some proteoglycans may exhibit a single GAG, others may exhibit hundreds, with each individual GAG exhibiting structural variations due to the non templated nature of their synthesis [46, 47].

1.1.6 GPI anchors

Glycosylphosphatidylinositol-anchored proteins (GPI-AP) are proteins modified with a glycolipid **Figure 1.7**. These glycoconjugates are widely observed throughout eukaryotic life, with at least 150 human proteins identified as possessing glycolipid linked modification. Proteins displaying a GPI anchor are found attached to the plasma membrane by the lipid, with the GPI moiety covalently bound to the protein's C-terminus [48]. GPI anchored proteins carry out a range of roles, serving as receptors, enzymes, transcytosis receptors, transporters and protease inhibitors.

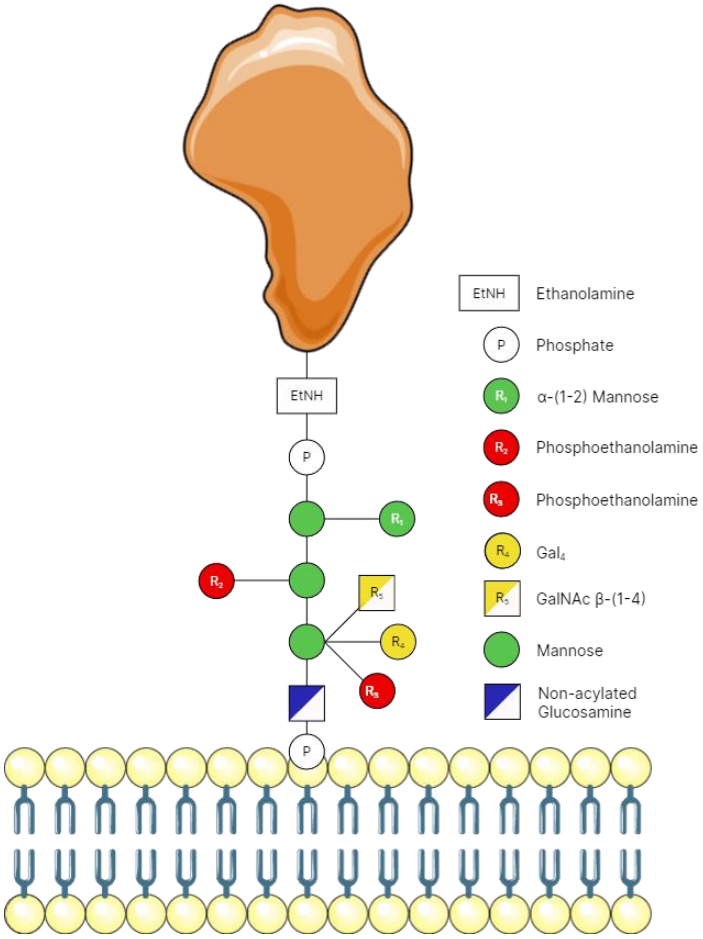


Figure 1.7 - Structure of a glycosylphosphatidylinositol-anchored protein (GPI-AP) showing the glucosamine/mannose core with multiple sites for both carbohydrate and fatty acid modification. (Parts of the figure were drawn by using pictures from Servier Medical Art. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

GPI anchors are assembled in the ER by a series of enzyme mediated reactions before being affixed to a protein molecule. This synthesis route can be divided into three sections beginning with the generation of a GPI precursor on the endoplasmic membrane. This is followed by its covalent attachment to nascent protein molecules via a multi-subunit transamidase complex, forming an ethanolamine phosphate bridge between carbohydrate and amino acid. Subsequent modification of both the lipid and carbohydrate side chains follows. Although mature GPI anchors exhibit high levels of structural heterogeneity, all share a common core of Manα1-2Manα1-6Manα1-4GlcNα1-6*myo*-inositol-1-*P*-lipid [49].

Mutations in the pathway responsible for the generation of GPI-anchored protein molecules have been linked to a group of conditions referred to as inherited GPI deficiency disorders (IGDs). Due to the wide range of functions that GPI anchored proteins serve, symptoms are varied between incidences. However, phenotypes including cognitive impairment, epilepsy and cardiac abnormalities have been observed [50].

1.1.7 Glycolipids

Glycolipids represent a structurally heterogeneous family of glycoconjugates, encompassing a wide variety of oligosaccharide structures covalently linked to hydrophobic lipid structures. Lipid linked carbohydrates exhibit wide variation in both lipid and carbohydrate structure. Glycolipids have been isolated in almost every living organism and can be divided into two main groups **Figure 1.8**. In cases where mono-, di- or trisaccharide carbohydrate structures are linked to diglyceride lipids, the resultant structure is referred to as a glycoglycerolipid [51]. Alternatively in cases were saccharyl structures are bound to ceramide lipids, the resultant molecule known as a glycosphingolipid.

The glycoglycerolipid structures shown in **Figure 1.9** are widely observed in the chloroplasts of photosynthesising organisms, ranging from plants to algae and bacteria. Both monogalactosyldiacylglycerol and digalactosyldiacylglycerol derivatives form over 60% of the thylakoid membrane lipids found in photosynthesizing organisms[52].

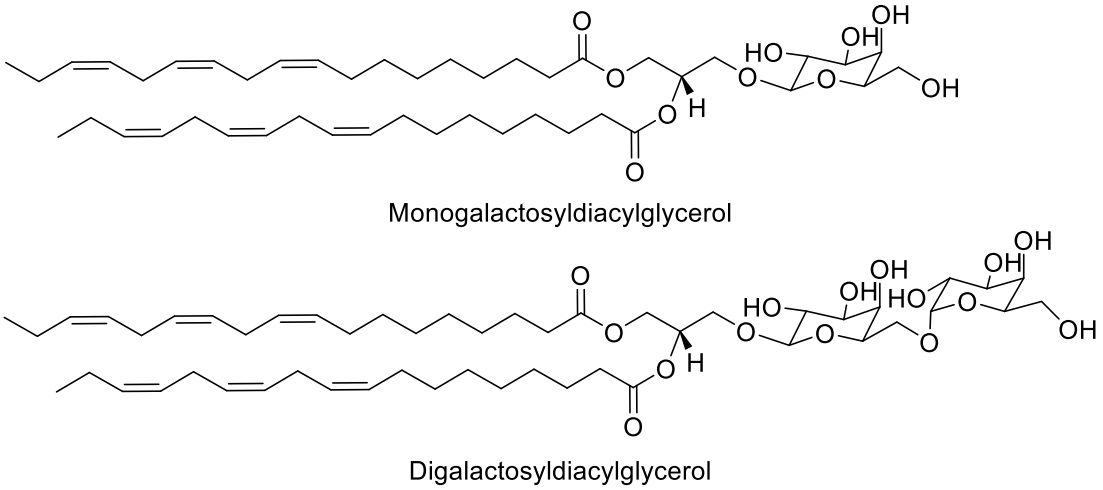


Figure 1.8 - Structures of both Monogalactosyldiacylglycerol (Top) and Digalactosyldiacylglycerol (Bottom)

Glycosphingolipids are more widely observed in higher eukaryotic life and can be further subdivided into neutral glycosphingolipids and acidic glycosphingolipids, including sulfo-sphingolipids and gangliosides [53]. Due to the large range of structural conformations exhibited by glycosphingolipid structures, multiple functions are displayed from receptors in the control cell growth and protein trafficking to cell differentiation [54, 55].

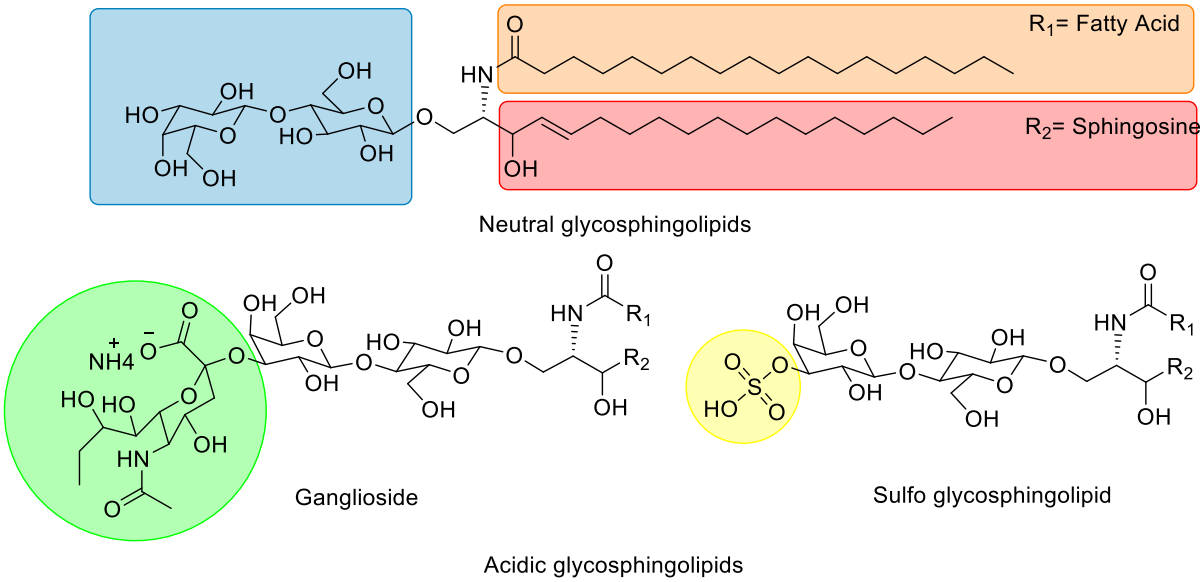


Figure 1.9 - Structure of a neutral Glycosphingolipid (Top) showing carbohydrate regions in blue as well as R1 fatty acid and R2 sphingosine structures that make up the lipid region of the molecule. Acidic glycosphingolipids take the form of sialylated gangliosides (left) and sulfo glycosphingolipids (Right).

Their synthesis begins on the cytoplasmic face of the ER, with the synthesis of ceramide before its transport onto the luminal face of the ER [56]. At this point, galactose-ceramide (GalCer) glycolipids are formed by the attachment of galactose to the newly formed ceramide lipid structure. This differs from glucose-ceramide (GlcCer), which is formed on the cytosolic face of the ER, before being flipped into the golgi following carbohydrate attachment [52]. Both GlcCer and GalCer then move into the Golgi for elongation by glycotransferases and sialylation by siaylalyltransferase. In the case of GalCer, sulfated groups may be added, resulting in the formation of sulfophinglolipids. The formation of GlcCer glycolipids can be seen in **Figure 1.10**.

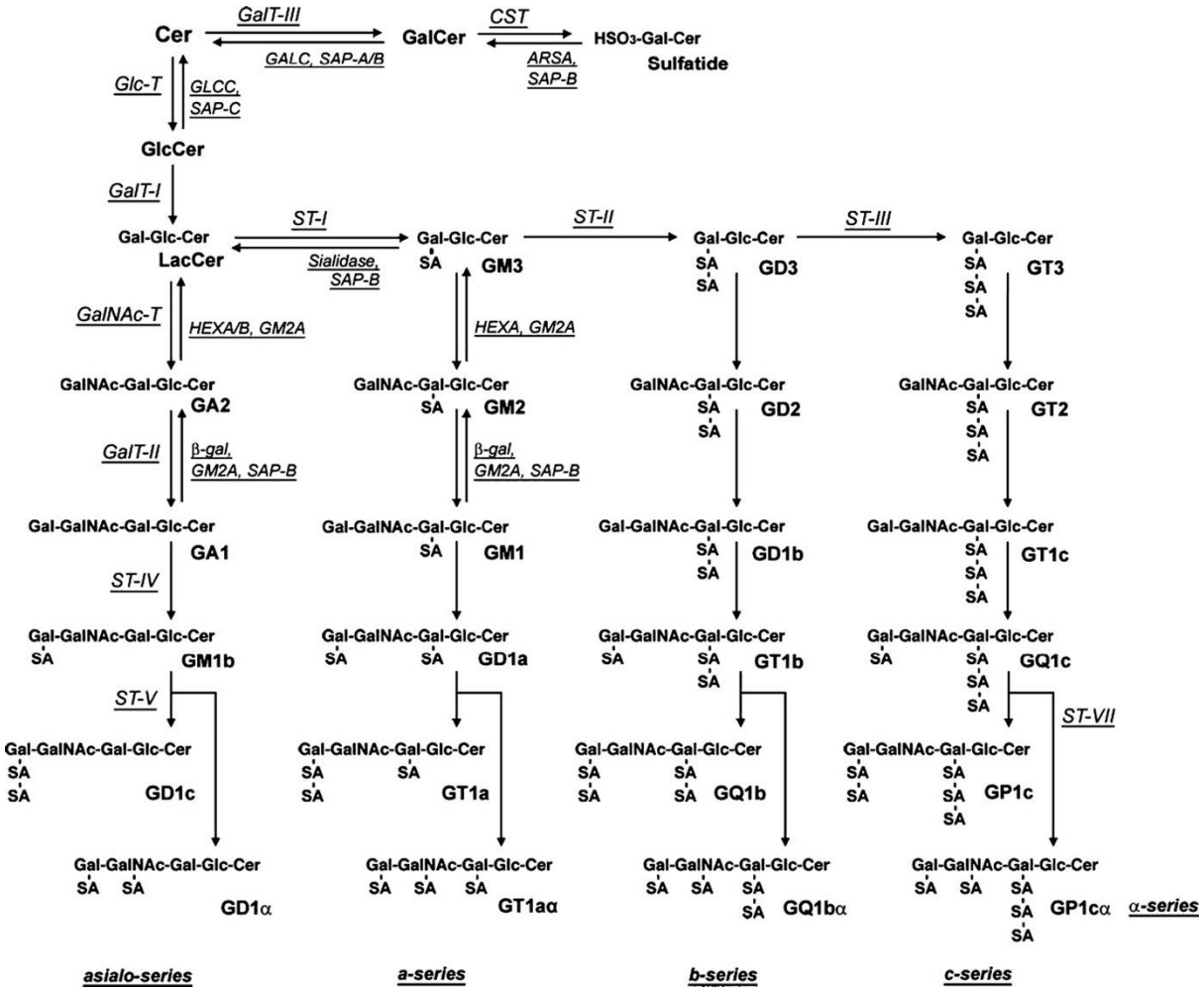


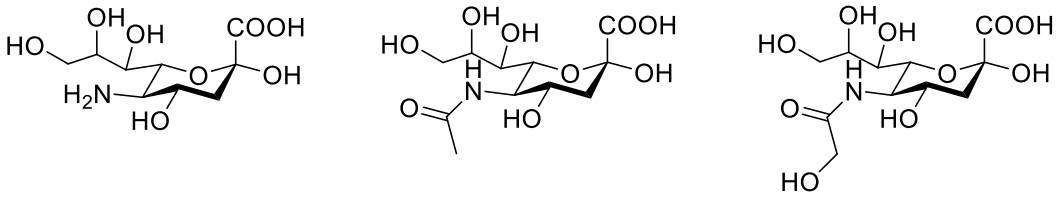
Figure 1.10 - Formation of neutral and sialyated glycosphingolipids from Glucose-Ceramide (GlcCer) and Sulfrosphingolipids from galactose-ceramide through enzymatic modification in the ER (Reprinted under creative commons license) [57]

1.2 Glycosylation of Protein based therapeutics

The previous sections outlined the significance of glycosylation upon protein and wider cellular function. Glycosylation plays a key role in cell-signalling, cell proliferation and growth, as well as cell-cell interactions and mucosal rheology. Consequently, glycosylation forms a significant part of the developmental stages of therapeutic drug design. It has been noted that both the degree and nature of glycosylation can impact the activity and serum half-life of a potential drug. Moreover, the linkage and structure of the attached sialylation has been shown to have a profound effect upon the safety of a recombinant protein based drug. The biopharmaceutical industry therefore aims to minimise the effects of such modifications by producing protein based drugs in mammalian cell lines. These include Chinese hamster ovary cells (CHO), mouse myeloma cells (MMC) or baby hamster kidney cells (BHK) [58-60].

Both *N* and *O*-linked glycosylation are heavily influenced by culture conditions thus affecting the activity and viability of the final drug molecule produced. Factors such as pH, temperature, oxygenation and culture medium have been shown to influence the type and structure of the glycan population ultimately affixed to the protein [61, 62]. Glycosylation is therefore defined as a critical quality control attribute for biotherapeutics produced from recombinant cell lines [63].

The stability of cell lines is therefore an important characteristic in selecting an adequate cell line for biotherapeutic drug synthesis. CHO cells represent the most heavily exploited cell line for the synthesis of protein based drugs, accounting for over 70% of the approved biotherapeutics currently on the market [64]. CHO cells have been repeatedly observed as being tolerant to changes in culture condition; in particular pH, oxygenation and temperature [65]. The increased tolerance and resultant product stability has resulted in CHO cell derived biotics having a high safety record [66]. This is further aided by the similarity of the glycan structures naturally produced by CHO cell lines. This is particularly prudent when compared to prokaryotic or plant based systems [67, 68]. Although recombinant CHO cell based systems possess increased similarity to human glycosylation, differences in sialylation produce glycosylation patterns differing from the 2, 3 and 2,6 linkages exhibited in human cells. This is because CHO cells do not naturally posses the 2,6 sialyltransferase enzyme. This results in incomplete sialylation and the production of almost exclusively 2,3 linkages. Furthermore, differences between hamster and human sialylation pathways can result in the addition of Neu5Gc **3**, a sialic acid not found in human glycan molecules.



1 **2** **3**

Figure 1.11 - Sialic acid structures present on both complex and hybrid N-glycans as well as Glycosphingolipids 1) Neuraminic acid 2) Neu5Ac 3) Neu5Gc

1.2.1 Immunogenicity of biotherapeutics

As well as their role in many significant cellular functions, glycans have also been shown to elicit an immunogenic response. This is particularly true in pathogen-host interactions, or as a result of interactions between proteins synthesized from non-human mammalian sources. In biotherapeutics the presence of Neu5Gc **3** (**Figure 1.11**) sialic acid represents a significant barrier to drug viability. This monosaccharide shares a structure similar to that of Neu5Ac **2**, a sialic acid naturally found on human glycans. The formation of **3** shown sees the modification of the acetyl group present in Neu5Ac **2**, which is converted to a glycol group by the action of CMP-*N*-acetylneuraminic acid hydroxylase (CMAH)[69].

Neu5Gc **3** is commonly found as a sialic acid affixed to many mammalian glycan structures. However, in higher mammals, the synthesis pathway displays alterations meaning Neu5Gc **3** is not naturally observed. As a result, biologics produced in lower mammalian cell lines such as CHO cells routinely display small amounts of Neu5Gc **3** as a result of CMAH expression [70]. The presence of Neu5Gc **3** has been shown to elicit a significant immune response in humans following the formation of an immune complex with anti-Neu5Gc [71, 72]. The result of this has been linked to inflammation and the potential onset of anaphylaxis [73].

Similar immunogenic effects have been observed in alpha-Gal containing biologics resulting from synthesis in cell lines producing the α-1,3-galactotransferase enzyme [74]. This enzyme is inactive in higher mammalian species but present in lines such as CHO and MMC [75] .

One such example of α-gal containing biotherapeutics is the monoclonal therapy Cetuximab®. Cetuximab® falls into a class of drugs known as monoclonal antibodies (mAbs) [76]. It is used in the treatment of skin cancer, non small cell lung cancer and metastatic colorectal cancer and acts by inhibiting epidermal growth factor. Since its approval in 2004, it has been heavily linked with hypersensitivity and immunogenic side effects, which have been connected to the glycan structures populating its surface. The synthetic route for Cetuximab® involves the use of MMC, which possesses the ability to produce active α-1,3-galactotransferase enzyme [77]. The result is a heterogeneous population of 21 different oligosaccharide structures, of which around 30% display α1,3 glycosidic linkages between galactose structures, as opposed to the Gal-β-1,4-Gal observed in higher mammalian cell lines [78, 79].

1.2.2 Effect of glycosylation on biotherapeutic efficacy

Aside from the impact of glycosylation upon the safety of biotherapeutics, glycosylation has been shown to impact both the activity and effectiveness of the biotherapeutic following administration. *N*-Linked glycans have been shown to influence glycoprotein bioavailability, solubility, stability and pharmacokinetic properties. By influencing glycan structure, the pharmaceutical industry aims to influence the length of time a biotic is active within the circulatory system. The presence of glycosylation has been shown to increase protein stability in response to temperature, pH and oxidative degradation, as well as reduce the likelihood of precipitation or aggregation [80-83]. The impact of glycosylation is also heavily involved in the circulatory time of biotherapeutic drugs, having a significant impact upon the pharmacokinetics. It has been widely observed that native sialylated glycoproteins exhibit increased circulatory lifetimes over non-glycosylated or partially glycosylated proteins. Further research showed that improperly glycosylated proteins were rapidly removed from solution [84, 85].

Glycosylation has been shown to increase the serum half-life of recombinantly generated mAB based therapies. However, divergence between lower mammalian cell lines such as CHO cells and native human cells results in the generation of non-human glycan species. This may lead to variations in both sialic acid structure and linkage type significantly effecting serum half-life [86]. These factors may result in increased rates of drug clearance via mechanisms such as phagocytic proteolysis. Variations in sialic acid addition and structure have been shown to significantly impact the serum half-life of biotherapeutics, with reduced clearance associated with increased presence of Neu5Ac **2** [87]. In this instance, the presence of increased sialylation reduces the presence of terminal galactose subunits mitigating against interactions with hepatocytic asialoglycoprotein receptors [88, 89]. Thus results in hepatic clearance of the drug by endocytosis [90, 91].

1.3 Analysis of glycans

The analysis of glycans represents a key procedure carried out within the development and production of biopharmaceuticals. The presence of incorrect glycoconjugate production can have a significant effect upon the safety and activity of a therapeutic drug, and glycosylation is denoted as a critical quality control attribute [92, 93]. Existing pharmaceutical workflows predominantly focus around the analysis of derivatised *N*-glycans following enzymatic release from protein sources. However,

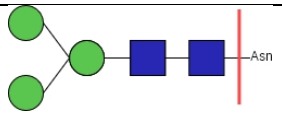
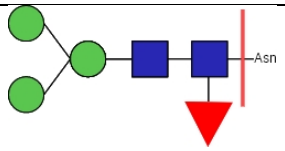
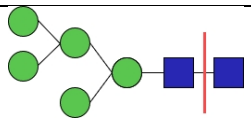
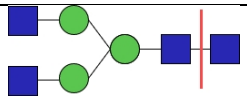
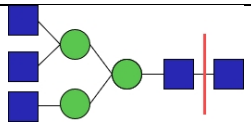
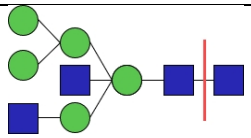
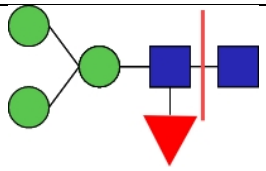
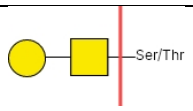
oligosaccharides can be released from both *N*- and *O*-linked sources by both enzymatic and chemical cleavage routes, and can then be subsequently derivatised by a range of chemical based methods.

Glycans are commonly affixed to other biomolecules, this provides a significant challenge to the analysis of the compound of interest. To simplify analysis, glycans are commonly released from their biomolecule and purified to remove protein debris [94]. These workflows differ significantly depending upon method variation, but can be broadly grouped into two categories with either chemical or enzymatic release used.

1.3.1 Enzymatic glycan release

Enzymatic glycan release is commonly chosen for the mild conditions in which glycan removal can be accomplished. The choice of enzyme differs with the type of glycan being targeted. Deglycosylation enzymes exhibit high substrate specificity to both glycosylation linkage and overall structure **Table 1.2**. Enzymatic cleavage can however be applied to *N*- and *O*-linked glycosylation in glycoproteins as well as proteoglycans, GPI anchors and glycolipids [95-99]. Enzymatic applications do have some limitations surrounding enzyme specificity with unwanted cleavage taking place, as well as specificity to the substrate, which leads to restricted or incomplete glycosylation [100, 101].

Table 1.2 - Enzymatic cleavage products for the release of both *N*- and *O*-linked glycan species

Glycosylation type	Enzyme	Type	Substrate	Reference
N-Linked	PNGase F	All *N*-glycan types not displaying 1,3 fucosylation		102, 103
	PNGase A	All *N*-glycan types regardless of 1,3 fucosylation		104
	Endoglycosidase F1	Oligomannose Hybrid at reduced rate		105
	Endoglycosidase F2	Oligomannose, Complex biantennary		105
	Endoglycosidase F3	Complex biantennary, Trimannosyl chitobiose		105
	Endoglycosidase H	Oligomannose, Hybrid type		106
	Endoglycosidase D	Paucimannose, with or without antennea		107
O-Linked glycans	Endo-α-N-acetylgalactosaminidase	*O*-Glycans displaying core 1 in mucin type glycans.		108

1.3.1.1 Intact glycan release

Existing biopharmaceutical and research workflows rely heavily upon the use of peptide *N*-glycosidase (PNGase) for the release of *N*-glycans, due to their ability to cleave intact asparagine linked *N*-glycans from a protein core. These strategies involve both PNGase A and F depending upon the glycan substrate being targeted. PNGases are used regularly and together can be used to release almost every *N*-linked oligosaccharide [94]. Glycan release by PNGase A and F occur at Asn residues. Cleavage occurs at the *N*-linked glycosidic bond between Asn and the first GlcNAc, provided the

primary structure exhibits an X-Asn-X with the asparagine residue being amide linked to another residue at both termini **Figure 1.12**. This therefore means that enzyme substrates can be no smaller than a tripeptide, with Asn as the central residue. Although PNGase F has been shown to cleave *N*-linked high mannose complex and hybrid type glycans, it is unable to perform glycan release at residues exhibiting an α(1→3)-fucose as part of the glycan's core [102]. These linkages are commonly found in samples derived from plant glycoproteins and remain resistant to PNGase F. PNGase A has been shown to cleave glycans exhibiting α(1→3)-fucose core, however it's use is limited by it's inability to cleave sialylated glycan species [94]. Although both PNGase A and F (**Figure 1.12**) are suitable for native glycan release, steric hindrance can significantly impact the rate and ultimate completion of sample de-glycosylation [109]. These factors can be removed, and glycan release accelerated, by the use of detergents such as sodium dodecyl sulphate and heat, to denature protein samples prior to PNGase addition.

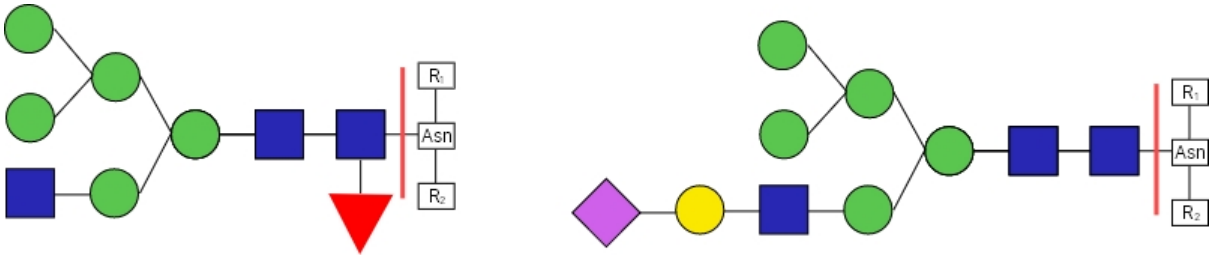


Figure 1.12 - Enzymatic cleavage substrates for PNGase A, possessing an α(1→3)-fucose core modification (Left) and PNGase F displaying terminal sialyation (Right)

1.3.1.2 Enzymatic release of native glycan fragments

Glycan release can also be carried out by the use of endoglycosidases. These enzymes are seldom used for intact *N*-glycan release due to their limited specificities and because they have been shown to cleave glycans at the second GlcNAc, leaving a single GlcNAc residue affixed to the protein backbone. Endoglycosidases H (endo H) and Endoglycosidases F (Endo F) enzymes can however be useful in native protein de-glycosylation, because they do not exhibit the same rate limitations as PNGases [110, 111].

Endo H enzymes can be particularly useful for the selective release of high mannose and hybrid type glycans, while leaving complex glycans intact [100]. Unlike PNGase, endoglycosidases are capable of cleaving glycans from single amino acids as well as dolichol-linked glycans that are yet to be affixed to their protein core [112]. The endo F family of enzymes also benefit from a lower sensitivity to protein conformation and therefore can be extremely useful in removal of almost all *N*-linked glycan

species. It should be noted however, that endo F enzymes are usually used for removal of specific glycans, because endo F subtypes exhibit high levels of specificity depending on glycan structure outlined in **Table 1.2**.

1.3.1.3 *O*-Linked glycan release

The release of *O*-linked glycans by enzymatic workflows presents a more significant challenge than for *N*-linked glycans, because the field of glycomics is yet to isolate a single enzyme for the release of *O*-linked oligosaccharides. This means that enzymic cleavage of complete *O*-linked glycans is currently not possible and provides a significant barrier to analysis. Enzymatic release of *O*-linked glycans can, however, be carried out via a process of sequential trimming. This process sees the use of both endo F enzymes as well as exoglycosidases to sequentially remove monosaccharides until the core Gal-β1,3-GalNAc remains attached to Ser/Thr. This core can then be removed by the use of *O*-glycanase (endo-α-*N*-acetylgalactosaminidase) [113], as an intact structure without further modification to the protein backbone. The activity of *O*-glycanase does not appear to show enhanced function following denaturation of the glycoprotein, however it's activity is severely hindered by the existence of any modification to the glycan's core, including the existence of mono-, di- or trisialylation [114 115, 116].

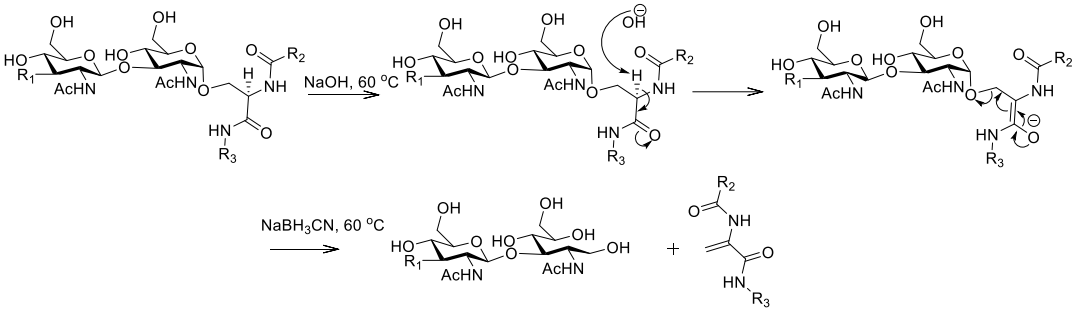
1.3.2 Chemical glycan cleavage

Chemical glycan release can broadly be defined as the removal of biologically significant oligosaccharides without the use of a biological element to aid the process. Although initially developed for the removal of *N*-glycans, these methods can now be used effectively for the release of multiple glycoconjugates including proteoglycans, GPI anchors and *O*-linked oligosaccharides [99, 117 118]. These methods exhibit lower levels of specificity to glycan substrate than enzymatic routes but commonly decreased yields are afforded without the use of harsh reaction conditions.

1.3.2.1 Alkaline β- elimination

The release of *O*-linked glycosidic linkages remains a challenge via enzymatic routes, and therefore their analysis relies heavily upon the use of chemical linked release. This can be achieved by the use of β-elimination under mild alkaline conditions. This workflow is widely used due to its reliance on easily acquirable stable reactants [119]. *O*-Glycosidic linkages between the glycan and protein backbone are readily hydrolysed in mild alkaline reaction conditions, requiring between 0.05M -0.1M sodium hydroxide to achieve glycan release over 16 h between 45-60 ^{o}C [120, 121]. However these workflows highlight a process known as peeling, which equates to the loss of monosaccharide subunits from intact glycans during the release process [122]. In an attempt to combat this, a reducing agent 1-2M sodium cyanoborohydride must be used [119]. The use of β-elimination shown in **Scheme 1.1** sees the

released glycan structure as a reduced alditol form of the molecule, with deacetylation of any GlcNAc residues present. These moieties must therefore undergo a process of reactylation with acetic anhydride to regenerate GlcNAc from GlcNH.

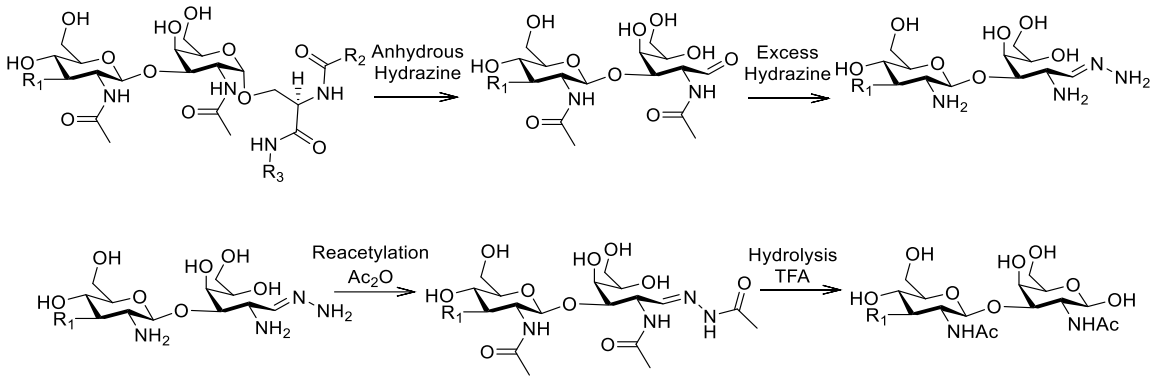


Scheme 1.1 - β-Elimination of glycans under alkaline conditions in the presence of a borohydride reducing agent generating released reducing glycan structures.

Alkaline β-elimination is seldom chosen as a glycan release method for mass spectrometric analysis applications, because the use of either sodium or potassium hydroxide results in significant salt concentrations present in the sample following glycan release. The presence of salt then interferes with analysis, causing suppression of target ions and a reduction in abundance recorded by the detectors [123]. Methods using ammonia in non-reducing conditions have attempted to reduce the presence of sodium or potassium ions in the sample. However de-glycosylation under these conditions can take up to 40 hours and results in significant 'peeling' of the *O*-glycan structure [124]. Although more commonly used for *O*-linked glycan release, this workflow can be utilised for the release of *N*-linked glycan species. This method does, however, require increased temperatures (up to 100 °C) and an increased concentration of base (1 M sodium hydroxide), making it unsuitable for the release of sialylated glycan species, and as such is rarely used in *N*-glycan release applications.

1.3.2.2 Hydrazine β-elimination

β-Elimination is also carried out utilising hydrazine to replace the alkaline conditions described above. This process outlined in **Scheme 1.2** sees glycans released from the protein backbone, forming a hydrazine containing glycan species, masking the carbohydrates reducing end [125]. This step also sees the deacetylation of GlcNAc residues, removing some of the glycans structure. The released glycan must therefore be re-acetylated, and its reducing end hydrolysed back to a hydroxyl functional group with TFA, before labelling can take place [126, 127]. This step results in a reducible oligosaccharide that can undergo further derivatisation, however the use of strong acids is prone to cause sample breakdown in acid labile regions [128].



Scheme 1.2 -Deglycosylation with hydrazine and subsequent reacetylation of GlcNAc residues with acetic anhydride.

1.3.3 Glycan derivatisation

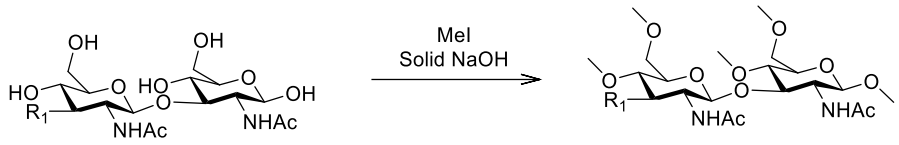
The release of glycans from their protein backbone aims to reduce the noise associated with non-target molecules within the sample during detection. However, the analysis of native glycans provides a significant analytical challenge. Glycans are known to be highly polar, non-fluorescent molecules, resulting in challenges in separation and analysis [123, 129]. In order to aid in the elucidation of these structures, derivatisation is commonly employed. These methods result in the introduction of physicochemical properties not exhibited by the glycan itself. For example a single fluorophore may be introduced, or more significant structural alterations involving two or more regions of modification. These typically aim to improve analysis by the addition of optical properties, or an increase in hydrophobicity to enhance ionisation.

1.3.3.1 Non-quantitative glycan derivatisation.

1.3.3.1.1 Per-*O*-methylation

The use of per-*O*-methylation in glycomics was developed to aid ionisation of highly polar oligosaccharides in mass spectrometric applications. The process sees the addition of methyl groups to hydroxyl groups present along the oligosaccharide (**Scheme 1.3**). The protocol was introduced by Ciccanu and Kerek in 1984 [130] and takes place in anhydrous DMSO, utilising methyl iodide and solid sodium hydroxide. This method is limited by the potential for oxidative degradation during the reaction. In order to increase yields, a modified protocol replaces totally anhydrous conditions in favour of slightly hydrated conditions, involving the addition of a small amount of water to DMSO [131]. The modified conditions reduce the risk of oxidative damage but pose a challenge in the analysis of low abundant glycan species due to multiple side reactions. This led to the development of a solid phase method, in which capillaries packed with sodium hydroxide beads are flushed with sample and

reagents. This allowed the derivatised sample to flow through, reducing the effects of side reaction and minimising oxidative damage [132].

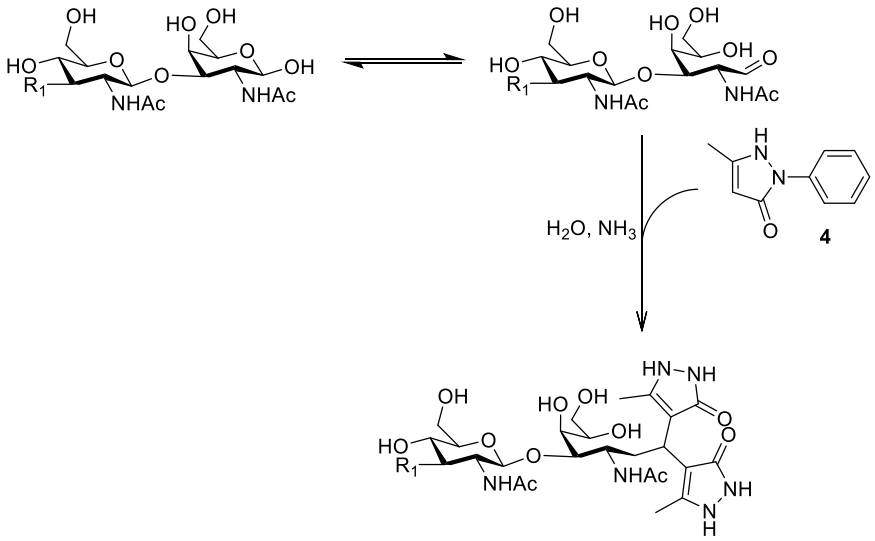


Scheme 1.3 -Glycan derivatisation by per-*O*-methylation of hydroxyl groups present along the glycans structure.

The increase in hydrophobicity as a result of per-*O*-methylation provides a significant benefit to analysis by mass spectrometry, showing a significant increase in ionisation efficiency in ESI (electrospray ionisation) [133]. Glycan modification by solid state per-*O*-methylation, has been shown to be effective for both neutral and sialylated glycans of high abundance. However, analysis of low abundance glycan species provides a challenge in infusion or MALDI-MS (Matrix Assisted Laser Desorption Ionisation) applications, where ionisation competition and source loading are important factors for efficiency [134].

1.3.3.1.2 Michael addition

Unlike per-*O*-methylation, Michael addition sees modification of a single site on the target glycan, targeting the reducing end of released carbohydrates. Michael addition takes place as a nucleophilic addition reaction between an α,β-unsaturated carbonyl compound and an active methylene donor (**Scheme 1.4**). By this mechanism, reagents such as 1-phenyl-3-methyl-5-pyrazolone (PMP) **4** or its derivative 1-(2-naphthyl)-3- methyl-5- pyrazolone **5**, have been applied to glycans [135, 136]. The derivatization takes place under alkaline conditions, with Michael donor molecules consecutively added to the glycan's reducing end in a ratio of two to one. Following the addition of the first label, a water molecule is released generating a α,β-unsaturated carbonyl compound, which is capable of undergoing a second label addition. This yields a 1,4-addition product and the addition of a second label. The use of PMP **4** and its derivatives have been shown to facilitate improved analysis by UV spectroscopy and mass spectrometry methods. The use of Michael addition exhibits some limitations for chromatographic separation, because the addition of two labelling molecules leads to separation mediated almost exclusively by the label, making differentiation of similar compounds challenging. The addition of multiple labels also results in a non-quantitative derivatisation, limiting the application of PMP **4** derived labels.

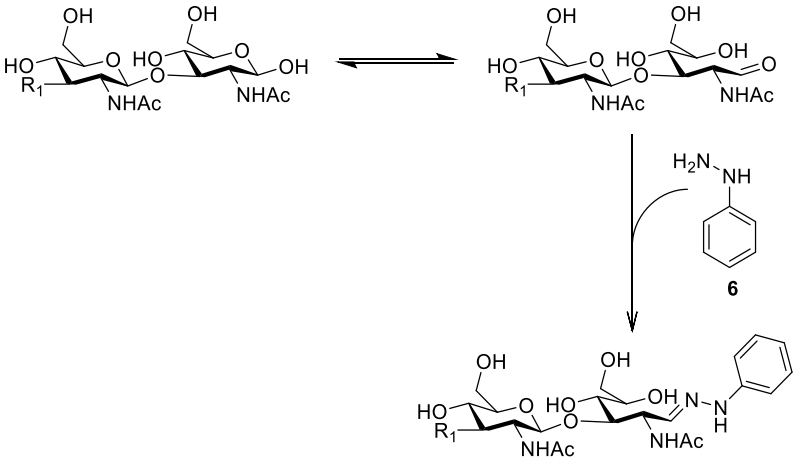


Scheme 1.4 -Glycan derivatisation by Michael addition and the addition of two molecules of PMP **4**

1.3.3.2 Quantitative glycan analysis.

1.3.3.2.1 Hydrazine labelling

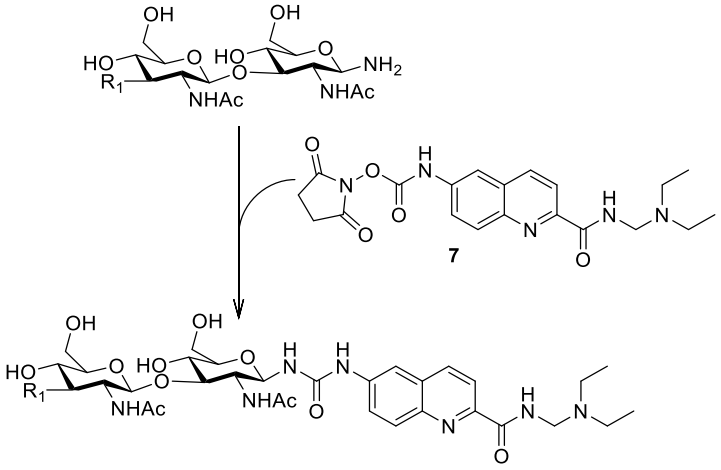
The conversion of oligosaccharides to phenylhydrazones was initially developed by Shinohara *et al* [137] and can be seen in **Scheme 1.5**. Since then, the use of hydrazine labelling has been shown to provide a powerful tool for the detection of low abundance oligosaccharides in both ESI and MALDI mass spectrometry applications. Hydrazine labelling is performed in acidified methanol at between 60-80 ^{o}C, with completion expected between 1-3 hours [138-140]. Lattova and Perreault demonstrated that the use of hydrazine labelling enabled the quantitative analysis of glycan species, reporting in up to a 15 fold increase in ionisation efficiency, which lead to significantly increased detection sensitivity [138].



Scheme 1.5 -Glycan derivatisation by reducing end modification with hydrazine labelling.

1.3.3.2.2 Activated carbamate labelling

In recent years, NHS-carbamate activated labelling compounds have become more widely used in the field of glycomics, and have been marketed as "rapid" labels [141]. These compounds contain an NHS-carbamate capable of derivatising amino sugars, resulting from basic de-glycosylation by PNGase F. The labelling reaction outlined in **Scheme 1.6** takes place in between 5-10 minutes using mild conditions, resulting in minimal loss of sialic acid and the attachment of a single label to the aminoglycans [142]. The resultant conjugate allows a quantitative glycan strategy, however high levels of sample purity are required to ensure amino acids and protein fragments are not labelled in the process [143]. Furthermore, glycosylamine type glycans exhibit instability when removed from high pH solutions, favouring the formation of a hemiacetal compound unable to be derivatised by NHS-carbamate labels.

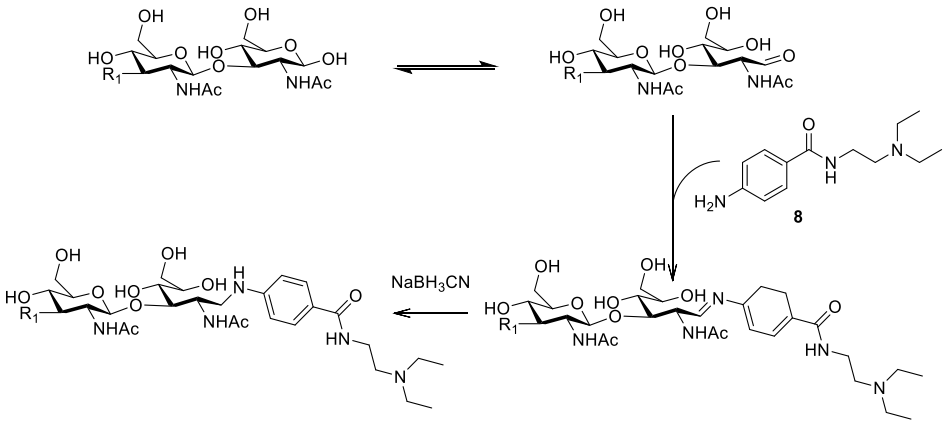


Scheme 1.6 -Activated carbamate labelling of glycosylamine containing carbohydrates with **7**.

1.3.3.2.3 Reductive amination

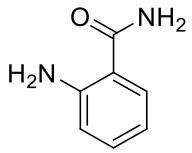
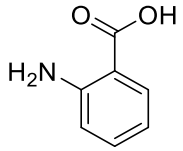
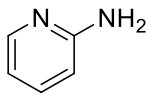
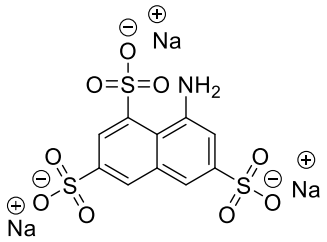
Reductive amination is the most widely used method of oligosaccharide derivatisation and sees the addition of a label to the glycan's reducing end. In this reaction, a primary amine containing label undergoes a condensation reaction with the aldehyde of a reducing sugar. The resultant imine is then reduced by a reducing agent, forming a stable secondary amine product. Reactions have been performed in DMSO acidified with acetic acid, however subsequent methods have sought to maintain sialylation using more mild conditions involving buffered methanol (MeOH) or tetrahydrofuran (THF). Significant optimisation has also been carried out surrounding the choice of reducing agents, with sodium cyanoborohydride ($NaBH_3CN$) being the most commonly used [144]. Its limitations commonly surround its toxicity, and methods using 2-picoline borane, have been developed as a non-toxic substitute [145]. Further developments have seen the use of sodium triacetoxyborohydrive [146] and borane-diethylamine [147], however these compounds are seldom used

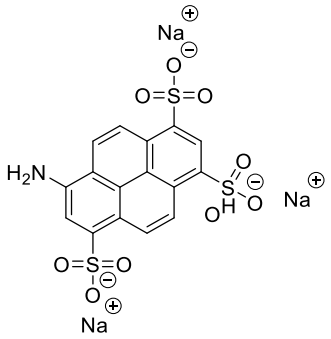
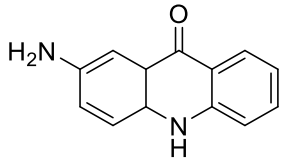
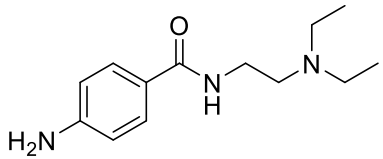
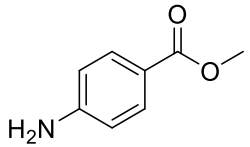
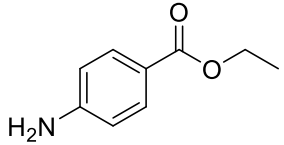
within routine analysis. **Scheme 1.7** outlines the process of reductive amination between GlcNAc residues and procainamide **8**.

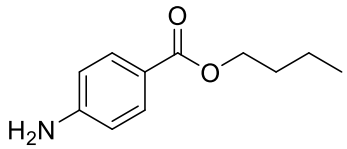
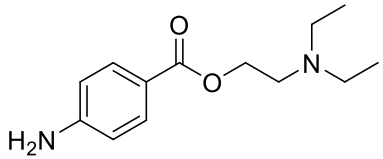
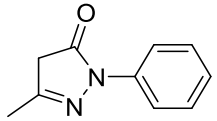
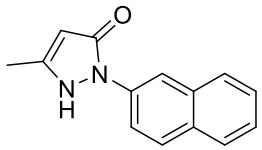
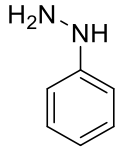
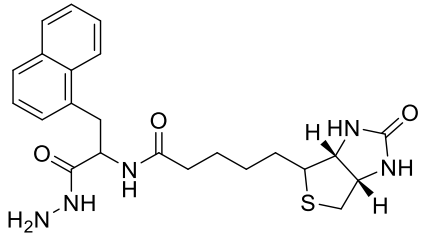


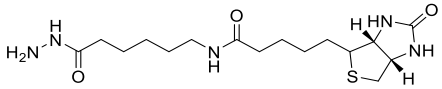
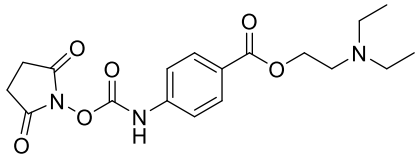
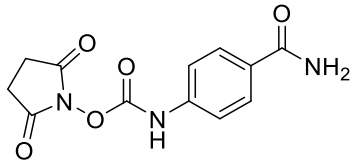
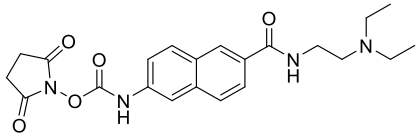
Scheme 1.7 - Derivatization of reducing carbohydrates with **8** by reductive amination in the presence of a reducing agent such as sodium cyanoborohydride.

Table 1.3 - Glycan labelling strategies for derivatisation of released carbohydrates following either chemical or enzymatic deglycosylation routes

Fluorescent label	Structure	Labelling mechanism	Substrate glycan	Data generated	Application	Ref
2-Aminobenzamide (2AB) **9**		Reductive amination	Reducing sugar	Quantitative	Widely used for structural determination of released glycans in commercial workflows due to a significant library of glycoform standards. Labelling imparts a single fluorophore to the target oligosaccharide, aiding in chromatographic separation and fluorescence analysis. Exhibits a low level of hydrophobicity and therefore less applicable to MS workflows than other compounds.	[144]
2-Anthranilic acid (2AA) **10**		Reductive amination	Reducing sugar	Quantitative	Used in HPLC and CE applications for separation of reducing glycans, for analysis by fluorescence detection. Exhibits a single negative charge in solution, making it suitable for negative mode MS analysis of charged glycans or positive mode MS of neutral glycans.	[148, 149]
2 Amino pyridine **11**		Reductive amination	Reducing sugar	Quantitative	Used in the analysis of reducing glycans by HPLC-MS, however compound must be crystalised prior to derivatisation resulting in a slower sample analysis time.	[150]
8-Aminonaphthalene 1,3,6-trisulfonic acid (ANTS) **12**		Reductive amination	Reducing sugar	Quantitative	Provides a fluorophore and is widely used in CE applications due to its permanent pH independent anionic charge. Its heavily sulphated structure makes it unsuitable for MALDI-TOF applications	[151]

8-Aminopyrene-1,3,6-trisulfonate (APTS) **13**		Reductive amination	Reducing sugar	Quantitative	Fluorescent label, widely used in CE analysis due to its negatively charged structure, however its charge leads to poor ionisation efficiency in MALDI-TOF applications.	[152]
2-Aminoacridone **14**			Reducing sugar	Quantitative	Exhibits a highly fluorescent aromatic region making it a powerful tool in the analysis of low abundance glycan species. Its hydrophobic structure also increases ionisation efficiency in both +ve and -ve ion mode ESI MS or MALDI-TOF	[153]
Procainamide (ProA) **8**		Reductive amination	Reducing sugar	Quantitative	Routinely used in existing biopharmaceutical analysis. Molecule exhibits high levels of fluorescence while exhibiting a hydrophobic region making it beneficial for separation by LC and detection by +ve mode ESI or MALDI-TOF MS.	[154, 155]
Methyl-4-aminobenzoate (ABME) **14**		Reductive amination	Reducing sugar	Quantitative	Not commonly applied in commercial glycan analysis due to poor ionisation efficiency and rapid elution in RP-HPLC separations.	[156]
Ethyl-4-aminobenzoate (ABEE) **16**		Reductive amination	Reducing sugar	Quantitative	Less commonly applied in commercial glycan analysis but exhibits increased retention over ABBE **17**and improved ionisation efficiency.	[156]

Butyl-4-amino benzoate (ABBE) **17**		Reductive amination	Reducing sugar	Quantitative	Rarely used commercially, however its increased chain length showed increased resolution in LC separation, as well as increased sensitivity over ABEE **16** in -ve mode MS.	[156]
Procaine **18**		Reductive amination	Reducing sugar	Quantitative	Less commonly used in commercial analysis **18** shares structural similarities with **8**. The molecule exhibits high levels of fluorescence while exhibiting a hydrophobic region making it beneficial for separation by LC and detection by +ve mode ESI or MALDI-TOF MS.	[157]
1-Phenyl-3-methyl-5-pyrazolone (PMP) **4**		Michael addition	Reducing sugar	Qualitative	The most commonly used Michael addition labelling takes place at a high pH, making it unsuitable for charged glycans but it is effective in both LC and MS applications.	[158, 159]
1-(2-Naphthyl)-3-methyl-5-pyrazolone (NMP) **5**		Michael addition	Reducing sugar	Qualitative	Similarly effective to its PMP **4** derivative, **5** exhibits enhanced sensitivity to ultraviolet (UV) detection when compared to other Michael addition derivatives.	[135]
Phenyl hydrazine **6**		Hydrazide labelling	Glycosyclamine	Quantitative	Shown to be highly effective in glycan analysis by both ESI and MALDI applications. Single label addition means hydrazine label is quantitative.	[160]
Biotinyl-L-3-(2-naphthyl)-alanine hydrazide (BNAH) **19**		Hydrazide labelling	Glycosyclamine	Quantitative	Displays effective UV and fluorescent properties even when used at low equivalencies. The generation of biotinyl-oligosaccharides allows protein carbohydrate interaction to be explored. Alanine substitutes for β-glycosylasparaginyl core required for some binding interactions to occur.	[139]

6-(Biotinyl)-aminocaproyl hydrazide (BACH) **20**		Hydrazide labelling	Glycosyclamine	Quantitative	Displays similar analysis characteristics to **19** with good UV and fluorescence outputs. **20** does not exhibit any alanine modification and is more commonly used for surface activation applications.	[139]
InstantPC **21**		Carbamate labelling	Glycosyclamine	Quantitative	Shares high levels of structural similarity with compound **8** (ProA) and has been shown to exhibit high levels of sensitivity in both fluorescence and MS analysis over a significantly shorter workflow.	[141]
InstantAB **22**		Carbamate labelling	Glycosyclamine	Quantitative	**22** shares a close structural connection to 2AB **9** and has been shown to be effective in fluorescence analysis of glycosyclamine containing glycans over a fast workflow.	[161]
Rapiflour-MS **7**		Carbamate labelling	Glycosyclamine	Quantitative	The most widely used of the rapid labels, the label contains both quinoline and a tertiary amine tail providing improved separation by LC methods as well as efficient ionisation in both ESI and MALDI-TOF MS techniques	[143]

1.3.4 Glycan Purification

The derivatisation of glycans requires reactions that yield a high proportion of labelled glycans. In order to facilitate this, high equivalencies of label are combined with released glycans. This is necessary due to the low concentration of glycans within released samples. The excess derivatisation agents, although essential to the labelling process, can provide noise in both separation and analysis. It is therefore essential that excess reagents are removed before the separation and analysis of derivatised glycans. The method of purification depends heavily upon the analytical workflow used. These can be grouped into six main categories; solid phase extraction (SPE) [145, 162, 163], liquid-liquid extraction [130], gel filtration [164], paper chromatography [165], precipitation [154] and dialysis [166]. These purification strategies and their applications are summarised in **Table 1.4**.

Reductive amination and hydrazine labelling reagents share similar structures and polarity, and are frequently purified by SPE [126]; while the significant polarity change resulting from per-*O*-methylation is more commonly purified by solvent-solvent extraction or precipitation [166]. These techniques are also applied to the products of michael addition, because the dual label addition results in highly non polar regions, making derivatised glycans more amenable to the organic phase [159]. Paper chromatography is rarely used in commercial settings but can be applied to both reductive amination and hydrazine labelling workflows. Purification by precipitation, although effective, is also seldom used commercially. However, it can be of use in methods where protein contaminants will damage product yields, such as carbamate labelling. Dialysis is routinely used for the clean-up and desalting of samples destined for analysis by mass spectrometry. However, it's limitation surrounds the techniques ability to remove both large and small molecular weight contaminants without damaging purified product yields [90]. Although all six of the techniques outlined in **Table 1.4** aim to provide a route for the purification of biologically derived oligosaccharides, SPE is frequently chosen for commercially viable workflows due to its fast sample turnaround and wide range of stationary phases available.

Table 1.4 - Glycan purification techniques for both derivatised and underivatized glycan samples.

Purification workflow	Material	Labelling workflow	Conditions	Reference
SPE	Amide-2 HILIC	Reductive amination	Condition- MeOH Load- 95% ACN Wash- 3-5% ACN Elute - H_2O	[167, 168]
SPE	Cellulose HILIC	Reductive amination	Condition- H_2O Load – 80% ACN Wash- 80% ACN Elute- H_2O	[169]
SPE	Discovery Glycan amide HILIC	Reductive amination	Condition- MeOH Load- 95% ACN Wash- 3-5% ACN Elute - H_2O	[170]
SPE	PGC	Reductive amination	Condition- H_2O Load – 80% ACN Wash- 80% ACN Elute- H_2O	[171]
SPE	RP, C_{18}	Per-*O*-methylation	Condition- MeOH or EtOH Load - H_2O Wash - H_2O Elute - 80-90% ACN *TFA or acetic acid may be used as a substitute pH modifier	[162, 172]
SPE	RP, C_{18}	Hydrazine labelling	Condition- H_2O Load – 80% ACN Wash- 40% ACN	[140]

			Elute- H_2O	
SPE	Oasis HLB	Reductive amination	Condition- 95% ACN Load – 95% ACN Wash- 95% ACN Elute- H_2O	[167]
SPE	HILIC DPA-6s	Reductive amination	Zip-tip method Condition- 20% ACN Condition- 95% ACN Load – 95% ACN Wash- 80% ACN Elute- 20% ACN (aq)	[168]
Gel filtration	Toyopearl HW-40F	Reductive amination	Condition- H_2O Load – H_2O Wash- H_2O Elute- H_2O + 50 g over 25 fractions	[173]
Gel filtration	Sephadex G15	Reductive amination	Samples were eluted with 10mM ammonium bicarbonate	[174]
Solvent-solvent extraction	Water/Ethyl Acetate	Hydrazine labelling	Purified glycans exhibit increased solubility in the aqueous phase while less polar labelling compounds are more soluble in the organic. Aqueous partition is therefore collected and concentrated.	[138]
Solvent-solvent extraction	Water/chloroform	Michael addition	Purified glycans exhibit increased solubility in the organic phase due to hydrophobic label addition. Organic portion containing purified glycans is therefore collected and concentrated.	[175, 176]

Solvent-solvent extraction	Water/Chloroform	Per-*O*-methylation	Purified glycans exhibit increased solubility in the organic phase due to increased hydrophobicity of methylation. Organic portion containing purified glycans is therefore collected and concentrated.	[130, 177]
Solvent-solvent extraction	Water/Dichloromethane	Per-*O*-methylation	Purified glycans exhibit increased solubility in the organic phase due to increased hydrophobicity of methylation. Organic portion containing purified glycans is therefore collected and concentrated.	[178]
Dialysis	1000 Da membrane	Per-*O*-methylation	Sample was dialysed against H_2O	[166]
Paper Chromatography	Whatman 3M filter paper	Reductive amination	Isotope labelled aniline was washed with ACN before being eluted with H_2O	[179, 180]
Precipitation	Acetone	Reductive amination	Samples were precipitated out in anhydrous acetone.	[154]

The limitations in glycan purification typically surround the high polarity resulting from glycan structure. The introduction of a derivatising agent and the inferred increase in hydrophobicity is necessary for the sample to adsorb onto most SPE based packing materials with sufficient retention to undergo reliable purification [181].

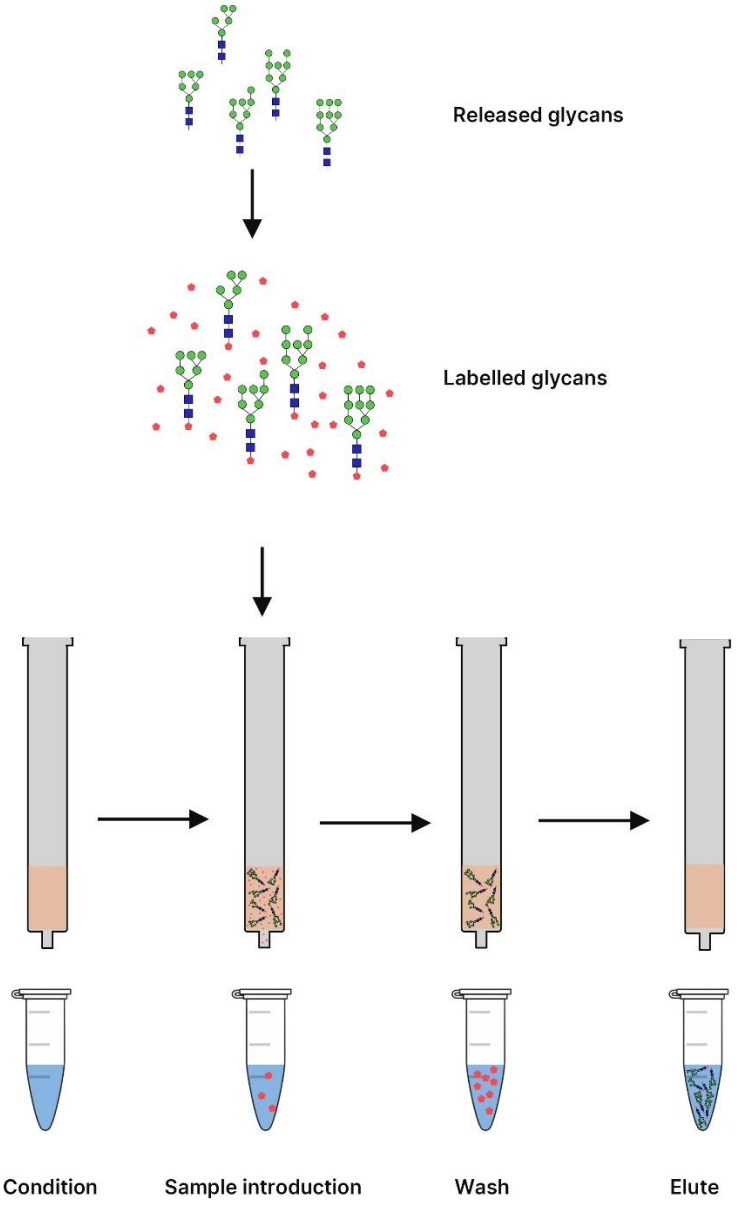


Figure 1.13 - SPE glycosylation workflow for the purification of both derivatised and underivatized glycan samples.

Of the purification methods summarised in **Table 1.4**, SPE based purifications represent one of the most commercially scalable workflows for glycan purification. This is due to the varied range of sorbent chemistry available and the speed at which purification can be carried out. The workflow outlined in **Figure 1.13** shows the process in which cartridges containing sorbent are conditioned, loaded and washed to generate a purified sample which can be eluted.

Reversed phase (RP) separations on C18 [182] are not routinely used for the purification of released glycans labelled by reductive amination due to challenges with high label polarity. Instead, it is more widely applied to samples derivatised by hydrazine or per-*O*-methylation. It should however be noted that improved retention can be achieved in PGC [183]. The retention of oligosaccharides on PGC was originally applied to non-derivatised glycans [184 185], however adequate purification has also been achieved for per-*O*-methylated samples [186]. PGC based separations display limitations for labelled glycan applications due to the retention of free label in eluted samples, although these only affect label-based derivatisation workflows. The high cost of PGC also limits its application in high throughput analysis workflows. Although significantly cheaper, the use of C18 based separations have been shown to be label mediated, with analyte retention being heavily dependent upon the addition of label or increase in hydrophobicity [140, 187].

The use of Hydrophilic interaction liquid chromatography (HILIC) based SPE separations has been proven to show significant advantages over reversed phase purifications. In HILIC-mode separation, glycans are retained based upon their hydrophilic properties, allowing glycans to be retained while the more hydrophobic labelling reagents are removed. HILIC based separation techniques commonly use polyamide sorbents. However, the use of cellulose and cotton wool-based systems have also been shown to be effective methods of glycan retention, while removing excess reagents from derivatisation [188, 189].

1.3.5 Glycan separation techniques

Purified glycan samples rarely contain a single glycoform, with purified samples of glycans from a single site exhibiting high levels of heterogeneity [190]. The variance in structure exhibited between glycoforms within a sample can range from differences in size and charge to linkage or variation in monosaccharide composition. This presents a significant challenge in the characterisation of glycan structures by many analytical techniques. In order to aid and simplify the analysis of individual glycan structures, separation techniques are commonly employed. Although a variety of methodologies

have been previously published, analytical scale glycan analysis is focused around the use of liquid chromatography and capillary electrophoresis [92, 191, 192].

1.3.5.1 High performance liquid chromatography

High-Performance Liquid Chromatography (HPLC) has been widely applied for the separation of glycan samples [92, 138, 148, 149, 155, 169]. Separation by HPLC is dependant upon the chromatography mode chosen for the application. In glycan analysis, these can be grouped into three main categories: normal phase (NP), reversed phase and ion-exchange.

1.3.5.1.1 Normal Phase glycan separation

NP mode separation was among the first separation mechanisms applied to HPLC. A polar stationary phase is run with a non-polar mobile phase. Normal phase chromatography traditionally takes place on a silica column, however amine and cyano bonded columns are also routinely used. The interaction between polar column packing and polar analytes results in retention of samples in order of increasing polarity. In glycan analysis, this results in glycan elution in order of increasing size. The decreasing hydrophobicity exhibited by larger carbohydrates leads to increased stationary phase interaction and increased retention. Packed silica based NP-chromatography relies upon highly non-polar solvents and experiences issues surrounding analyte solubility in underivatised applications [193]. This is negated by the use of a derivatisation agent and has been widely used for the separation of fluorescently labelled glycans [194].

1.3.5.1.2 Reversed phase chromatography

Reversed phase chromatography represents one of the most common analytical techniques applied to liquid chromatography. Unlike NP chromatography, RP based separations see the use of a nonpolar stationary phase and a polar mobile phase. RP separation commonly takes place on C18 columns but can be applied to a number of hydrocarbon-based sorbents, as well as graphitised carbon. Separation is therefore dependent upon hydrophobic interaction between analyte and stationary phase, resulting in increased retention of more hydrophobic molecules, while polar compounds elute earlier. The lack of hydrophobicity exhibited by underivatised released glycan samples results in little to no interaction with the stationary phase. This manifests as poor retention and minimal separation, because glycans simply pass through column voids and are eluted with the solvent front [195]. RP separation of underivatized glycans is therefore unachievable and instead the addition of a hydrophobic label must be applied. This aims to provide sufficient hydrophobicity to

infer sufficient stationary phase interaction, to generate separation and retention. RP based separation is therefore suited to glycan derivatisation workflows where a significant increase in hydrophobicity is achieved, such as per-*O*-methylation or Michael addition [196]. Separation of more polar derivatisation workflows such as reductive amination can be achieved by the use of polar encapped stationary phases [197, 198]. These exhibit polar regions containing amide or exposed silica to improve interactions with polar analytes and maintain packing bed polarity during the separation. This aims to increase retention of polar analytes, while still providing interaction for non-polar compounds. Alternatively, the use of mobile phase modifiers can aid separation of polar or charged by ion-pairing. This relies on the ionic interaction between an ion pairing reagent such as triethylammonium acetate and an analyte with both an ionic and non-polar region. In derivatised glycans, a neutral complex is formed between an ion pair reagent and label, increasing stationary phase interaction and increasing retention [199].

1.3.5.1.3 Porous graphitized carbon chromatography

A close variant of C18 reversed phase chromatography is the use of a porous graphitised carbon (PGC) as a stationary phase. PGC brings benefits over silica bonded stationary phases, exhibiting increased stability over a wide range of pH and temperature conditions [200, 201]. Additionally, due to the column's construction, the risk of solid phase shedding is significantly reduced, resulting in more stable separation [202]. Separation on PGC stationary phases relies on a combination of hydrophobic interaction, adsorption and ionic interaction, resulting in similar separation orders to those observed on reversed phase. In order to achieve this, separations are carried out with gradients of organic against aqueous solvents containing mobile phase modifiers such as ammonium bicarbonate, or other small salts [203]. PGC based chromatographic methods have been shown to be highly effective in separating underivatized glycan samples released from both sialylated [204, 205] and unsialylated *N*- and *O*-linked samples [206-208].

1.3.5.1.4 Hydrophilic interaction liquid chromatography

A close variant of both normal phase and reversed phase separation modes is the use of HILIC. To achieve separation, HILIC mode separation relies upon hydrophilic interactions between polar analytes and a polar stationary phase [209]. Elution is then mediated by mobile phase polarity, following an aqueous gradient similar to that used in reversed phase methods. This results in separation in order of increasing polarity, similar to that of NP separations. Stationary phases commonly consist of a silica-based column beds modified with amide, diol or cyano groups, in order to introduce stationary phase polarity. HILIC chromatography has been shown to be highly effective at separating both derivatised and underivatized released glycan samples bearing sialic acid modification [92, 170, 210-212].

1.3.5.1.5 Anion exchange chromatography

The use of anion exchange chromatography represents a widely applied technique in the separation of naturally derived glycan conjugates. Anion exchange separation is a form of ion exchange chromatography and separates analytes on the basis of charge. Resin based stationary phases are routinely used, and are typically modified diethyl-aminoethyl groups, providing a net positive charge. Separation occurs at neutral or slightly alkaline pH, resulting in deprotonation of sulphates, phosphates and carboxylic acids, which results in interaction with the positive stationary phase [213].

In glycan analysis, weak anion exchange (WAX) is routinely used to separate underivatized oligosaccharide samples based upon their charge. Almost all *N*-linked glycans exhibit a negative charge across the molecule. This results from components such as sialic acid, sulfates, or phosphates. Increased levels of negative charge therefore result in increased retention. WAX separation is therefore capable of separating between differing mono-, di-, tri- or tetra-sialylated structures in underivatized samples due to their differing negative charges [214]. A variation of this method is strong anion exchange chromatography, which sees increased charge across the stationary phase and the use of small ionic salts as mobile phase modifiers. This results in separation based upon ionic interaction, because retention is based on displacement by competing salts. This method is less commonly used on released glycan samples and is more likely to be applied to the separation of GAGs [215].

1.3.5.1.6 High Performance Anion Exchange Chromatography

Anion exchange chromatography can also be performed at high pH; this is often referred to as high performance anion exchange chromatography (HPAEC). The application of high pH aims to separate compounds that possess weak acid properties on a highly polar pH stable stationary phase. Carbohydrates not possessing sialic acid modification are typically weak acids with pK_a values ranging from 12-14 [216]. In high pH alkaline conditions this prompts partial dissociation of hydroxyl groups, resulting in multiple negative charges capable of interacting with stationary phase charges resulting in retention. This leads to selective carbohydrate elution, largely influenced by saccharide polymerisation and isomerism, due to the number of ionisable hydroxyl groups present on the structure [217]. Selectivity is highly dependant upon stationary phase structure, with ethylvinylbenzene/divinylbenzene ammonium microbead columns providing good separation for larger carbohydrates. While vinylbenzylchloride-divinylbenzene based columns display high selectivity for mono- or disaccharide separations, in carbohydrate applications elution is typically carried out isocratically at 10-20 mM aqueous sodium hydroxide [218].

1.3.5.1.7 2-Dimensional mixed mode anion exchange HILIC separation

In reality, anion exchange chromatography exhibits limitations in resolution and run repeatability. To overcome this, 2-dimensional separation techniques have been developed. The technique sees analytes separated by HILIC chromatography before individual co-eluting peaks in the first dimension are sent on to a further anion exchange column to undergo additional separation [219]. These methods typically see the use of amide or zwitterionic based HILIC separations in the first dimension coupled via a separating valve to diethylaminoethyl functionalised resin anion exchange columns [220, 221]. The result is high resolution separation of both derivatised and underivatized samples released from almost any biological carbohydrate source [136, 221].

1.3.5.2 Capillary electrophoresis

Separation of glycans can also be performed by the use of electrophoretic separation. In the case of capillary electrophoresis (CE), analytes are separated based upon their electrophoretic mobility to voltage applied to a capillary [222]. Unlike LC applications, CE does not rely upon analytes' interactions with stationary or mobile phase and instead separates based upon size and charge in response to the applied voltage [223]. In instances where analytes share similar charge, separation is typically dependent upon size, with larger molecules exhibiting larger field strengths and greater mobility than smaller molecules. In instances where molecular size is similar, but analytes display differing charges such as sialylation or sulfation, separation is dependent upon the charge of the molecules, with greater charge exhibiting higher mobility. CE applications display benefits over HPLC for some glycan separations, in that there is no variability in mobile phase, and runtimes are quoted as being significantly shorter than LC-HILIC based methods [224, 225].

1.3.6 Glycan sample detection

Glycan derivatisation not only aims to impart favourable characterises for the separation of glycans but also to aid in their detection. While some labelling compounds aim to impart optically active regions to the target glycan, others exhibit regions of increased hydrophobicity or permanent charge designed for mass spectrometry-based methods. A variation in detection methods is therefore seen across the field of glycan analysis, with some methods providing in depth structural analysis of

samples, while others require the use of an external retention standard to differentiate between glycoforms within a sample.

1.3.6.1 Optical glycan detection

The addition of an optically active region to released glycan samples forms the basis for many glycan profiling workflows. Following separation, derivatised glycans are pumped through a flow cell and exposed to a monochromatic electromagnetic radiation source. In ultraviolet applications, detectors then measure the light absorbed by the sample, while fluorescence detectors measure the emission of a particular wavelength from the sample.

1.3.6.1.1 Ultraviolet detection of glycans

The addition of a chromophore or fluorophore is common in many commercially used glycan labels. UV detection is commonly employed as a detection method for samples separated by HPLC. Detection usually takes place on a diode array or variable wavelength detector. Light sources typically consist of xenon or halogen lamps, which pass light through a filter or monochromator, producing a light source of a single wavelength [226]. This is then used to irradiate samples passing through the flow-cell. Photodetectors then measure the intensity of light passing through the sample, with increased absorbance indicating the presence of analytes in the sample. As each molecule irradiated absorbs the same number of photons, the concentration of analyte is therefore proportional to the absorbance. This allows for detector response quantification when compared to external standards of known concentration. As analysis takes place over a single wavelength, UV detection can display limitations with non-specificity, because multiple compounds may display absorbance at similar wavelength [227]. UV detection is therefore commonly coupled to MS based detection for further structural elucidation and qualification [195].

1.3.6.1.2 Fluorescence detection of glycans

While UV based detection methods seek to measure absorbance at a given wavelength, the use of fluorescent light detection relies upon the excitation of electrons from their ground state to a higher vibrational state, following the absorption of a photon of the correct wavelength [226]. The increased molecular collisions associated with increased energy state cause electrons to drop back to a ground state, resulting in the emission of a photon. The intensity of the light emitted is then detected resulting in data that can be processed by analysis software. This brings benefits over UV based detection methods, because analytes must be active at both excitation and emission wavelengths.

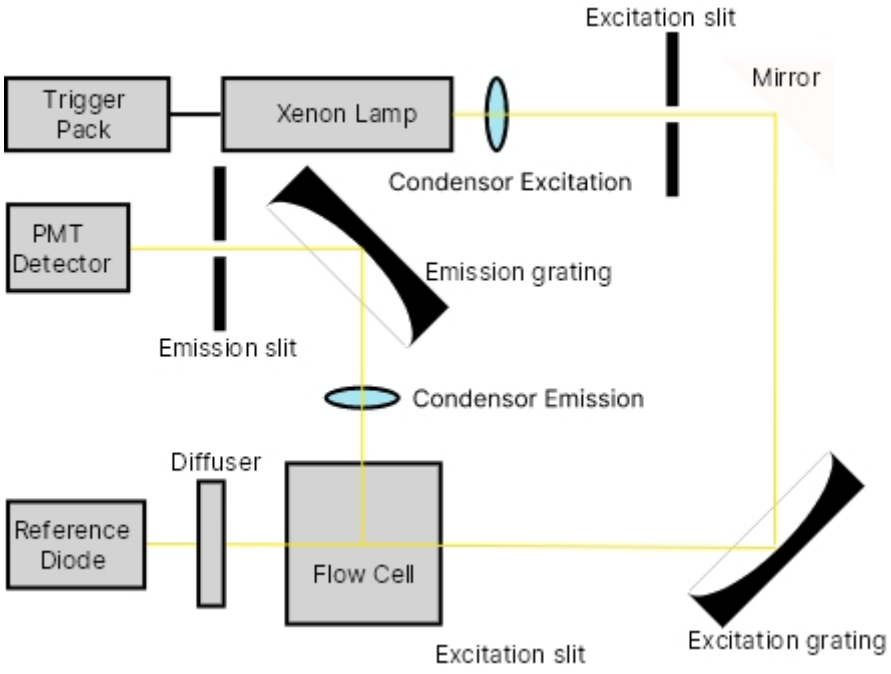


Figure 1.14 - Outline of a fluorescence detector incorporating a liquid flow cell

In fluorescence, a xenon lamp emits electromagnetic radiation which is then filtered by a monochromator to obtain the chosen analysis wavelength [226]. This is then passed though the flow-cell, irradiating the sample. This results in excitation of any analyte irradiated in the flow-cell. Detectors then pick up emission wavelengths produced by the sample (**Figure 1.14**). Fluorescence detection has been shown to provide superior sensitivity of between 100-1000x that of UV detection methods [228].

FLD detection is widely applied to the detection of glycans separated by chromatographic methods but has also been shown to be effective in CE applications. These applications differ from those applied to LC, in that excitation is carried out by a laser instead of xenon [229]. This results in high levels of sensitivity over very small sample volumes as exhibited by CE separation [230, 231].

1.3.6.2 Detection of glycans by mass spectrometry (MS)

MS based detection techniques allow for the detection and structural determination of both derivatised and underivatised carbohydrate samples. MS relies on the ionisation of analytes, resulting in the formation of charged adducts. The detection of such charged species is measured in

relative abundance against mass/charge ratio (M/Z). This generates a spectrum unique to each compound being analysed.

The analysis of carbohydrate samples is commonly carried out by 'soft' ionisation methods such as electrospray ionisation (ESI-MS) and matrix assisted laser desorption ionisation (MALDI-MS). These techniques aim to achieve efficient ionisation, while imparting little excess energy. This aims to reduce decay associated with the loss of labile groups, such as sialic acid or fucose moieties [232].

1.3.6.2.1 Electrospray ionisation (ESI)

ESI is used for the ionisation of samples dissolved in liquids in the gaseous phase (**Figure 1.15**). This is achieved in three stages. The sample solvent is first nebulised to a fine spray of charged droplets from a highly charged needle sheathed with nitrogen. The cloud of droplets then moves down a pressure gradient towards the mass analyser region of the spectrometer. Increased source temperature combined with nitrogen drying gas then evaporates the solvent, forming droplets with ever increasing surface charge density [233]. The increase in charge density eventually results in a process known as coulomb fission, whereby droplets break apart to form smaller droplets. This process is repeated as solvent is removed from the droplets. Eventually droplets then move into a heated dielectric capillary (100-300 ^{o}C), at which point the electric field generated is energetically capable of ejecting ions at the droplets surface, to be ejected into the gaseous phase, leading to total desolvation [234].

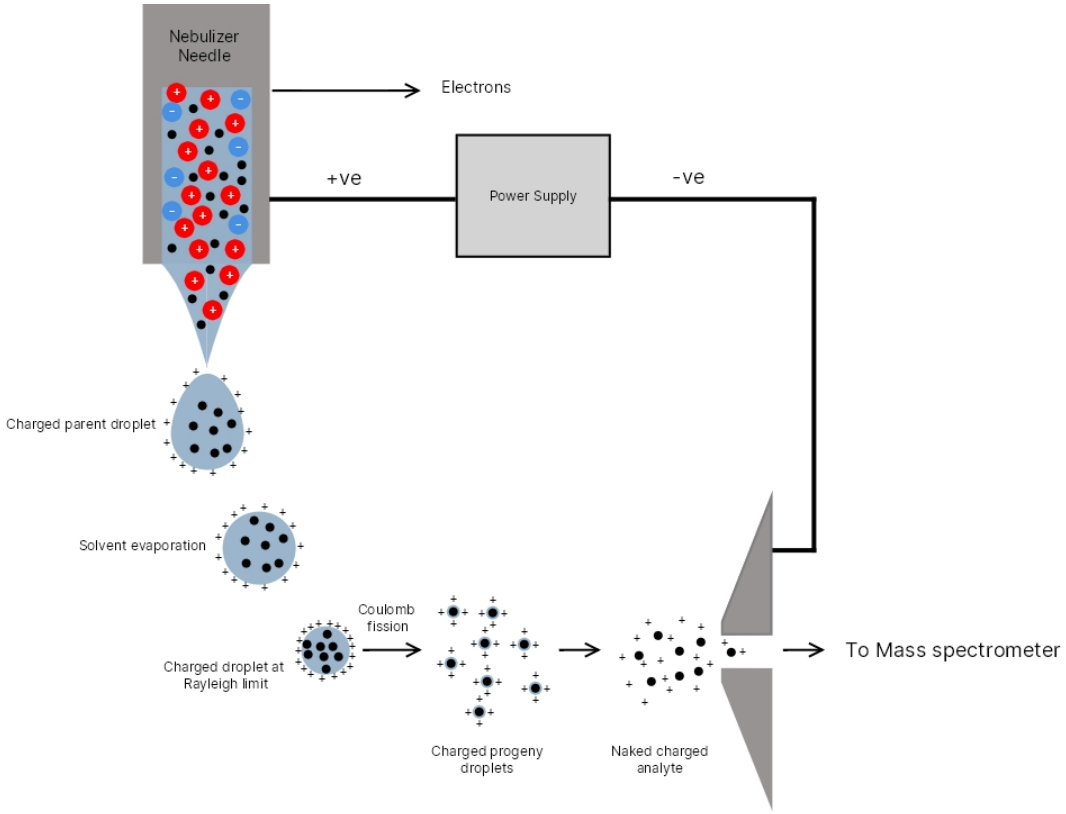


Figure 1.15 - ESI ionisation mechanism showing nebulised liquid droplets undergoing coulomb fission to form small charged droplets and ultimately charged analyte species.

Electrospray ionisation (ESI) generates a variety of singularly or multiply charged adducts $(M+nH)^{n+}$ which undergo acceleration before being analysed in the MS [232]. ESI allows for both liquid chromatography linked online detection, as well as direct infusion offline analysis routes. ESI ionisation is typically employed in quadrupole or Orbitrap mass analysis.

1.3.6.2.2 Matrix assisted laser desorption ionisation (MALDI)

MALDI differs significantly from ESI, in that ion formation is facilitated using high intensity laser irradiation. In addition, the process relies upon the use of a matrix capable of absorbing wavelengths achievable by the irradiation source. Sample preparation sees the sample mixed with the matrix on a target plate at small volumes (<1 µl) [235, 236]. The choice of matrix and solvent can have a significant impact upon the quality of data extracted form a MALDI experiment and there are several which have been applied to glycan analysis, outlined in **Table 1.5**.

Table 1.5 - MALDI matrix constituents for the analysis of released glycans.

Sample structure	Sample	Matrix	Analysis polarity	Reference
Neutral glycans	Released human milk oligosaccharides, hydrazine labelled	2,5-Dihydroxybenzoic acid (2,5-DHB)	Positive ion mode	[187, 237]
Acidic glycans	Released *N*-glycans bearing sialic acid esterification	2,4,6-Trihydroxyacetophenone (THAP)	Negative ion mode	[238]
Acidic glycans	Released *N*-glycans derivatised by Michael addition	4-Chloro-α-cyanocinnamic acid (Cl-CCA)	Negative ion mode	[189]
Neutral	Released *N*-glycans from mouse brain tissue	α-Cyano-4-hydroxycinnamic acid (CHCA)	Positive ion mode	[235, 239]

The spot is then dried, forming a crystal on a target plate surface, which is then irradiated by a pulsed laser (20-100 Hz). Energisation of matrix/analyte co-crystal then sees both matrix and analyte ejected from the spot into the gas phase (**Figure 1.16**). Here, charged matrix molecules transfer charged protons to the analyte, resulting in sample ionisation [240].

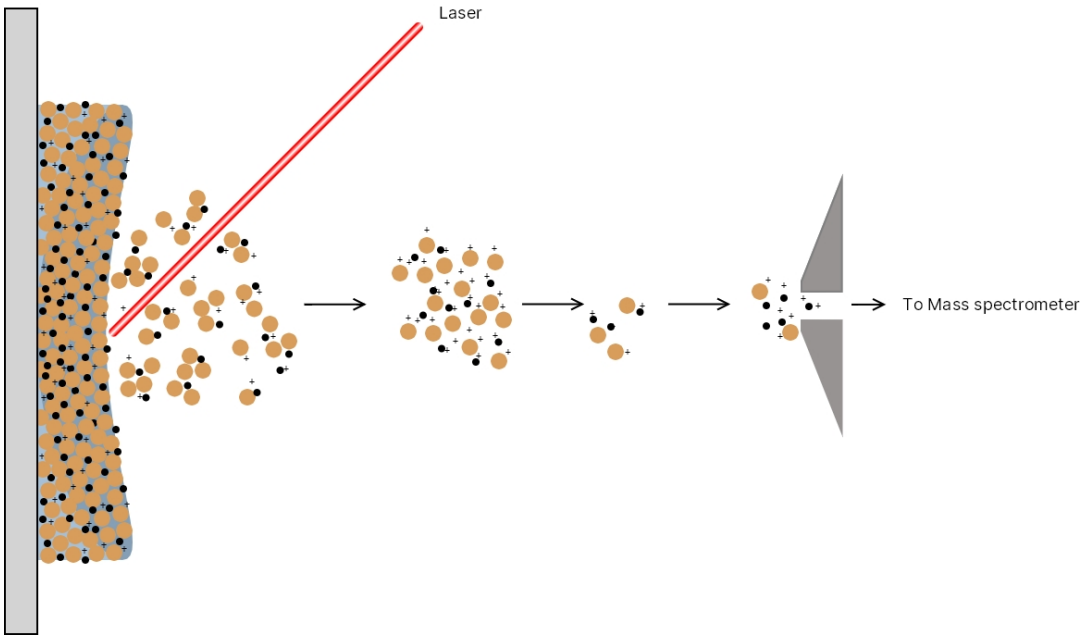


Figure 1.16 - MALDI matrix ionisation mechanism showing laser ionisation of matrix constituents resulting in charged analyte species entering the front of the MS.

MALDI ionisation techniques show benefit against ESI in being able to ionise high salt or contaminated samples with reduced suppression. However, matrix and crystallisation conditions can have a significant impact upon the accuracy of data acquired through the method. MALDI is most commonly attached to time of flight-based mass spectrometers, for the offline analysis of oligosaccharides and other large molecules [235].

1.3.6.3 Mass analysers

1.3.6.3.1 Time of flight

Time of flight (TOF) mass analysers separate ions based upon the time taken to travel down a field free flight tube. Following ionisation, ions are accelerated by an electric current before moving into a field free region or flight tube. Ions are then separated by m/z, with smaller compounds traveling faster than those with a larger m/z [241]. Ions displaying a smaller m/z therefore reach the detector first, allowing for differentiation between ion masses as shown in **Figure 1.17**. Modern TOF analysers typically seek to improve resolution by the use of a reflectron. This aims to refocus ions, reducing the likelihood of differences in kinetic energy between ions of the same m/z [242].

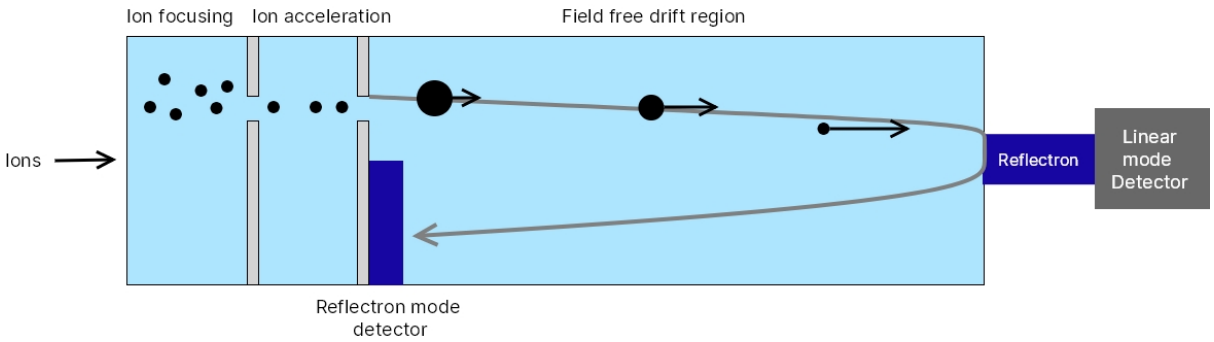


Figure 1.17 - Time of flight MS mass separation of charged analytes though a field free drift region. Heavier analytes travel slower through the drift region while smaller analytes travel faster reaching the detector first.

Time of flight mass analysis is widely used for the analysis of carbohydrates due to its ability to separate oligosaccharides over a wide range of intact masses. In addition, tandem MS/MS allows for fragmentation that can be analysed in the same spectrum without the requirement for multiple analysis windows [243, 244].

1.3.6.3.2 Quadrupole mass analysers.

Quadrupole MS consists of four parallel rods, each displaying an alternating radio frequency potential and direct current. The process outlined in **Figure 1.18** shows that opposing rods share the same polarisation, while those perpendicular display the inverse. Ions travel down the quadrupole in a corkscrew motion in response to oscillating polarisation of the rods. Quadrupoles serve as both a mass filter and separator, with the destination of ions being dependent upon two main factors: the radio frequency of opposing rods and the m/z of the analyte. Alteration of the electric field generated can, therefore, determine if analytes pass down the quadrupole or are removed before arriving at the detector [245]. In positive mode, ions displaying a positive charge will be held in the centre of the quadrupole by repulsion. However, oscillation polarisation will see ions attracted toward the negative regions of the quadrupole before the polarity changes. Ions with a smaller m/z will travel faster and therefore travel a longer distance to reach the detector. This also results in ions of m/z lower than the mass threshold colliding with the quadruple, filtering them from detection.

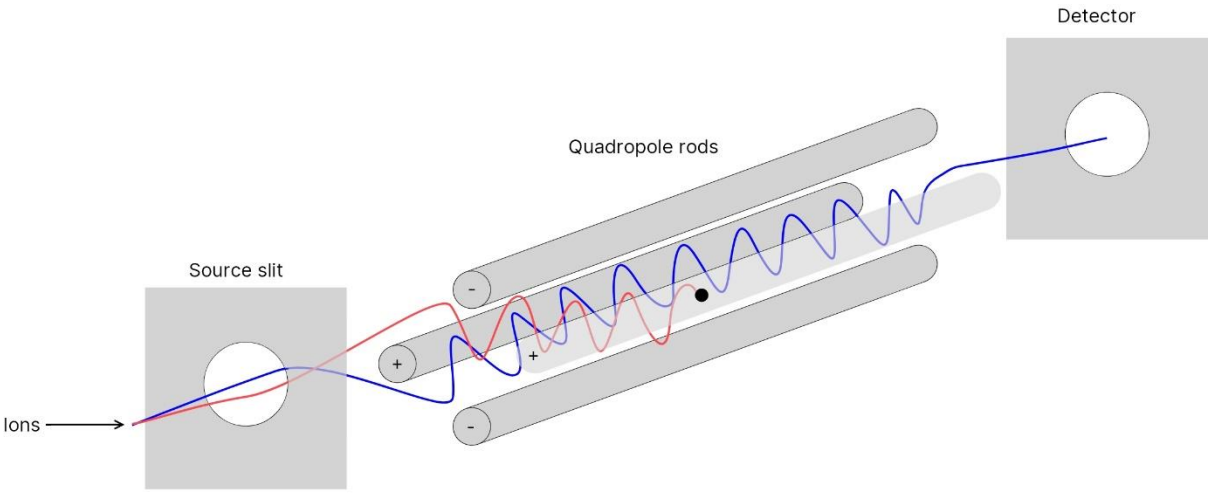


Figure 1.18 - Quadrupole MS mass separation. Charged analytes move though polarity switching magnets. Larger charged species move more slowly through the quadrupole while smaller species move more rapidly. Quadrupoles are also used as mass filters with excluded masses colliding with magnets and not making it to the detector.

Quadrupoles are seldom used as a single form of mass analysis and are routinely used in conjunction with other quadrupoles and the formation of a triple quadrupole, enabling the quantification by MS/MS [246]. These rely upon analysis of fragmentation of defined species occurring between mass analysers for both quantification and qualification. Additionally, they can be combined with time of flight spectrometers in an instrument known as a QTOF, where they aim to improve quantification of analytes before entering the TOF.

1.4 Aims

The research described in this thesis aimed to develop and implement tools for the analysis of glycans, and elucidation of carbohydrate structure and function. This was achieved by the development and synthesis of a suite of multifunctional glycan labels. These labels were designed to be capable of quantitatively derivatising released glycans, while maintaining a second region for covalent reactions. This multifunctional handle was designed in order to facilitate site-specific modification by bioorthogonal linking reactions, to form novel glycoconjugates. Following the generation of a suite of multifunctional labelling compounds, the analysis characteristics for each of the modified labels was evaluated.

HILIC-FLD analysis was used to compare retention time and labelling efficiency for both commercially available labels, as well as their modified derivatives. This was achieved through the use of model

carbohydrate standards such as dextran hydroxylate or maltoheptose (GU7), before verifying their activity with released biologically derived carbohydrate samples. This enabled the effectiveness of each of the modified multifunctional labelling compounds as glycan labels to be evaluated in comparison with existing pharmaceutical industry benchmarks.

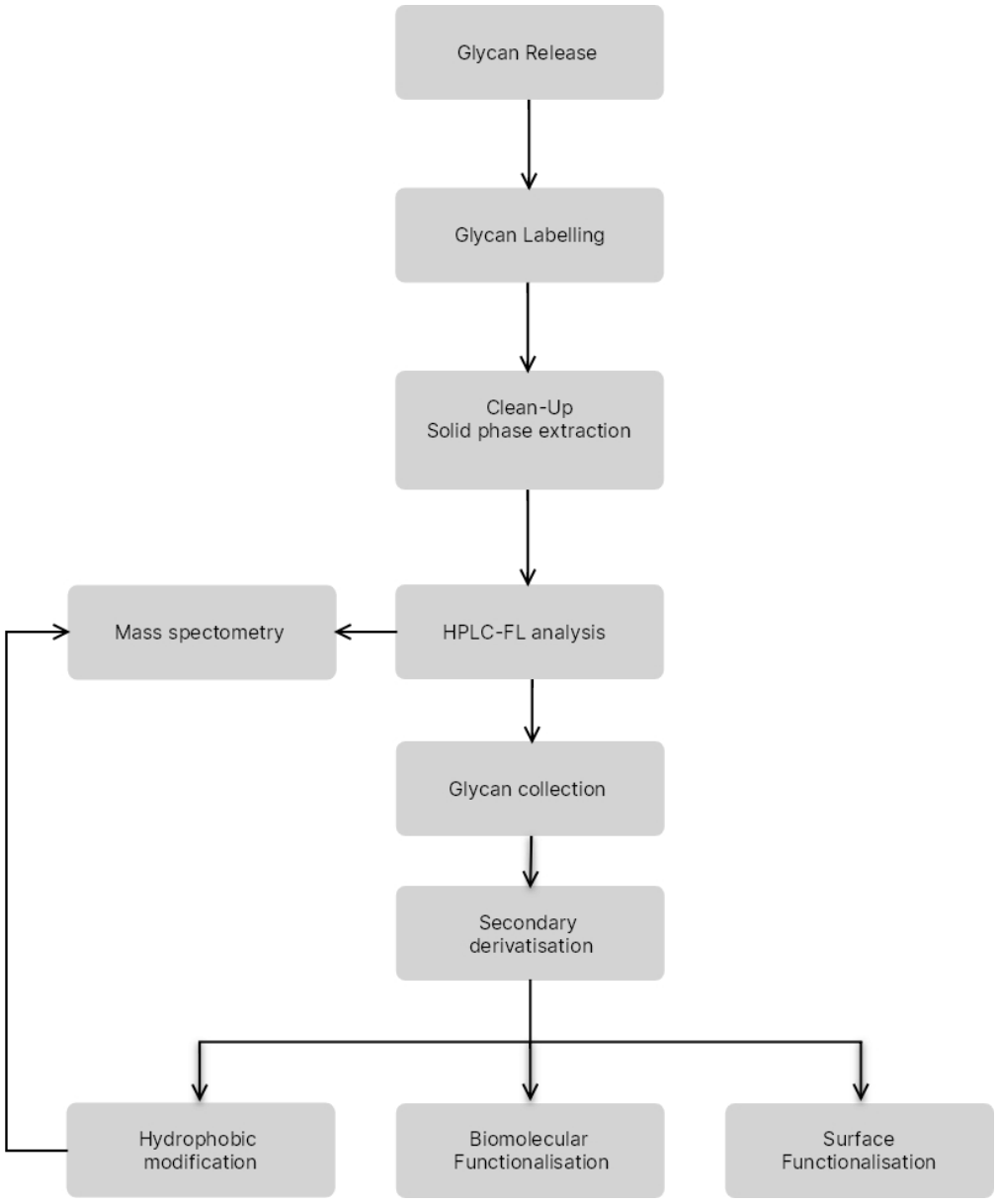


Figure 1.19 - Workflow for the application of a multifunctional glycan label. Glycans are initially labelled in line with existing pharmaceutical workflows before oligosaccharides of interest are collected and rederivatised to improve analysis characteristics or surface activation.

The ability of multifunctional labels to form glycoconjugates was then evaluated as a tool to improve ionisation efficiency in mass spectrometric analysis of low abundance glycans. This was achieved

initially by the collection of glycans of low abundance, in line with currently available commercial glycan profiling workflows, such as that outlined in **Figure 1.19** stage one. Collected glycans then underwent further covalent modification with the addition of a nonpolar or charged group, in order to investigate their effect on the relative abundance observed in analysis.

Finally, the investigation sought to determine the effect of glycosylation on protein structure and function. This was achieved by the site selective activation of collected carbohydrates to modified lysine residues present within the target protein. This modified glycoprotein conjugate was characterised by RP-LCMS before being investigated for activity and functionality.

1.5 References

1. C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, **15**, 346-366.
2. Y. Y. Zhao, M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume and N. Taniguchi, *Cancer. Sci.*, 2008, 99.
3. M. R. Wormald and R. A. Dwek, *Structure*, 1999, **7**, 155-160.
4. H. J. Tarbet, C. A. Toleman and M. Boyce, *Biochemistry*, 2018, **57**, 13-21.
5. J. Nilsson, A. Halim, A. Grahn and G. Larson, *Glycoconjugate J.*, 2013, **30**, 119-136.
6. F. Schwarz and M. Aebi, *Curr. Opin. Struct. Biol.*, 2011, **21**, 576-582.
7. A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill and T. Kinoshita, *Glycobiology*, 2015, **25**, 1323-1324.
8. C. Y. Liew, C.-C. Yen, J.-L. Chen, S.-T. Tsai, S. Pawar, C.-Y. Wu and C.-K. Ni, *Commun. Chem.*, 2021, **4**, 92.
9. Y.-M. She, R. Y. Tam, X. Li, M. Rosu-Myles and S. Sauvé, *Anal. Chem.*, 2020, **92**, 14038-14046.
10. T. J. Boltje, T. Buskas and G.-J. Boons, *Nat. Chem.*, 2009, **1**, 611-622.
11. H. S. Overkleeft and P. H. Seeberger, *Chemoenzymatic Synthesis of Glycans and Glycoconjugates*, 2015.
12. B. Imperiali and S. E. O'Connor, *Curr. Opin. Chem. Biol.*, 1999, **3**, 643-649.
13. R. Strasser, G. Seifert, M. S. Doblin, K. L. Johnson, C. Ruprecht, F. Pfrengle, A. Bacic and J. M. Estevez, *Front. Plant. Sci.*, 2021, **12**.
14. E. Bieberich, *Adv. Neurobiol.*, 2014, **9**, 47-70.
15. S. C. Pakhrin, K. F. Aoki-Kinoshita, D. Caragea and D. B. Kc, *Molecules*, 2021, **26**.
16. G. Taherzadeh, A. Dehzangi, M. Golchin, Y. Zhou and M. P. Campbell, *Bioinformatics*, 2019, **35**, 4140-4146.
17. J. S. Rush, *Lipid Insights*, 2015, **8s1**, 45-53.
18. R. G. Spiro, *Glycobiology* 2002, **12**, 43-56.
19. Y. Harada, Y. Ohkawa, Y. Kizuka and N. Taniguchi, *Int. J. Mol. Sci.,* 2019, **20**, 6074.
20. S. R. Hamilton, H. Li, H. Wischnewski, A. Prasad, J. S. Kerley-Hamilton, T. Mitchell, A. J. Walling, R. C. Davidson, S. Wildt and T. U. Gerngross, *Glycobiology*, 2005, **15**, 615-624.
21. A. Driouich, P. Gonnet, M. Makkie, A. C. Laine and L. Faye, *Planta*, 1989, **180**, 96-104.
22. Y. Kizuka and N. Taniguchi, *Biomolecules*, 2016, **6**.
23. A. Yadav, Q. Vagne, P. Sens, G. Iyengar and M. Rao, *eLife*, 2022, **11**.
24. Z. D'Souza, F. T. Sumya, A. Khakurel and V. Lupashin, *Cells*, 2021, **10**.
25. A. Frappaolo, A. Karimpour-Ghahnavieh, S. Sechi and M. G. Giansanti, *Cells*, 2020, **9**.
26. I. Nishikawa, Y. Nakajima, M. Ito, S. Fukuchi, K. Homma and K. Nishikawa, *Int. J. Mol. Sci.*, 2010, **11**, 4991-5008.
27. C. Robbe, C. Capon, B. Coddeville and J. C. Michalski, *Biochem. J.*, 2004, **384**, 307-316.
28. C. Schäffer, M. Graninger and P. Messner, J. *Proteomics.*, 2001, **1**, 248-261.
29. S. Wang, C. Chen, M. R. Gadi, V. Saikam, D. Liu, H. Zhu, R. Bollag, K. Liu, X. Chen, F. Wang, P. G. Wang, P. Ling, W. Guan and L. Li, *Nat. Commun.*, 2021, **12**, 3573.
30. I. Brockhausen, *EMBO. Rep.*, 2006, **7**, 599-604.
31. M. A. Hollingsworth and B. J. Swanson, *Nat. Rev. Cancer.*, 2004, **4**, 45-60.
32. D. J. Thornton, K. Rousseau and M. A. McGuckin, *Annu. Rev. Physiol.*, 2008, **70**, 459-486.
33. A. Guzman-Aranguez and P. Argüeso, *Ocul. Surf.*, 2010, **8**, 8-17.

34. P. Argüeso and I. K. Gipson, *Exp. Eye Res.*, 2001, **73**, 281-289.
35. N. Afratis, C. Gialeli, D. Nikitovic, T. Tsegenidis, E. Karousou, A. D. Theocharis, M. S. Pavão, G. N. Tzanakakis and N. K. Karamanos, *Febs. j.*, 2012, **279**, 1177-1197.
36. R. Raman, V. Sasisekharan and R. Sasisekharan, *Chem. Biol.*, 2005, **12**, 267-277.
37. K. Prydz, *Biomolecules*, 2015, **5**, 2003-2022.
38. G. Ghiselli, *Med. Res. Rev.*, 2017, **37**, 1051-1094.
39. R. C. Gupta, R. Lall, A. Srivastava and A. Sinha, *Front. Vet. Sci.*, 2019, **6**, 192.
40. P. Ghosh and D. Guidolin, *Semin. Arthritis. Rheum.*, 2002, **32**, 10-37.
41. A. G. Ogston and J. E. Stanier, *J. Physiol.*, 1953, **119**, 244-252.
42. V. H. Pomin and B. Mulloy, *Pharmaceuticals*, 2018, vol. 11.
43. R. Raman, S. Raguram, G. Venkataraman, J. C. Paulson and R. Sasisekharan, *Nat. Methods*, 2005, **2**, 817.
44. R. Sasisekharan and G. Venkataraman, *Curr. Opin. Chem. Biol.*, 2000, **4**, 626-631.
45. K. M. Arnold, S. J. Capuzzi, Y. Xu, E. N. Muratov, K. Carrick, A. Y. Szajek, A. Tropsha and J. Liu, *Pharmaceuticals*, 2017, **10**.
46. P. Yu, C. S. Pearson and H. M. Geller, *Trends Neurosci.*, 2018, **41**, 47-61.
47. E. Schaberg, U. Theocharidis, M. May, K. Lessmann, T. Schroeder and A. Faissner, *Front. Cell Dev. Biol.*, 2021, **9**.
48. T. Kinoshita, *J. Lipid Res.*, 2016, **57**, 4-5.
49. A. R. Borges, F. Link, M. Engstler and N. G. Jones, *Front. Cell Dev. Biol.*, 2021, **9**.
50. K. Bellai-Dussault, T. T. M. Nguyen, N. V. Baratang, D. A. Jimenez-Cruz and P. M. Campeau, *Clin. Genet.*, 2019, **95**, 112-121.
51. J. Jouhet, E. Maréchal and M. A. Block, *Prog. Lipid Res.*, 2007, **46**, 37-55.
52. J. Kopitz, in *Glycoscience*, 1996, pp. 163-189.
53. P. Singh, Y. D. Paila and A. Chattopadhyay, *J. Neurochem.*, 2012, **123**, 716-724.
54. S. Lahiri and A. H. Futerman, *Cell. Mol. Life Sci.*, 2007, **64**, 2270-2284.
55. M. Bektas and S. Spiegel, *Glycoconjugate J.*, 2004, **20**, 39-47.
56. A. Carruthers and E. M. Carey, *J. Neurochem.,* 1983, **41**, 22-29.
57. R. K. Yu, Y. Nakatani and M. Yanagisawa, *J. Lipid Res.*, 2009, **50**, 440-S445.
58. C. Wang, X. He, B. Zhou, J. Li, B. Li, W. Qian, S. Hou, H. Wang, Y. Shi and Y. Guo, *MAbs*, 2011, **3**, 67-75.
59. J. Y. Kim, Y.-G. Kim and G. M. Lee, *Appl. Microbiol. and Biot.*, 2012, **93**, 917-930.
60. M. Hasani, R. Golhosseini, S. M. Azimi, M. Ardjmand, H. Mahravani and S. S. Parizi, *J. Med. Life.*, 2020, **13**, 241-248.
61. J. H. Nam, F. Zhang, M. Ermonval, R. J. Linhardt and S. T. Sharfstein, *Biotechnol. Bioeng.*, 2008, **100**, 1178-1192.
62. M. Spearman, J. Rodriguez, N. Huzel, K. Sunley and M. Butler, *Cell Technology for Cell Products*, 2007, 71-75.
63. A. Delobel, *Mass Spectrometry of Glycoproteins: Methods and Protocols*, 2021, pp. 1-21.
64. M.-E. Lalonde and Y. Durocher, *J. Biotechnol.*, 2017, **251**, 128-140.
65. H. Dahodwala and K. H. Lee, *Curr. Opin. Biotech.*, 2019, **60**, 128-137.
66. P. Hossler, S. F. Khattak and Z. J. Li, *Glycobiology* 2009, **19**, 936-949.
67. C. Schäffer and P. Messner, *FEMS Microbiol. Rev.*, 2017, **41**, 49-91.
68. L. D. Bennett, Q. Yang, B. R. Berquist, J. P. Giddens, Z. Ren, V. Kommineni, R. P. Murray, E. L. White, B. R. Holtz, L.-X. Wang and S. Marcel, *Int. J. Mol. Sci.*, 2018, **19**.

69. D. Ghaderi, M. Zhang, N. Hurtado-Ziola and A. Varki, *Biotechnol. Genet. Eng.*, 2012, **28**, 147-176.
70. C. H. Hokke, A. A. Bergwerff, G. W. K. van Dedem, J. van Oostrum, J. P. Kamerling and J. F. G. Vliegenthart, *FEBS lett.*, 1990, **275**, 9-14.
71. D. Ghaderi, R. E. Taylor, V. Padler-Karavani, S. Diaz and A. Varki, *Nat. Biotechnol.*, 2010, **28**, 863-867.
72. V. Padler-Karavani, H. Yu, H. Cao, H. Chokhawala, F. Karp, N. Varki, X. Chen and A. Varki, *Glycobiology*, 2008, **18**, 818-830.
73. S. Yehuda and V. Padler-Karavani, *Front. Immunol.*, 2020, **11**, 21.
74. R. Rao and D. Verthelyi, in *The Science and Regulations of Naturally Derived Complex Drugs*, Springer, 2019, pp. 219-244.
75. C. J. Bosques, B. E. Collins, J. W. Meador, H. Sarvaiya, J. L. Murphy, G. DelloRusso, D. A. Bulik, I. H. Hsu, N. Washburn, S. F. Sipsey, J. R. Myette, R. Raman, Z. Shriver, R. Sasisekharan and G. Venkataraman, *Nat. Biotechnol.*, 2010, **28**, 1153-1156.
76. C. H. Chung, B. Mirakhur, E. Chan, Q.-T. Le, J. Berlin, M. Morse, B. A. Murphy, S. M. Satinover, J. Hosen, D. Mauro, R. J. Slebos, Q. Zhou, D. Gold, T. Hatley, D. J. Hicklin and T. A. E. Platts-Mills, *New Engl. J. Med.*, 2008, **358**, 1109-1117.
77. A. Mangla and N. Agarwal, *J. Oncol. Pract.*, 2019, **15**, 679-680.
78. D. Ayoub, W. Jabs, A. Resemann, W. Evers, C. Evans, L. Main, C. Baessmann, E. Wagner-Rousset, D. Suckau and A. Beck, *MAbs*, 2013, **5**, 699-710.
79. J. B. Goh and S. K. Ng, *Crit. Rev. Biotechnol.*, 2018, **38**, 851-867.
80. B. D. Weintraub, B. S. Stannard and L. Meyers, *Endocrinology*, 1983, **112**, 1331-1345.
81. M. Baudys, T. Uchio, D. Mix, D. Wilson and S. W. Kim, *J. Pharm. Sci.*, 1995, **84**, 28-33.
82. E. Uchida, K. Morimoto, N. Kawasaki, Y. Izaki, A. Abdu Said and T. Hayakawa, *Free Radical Res.*, 1997, **27**, 311-323.
83. E. Tsuda, G. Kawanishi, M. Ueda, S. Masuda and R. Sasaki, *Eur. J. Biochemistry*, 1990, **188**, 405-411.
84. R. J. Solá and K. Griebenow, *BioDrugs*, 2010, **24**, 9-21.
85. P. Zhang, S. Woen, T. Wang, B. Liau, S. Zhao, C. Chen, Y. Yang, Z. Song, M. R. Wormald, C. Yu and P. M. Rudd, *Drug Discov. Today*, 2016, **21**, 740-765.
86. M. Mital and B. McCormack, *Cell Mol. Life Sci.*, 2000, **57**, 1964-1969.
87. M. J. Koury, *Trends Biotechnol.*, 2003, **21**, 462-464.
88. J. W. Baynes and F. Wold, *J. Biol. Chem.*, 1976, **251**, 6016-6024.
89. G. Ashwell and A. Morell, *Biochem. Soc. Symp.*, 1974, 117-124.
90. A. G. Morell, R. A. Irvine, I. Sternlieb, I. H. Scheinberg and G. Ashwell, *J. Biol. Chem.*, 1968, **243**, 155-159.
91. V. Gross, P. C. Heinrich, D. vom Berg, K. Steube, T. Andus, T. A. Tran-Thi, K. Decker and W. Gerok, *Eur. J. Biochem.*, 1988, **173**, 653-659.
92. M. Melmer, T. Stangler, M. Schiefermeier, W. Brunner, H. Toll, A. Rupprechter, W. Lindner and A. Premstaller, *Anal. Bioanal. Chem.,* 2010, **398**, 905-914.
93. Z. Segu, T. Stone, C. Berdugo, A. Roberts, E. Doud and Y. Li, *MAbs*, 2020, **12**, 1750794.
94. D. A. Fischler and R. Orlando, *J. Biomol. Tech.*, 2019, **30**, 58-63.
95. R. I. G. Morrison, *Connective Tissues: Biochemistry and Pathophysiology*, eds. R. Fricke and F. Hartmann, Springer Berlin Heidelberg, Berlin, Heidelberg, 1974, pp. 150-157.
96. K. Shimada and T. Ozawa, *Thromb. Res.*, 1985, **39**, 387-397.

97. M. Nakano, S. Sabido-Bozo, K. Okazaki, A. Aguilera-Romero, S. Rodriguez-Gallardo, A. Cortes-Gomez, S. Lopez, A. Ikeda, K. Funato and M. Muñiz, *Plos One*, 2021, **16**, e0257435.
98. S. Albrecht, S. Vainauskas, H. Stöckmann, C. McManus, C. H. Taron and P. M. Rudd, *Anal. Chem.*, 2016, **88**, 4795-4802.
99. H. Wilkinson and R. Saldova, *J. Proteome. Res.*, 2020, **19**, 3890-3905.
100. T. H. Plummer, Jr., J. H. Elder, S. Alexander, A. W. Phelan and A. L. Tarentino, *J. Biol. Chem.*, 1984, **259**, 10700-10704.
101. N. Takahashi, *Biochem. Bioph. Res. Co.*, 1977, **76**, 1194-1201.
102. V. Tretter, F. Altmann and L. MÄRz, *Eur. J. Biochem.*, 1991, **199**, 647-652.
103. Y. Huang and R. Orlando, *J. Biomol. Tech.*, 2017, **28**, 150-157.
104. S. Yan, J. Vanbeselaere, C. Jin, M. Blaukopf, F. Wöls, I. B. H. Wilson and K. Paschinger, *Anal. Chem.*, 2018, **90**, 928-935.
105. R. B. Trimble and A. L. Tarentino, *J. Biol. Chem.*, 1991, **266**, 1646-1651.
106. H. H. Freeze and C. Kranz, *Curr. Protoc. Immunol.*, 2008, **83**, 8.15.11-18.15.26.
107. H. H. Freeze and C. Kranz, *Curr. Protoc. Mol. Biol.*, 2010, **Chapter 17**, Unit 17.13A.
108. D. Koutsioulis, D. Landry and E. P. Guthrie, *Glycobiology*, 2008, **18**, 799-805.
109. Z. Szabo, A. Guttman and B. L. Karger, *Anal. Chem.*, 2010, **82**, 2588-2593.
110. L. Cao, J. K. Diedrich, Y. Ma, N. Wang, M. Pauthner, S.-K. R. Park, C. M. Delahunty, J. S. McLellan, D. R. Burton, J. R. Yates and J. C. Paulson, *Nat. Protoc.*, 2018, **13**, 1196-1212.
111. A. Bunyatratchata, A. L. Parc, J. M. L. N. de Moura Bell, J. L. Cohen, H. Duman, A. Arslan, M. Kaplan, D. Barile and S. Karav, *Enzyme Microb. Tech.*, 2023, **162**, 110138.
112. X. Song, H. Ju, Y. Lasanajak, M. R. Kudelka, D. F. Smith and R. D. Cummings, *Nat. Methods.*, 2016, **13**, 528-534.
113. M. Fukuda and A. Kobata, *Glycobiology: A Practical Approach* Oxford university press, Oxford, 1994.
114. P. Påhlsson, D. P. Blackall, M. Ugorski, M. Czerwinski and S. L. Spitalnik, *Glycoconjugate. J.*, 1994, **11**, 43-50.
115. S. Saito, S. B. Levery, M. E. Salyan, R. I. Goldberg and S. Hakomori, *J. Biol. Chem.*, 1994, **269**, 5644-5652.
116. R. G. Spiro and V. D. Bhoyroo, *J Biol Chem*, 1974, **249**, 5704-5717.
117. D. B. Diep, K. L. Nelson, S. M. Raja, E. N. Pleshak and J. T. Buckley, *J. Biol. Chem.*, 1998, **273**, 2355-2360.
118. M. Nakano, S. Sabido-Bozo, K. Okazaki, A. Aguilera-Romero, S. A.-O. Rodriguez-Gallardo, A. Cortes-Gomez, S. Lopez, A. Ikeda, K. Funato and M. A.-O. Muñiz, *Plos One*, 2021, **16**.
119. M. Fukuda, Glycoproteins and Glycopeptides: *Curr. Protoc. Mol. Biol.*, 2001, vol 31, ch 17, 15.4-15.9.
120. Y. Huang, Y. Mechref and M. V. Novotny, *Anal. Chem.*, 2001, **73**, 6063-6069.
121. Y. Miura, K. Kato, Y. Takegawa, M. Kurogochi, J.-i. Furukawa, Y. Shinohara, N. Nagahori, M. Amano, H. Hinou and S. I. Nishimura, *Anal. Chem.*, 2010, **82**, 10021-10029.
122. Y. Goso, T. Sugaya, K. Ishihara and M. Kurihara, *Anal. Chem.*, 2017, **89**, 8870-8876.
123. J. Zaia, *Omics.*, 2010, **14**, 401-418.
124. G. Yu, Y. Zhang, Z. Zhang, L. Song, P. Wang and W. Chai, *Anal. Chem.*, 2010, **82**, 9534-9542.
125. T. Patel, J. Bruce, A. Merry, C. Bigge, M. Wormald, R. Parekh and A. Jaques, *Biochemistry*, 1993, **32**, 679-693.

126. A. Kameyama, S. K. Dissanayake and W. W. Thet Tin, *Plos One*, 2018, **13**, e0196800-e0196800.
127. I. Turyan, X. Hronowski, Z. Sosic and Y. Lyubarskaya, *Anal. Biochem.*, 2014, **446**, 28-36.
128. K. R. Anumula, *Anal. Biochem.*, 2008, **373**, 104-111.
129. K. Xiao, Y. Han, H. Yang, H. Lu and Z. Tian, *Anal. Chim. Acta*, 2019, **1091**, 1-22.
130. I. Ciucanu and F. Kerek, *Carbohyd. Res.*, 1984, **131**, 209-217.
131. I. Ciucanu and C. E. Costello, *J. Am. Chem. Soc.*, 2003, **125**, 16213-16219.
132. P. Kang, Y. Mechref, I. Klouckova and M. V. Novotny, *Rapid Commun. Mass Sp.*, 2005, **19**, 3421-3428.
133. S. Zhou, K. M. Wooding and Y. Mechref, *Methods Mol. Biol.*, 2017, **1503**, 83-96.
134. Y. Hu and Y. Mechref, *Electrophoresis*, 2012, **33**, 1768-1777.
135. J. You, X. Sheng, C. Ding, Z. Sun, Y. Suo, H. Wang and Y. Li, *Anal. Chim. Acta*, 2008, **609**, 66-75.
136. G. Zauner, A. M. Deelder and M. Wuhrer, *Electrophoresis*, 2011, **32**, 3456-3466.
137. Y. Shinohara, H. Sota, M. Gotoh, M. Hasebe, M. Tosu, J. Nakao, Y. Hasegawa and M. Shiga, *Anal. Chem.*, 1996, **68**, 2573-2579.
138. E. Lattova and H. Perreault, *J. Chromatogr. B*, 2003, **793**, 167-179.
139. C. Leteux, R. A. Childs, W. Chai, M. S. Stoll, H. Kogelberg and T. Feizi, *Glycobiology*, 1998, **8**, 227-236.
140. P. Kapkova, *Rapid Commun. Mass Sp.*, 2009, **23**, 2775-2784.
141. S. Vainauskas, C. Kirk, L. Petralia, E. Guthrie, E. McLeod, A. Bielik, A. Luebbers, J. Foster, C. Hokke, P. Rudd, X. Shi and C. Taron, *Protein Expres. Purif.*, 2018, **8**.
142. S. Zhou, L. Veillon, X. Dong, Y. Huang and Y. Mechref, *Analyst*, 2017, **142**, 4446-4455.
143. T. Keser, T. Pavić, G. Lauc and O. Gornik, *Front. Chem.*, 2018, **6**, 324-324.
144. J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Anal. Biochem.*, 1995, **230**, 229-238.
145. L. R. Ruhaak, E. Steenvoorden, C. A. Koeleman, A. M. Deelder and M. Wuhrer, *Proteomics*, 2010, **10**, 2330-2336.
146. D. S. Dalpathado, H. Jiang, M. A. Kater and H. Desaire, *Anal. Bioanal. Chem*, 2005, **381**, 1130-1137.
147. D. Locke, C. G. Bevans, L.-X. Wang, Y. Zhang, A. L. Harris and Y. C. Lee, *Carbohyd. Res.*, 2004, **339**, 221-231.
148. J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Anal. Biochem.*, 1995, **230**, 229-238.
149. K. R. Anumula, *Anal. Biochem.*, 1994, **220**, 275-283.
150. Y. Kishimoto, F. Okada, T. Maesako, S. Yamamoto, M. Kinoshita, T. Hayakawa and S. Suzuki, *J. Chromatogr. A*, 2020, **1625**, 461194.
151. J. Lemoine, M. Cabanes-Macheteau, M. Bardor, J. C. Michalski, L. Faye and P. Lerouge, *Rapid Commun. Mass Sp.*, 2000, **14**, 100-104.
152. H. J. Danyluk, L. K. Shum and W. F. Zandberg, *Methods Mol. Biol.*, 2017, **1588**, 223-236.
153. Y. Lang, Y. Zhang, C. Wang, L. Huang, X. Liu, N. Song, G. Li and G. Yu, *Front. Chem.*, 2021, **9**.
154. M. Pabst, D. Kolarich, G. Pöltl, T. Dalik, G. Lubec, A. Hofinger and F. Altmann, *Anal. Biochem.*, 2009, **384**, 263-273.
155. R. P. Kozak, C. B. Tortosa, D. L. Fernandes and D. I. R. Spencer, *Anal. Biochem.*, 2015, **486**, 38-40.
156. D. Schmid, B. Behnke, J. Metzger and R. Kuhn, *Biomed. Chromatogr.*, 2002, **16**, 151-156.

157. H. Zhang, Z. Wang, J. Stupak, O. Ghribi, J. D. Geiger, Q. Y. Liu and J. Li, *Proteomics*, 2012, **12**, 2510-2522.
158. J. A. Saba, X. Shen, J. C. Jamieson and H. Perreault, *Rapid Commun. Mass Sp.*, 1999, **13**, 704-711.
159. G. Zauner, C. A. M. Koeleman, A. M. Deelder and M. Wuhrer, *B. B. A. Gen. Subjects*, 2012, **1820**, 1420-1428.
160. E. Lattová, J. Skřičková and Z. Zdráhal, *Anal. Chem.*, 2019, **91**, 7985-7990.
161. R. Duke and C. Taron, *Biopharm. Int.*, 2015, **28**, 59-64.
162. W. Morelle and J.-C. Michalski, *Nat. Protoc.*, 2007, **2**, 1585-1602.
163. L. Royle, M. P. Campbell, C. M. Radcliffe, D. M. White, D. J. Harvey, J. L. Abrahams, Y.-G. Kim, G. W. Henry, N. A. Shadick, M. E. Weinblatt, D. M. Lee, P. M. Rudd and R. A. Dwek, *Anal. Biochem.*, 2008, **376**, 1-12.
164. H. Nakagawa, M. Hato, Y. Takegawa, K. Deguchi, H. Ito, M. Takahata, N. Iwasaki, A. Minami and S. Nishimura, *J. Chromatogr. B*, 2007, **853**, 133-137.
165. L. Royle, T. S. Mattu, E. Hart, J. I. Langridge, A. H. Merry, N. Murphy, D. J. Harvey, R. A. Dwek and P. M. Rudd, *Anal. Biochem.*, 2002, **304**, 70-90.
166. Y. Guan, M. Zhang, J. Wang and H. Schlüter, *J. Proteome Res.*, 2021, **20**, 2914-2922.
167. K. R. Anumula and S. T. Dhume, *Glycobiol*, 1998, **8**, 685-694.
168. B. D. Prater, K. R. Anumula and J. T. Hutchins, *Anal. Biochem.*, 2007, **369**, 202-209.
169. L. R. Ruhaak, C. Huhn, W.-J. Waterreus, A. R. de Boer, C. Neusüss, C. H. Hokke, A. M. Deelder and M. Wuhrer, *Anal. Chem.*, 2008, **80**, 6119-6126.
170. C. Manz, M. Götze, C. Frank, A. Zappe and K. Pagel, *Anal. Bioanal. Chem.,* 2022, **414**, 5023-5031.
171. A. Gyorgypal and S. P. S. Chundawat, *Anal. Chem.,* 2022, **94**, 6986-6995.
172. M. Lei, Y. Mechref and M. V. Novotny, *J. Am. Soc. Mass Spectr.*, 2009, **20**, 1660-1671.
173. J. Schwarzer, E. Rapp and U. Reichl, *Electrophoresis*, 2008, **29**, 4203-4214.
174. Y. Kita, Y. Miura, J.-i. Furukawa, M. Nakano, Y. Shinohara, M. Ohno, A. Takimoto and S.-I. Nishimura, *Mol. Cell Proteomics*, 2007, **6**, 1437-1445.
175. J. A. Saba, X. Shen, J. C. Jamieson and H. Perreault, *J. Mass Spectrom.*, 2001, **36**, 563-574.
176. X. Yang, Y. Zhao, Y. Ruan and Y. Yang, *Biol. Pharm. Bull.*, 2008, **31**, 1860-1865.
177. M. Guillard, J. Gloerich, H. J. C. T. Wessels, E. Morava, R. A. Wevers and D. J. Lefeber, *Carbohyd. Res.*, 2009, **344**, 1550-1557.
178. J. C. Botelho, J. A. Atwood Iii, L. Cheng, G. Alvarez-Manilla, W. S. York and R. Orlando, *Int. J. Mass Spectrom.*, 2008, **278**, 137-142.
179. B. Xia, C. L. Feasley, G. P. Sachdev, D. F. Smith and R. D. Cummings, *Anal. Biochem.*, 2009, **387**, 162-170.
180. A. H. Merry, D. C. A. Neville, L. Royle, B. Matthews, D. J. Harvey, R. A. Dwek and P. M. Rudd, *Anal. Biochem.*, 2002, **304**, 91-99.
181. S. Yang and H. Zhang, *Proteomics Clin. Appl.*, 2012, **6**, 596-608.
182. J. Bones, S. Mittermayr, N. O'Donoghue, A. s. Guttman and P. M. Rudd, *Anal. Chem.*, 2010, **82**, 10208-10215.
183. Q. Zhang, H. Li, X. Feng, B.-F. Liu and X. Liu, *Plos One*, 2014, **9**, e94232.
184. N. H. Packer, M. A. Lawson, D. R. Jardine and J. W. Redmond, *Glycoconjugate J.*, 1998, **15**, 737-747.
185. P. H. Jensen, N. G. Karlsson, D. Kolarich and N. H. Packer, *Nat. Protoc.*, 2012, **7**, 1299-1310.

186. K. Stavenhagen, D. Kolarich and M. Wuhrer, *Chromatographia*, 2015, vol. 78, pp. 307-320.
187. W. Morelle, V. Faid, F. Chirat and J.-C. Michalski, in *Glycomics: Methods and Protocols*, eds. N. H. Packer and N. G. Karlsson, Humana Press, Totowa, NJ, 2009, pp. 3-21.
188. Y. Ohta, K. Kameda, M. Matsumoto and N. Kawasaki, *Mass Spectrom.*, 2017, **6**, A0061.
189. M. H. Selman, M. Hemayatkar, A. M. Deelder and M. Wuhrer, *Anal. Chem.*, 2011, **83**, 2492-2499.
190. P. Gagneux, V. Panin, T. Hennet, M. Aebi and A. Varki, *Essentials of Glycobiology. 4th edition*, 2022.
191. S. Kamoda, R. Ishikawa and K. Kakehi, *J. Chromatogr. A*, 2006, **1133**, 332-339.
192. Y. Zhu, X. Liu, Y. Zhang, Z. Wang, Y. Lasanajak and X. Song, *Bioconjugate Chem.*, 2018, **29**, 3847-3855.
193. K. Naresh, F. Schumacher, H. S. Hahm and P. H. Seeberger, *Chem. Commun.*, 2017, **53**, 9085-9088.
194. G. R. Guile, P. M. Rudd, D. R. Wing, S. B. Prime and R. A. Dwek, *Anal. Biochem.*, 1996, **240**, 210-226.
195. G. C. Vreeker and M. Wuhrer, *Anal. Bioanal. Chem.*, 2017, **409**, 359-378.
196. J. Delaney and P. Vouros, *Rapid Commun. Mass Sp.*, 2001, **15**, 325-334.
197. X. Dong, S. Zhou and Y. Mechref, *Electrophoresis*, 2016, **37**, 1532-1548.
198. J. G. Wilhelm, M. Dehling and F. Higel, *Anal. Bioanal. Chem.*, 2019, **411**, 735-743.
199. Y. Yoshinaka, Y. Ueda and S. Suzuki, *J. Chromatogr. A*, 2007, **1143**, 83-87.
200. T. E. Bapiro, F. M. Richards and D. I. Jodrell, *Anal. Chem.*, 2016, **88**, 6190-6194.
201. L. Monser and F. Darghouth, *J.Pharm. Biomed. Anal.*, 2000, **23**, 353-362.
202. M. Melmer, T. Stangler, A. Premstaller and W. Lindner, *J. Chromatogr. A*, 2011, **1218**, 118-123.
203. M. Pabst and F. Altmann, *Anal. Chem.*, 2008, **80**, 7534-7542.
204. R. P. Estrella, J. M. Whitelock, N. H. Packer and N. G. Karlsson, *Anal. Chem.*, 2007, **79**, 3597-3606.
205. C. Young, M. R. Condina, M. T. Briggs, E. S. X. Moh, G. Kaur, M. K. Oehler and P. Hoffmann, *Front. Chem.*, 2021, **9**.
206. M. Pabst, S. Q. Wu, J. Grass, A. Kolb, C. Chiari, H. Viernstein, F. M. Unger, F. Altmann and S. Toegel, *Carbohyd. Res*, 2010, **345**, 1389-1393.
207. H. Ito, K. Yamada, K. Deguchi, H. Nakagawa and S. I. Nishimura, *Rapid Commun. Mass Sp.*, 2007, **21**, 212-218.
208. N. Kawasaki, M. Ohta, S. Hyuga, O. Hashimoto and T. Hayakawa, *Anal. Biochem.*, 1999, **269**, 297-303.
209. B. Buszewski and S. Noga, *Anal. Bioanal. Chem.*, 2012, **402**, 231-247.
210. P. Kozlik, R. Goldman and M. Sanda, *Anal. Bioanal. Chem.*, 2018, **410**, 5001-5008.
211. B. L. Duivelshof, S. Denorme, K. Sandra, X. Liu, A. Beck, M. A. Lauber, D. Guillarme and V. D'Atri, *Pharmaceutics*, 2021, **13**.
212. Y. Huang, Y. Nie, B. Boyes and R. Orlando, *J. Biomol. Tech.*, 2016, **27**, 98-104.
213. C. Boone and J. Adamec, in *Proteomic Profiling and Analytical Chemistry (Second Edition)*, eds. P. Ciborowski and J. Silberring, Elsevier, Boston, 2016, pp. 175-191.
214. D. L. Chance and T. P. Mawhinney, *Carbohyd. Res*, 1996, **295**, 157-177.
215. M. Lei, M. V. Novotny and Y. Mechref, *J. Am. Soc. Mass Spectr.*, 2010, **21**, 348-357.
216. C. Corradini, A. Cavazza and C. Bignardi, *Int. J. Carbohydr. Chem.*, 2012, **2012**.

217. T. J. Paskach, H.-P. Lieker, P. J. Reilly and K. Thielecke, *Carbohyd. Res.*, 1991, **215**, 1-14.
218. C. Corradini, A. Cavazza and C. Bignardi, *Int. J. Carbphydr. Chem.*, 2012, **2012**, 487564.
219. S. Schneider, E. Naegele and S. Krieger, *Methods Mol. Biol.*, 2021, **2271**, 221-235.
220. K. Deguchi, T. Keira, K. Yamada, H. Ito, Y. Takegawa, H. Nakagawa and S.-I. Nishimura, *J. Chromatogr. A*, 2008, **1189**, 169-174.
221. J. Bones, N. McLoughlin, M. Hilliard, K. Wynne, B. L. Karger and P. M. Rudd, *Anal. Chem.*, 2011, **83**, 4154-4162.
222. L. A. Holland, N. P. Chetwyn, M. D. Perkins and S. M. Lunte, *Pharm. Res.*, 1997, **14**, 372-387.
223. G. Lu, C. L. Crihfield, S. Gattu, L. M. Veltri and L. A. Holland, *Chem. Rev.*, 2018, **118**, 7867-7885.
224. K. D. Lukacs, *Theory, instrumentation, and applications of capillary zone electrophoresis*, The University of North Carolina at Chapel Hill, 1983.
225. D. Reusch, M. Haberger, T. Kailich, A.-K. Heidenreich, M. Kampe, P. Bulau and M. Wuhrer.
226. M. Swartz, *J. Liq. Chromatogr. R. T.*, 2010, **33**, 1130-1150.
227. L. Antonov and D. Nedeltcheva, *Chem. Soc. Rev.*, 2000, **29**, 217-227.
228. J. D. Stuart, *J. Am. Chem. Soc.*, 2005, **127**, 9310-9310.
229. J. W. Daily, *Prog. Energ. Combust.,* 1997, **23**, 133-199.
230. Á. Szekrényes, S. S. Park, M. Santos, C. Lew, A. Jones, T. Haxo, M. Kimzey, S. Pourkaveh, Z. Szabó and Z. Sosic, *MAbs*, 2016, **8**, 56-64.
231. M. Hamm, Y. Wang and R. R. Rustandi, *Pharmaceuticals*, 2013, **6**, 393-406.
232. L. Han and C. E. Costello, *Biochemistry*, 2013, **78**, 710-720.
233. L. Konermann, E. Ahadi, A. D. Rodriguez and S. Vahidi, *Anal. Chem.*, 2013, **85**, 2-9.
234. S. Banerjee and S. Mazumdar, *Int. J. Anal. Chem.*, 2012, **2012**, 282574.
235. R. R. Drake, T. W. Powers, E. E. Jones, E. Bruner, A. S. Mehta and P. M. Angel, *Adv. Cancer Res.*, 2017, **134**, 85-116.
236. C. Jurinke, P. Oeth and D. van den Boom, *Mol. Biotechnol.*, 2004, **26**, 147-163.
237. Y. Zhang, B. Wang, W. Jin, Y. Wen, L. Nan, M. Yang, R. Liu, Y. Zhu, C. Wang, L. Huang, X. Song and Z. Wang, *Anal. Chim. Acta*, 2019, **1048**, 105-114.
238. Y. Peng and X. Xu, *J. Nat. Sci.*, 2014, **19**, 245-252.
239. E. Stephens, S. L. Maslen, L. G. Green and D. H. Williams, *Anal. Chem.*, 2004, **76**, 2343-2354.
240. M. Karas and R. Krüger, *Chem. Rev.*, 2003, **103**, 427-440.
241. B. A. Mamyrin, V. I. Karataev, D. V. Shmikk and V. A. Zagulin, *J. Exp. Theor. Phys.*, 1973, **37**, 45.
242. T. Handa, T. Horio, M. Arakawa and A. Terasaki, *Int. J. Mass Spectrom.*, 2020, **451**, 116311.
243. R. A. Gomes, C. Almeida, C. Correia, A. Guerreiro, A. L. Simplício, I. A. Abreu and P. G. Alves, *Plos One*, 2019, **14**, e0219156.
244. M. Baerenfaenger and B. Meyer, *J. Proteome Res.*, 2018, **17**, 3693-3703.
245. P. E. Miller and M. B. Denton, *J. Chem. Educ.*, 1986, **63**, 617.
246. R. A. Yost and C. G. Enke, *Anal. Chem.*, 1979, **51**, 1251-1264.

Chapter 2- Synthesis of a suite of multifunctional glycan labels

2.1 Introduction

Analytical methodologies for both the quantitative and qualitative profiling [1,2] of released glycan samples are well established procedures. In modern glycan analysis workflows, a derivatisation agent is routinely employed to impart beneficial analytical characteristics to the target oligosaccharide. In most applications the use of derivatisation is required due to the low concentration of glycan material released from biologically derived samples, combined with the lack of fluorescence and high polarity exhibited by oligosaccharide compounds. The workflows outlined in the previous chapter represent the most effective and commercially scalable methods of glycan derivatisation, however aspects such as labelling efficiency and clean-up recovery can have a significant impact upon the sample concentration and observed glycan structures present in analysis by Fluorescent Light Detection (FLD) or Mass Spectrometry (MS) [3].

The diverse structural characteristics displayed between glycan labels can make specific derivatisation workflows more suited to a particular application than others [4]. The choice of glycan derivatisation can therefore have a significant bearing upon the sensitivity and reliability of an analysis method. The use of labels such as 2-anthranilic acid (2AA) **10** or 2-aminobenzamide (2AB) **9** give rise to a change in optical properties by the addition of a fluorescent label. Other methods such as per-*O*-methylation lead to multisite modification aiming to reduce glycan polarity and increase ionisation efficiency in MS applications. Although the use of quantitative approaches such as reductive amination or carbamate labelling provide efficient labelling yields, challenges around limit of detection and sample separation are still observed.

Much of the work in the field of quantitative glycomic analysis and profiling has sought to investigate specificities that exist between different glycan labelling strategies, as it has previously been noted that may glycan labels occupy analytical niches exhibiting superior characteristics in a particular application while proving inadequate for others. This is commonly seen in samples where glycans are of low abundance and losses from derivatisation result in glycoforms that are below the limit of detection [3, 5].

The effects of derivatisation of glycoforms with procaine **18** have been analysed by positive mode ESI-MS and MALDI-TOF MS [6, 7]. In these studies (which employed maltohexose as a model

carbohydrate) the authors observed more than 1000-fold increase in signal intensity over underivatized glycan demonstrating the potential of more hydrophobic derivatisation agents. Similarly, Pabst *et al* [8] analysed the detector response of non-sialylated glycans derived from bovine fibrin labelled with 15 different derivatisation agents, including Procaine **18**, Procainamide **8**, 2AB **9** and 2AA **10**. Procaine **18** was found to be significantly less effective compared to procainamide **8** when the derivative was analysed by positive mode ESI-MS. Moreover, the study found that procaine derivatives showed significantly reduced resolution compared to both **10** and **9** when separating fetuin glycans by NP-HPLC. Significantly, although **10** and **9** provided lower intensity signals at analysis level, both compounds showed significantly increased resolution in chromatographic separations. Although **10**, **18** and **8** ionised poorly in negative mode, **10** provided a sufficient ionisation due to its one negative charge.

Previous work carried out by Yoshino *et al*, [6] Takao *et al*, [7] and Pabst *et al*,[8] goes a long way to demonstrating the application specificities occupied by carbohydrate derivatisation agents. Work carried out by [9] sought to compare the separation efficiency and suitability of **9** and **8** for glycan profiling, when labelling IgG glycans from human and murine sources. In these experiments, separation and analysis was carried out by HPLC-FLR. It was observed that **9** exhibited both greater resolution and improved peak shape for major glycan peaks. However, analysis of procainamide labelled samples lead to an increased number of minor peaks being resolved due to higher signal intensities. Analysis by HPLC-Q-TOF also suggested a greater efficiency of ionisation for procainamide labelled samples - the total ion chromatograms (TIC) show a higher signal intensity for procainamide labelled samples than those labelled with 2AB. Similar results were also observed by [10]; in this study the authors report a significantly improved base peak chromatogram (BPC) for **8** labelled IgG samples over that of **9** derivatised glycans.

The existing research surrounding commonly used reductive amination labels **10**, **9**, **18** and **8** focuses heavily around the separation and ionisation characteristics but many authors fail to mention the structural differences which lead to the observed changes in fluorescence and ionisation efficiency [11]. Both **10** and **9** possess an aromatic ring with an *ortho*-amine. In both cases this is the site of reductive amination for carbohydrate derivatisation [12]. It is well documented that 2AB provides shelf stable labelling solution with excellent fluorescence; this stability is centred around the compound's amide functional group also present on the ring [13]. Conversely although **10** has the same basic structure, it exhibits a carboxylic acid instead of the amide present on **9** [14]. This makes **10** less stable and it has been documented that the label undergoes photolysis when exposed to ultraviolet light.

Research carried out by Kamoda *et al.*[15] alluded to the beneficial effect of **10** labelled glycans when analysed by MS. Further research commented that when analysed in negative ion mode, **10** exhibited a three fold increase in relative abundance over **9** and **8** [8]. This is due to the one negative charge exhibited by the carboxylic acid in its deprotonated state. However analysis in positive ion mode revealed **10** failed to significantly enhance ionisation efficiency when compared to **9** [8]. Furthermore, the less polar label procaine showed high ionisation efficiency when directly compared to **10** and **9**. Researchers speculated that this improvement was largely down to the tertiary amine tail exhibited by procaine. This is largely backed up by further work concerning the increased efficiency of procainamide **8** [9, 10, 16].

The detection and quantification of glycans provides a significant analytical challenge both due to their low concentration but also due to their inability to efficiently ionise. Existing glycan labelling workflows provide routes for the analysis of these structures in high concentrations, however the analysis of low abundance glycans (<170 pMol per sample) still presents a significant analytical challenge [17] . Moreover, as the intricate structure of many biologically derived *N*-glycans are challenging to synthesise there is a requirement for a labelling strategy that not only facilitates the analysis of low abundance glycans, but also provides a route for further glycoconjugate development.

2.2 Aims and objectives

This chapter will aim to synthesise a suite of multifunctional labels capable of not only improving the analysis of low abundance glycans, but also undergoing selective post analysis processing to form novel glycoconjugates to probe glycan structure and interaction.

It was hypothesised that multifunctional labels, can be covalently attached to the target glycan and will exhibit activity to undergo further reactions. This could be employed to introduce further hydrophobicity in the for enhanced analysis by MS applications. Alternatively target glycans could undergo surface or biomolecular activation. This would aim to generate combinatorial array surfaces for the investigation of glycan binding or synthetic glycoprotein aimed at investigating the effect of glycosylation on protein structure.

The work outlined in this chapter details the design and optimisation of the synthesis of a suite of multifunctional labels. These will posses the ability to react with reducing carbohydrates through reductive amination while also maintaining a region for selective covalent binding following separation and analysis by HPLC-FL. These labels must prove reliable and sensitive as stand-alone

glycan labelling strategies while also maintaining additional functionality to undergo chemoselective reactions following the initial analysis.

This will be achieved initially by the collection of glycans of low abundance in line with currently available commercial glycan profiling workflows, before samples undergo a second derivatisation. This step will either see the addition of nonpolar or charged groups to facilitate their analysis by MS, or the activation of the biomolecule of interest to investigate the interaction of labelled carbohydrates on structure and function.

2.3 Rational for the development of multifunctional glycan labels

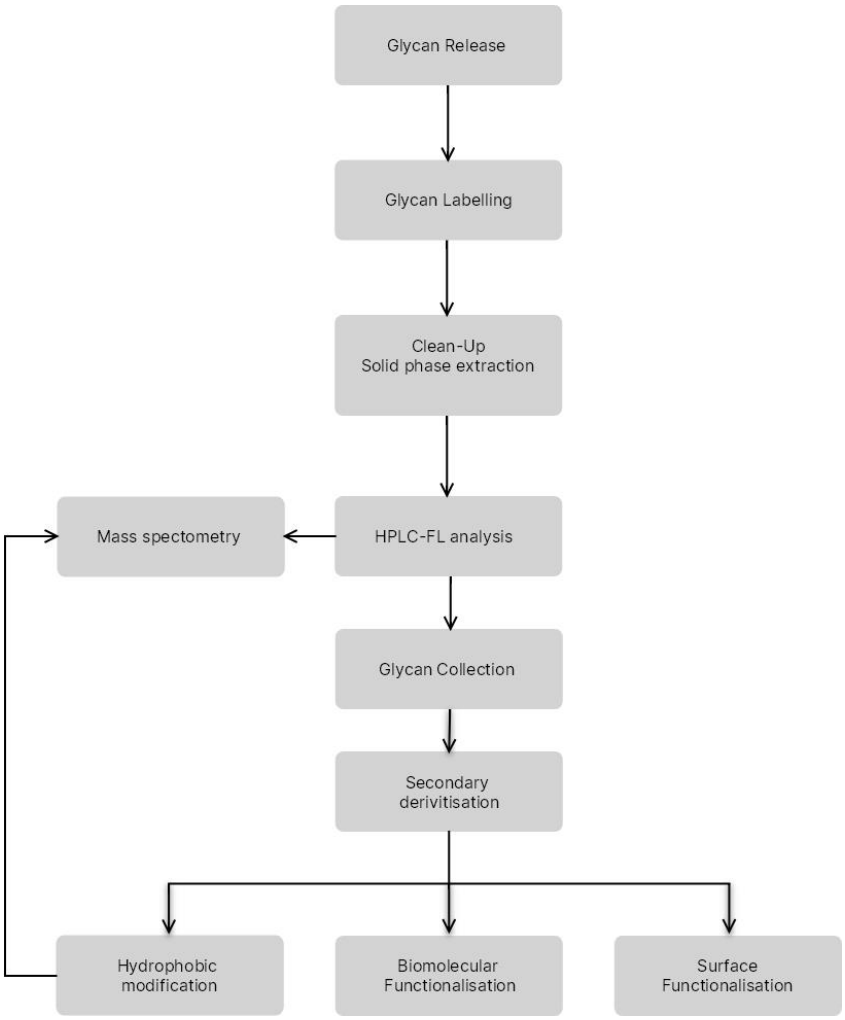


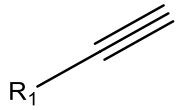
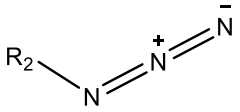
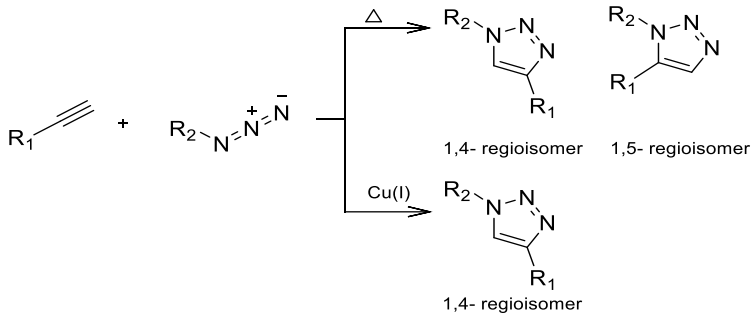
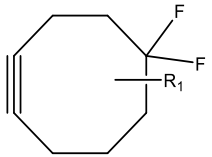
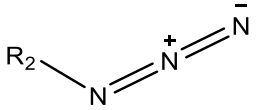
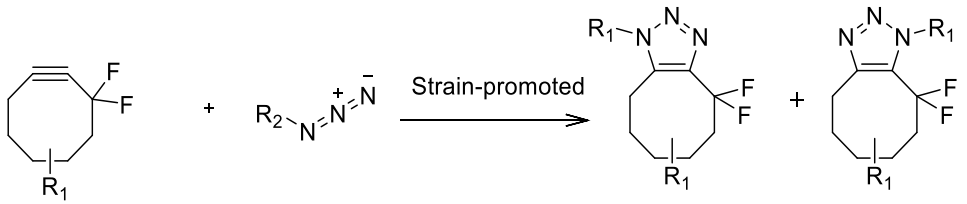
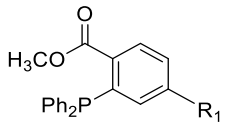
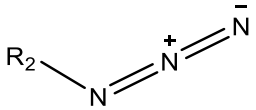
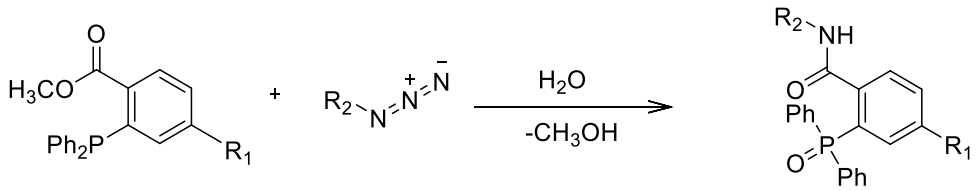
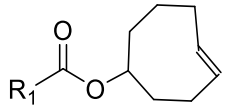
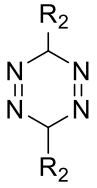
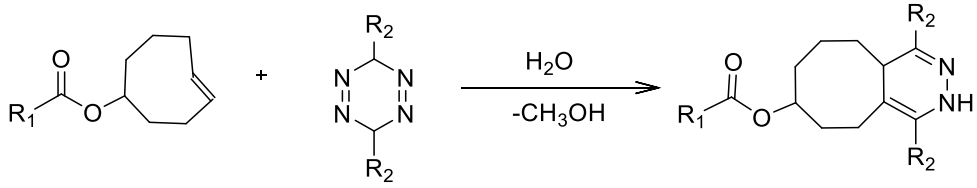
Figure 2.1 - Proposed workflow for reductive amination glycan labels bearing a multifunctional handle

Existing glycan analysis workflows for the profiling and structural elucidation of biologically derived carbohydrates seek to modify released glycans with hydrophobic or fluorescent labels to facilitate analysis. This is typically achieved through reducing end modification followed by separation and analysis by optical or MS routes. The development of multifunctional labels outlined in this chapter would seek to formulate glycan labels bearing additional covalent functionality. A proposed workflow outlined in **Figure 2.1** would see derivatised carbohydrates collected following optical analysis, enabling the use of additional downstream functionality. This could be applied to the generation of neo-glycolconjugates or the addition of further hydrophobicity to facilitate analysis by MS.

2.3.1 Rationale for multifunctional linker chemistry

The addition of a functional handle aims to add additional chemoselective reactivity following the attachment of the glycan label by reductive amination. The functional handle must be able to undergo successful and efficient reactions in polar or aqueous conditions due to glycans lacking solubility in nonpolar solvents. The functional handle should also remain stable throughout analysis and should not display reactivity towards underivatized carbohydrates in the initial derivatisation. Moreover the addition of a functional handle must remain selective in reaction substrate allowing for site specific modification in downstream applications. Bioorthogonal chemistry has developed several systems outlined in **Table 2.1** which could be considered suitable for the application.

Table 2.1-Potential routes for the addition of bioorthogonal multifunctionality

Reaction	Reactant 1	Reactant 2	Mechanism	Reference
Copper-catalysed azide-alkyne cycloaddition (CuAAc)				18
Strain promoted azide alkyne cycloaddition				19
Staudinger ligation				20
Inverse electron-demand Diels-Alder reaction				21

The reaction occurring between azide and alkyne is routinely used for the selective addition of terminal alkynes to azide containing substrates. This approach was initially pioneered by Huisgen *et al* in a processes known as Huisgen's cycloaddition; the reaction results in a 1,2,3-triazole product containing both 1,4 and 1,5 regioisomers of the product [22-24]. These reactions were largely deemed unsuitable for many biological applications because reaction times range from overnight to several days at temperatures exceeding 100 ^{o}C [18, 25-27].

Variations of this method however have proved powerful in bioorthogonal applications. These reactions rely upon the use of a Cu(I) catalyst [28]. The use of a catalysed system leads to a significant reduction in reaction time and temperature with significantly increased yields observed over shortened reaction time <6hr [29]. Furthermore, the application of Cu(I) results in the generation of the 1,4 isomer only as shown in **Table 2.1**.

Although a highly efficient reaction, CuAAc is limited for *in vivo* use due to the cytotoxicity of Cu(I) meaning that significant copper chelation is required to reduce the interaction of copper with proteins and other biomolecules present in the cell. The use of a strain promoted azide alkyne cycloaddition provides an alternative system capable of producing high yielding cycloaddition in bioorthogonal conditions. In this reaction, the alkyne forms part of a strained system in which the alkyne is forced to adopt a bond angle of less than the desired 180^{o}. This results in higher potential energy across the bond increasing the reactivity of the alkyne moiety over CuAAc which utilises a much more stable terminal alkyne. Unlike its copper catalysed derivative, the increased reactivity exhibited by SPAAC systems does not require the use of a catalyst, making it a popular choice for *in vivo* cellular imaging.

In addition to the use of Strain promoted Alkyne Azide cycloaddition SPAAC, the Staudinger ligation is also a widely performed bioorthogonal tool for the introduction of modifications to biologically significant molecules. The reaction outlined in **Table 2.1** shows the reaction in which azides are converted to primary amines upon exposure to a phosphine, which is then capable of reacting with an adjacent carbonyl group forming a stable secondary amide bond between both reactants. These reactions have been successfully carried out *in vivo*, however the rapid oxidation and enzyme breakdown of phosphine provides a major disadvantage as it leads to reduced reaction rate resulting in decreased product yields.

The use of an Inverse Electron-Demand Diels-Alder reaction (IEDDA) is also a popular choice for both *in vivo* and *in vitro* modification of biomolecules, because it has been shown to proceed faster in water than in organic solvents. IEDDA sees the reaction between an electron rich dienophile such as *trans*-cyclooctene or norbornene and an electron poor diene such as tetrazine. The resulting

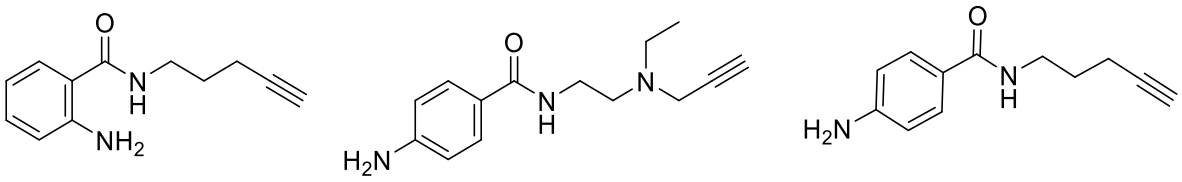
structure is a [4+2] cycloadduct and the formation of a six membered dihydropyrazine cycloadduct with N_2 as the only significant biproduct. This application is often referred to as tetrazine ligation and is widely applied to *in vitro* biomolecule modification although seldom applied to *in vivo* applications.

2.3.1.1 Evaluation of multifunctional linking reactions

The four methods outlined in **Table 2.1** provide a selection of the most widely used bioorthogonal systems used for the chemical modification of biologically significant molecules. All have benefits and drawbacks but individually present high yielding product formation in aqueous conditions over a wide range of pH values. In addition, these reactions produce unreactive by-products which can be easily purified without the requirement of instrumental separation. CuAAc, SPAAC and Staudinger ligation all benefit from the use of a functionalised azide. This azide moiety is not present in Nature but is highly stable at physiological conditions. Azides are also relatively easily incorporated into target molecules allowing selective modification while having minimal impact upon the target molecule due to their small size. Tetrazine ligation meanwhile requires the addition of a large electron withdrawing group increasing the potential for further structural modification[21]. CuAAc[18] , SPAAC[19], IEDDA[21] and Staudinger ligation [20] all provide robust efficient one-pot systems displaying the correct chemistry for a multifunctional handle. However due to the low analyte recovery retrieved from biological samples and speed at which derivatisation should be carried out, any mechanism considered as an addition provide fast efficient product formation. IEDDA and CuAAc provide the fastest reaction rates with 100-1000 $M^{-1}s^{-1}$ and 10-100 $M^{-1}s^{-1}$ respectively[18, 21, 28]. These systems provide significantly faster rates of reaction than Staudinger ligation which has been shown to be at least 25 times slower than CuAAc[20]. The use of copper free SPAAC is also hampered by reduced reaction rates typically exhibiting between 0.0012-0.96 $M^{-1}s^{-1}$ in low concentration[19]. Considering the benefits of each labelling mechanism presented in the previous section, CuAAc was chosen for the generation of multifunctional labels due to its favourable reaction rate and high product yield. Moreover, it was expected that the relatively minor modification to the compound's structure would not significantly alter the compound's fluorescence or polarity in such a way as to impact its ability to function as a successful glycan label.

2.3.2 Structural development of multifunctional labels

In order to align with widely utilised commercial workflows, labels **25**, **26** and **27** (**Figure 2.2**) were designed as derivatives of currently adopted reductive amination labels. They were designed to include additional functionality to allow subsequent selective CuAAc reaction with azide containing reaction substrates in order to facilitate the analysis of carbohydrates of low abundance.



25 **26** **27**

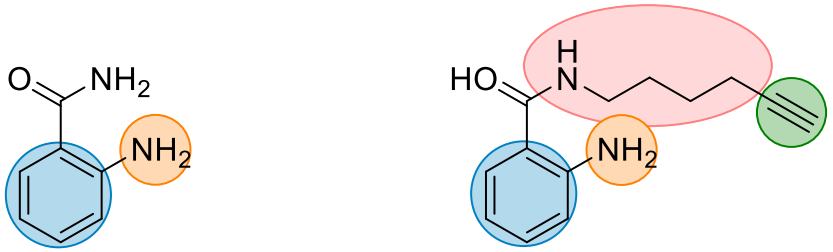
Figure 2.2- Proposed structures for multifunctional labels **25**, **26** and **27**

The multifunctional glycan labels **25**, **26** and **27** represent structures capable of derivatising reducing carbohydrates in line with existing pharmaceutical workflows. In addition, the terminal alkyne exhibited by **25**, **26** and **27** provides a region capable of downstream covalent activity.

2.3.2.1 Rationale for the development of multifunctional label 25

Derivative **25** shares close structural similarity to 2AB **9**, a label routinely used by industrial workflows for the profiling of glycans by HILIC-Fl analysis due to its superior peak resolution over less polar labels such as **8** [9] [8]. **25** was designed with modification at the *ortho*-amide group, thus the addition of a pentyne chain containing a terminal alkyne facilitates the ability to undergo CuAAC. maintaining the label's primary amine allows for reductive amination. The addition of a pentyne chain was chosen to minimise the size of the modification and the impact on the compound's fluorescence while reducing the steric effects of the label and its ability to label released oligosaccharides [18]. The introduction of a region of reduced polarity in the form of an aliphatic region also aims to address the limitations of **9** in MS applications where it has been noted that its polarity has a detrimental effect upon its ability to efficiently ionise [30]. It should be noted however that any modification cannot significantly influence the compound's partition coefficient. This would

result in a reduction in separation resolution by HILIC mode separation limiting its use in existing glycan analysis workflows.



9 **25**

Figure 2.3- Structure of 2AB **9** (Left) and its multifunctional alkyne containing derivative **25** (Right) highlighting important moieties of interest. Orange regions represent primary amines capable of covalently bonding to reducing carbohydrates, Blue represents regions responsible for the fluorescent properties of each label, red regions are areas of additional hydrophobicity and green areas represent the regions of the molecules involved in CuAAc.

The structural similarities between **9** and **25** can be seen in **Figure 2.3**. **9** and **25** share two main regions making them suitable for glycan labelling. Both share a primary amine capable of covalently bonding to target oligosaccharides *via* reductive amination as well as a region of delocalisation highlighted in blue capable of fluorescence under analysis conditions. While **25** maintains many of the characteristics of **9** the addition of a multifunctional linker serves to provide two benefits. In MS applications the presence of an increased aliphatic region highlighted in red serves to increase ionisation efficiency while the terminal alkyne shown in green facilitates the use of CuAAc for downstream functionalisation.

2.3.2.2 Rationale for the development of multifunctional label 26

The proposed structure for label **26** outlined in **Figure 2.4** also shares may structural similarities with the commercial glycan label **8**. ProA is used predominantly for glycan analysis within MS applications where it has been shown to have superior ionisation efficiency exhibiting an increase in signal intensity of 50% over glycans from the sample labelled with **10**.

The modification of **8** was designed to exhibit similar functionality to **25** while maintaining the analytical characteristics for MS analysis. This was carried out by the addition of a terminal alkyne moiety to the tertiary amine region of the molecule. The modification sees replacement of one of the hydrophobic tails exhibited by **8** with a propargyl group capable of undergoing CuAAC. Due to

the beneficial physicochemical properties exhibited by **8** it was important that any modification did not significantly alter the polarity or size of the molecule as it was predicted that this would have a significant impact upon the partition coefficient and ionisation efficiency.

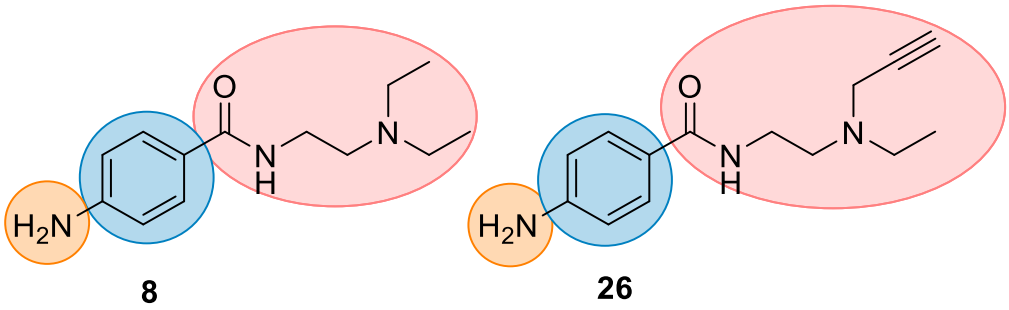


Figure 2.4 -Structure of ProA **8** (Left) and its multifunctional alkyne containing derivative **26** (Right) both highlighting important moieties of interest. Orange regions represent primary amines capable of covalently bonding to reducing carbohydrates, Blue represents regions responsible for the fluorescent properties of each label, red regions are areas of additional hydrophobicity and green areas represent the regions of the molecules involved in CuAAc.

Similarities between multifunctional label **26** and **8** can be seen in **Figure 2.4**. As both labels fall within existing workflows for the reductive amination of glycans, both exhibit a primary amine shown in orange as well as a region of delocalization capable of producing fluorescence under analytical conditions. What separates **8** from fluorescence profiling labels such as **9** is its tertiary amine tail shown in red, which has been shown to provide a region of hydrophobicity increasing ionisation efficiency. These characteristics are retained in multifunctional label **26** but with the addition of a terminal alkyne highlighted in green capable of undergoing CuAAc.

2.3.2.3 Rationale for the synthesis of multifunctional labels 27

Label **27**, a compound previously developed members of the S. Allman research group [31], is a close derivative of the glycan label **17** and is derived from *p*-aminobenzoic acid (**Figure 2.5**). The presence of an alkyl chain in both **17** and multifunctional label **27** outlined in red serves to improve ionisation in MS applications [8]. The addition of a terminal alkyne circled in green facilitates the ability of the label to undergo CuAAc reactions following separation and analysis by HPLC-FL. The sensitivity has been compared to **9** in MS applications however its alkyne is anticipated to be the least sterically hindered of the three labels in the suite.

.

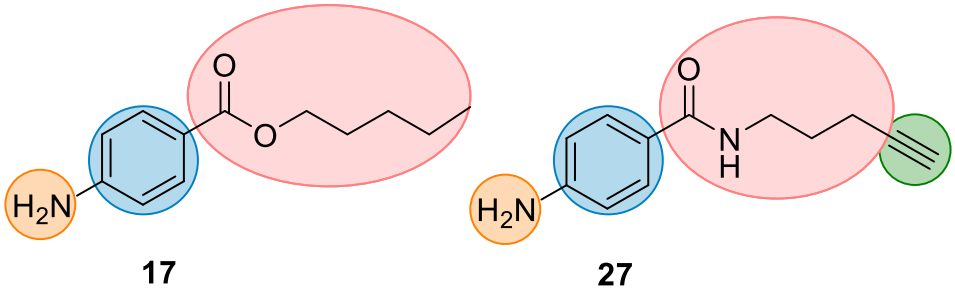


Figure 2.5 -Structure of ABBE **17** (Left) and its multifunctional alkyne containing derivative **27** (Right) both highlighting important moieties of interest. Orange regions represent primary amines capable of covalently bonding to reducing carbohydrates, Blue represents regions responsible for the fluorescent properties of each label, red regions are areas of additional hydrophobicity and green areas represent the regions of the molecules involved in CuAAc.

2.4 Results and discussion

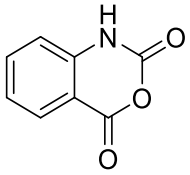
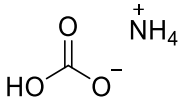
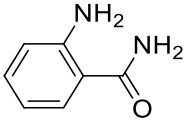
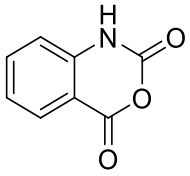
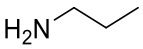
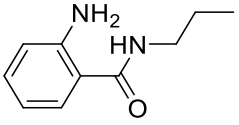
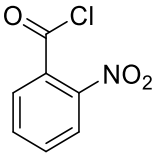
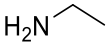
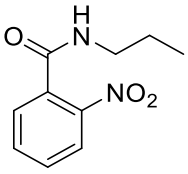
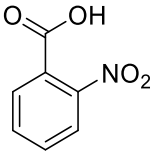
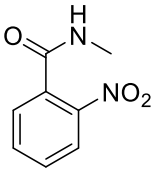
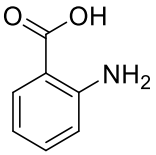
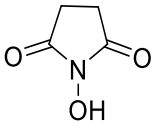
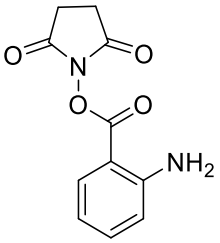
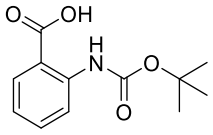
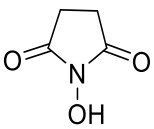
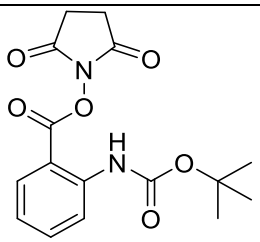
2.4.1 Development of a synthetic pathway for multifunctional label 25

The modification of *o*-aminobenzoic acid derivatives **10** is well documented in existing literature [32, 33]. This serves to highlight and define a potential synthetic route for the generation of **25**. **Table 2.2** sets out a series of conditions that have been used in the development of *o*-aminobenzoic acid derivatives.

The synthesis of **9** outlined in **Table 2.2** entry 1 involves the breakdown of isatoic anhydride in the presence of excess ammonium bicarbonate resulting in the formaserion of a stable primary amide. Subsequent modifications of **9** have sought to apply this approach to the generation of products such as that outlined in entry 2. These have shown to be less effective than the initial method with significantly harsher reaction conditions required to achieve comparable yields. Subsequent modifications have sought to formulate a similar product via the use of nitrobenzoic acid derivatives. The scheme shown in **Table 2.2** shows the formation of an amide product resulting from the reaction between carboxylic acid and amine containing linker. These reactions are hampered due to the reactivity of carboxylic acids and the formation of unreactive carboxylate moiety with product formation only favoured by the application of heat in excess of 100 °C to facilitate the removal of water [34]. Similar approaches have seen the use of an acid chloride such as that outlined in entry 3. The addition of a primary amine containing linker results in the nucleophilic addition of reactant 2 yielding an amide containing product and HCl. These reactions, although high yielding, present unsuitable conditions for the generation of **25** on account of the low pH and high temperature required for their generation. These reactions also commonly see the use of starting materials

containing nitro moieties which require reduction to afford primary amines in order to be beneficial as glycan labels. These reactions typically require catalytic hydrogenation with platinum (IV) oxide or temperatures in excess of 100 °C which would likely cause degradation of the anticipated structure of multifunctional label **25**.

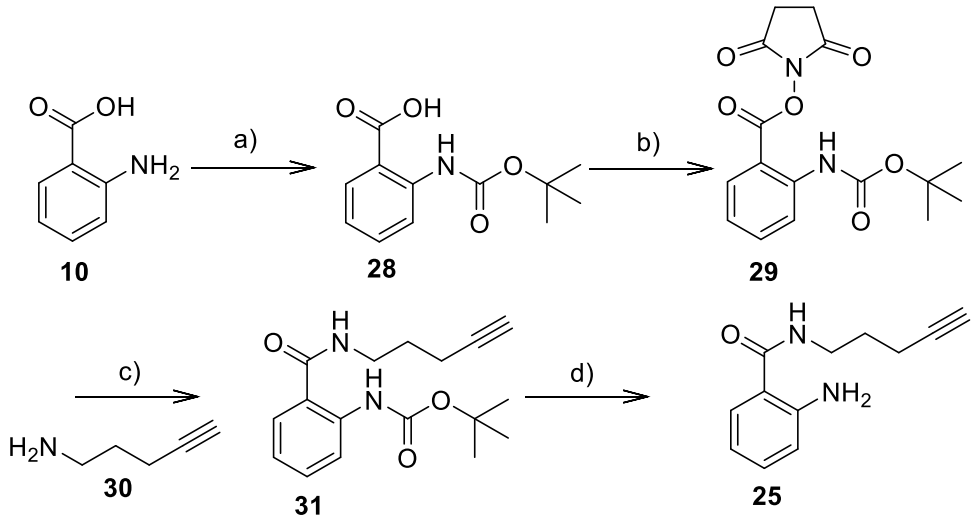
Table 2.2 - Potential synthetic routes for the synthesis of multifunctional label **25**

Entry	Reactant 1	Reactant 2	Product	Conditions	Reference
1				1,4 dioxane 60 °C 0.5h (91%)	[35]
2				1,4 dioxane, 70 °C, 2h (90%)	[36] [37]
3				TEA, DCM, 0 °C, 12h (60%)	[38]
4				DMF, Cs_2CO_3, 150 °C, 6h (20%)	[39]
5				DCM, 20 °C, 3h (8%)	[32]
6				DCC, THF, 20 °C, 24h (90%)	[33]

Instead, the milder conditions represented in line 5 of **Table 2.2** show the use of *N*-hydroxysuccinamate as a leaving group negating the requirement for high temperatures traditionally required to favour reactions with carboxylic acid groups. Early attempts at these reactions saw low yields and the formation of non-useful by-products. A development of this method shown in entry 6 saw the use of amine protecting groups such as *tert*-butyloxycarbonyl group (Boc) to reduce the formation of non useful by-products. These reactions saw a significant increase in yield when dicyclohexyl carbodiimide was added to the reaction mixture. The use of DCC coupling significantly decreases carboxylate formation when combined with NHS to form a stable reactive ester capable of acting as an efficient leaving group when exposed to an amine containing reactant.

2.4.2 Synthesis of multifunctional label 25

After consideration of the available literature in line with the aims for the structure of multifunctional label **25**, a synthetic route was designed to both maximise available yield but also provide conditions applicable for the addition of a terminal alkyne containing linker.



Scheme 2.1 - Proposed synthetic route for the production of multifunctional label **25** a) Boc_2O, EtOAc, 75% b) DCC, NHS, THF 95% c) **30**, DCM,TEA, 74% d) TFA, DCM 95%.

The scheme outlined above (**Scheme 2.1**) shows a reaction pathway beginning with the already commonly used glycan label 2AA **10**. Synthesis of multifunctional label **25** begins with the protection of **10**. This stage was carried out on a 200 mg scale with reactants initially being cooled to 0 ^{o}C before addition of the Boc_2O before the reaction was slowly heated to 50 ^{o}C overnight. Following this,

product **28** was extracted using ethyl acetate and dried to afford a white crystalline powder in a yield of 75%. Following characterisation by ^{1}H NMR spectroscopy the generation of **28** was confirmed by the presence of a new singlet at 1.54 ppm. This was confirmed by a 100 Da increase in mass when analysed by MS constant with the addition of a Boc protecting group. This product was taken forward to stage 2.

Following the successful generation **28** 245 mg of the protected product was dissolved in dry THF and stirred with DCC before NHS was added and the solution was stirred overnight. This workflow is an adaption of the work carried out by Mickle *et al*[33] and resulted in the formation of a white precipitate similar to that characterised as DCU in previous studies. Following verification of reaction completion by TLC analysis, the precipitate was removed by filtration and the organic phase concentrated before being purified on silica by flash column chromatography. This yielded a white crystalline powder which was confirmed by the formation of a new singlet a 2.92 ppm following ^{1}H NMR spectroscopy. Product **28** was yielded at 95%.

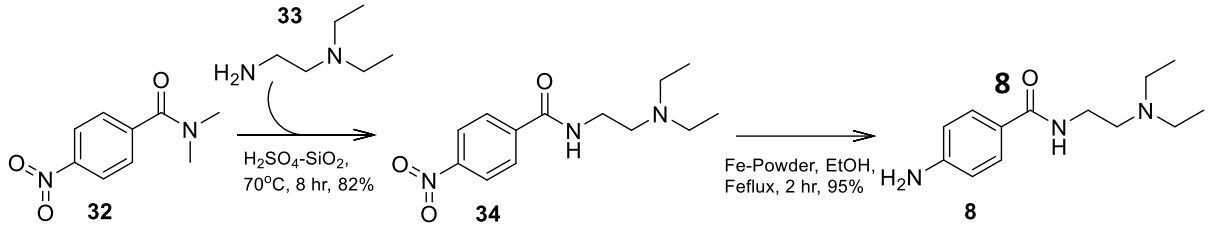
The successful synthesis of **28** produced a stable reactive intermediate of which 200 mg was dissolved in DCM and stirred at 25 ^{o}C for 5 minutes under an inert atmosphere. This was followed by the addition of pent-4-yn-1-amine **30** dissolved in a solution of DCM and TEA. This mixture was stirred overnight, and its progress was monitored by TLC analysis. Following reaction completion, the mixture was concentrated under reduced pressure and dry loaded onto silica before being purified by flash column chromatography. This yielded an off white solid in 74% yield which crystalised rapidly once concentrated. This was characterised as **31** by ^{1}H NMR spectroscopy by the disappearance of the singlet at 2.92 ppm and the formation of multiple peaks, notably a singlet at 2.04 ppm indicating the addition of an alkyne functional group.

This product was then solubilised in DCM and acidified with a 1:6 v/v ratio of TFA to facilitate the removal of the Boc protecting group and was stirred at room temperature for 1.5 h with periodic monitoring by TLC analysis. Following completion of the reaction as determined by TLC analysis, the solution was dried under reduced pressure and redissolved in MeOH containing amberlite IRA-57 ion exchange resin. Following neutralisation, the amberlite was removed by filtration and the product was concentrated and purified by flash column chromatography to yield a pale-yellow solid. This was characterised by 1H NMR by the removal of a high intensity singlet previously present at 1.51ppm confirming the removal of the Boc protecting group and the successful synthesis of **25** at 99% yield.

2.4.3 Development of a synthetic pathways for multifunctional label 26

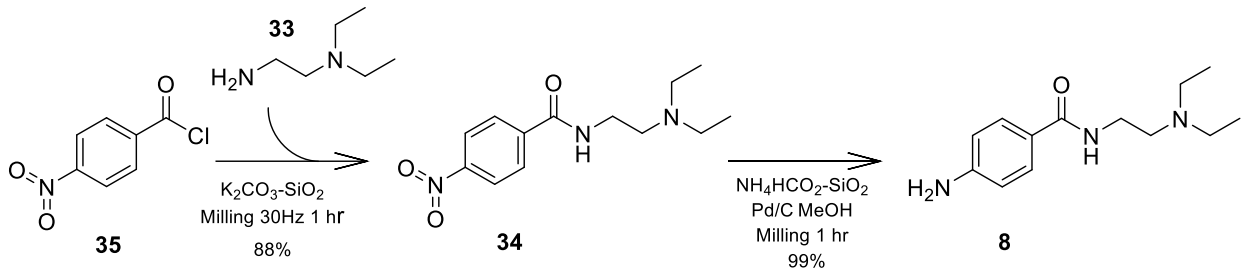
The aim for this section of work was to generate a suite of multifunctional glycan labels compatible with existing pharmaceutical analysis workflows. The generation of **25** in the previous section provided a stable multifunctional derivative of **9**. The synthesis of **26** was designed to fall within the current pharmaceutical workflow for ProA **8**, a widely used glycan labelling tool for both fluorescence and MS applications.

Existing synthesis routes for the generation of **8** on a commercial scale fall into two distinct routes. In both cases the synthesis begins with the addition of a tertiary amine containing tail and the generation of a secondary amide.



Scheme 2.2 -Synthesis of ProA 8 performed by Rasheed *et al*.

The synthesis route shown in **Scheme 2.2** sees the generation of a secondary amide in a metal free approach. This route was designed by Rasheed *et al*[40] and sees the formation of **8** from *N*-methyl-4-nitrobenzamide **32** at 95% yield while using mild conditions and stirring over 12 h.

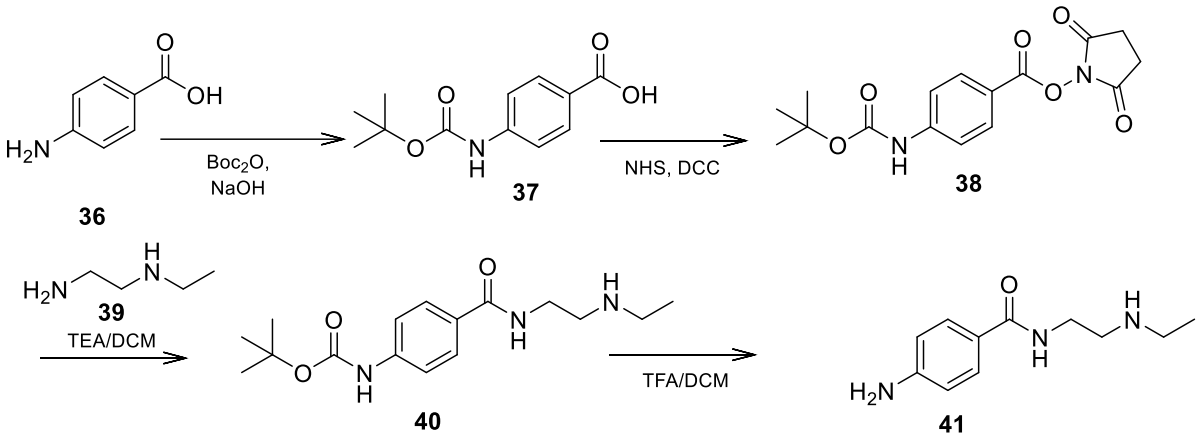


Scheme 2.3 -Synthesis of ProA 8 performed by (Portada et al., 2018)

The route outlined in **Scheme 2.3** was initially developed by Baltzly and Ferry, [41] however the approach was optimised to improve its commercial scalability [42]. This workflow sees the solid phase

synthesis of **8** from 4-nitrobenzoyl chloride **35** over 1 h with continuous milling at 30 Hz yielding 88% product formation.

Both the approaches of Portada *et al*, and Rasheed *et al*, sought to generate **8** from *p*-nitrobenzene derivatives and require hydrogenation to form the primary amine present on **8**. A 95% yield was reported by Rasheed *et al*, following a 2h reaction catalysed by Fe Powder in EtOH. The approach of Portada *et al*, followed a solid phase approach and saw increased yields of 99% for the generation of **8**.

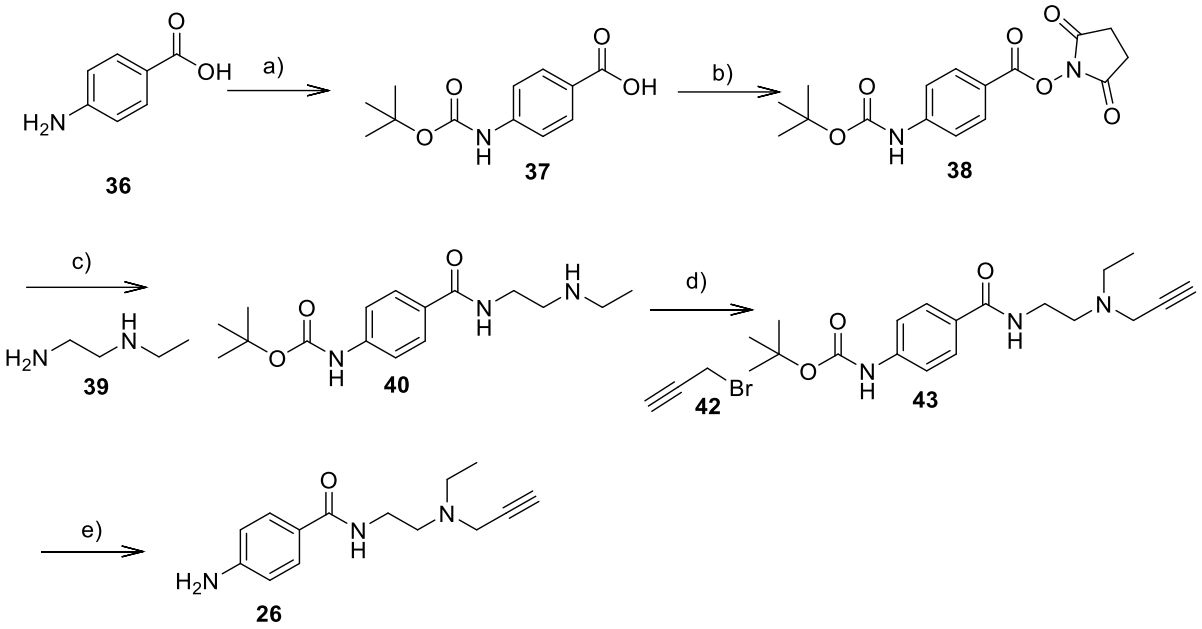


36 **37** **38** **39** **40** **41**

Scheme 2.4 -Synthesis of ProA **8** derivative **41** performed by Chu *et al*.

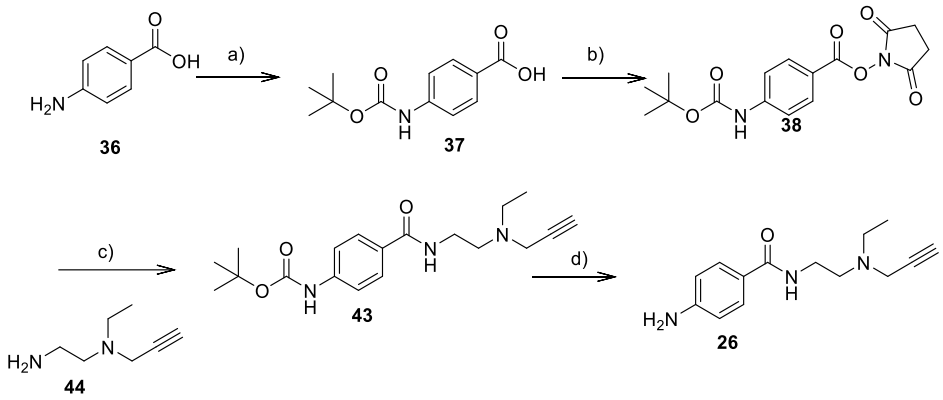
An alternative approach (**Scheme 2.4**) sought to develop a ProA **8** derivative following a similar route to that applied to the synthesis of multifunctional label **25** [43]. Unlike the synthetic routes detailed in **Scheme 2.2** and **Scheme 2.3** the synthesis begins with protection of the primary amine at the *para*-position of **36**. The route then proceeds with the use of an NHS-intermediate in the form of **38** to aid amide formation and the formation of a protected ProA derivative **40**. The compound's primary amine is then reinstated by the use of a mildly acidic deprotection yielding **41** in high yield.

After consideration of **Scheme 2.2** and **Scheme 2.4** two possible synthetic routes where designed. Both routes outlined in **Scheme 2.5** and **Scheme 2.6** were developed around two distinct stages: formation of an amide bond at achievable conditions conducive to alkyne stability as well as the addition of the alkyne itself.



Scheme 2.5 - Proposed synthetic route for the production of multifunctional label **26** via a five-stage stepwise approach. A) Boc_2O, TEA, Dioxane (aq) b) DCC, NHS,THF c) **39**, TEA, DCM d)**42**, K_2CO_3, ACN e)TFA:DCM.

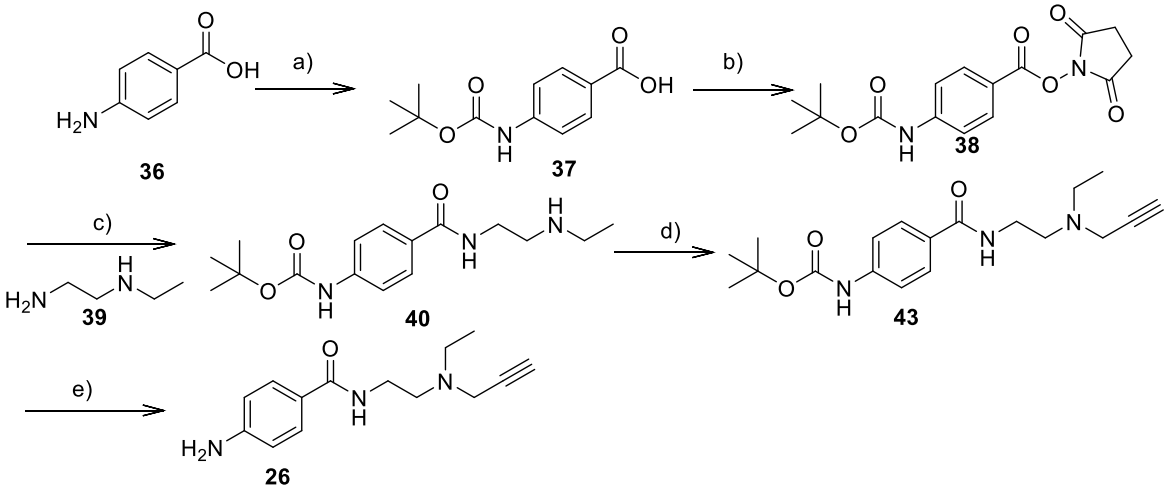
The synthesis route outlined in **Scheme 2.5** follows a similar approach for amide formation with the multifunctional alkyne addition occurring later in the synthesis. This synthesis route also reflects work initially carried out by Allen *et al* [31], who sought to develop this derivative *via* the use of a Boc-protected NHS intermediate. This yields a protected form of the compound and aims to increase product yield while reducing the formation of by-products. In this approach the addition of an alkyne is carried out by the use of propargyl bromide resulting in the formation of a tertiary amine and the addition of an alkyne to the aliphatic region of the molecule. Similar to the approach outlined in **Scheme 2.4** this yields a protected form of multifunctional label **43**. The removal of Boc then takes place under mild acidic conditions to reinstate the primary amine resulting in an active form of multifunctional label **26**.



Scheme 2.6 - An alternative synthetic route for the production of multifunctional linker **26** over a four-stage process. a) Boc_2O, TFA, 1,4 Dioxane (aq) b) DCC, NHS, THF c) **44**, TEA, DCM d) TFA:DCM

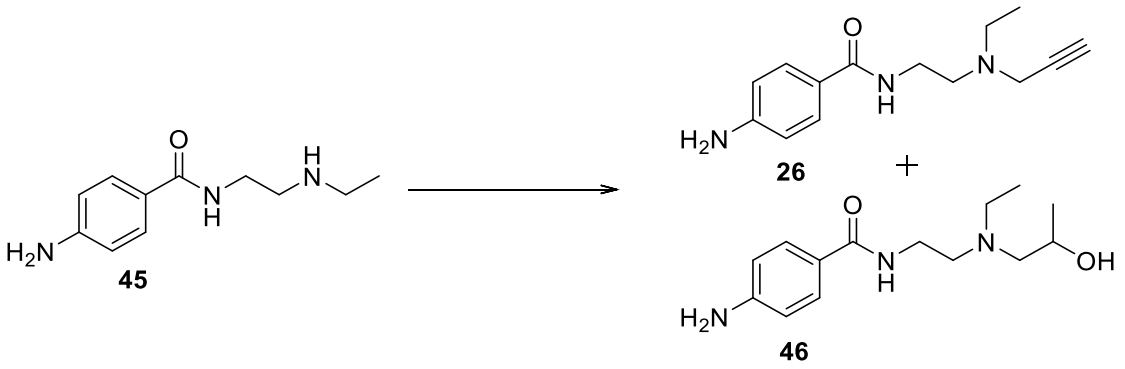
In contrast to **Scheme 2.5** the alternative synthesis route outlined in **Scheme 2.6** seeks to form the aliphatic region of the molecule prior to amide formation [42, 43]. This approach initially sees the selective Boc protection of the primary amine present on *N,N*-ethylethylenediamine **39** before the addition of the propargyl arm *via* nucleophilic substation with propargyl bromide **41**. This forms a protected form of the aliphatic region which can now be amidated to form a protected form of **44** which can be deprotected with the use of TFA/DCM. This synthetic route benefits over an *in situ* approach as the synthesis of **38** and **44** can take place simultaneously reducing the time required for the synthesis. This synthetic route sees the addition of the propargyl group take place prior to the addition of the aliphatic section of the molecule.

2.4.4 Synthesis of Multifunctional label 26



Scheme 2.7 - Initial synthesis of multifunctional label **26**. a) Boc_2O, TEA, 1,4 Dioxane (aq), room temperature (RT) over night (O.N) stirring (99%) b) DCC, NHS, DCM, THF, RT, ON stirring (79%) c) TEA, DCM, RT, 10 h (70%). d) **42**, K_2CO_3, ACN, RT 50 h 40% e) TFA:DCM 1hr (99%)

The conditions outlined in **Scheme 2.7** have previously been attempted for the production of **26**. While sections a-c in the synthesis route produces workable yields of >70% the addition of alkyne functionality in step d resulted in the formation of multiple inseparable by-products and a reduction in final product yield.



Scheme 2.8 – Potential reaction products from initial attempts to synthesise multifunctional label 26.

Scheme 2.8 outlines the addition of the propargyl group and concurrent formation of non-useful products as well as the production of **26**. Although challenging to separate by methods such as flash column chromatography the existence of by-product **46** was tentatively characterised by ^1H NMR spectroscopy and MS.

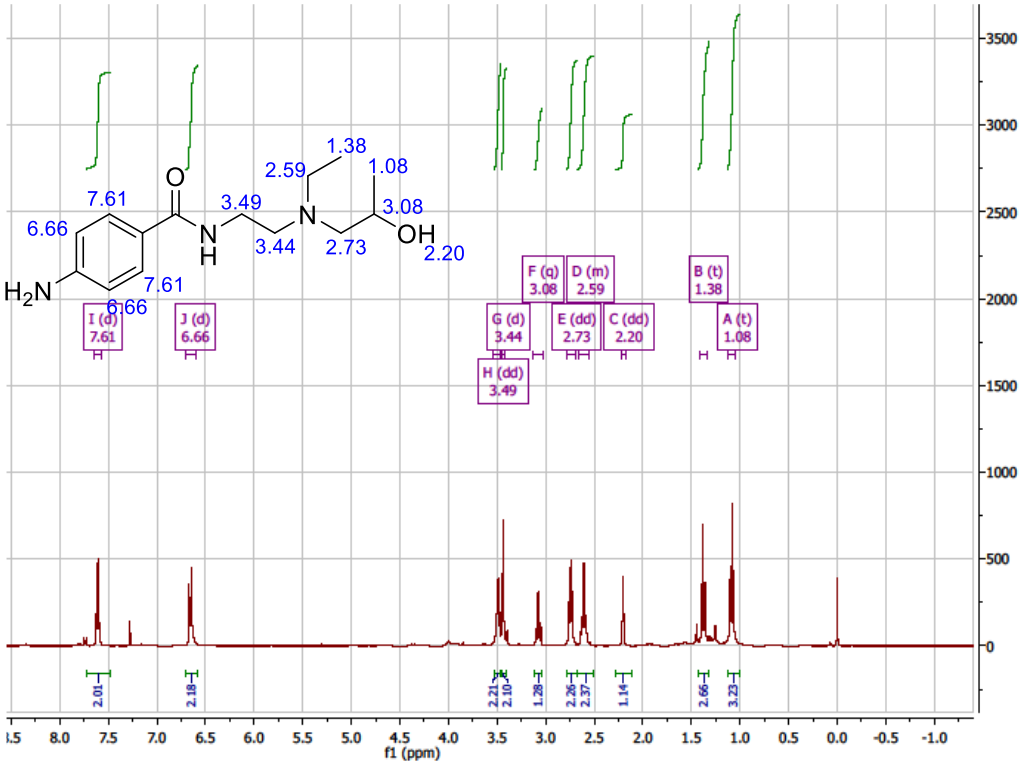
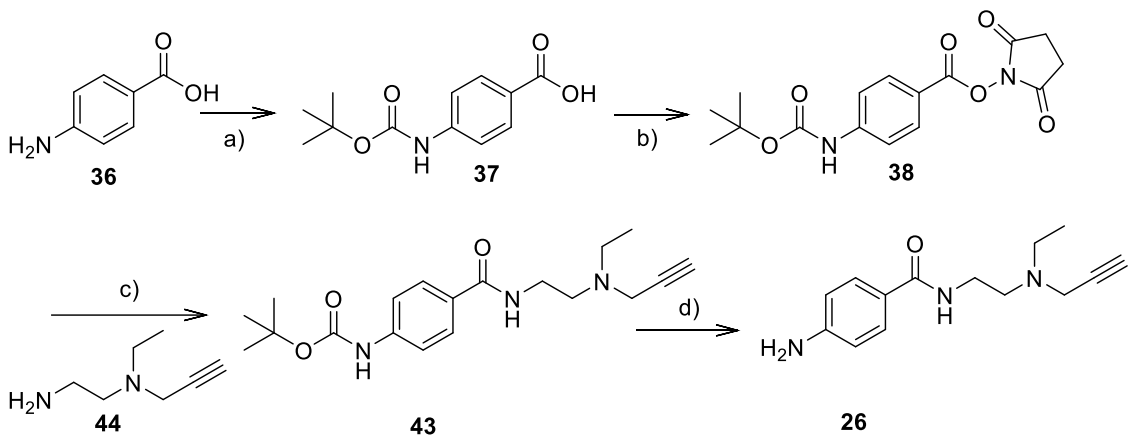


Figure 2.6 - NMR data for by-product **46** resulting from initial attempts to synthesise multifunctional label **26**

Figure 2.6 provides a spectrum for compound **46** following purification by flash column chromatography. The presence of a by-product such as that shown in **Scheme 2.8** is represented by triplet **A** presents at 1.08 ppm. This is backed up by HRMS data outlining a mass increase of 21Da consistent with the generation of by-product **46**. This presented a significant barrier to the use of **26** synthesised by the route outlined **Scheme 2.7** as impurities limit the label's ability to act quantitatively. The structure formed also saw the removal of any downstream capability from **26**. For these reasons the synthesis of **26** though the route outlined in **Scheme 2.7** was abandoned.

2.4.5 Optimised Synthesis of Multifunctional label 26

This alternative approach aimed to follow a similar reaction pathway to that used by Rasheed *et al*[40] and Hoang *et al* [44] in the initial development of **8**, starting from **36**. This requires an additional series of protections in order to preserve amine functionality in the complete label, followed by the addition of a multifunctional aliphatic handle **44**. Finally, the label is deprotected to yield **26.**



Scheme 2.9 - Optimised route for the synthesis of multifunctional label **26** via a four-step process. a) Boc_2O, TEA, 1,4 Dioxane (aq) (90%) b) DCC, NHS, DCM, THF (86%) c) **44**, THF 76% d) DCM:TFA

26 was synthesised from **36** (**Scheme 2.9**) in four steps. *p*-Aminobenzoic acid **36** was Boc-protected by stirring at room temperature overnight before being precipitated out in 25% HCl. This resulted in the formation of a white solid which was purified by flash column chromatography. The successful formation of **37** was then confirmed by ^{1}H NMR spectroscopy and the formation of a new singlet at 1.5 ppm (90%). **37** was stirred overnight with NHS and DCC forming **38** and activated NHS ester at a yield of 86% following purification by flash column chromatography and characterisation by ^{1}H NMR spectroscopy.

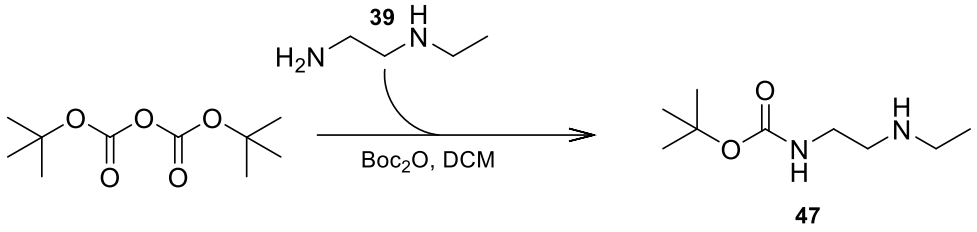
The synthesis of multifunctional handle **44** is shown in **Scheme 2.11** and **Scheme 2.12**and detailed in section 2.4.5.1 below. The synthesis begins with the selective Boc_2O protection of *N,N*-ethylethylenediamine **39**. This yielded **47** at 89% which was confirmed following characterisation by ^{1}H NMR and the generation of a new singlet at, 1.5 ppm. The newly formed Boc- *N,N*-ethylethylenediamine was then, basified and stirred with propargyl bromide to form the protected multifunctional linker **44**. The successful generation of **48** was confirmed by the formation of a new triplet at 2.2 ppm indicating the presence of an alkyne functional group at a yield of 74%. Deprotection was carried out in DCM/TFA, before the reaction was concentrated and resuspended in acetonitrile (ACN) and neutralised with Amberlite IRA-57. The successful deprotection of **48** was confirmed by the removal of the singlet at 1.5 ppm in ^{1}H NMR spectroscopy yielding **44** at 90%.

The deprotected linker **44** was then stirred with **38** in tetrahydrofuran (THF) overnight to yield a protected form of multifunctional label **26** which was then purified by flash column chromatography. To yield a pale brown oil which was characterised as **43** in 76% yield following ^{1}H NMR spectroscopy with the formation of a triplet at 2.21 ppm indicating the addition of alkyne and the removal of a singlet at 2.86 ppm consistent with the loss of NHS.

Deprotection was carried out by the use of TFA in DCM. This solution was concentrated and resuspended in ACN before being neutralised with amberlite IRA-57 ion-exchange resin. The neutralised solution was concentrated and purified by flash column chromatography to yield multifunctional label B as pale brown solid. This was later characterised by ^{1}H NMR and identified as product **26** by the removal of the 1.51 ppm 9 proton singlet at 1.51 indicating the removal of Boc at a yield of 90%. This synthesis was the culmination of a series of developmental stages outlined in the sections below.

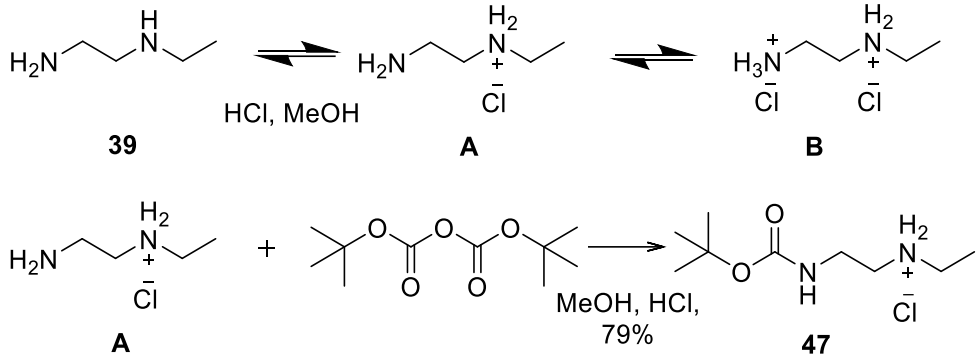
2.4.5.1 Synthesis of multifunctional linker 44

Due to the reactivity of primary amines, a protection step must take place before modification of any other region. The protection of *N,N*-ethylethylenediamine provides a particular challenge due to the presence of multiple amines suitable for Boc protection.



Scheme 2.10- Synthesis route for the selective protection of diamines in line with the approach of Krapcho et.al [45]

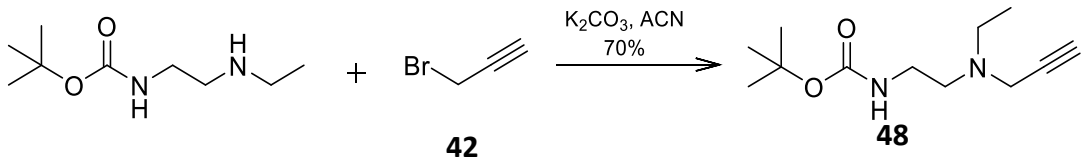
The route outlined in **Scheme 2.10** aims to limit dual protection by reducing the equivalents of di-*tert*-Butyl dicarbamate Boc_2O to amine. In small aliphatic amines, terminal amines form the most reactive region and are therefore the most likely to be protected reducing the likelihood of dual protection. Product yields resulting from this reaction are typically reported at between 40-50%[45].



Scheme 2.11 - Mono Boc protection of *N,N*-ethylethylenediamine **39**[46]. *N,N*-ethylethylenediamine **39** stirred in acidified MeOH, 15 minutes at 0°C, BOC_2O was then added and solution was stirred for 1.5 h. Following clean-up product **47** was isolated at a yield of 76%

The work of [46] set out to increase the potential yield of both symmetrical and non-symmetrical diamine protection. **Scheme 2.11** shows the addition of 50% aqueous acidified methanol at a concentration of 1:1 hydrogen chloride to diamine. This has the potential to form three species outlined in **Scheme 2.11**. At complete equilibrium **A** is likely to be the dominant species leading to the formation of an HCl salt at the secondary amine while leaving one free amine. **Scheme 2.11** demonstrates the successful selective protection of **39** by the addition of >1mol Boc_2O to a solution of *N,N*-ethylethylenediamine **39** following acidification with 1 Mol% HCl.

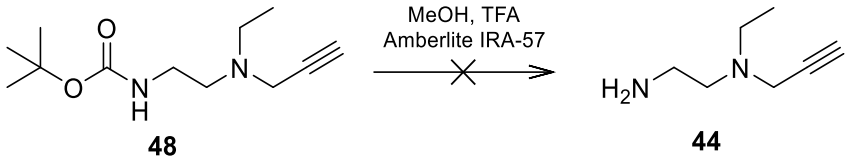
The crude reaction mixture was washed with diethyl ether then neutralised with saturated NaOH (3x5 ml). The selectively protected product was then extracted with DCM and concentrated to an oil which solidified upon cooling yielding compound **47** (80%).



Scheme 2.12 -Addition of terminal alkyne functionality via the use of propargyl bromide **42**.

The protected form of N,N-ethylethylenediamine 47 was a stable intermediate to which addition of the propargyl group could take place. This was attempted by the use of a dehalogenation reaction in which propargyl bromide 42 acts as an electrophile allowing for the selective addition of a propargyl group to the exposed secondary amine. The reaction shown in

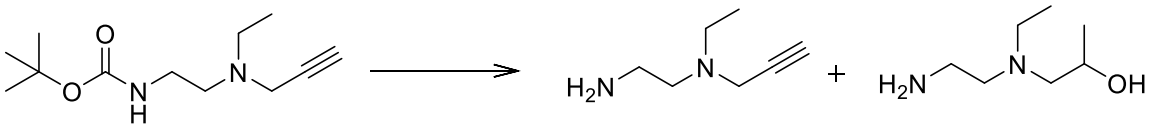
Scheme 2.12 outlines the reaction undergone by the Boc_2O protected amine linker to form **48** as a dark yellow oil characterised as a protected form of **44** by ^1H NMR and the formation of a 1 proton singlet at 3.21 ppm indicating the addition of an alkyne moiety in 70% yield.



48 **44**

Scheme 2.13 -Deprotection of multifunctional linker **44** utilising a TFA:MeOH system, did not yield any recoverable product following neutralisation by Amberlite ion exchange resin.

Following the addition of a propargyl group the reactivation of **48** was carried out. Initially this was carried out via the use of acidified methanol outlined in **Scheme 2.13**. However, following purification by preparative TLC, ^1H NMR spectroscopic analysis revealed that although deprotection was successful, methylation had occurred resulting in the tentstive structural assignment of compound **49** which was identified by NMR and shown in **Scheme 2.14** and **Figure 2.7**.



48 **44** **49**

Scheme 2.14 - Products of Boc deprotection at low pH under methanolic conditions

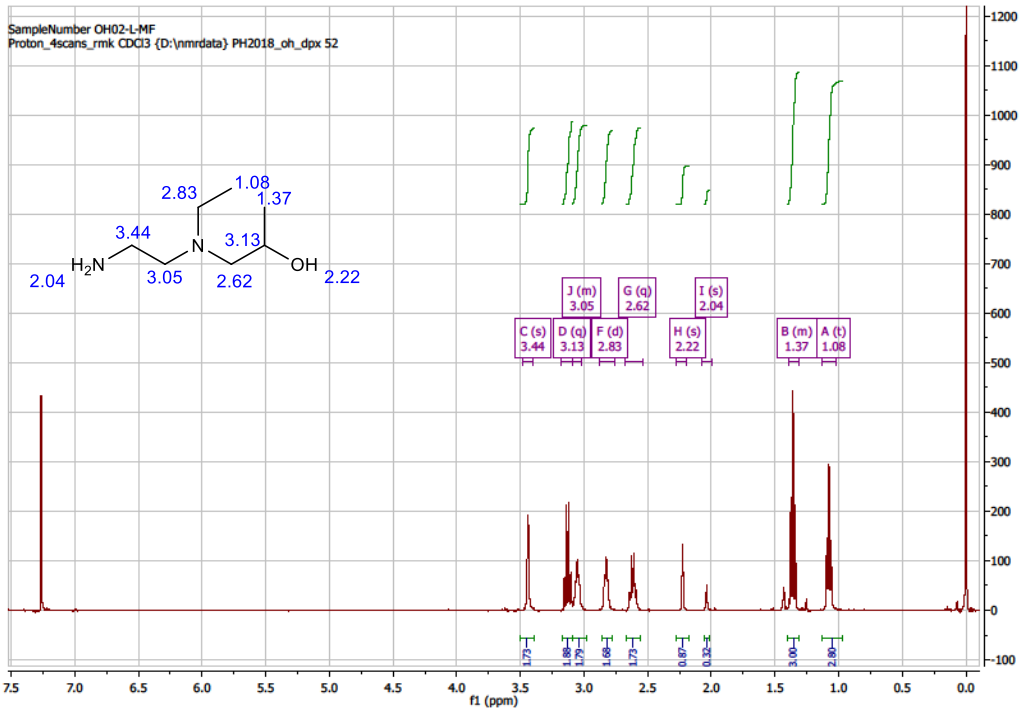


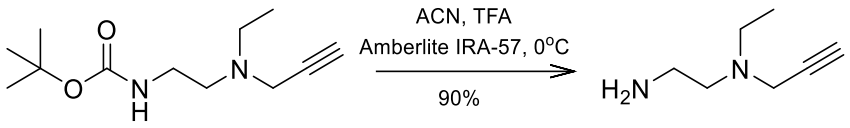
Figure 2.7 -^{1}H NMR spectrum for by-product **49** resulting from acidic deprotection by the use of acidified MeOH

In order to improve reaction yield and reduce the impact of by-products in the formation of multifunctional linker **44**, a series of test reactions was carried out. These contained a variety of conditions both containing HCl:MeOH routes as well as TFA:DCM systems (**Table 2.3**). Reactions were monitored by TLC analysis and it was shown that conditions 1-3 all provided deprotection at a 5 mg scale however upon scale up multifunctional handle **44** exhibited significant acid lability and compound breakdown was seen by the formation of multiple spots on TLC. Conditions 4-6 proved more reliable and facilitated deprotection of multifunctional handle **44** in <30 minutes. Furthermore, upon reaction completion and neutralisation with Amberlite IRA-57 in methanol, methylation was observed and analysis by ^{1}H NMR spectroscopy revealed methylation in the alkyl region of the molecule shown in **Figure 2.7** by the presence of an additional triplet at 1.37 ppm.

Table 2.3 -Test reactions for the deprotection of multifunctional linker **44** carried out on a 5 mg scale. Reactions were carried out using a range of acidic conditions and monitored by TLC, results were visulaised with ninhydrin stain.

Condition	1	2	3	4	5	6
Acid	HCl	HCl	HCl	TFA	TFA	TFA
Solvent	MeOH	MeOH	MeOH	DCM	DCM	DCM
Acid:Solvent ratio	1:2	1:5	1:10	1:1	1:10	1:20
Completion Time (TLC)	N/a	N/a	N/a	2 Minutes	15 Minutes	30 Minutes

Further development of this method in combination with work carried out by Babu G *et al*, in the use of ion-exchange resins in non-methanol phases provided an alternative method for the deprotection of **48** [47]. The resulting deprotection of **48** used Amberlite IRA-57 activated with ACN (**Scheme 2.15**) resulted in the formation of **44** at 90% yield. Characterisation was achieved by ^{1}H NMR spectroscopy which showed the removal of a 9H integration singlet at 1.5 ppm indicating the removal of a Boc group.

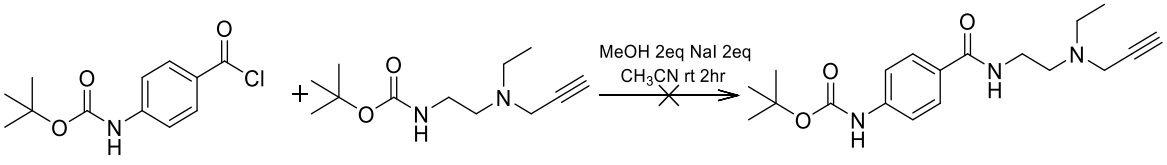


48 **44**

Scheme 2.15 - Deprotection of multifunctional linker **44**, by the use of acidified ACN at 0 °C, yielding **44** at 90%

2.4.5.2 Optimisation of Amide synthesis.

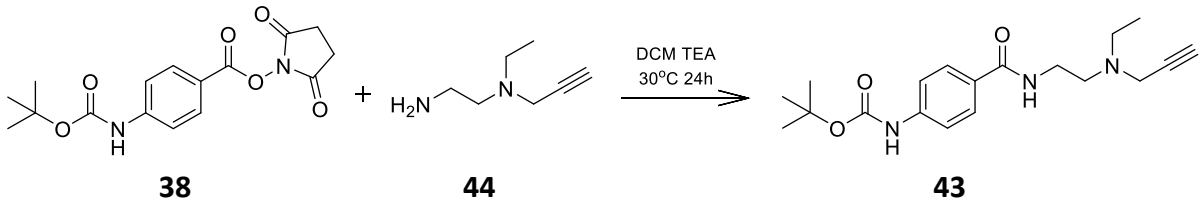
The formation of the amide linkage was carried out by the use of an activated NHS-ester bound to protected compound **37**. The approach aims to increase the product yield while minimising the chances of side reactions and the formation of by-products.



50 **48** **43**

Scheme 2.16 - Proposed synthetic route following the work of Nazih et.al [48] for the generation of an amide required for the formation of multifunctional label **26**.

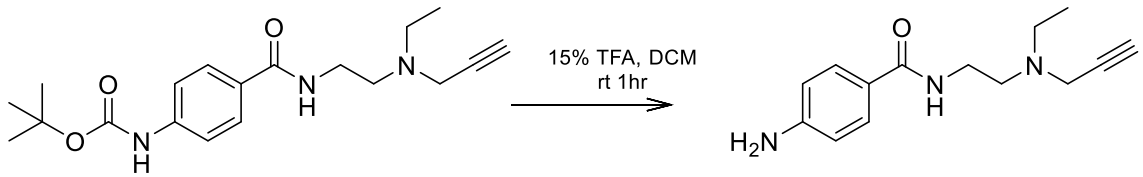
Scheme 2.16 was initially attempted as a one pot reaction between the protected form of multifunctional linker **44** and the protected alkyl chloride **50** [48]. However, the conditions outlined above experienced surprisingly low product yields of 30%. Furthermore, this approach also lead to the formation of multiple by-products due to the additional deprotection of **48.**



38 **44** **43**

Scheme 2.17 -Synthetic route for the addition of multifunctional linker **26** by the use of an activated NHS ester.

Instead, following the successful deprotection of multifunctional linker **44** the purified product was suspended in a solution of DCM (15 mL) basified with TEA (600 µl) and added to a flask containing a solution of the activated NHS ester **38** in DCM. The reaction mixture was stirred under nitrogen overnight at 30 °C yielding compound **43** which was characterised by ^{1}H NMR spectroscopy by the formation of an alkyne peak at 2.19 ppm as well as a quartet at 2.61 ppm at a yield of 70% (**Scheme 2.17**) .

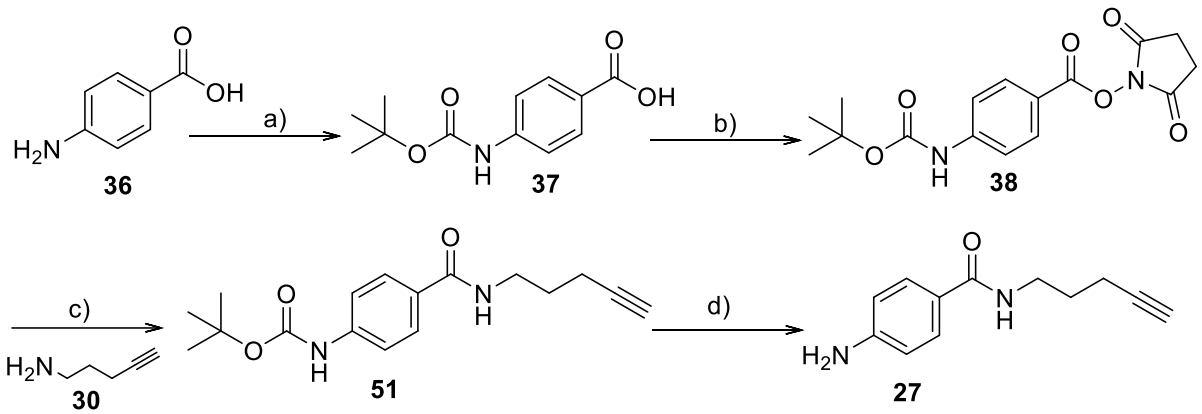


43 **26**

Scheme 2.18 – Deprotection of **43** and the successful production of multifunctional label **26** at 99% yield.

The protected procainamide derivative **43** was then deprotected under acidic conditions (15%TFA in DCM). Following the complete deprotection of **43** the newly formed activated multifunctional label was neutralised by the addition of Amberlite IRA-57 in ACN characterised by the removal of the 9H singlet at 1.51 ppm yielding **26** at 99% yield (**Scheme 2.18**).

2.4.6 Synthesis of multifunctional label 27



Scheme 2.19 -Synthesis of multifunctional label **27** in line with the route previously explored by members of the S. Allman research group over a four stage synthesis a) Boc_2O, 1,4-Dioxane, TEA 99% b) DCM:THF, DCC, NHS 79% c) DCM:TEA, linker **30** 89% d) DCM:TFA 88% .

The preparation of **27** outlined in **Scheme 2.19** aims to integrate with the synthetic workflows already outlined above. The four-step synthesis begins as with compound **26** with the Boc_2O protection of *p*-aminobenzoic acid **36** on a 1000 mg scale (99%). Similarly, following purification, **37** (1g) is then stirred overnight with NHS and DCC to afford the activated ester **38** (79%). This product provides a platform whereby amination with pent-4-yn-amine **30** can take place, forming the Boc protected form of multifunctional label **51** (89%). Deprotection was then carried out with the use of TFA in DCM which was then quenched with amberlite. Following purification, compound **27** was isolated as a pale brown solid (88%).

2.5 Conclusion

This chapter outlines the design and rationale for the production of three multifunctional labels **25**, **26** and **27** for the analysis of glycans by existing liquid chromatography workflows. This began with a comparison of available methods of bioorthogonal labelling techniques displaying high yields under aqueous conditions while maintaining stability in high organic conditions associated with reductive amination labelling. The chapter then moved on to plan and develop commercially achievable synthesis routes for the production of each of the three multifunctional labels. This was followed by the successful synthesis of **25** over a four-step pathway resulting in a total product yield of 50%. The synthesis of both multifunctional labels **26** and **27** followed displaying yields of 54% and 61% respectively. The additional functionality present on labels **25**, **26** and **27** take the form of a terminal

alkyne capable of undergoing CuAAc functionalisation to azide containing structures. The use of azide-alkyne functionalisation will provide a powerful tool in the elucidation of low abundant glycan structures by the addition of hydrophobic regions to aid ionisation and the formation of bespoke glycan derivatisation labels.

2.6 Experimental

2.6.1 General Experimental

IR was performed by attenuated total reflectance infra-red spectroscopy on a Perkin Elmer 100 FTIR spectrometer. Absorption maxima were recorded in wavenumbers (cm^{-1}) and classified as S strong, M medium, W weak, B Broad, N narrow. Only signals representing functional groups were reported, C-H absorption as well as the fingerprint region were not listed.

High resolution mass spectra (HRMS) were obtained by either LCMS or direct infusion on a Thermofisher Scientific Orbitrap XL mass spectrometer in both ESI+ and ESI- modes.

Melting points were recorded digitally by the use of Cole-Palmer Stuart digital melting point apparatus. Reaction monitoring was carried out by thin layer chromatography (TLC) on Merck aluminium backed silica get 60 F254 nm plates. Visualisation was carried out by UV (λ = 254 nm or 302 nm) combined with either potassium permanganate or ninhydrin staining.

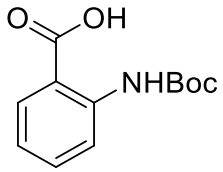
Proton nuclear magnetic resonance (δ_H) spectra were recorded on both a Bruker DPX400 (400 MHz) spectrometer and a Bruker Nano400 (400 MHz) spectrometer. Proton spectra were assigned by COSY were appropriate. Carbon nuclear magnetic resonance (δ_c) spectra were recorded on a Bruker DPX400 (100.7MHZ) spectrometer and were assigned using HSQC and HMBC when necessary. Chemical shifts were quoted on the δ-scale in parts per million (ppm) using the following abbreviations for splitting patterns: s, singlet; d, doublet; dd, doubledoublet; t, triplet; q, quartet; m, multiplet. Coupling constants or J-values were quoted in Hz and rounded to the nearest 0.5 Hz. All NMR solvents were purchased from Merck or Fluorochem.

Anhydrous solvents including acetonitrile, methanol and ethyl acetate were purchased from Acros organics in Acro-seal bottles stored over molecular sieves. All other solvents were of HPLC grade or higher and were purchased from Fisher Scientific, Merck, Alfa Aeser, Acros organics and used

without further purification. All reagents were purchased through Fisher scientific or Merck and were used as supplied.

2.6.2 Synthesis of multifunctional glycan labels

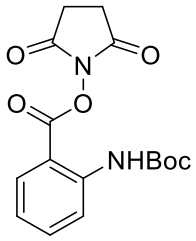
Synthesis of 2-((*tert*-butoxycarbonyl)amino)benzoic acid (28) [33]



2-Anthranilic acid **10** (200 mg, 1.45 mMol) was dissolved in DMF (1.5 mL) and heated to 50 °C. Triethylamine (TEA) (3 mL) and di-*tert*-butyl-dicarbonate Boc_2O (450 mg 2.06 mMol) were added. The reaction mixture was stirred overnight at 50 °C. Following overnight stirring the solvent was evaporated under reduced pressure and the product was resuspended in ethyl acetate. The organic phase was then washed with ice cold sulphuric acid (0.01N, 3 x 5 mL) followed by water (3 x 5 mL). The organic phase was dried with magnesium sulphate, filtered and concentrated under reduced pressure. The resulting white crystalline solid was used without further purification and characterised as product **28** (257 mg, 75% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.47 (d, *J* = 8.5 Hz, 1H, Ar H), 8.09 (d, *J* = 8.0 Hz, 1H, Ar H), 7.56 (t, *J* = 8 Hz, 1H, Ar H), 7.03 (t, *J* = 7.5 Hz, 1H, Ar H), 1.54 (s, 9H).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 152.80 (C=OOH), 150.93(C=O) , 145.35 (NCO), 142.94 (CH Ar) ,140.76 (CH Ar), 135.51(CH Ar), 131.81(CH Ar), 121.30 (CH Ar), 118.90(C-($3CH_3$), 28.33 ($3CH_3$).**IR:** V_{max}/cm^{-1}: 3200 M (C-H, stretch), 3000 VB (O-H stretch), 1700 S (C=O, stretch), 1500 M (C-C, stretch), 1100 S (C-O, Stretch), 800 M (O-H, Bend) **HRMS:** $C_{12}H_{15}NO_4$ Expected [M+Na] 260.1001 found: 260.0893[M+Na] **MP** 182-184 °C Literature MP 148-151 °C[33].

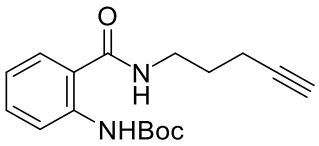
Synthesis of 2,5-dioxopyrrolidin-1-yl 2-((*tert*-butoxycarbonyl)amino)benzoate (29)[33]



Compound **28** (245 mg, 1.035 mmol) and NHS (130mg 1.138 mmol) were dissolved in THF (2.5 mL) and the mixture stirred at 25 °C until **28** had dissolved. DCC (234 mg 1.138 mmol) was added and the reaction mixture stirred at 25 °C overnight. DCU was removed by filtration and the precipitate washed with THF. The reaction mixture was then evaporated to dryness before being resuspended in the mobile phase and purified by flash column chromatography (EtOAc: Petroleum ether 50:50). Compound **29** was isolated as a pale yellow oil which crystallised at -20 °C. (327 mg, 95% yield)

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.43 (s, 1H N-H), 8.44 (d, J = 9.0 Hz, 1H Ar CH), 8.10 (d, J = 8.0 Hz, 1H Ar CH), 7.56 (t, J = 7.5 Hz, 1H) Ar CH, 7.00 (t, J = 7.5 Hz, 1H Ar CH), 2.86 (s, 4H 2 x CH2), 1.44 (s, 9H Boc).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 169.16 (2,C=O), 163.05 (C=OON) , 152.37 (C=O), 143.52 (CH Ar), 136.77 (CH Ar) , 131.16 (CH Ar), 121.53(CH Ar), 119.10 (CH Ar), 109.31 (CH Ar), 81.19 (C-3CH_3), 28.25 (3CH_3), 25.70 (2 CH_2). **IR:** V_{max}/cm^{-1}: 3300 M (N-H, stretch), 3000 M (C-H, stretch), 1700 S (C=O, stretch), 1500 M (C-C, stretch), 1100 S (C=O, stretch), 700 SB (N-H, wing). **HRMS:** $C_{16}H_{18}N_2O_6$ Expected: 357.1164 [M+Na] found: 357.1057 [M+Na]. **MP** 158-162 °C.

Synthesis of *tert*-butyl (2-(pent-4-yn-1-ylcarbamoyl)phenyl)carbamate (31)

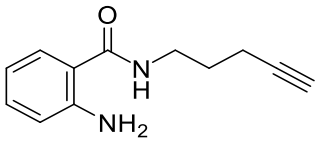


Product **29** (200 mg, 0.6 mmol) was dissolved in DCM (15 mL) and stirred at 25 °C. This was followed by the addition of 4-pentyl-1-amine **30** (138 mg) and TEA (600 µL). The solution was then stirred overnight. The reaction mixture was concentrated under reduced pressure before being dry-loaded and purified by flash column chromatography (EtOAc: Petroleum ether 50:50). The title compound **31** was obtained as a white solid (133 mg, yield 74%)

^{1}H NMR (400 MHz, $CDCl_3$) δ 10.19 (s, 1H N-H), 8.39 (t, J = 8.0 Hz, 1H Ar CH), 7.53 – 7.38 (m, 2H Ar CH), 7.01 (t, J = 7.5 Hz, 1H Ar CH), 6.48 (s, 1H N-H), 3.61 (q, J = 6.5 Hz, 2H CH_2), 2.38 (td, J = 7.0, 2.5 Hz, 2H CH_2), 2.07 (t, J = 2.5 Hz, 1H ≡C*H*), 1.91 (p, J = 7.0 Hz, 2H CH_2), 1.55 (s, 9H Boc).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 169.06 (C=ONH), 153.15(C=O), 140.33(C-NH Ar), 132.51 (CH Ar), 126.41 (CH Ar) ,

121.36 (CH Ar), 119.95 (CH Ar), 119.79 (CH Ar), 83.48 (C≡C), 80.28 (C-CH_3), 69.59 (C≡C-H), 39.27 (NCH_2), 28.37 ($3CH_3$) , 27.79 (CH_2), 16.35 (CH_2). **IR:** V_{max}/cm^{-1} 3300 M (N-H, stretch), 3200 NS (C-H, stretch), 1700 S (C=O, stretch), 1500 M (C-C, stretch), 1200 V (C-N, stretch), 1100 S (C=O, stretch), 700 SB (N-H, wing). **HRMS**: $C_{17}H_{22}N_2O_3$ Expected 325.1631 [M+Na], Found: 325.1515 [M+Na]. **MP** 98-102 ^oC

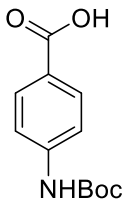
Synthesis of 2-amino-*N*-(pent-4-yn-1-yl)benzamide (25)



Product **31** (135 mg, 0.44 mmol) was dissolved in TFA:DCM 1:6 v/v(2 mL) and left to stir at room temperature for 1.5 h. The reaction mixture was quenched with Amberlite® IRA-67 (Sigma Aldrich), filtered and purified by flash column chromatography (EtOAc: Petroleum ether 50:50).Title compound was obtained as a yellow crystalline structure **25** (85 mg, yield 95%)

^{1}H NMR (400 MHz, MeOD) δ 7.32 (d, *J* = 8.0 Hz, 1H, Ar CH), 7.07 (t, *J* = 7.5 Hz, 1H, Ar CH), 6.64 (d, *J* = 8.0 Hz, 1H, Ar CH), 6.52 (t, *J* = 7.5 Hz, 1H, Ar CH), 3.32 (t, *J* = 7.0 Hz, 2H, CH_2), 2.20 – 2.13 (m, $3HCH_2$≡CH), 1.71 (q, *J* = 7.0 Hz, 2H, CH_2).**^{13}C NMR (101 MHz, MeOD)** δ 172.13 (C=ONH), 150.17 (C=O), 133.08 (C-NH_2 Ar), 129.01 (CH Ar), 118.16 (CH Ar), 117.87 (CH Ar), 117.43 (2CH Ar), 84.34 (C≡C), 69.99 (C≡C-H), 39.77 (CH_2) 29.60 (CH_2), 16.83.**IR:** V_{max}/cm^{-1} 3300 M (N-H, stretch), 3200 NS (C-H, stretch), 3000 M (C-H, stretch), 1700 S (C=O, stretch), 1200 V (C-N, stretch), 700 SB (N-H, wing). **HRMS:** $C_{12}H_{14}N_2O$ Expected: 203.1106 [M+H] found: 203.1179 [M+H]. **MP** 100-105 ^oC

4-(*tert*-Butoxycarbonylamino)benzoic acid (37)[49]

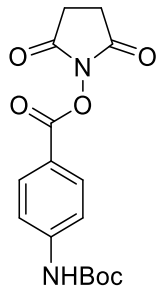


p-Aminobenzoic acid **36** (1000 mg, 7.29 mmol) was dissolved in water (8 mL) and 1,4 dioxane (8 mL). The resultant solution was cooled to 0 ^oC and basified using triethylamine (1033 mg, 10.20 mmol 1.4eq). A solution of Boc_2O (2226.05 mg, 10.20 mmol, 1.4eq) in 1,4 dioxane (8 mL) was added dropwise to the reaction mixture. The reaction was stirred overnight and allowed to warm to RT

after 4 hours. After overnight stirring the product was precipitated from solution as a white solid using 2M HCl (8 mL). The solid product was filtered out and washed with a large amount of cool 2M HCl, followed by water and petroleum ether. Yielding white solid **37** (1745 mg, 7.36 mmol, 99%)

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.03 (d, *J* = 8.5 Hz, 2H, 2xAr CH), 7.46 (d, *J* = 8.0 Hz, 2H, 2xAr CH), 1.53 (s, 9H, Boc).**^{13}C NMR (101 MHz, DMSO-d6)** δ 167.48 (-C=OOH), 153.01 (NCO), 144.24 (CH Ar), 130.82 (CH Ar), 124.43 (CH Ar), 80.12 (C-($3CH_3$), 117.68 (CH Ar), 28.51 ($3CH_3$). **IR: V_{max}/cm^{-1}** 3200, M (C-H, Ar stretch), 3000 VB (O-H stretch), 1700 S (C=O stretch), 1500 M (C-C, Ar stretch) 1100 S (C-O, Stretch) 800 SB (N-H, Wing). **HRMS** $C_{12}H_{15}NO_4$ Expected [M+Na] 260.0899, found 260.089 [M+Na]. **MP** 190-193 °C

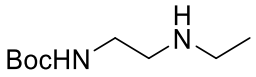
2,5-Dioxopyrrolidin-1-yl 4-(*tert*-butoxycarbonylamino)benzoate (38)



37 (1000 mg, 4.21 mMol) was dissolved in DCM (10 mL) and THF (4 mL) and the flask was flushed with N_2. DCC (870.09 mg, 4.217 mmol, 1eq) was added along with NHS (485.30 mg, 4.217 mmol, 1eq). The reaction was stirred at RT overnight and a white precipitate formed. The DCU precipitate was removed from the completed reaction mixture by filtration before the remaining solution was concentrated under reduced pressure. Purification was carried out by flash column chromatography on a 10 cm silica column with ethylacetate: petroleum ether 1:1. This afforded the product **38** as a white solid (1113 mg, 3.33 mmol, 79%)

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.05 (d, *J* = 9.0 Hz, 2H, 2xAr CH), 7.50 (d, *J* = 8.5 Hz, 2H, 2xAr CH), 2.90 (s, 4H 2x CH_2), 1.53 (s, 9H, Boc).**^{13}C NMR (101 MHz, $CDCl_3$)** δ ppm 169.60 (2 ONC=O), 161.53 (Ar-COO), 152.08 (C=O), 144.82 (CH Ar), 132.28 (CH Ar), 118.88 (CH Ar), 117.66 (CH Ar), 81.78 (C-CH_3), 28.42 (3 CH_3), 25.85 (2 CH_2). **IR:** V_{max}/cm^{-1} 3200 M (C-H, Ar stretch), 1700 S (C=O, Stretch), 1500 M (C-C, Ar stretch), 1200 S (C-O, Stretch). **HRMS** $C_{16}H_{18}N_2O_6$ expected [M+Na] 357.1165 found: 357.1061 [M+Na]. **MP** 170-171 °C

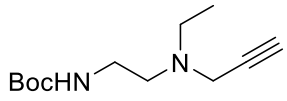
***tert*-Butyl (2-(ethylamino)ethyl)carbamate (47)[46]**



Acidified methanol was made by stirring MeOH (3.5 mL) with cHCl (779 µL) for 15 minutes. The solution was then cooled to 0 °C and *N,N*-ethylethylenediamine **39** (837 mg, 9.50 mmol) was added. The mixture was stirred for a further 15 minutes before water (1 ml) was added. This solution was stirred for a further 30 minutes. A solution of Boc_2O (2071 mg, 9.49 mmol, 1eq) in MeOH was added to the reaction mixture and stirred for 1.5 hours. The reaction mixture was concentrated under reduced pressure and resuspended in diethylether (30 ml) and neutralised with 2M NaOH (2x50 mL). The aqueous phase was collected and washed with DCM (4x30 mL). Concentration of the organic phase revealed a yellow oil **47** which crystallised when cooled (1369.18 mg, 7.277 mmol,76%).

^{1}H NMR (400 MHz, MeOD) δ 3.09 (t, J = 6.5 Hz, 2H CH_2), 2.55 (dt, J = 14.5, 7.0 Hz, 4H 2 x CH_2), 1.35 (s, 9H Boc), 1.03 (t, J = 7.0 Hz, 3H CH_3).**^{13}C NMR (101 MHz, MeOD)** δ 157.13 (C=O), 99.68 (C-3CH_3), 78.65 (CH_2), 43.05 (CH_2), 39.38 (CH_2), 27.34 (3CH_3), 13.37 (CH_3). **IR:** V_{max}/cm^{-1} 3400 M (N-H Stretch) 2900 M (C-H) 1600 M (N-H stretch) **HRMS** $C_9H_{20}N_2O_2$ expected [M+H] 189.1523 found: **MP**<25°C

***tert*-butyl (2-(ethyl(prop-2-yn-1-yl)amino)ethyl)carbamate (48)**

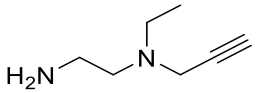


Compound **47** (304 mg, 1.61 mmol) was added to a round bottom flash along with potassium carbonate (446.38 mg, 3.23 mmol, 2 eq). The flask was flushed with N_2 and dry ACN was added, and the solution was stirred for 5 minutes. The flask was kept in the dark for the duration of the rection time. Propargyl bromide (237.35 mg, 1.77 mMol 1.1eq) was added dropwise wise maintaining an inert atmosphere. The resultant mixture was stirred at RT for 36 hours after which time potassium carbonate was filtered off and the reaction mixture was concentrated. The crude reaction mixture

was then dry loaded onto a 15 cm silica column with ethyl acetate and petroleum ether (1:1). Fraction containing **48** were pooled and concentrated under reduced pressure, product was isolated as a yellowish oil (280 mg, 1.23 mmol 77%)

^{1}H NMR (400 MHz, MeOD) δ 3.34 (d, J = 2.5 Hz, 2H CH_2), 3.21 (dt, J = 3.0, 1.5 Hz, 1H ≡C*H*), 3.06 (t, J = 7.0 Hz, 2H CH_2), 2.50 (q, J = 7.0 Hz, 4H 2 x CH_2), 1.33 (s, 9H Boc), 0.97 (t, J = 7.0 Hz, 3H CH_3).**^{13}C NMR (101 MHz, MeOD)** δ 156.99 (C=O) , 78.64 (C-3CH_3) , 77.28 (C≡C), 73.38 (C≡C-H), 52.32 (CH_2) , 47.12 (CH_2), 40.67 (CH_2), 37.65 (CH_2), 27.36 (3CH_3), 11.18 (CH_3). **IR:** V_{max}/cm^{-1} M 3400 (N-H Stretch), M 2900 (C-H Stretch) M 1400 (C-H Bend, Rock) M 1100 (C-N Stretch). **HRMS** $C_{12}H_{22}N_2O_2$ expected [M+H] 226.1681 found: 226.1742. **MP**<25^{o}C

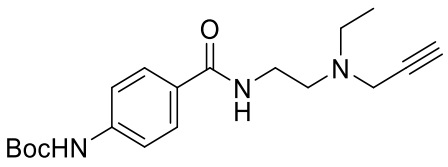
N1-ethyl-N1-(prop-2-yn-1-yl)ethane-1,2-diamine (44)



Compound **48** (200 mg, 0.84 mmol) in DCM (4.5 mL) was added to a round bottom flask and stirred at >10 ^{o}C. TFA (1.5 mL) was added dropwise over a 5-minute period while stirring continued. The reaction was monitored by TLC. After 30 minutes deprotection was deemed complete and the reaction mixture was concentrated under reduced pressure. The reaction mixture was reconstituted in ACN (10 mL) and stirred with amberlite IRA-57 (10 g) washed in ACN. Neutrality was achieved after stirring for 1 hour and the Amberlite was removed by filtration. The resulting solution was then concentrated to reveal a dark brown oil **44** (74 mg, 0.58 mmol, 70%)

^{1}H NMR (400 MHz, D2O) δ 4.10 (d, J = 2.5 Hz, 2H CH_2), 3.58 – 3.53 (m, 2H CH_2), 3.41 – 3.31 (m, 4H 2x CH_2), 3.08 (t, J = 2.5 Hz, 1H ≡C*H*), 1.89 (s, 2H NH_2), 1.26 (q, J = 7.5 Hz, 3H CH_3).**^{13}C NMR (101 MHz, D_2O)** δ 80.03 (CH_2-**C**≡C*H*), 71.24 (C≡**C**H), 49.37 (CH_2-**C**H_2-N), 48.79 (N-**C**H_2-CH_3), 41.98 (N-**C**H_2-CH), 34.02 (NH_2-**C**H_2-CH_2), 8.80(CH_2-**C**H_3).**IR:** V_{max}/cm^{-1} M 3400 (N-H Stretch), M 2900 (C-H Stretch) M 1400 (C-H Bend, Rock) M 1100 (C-N Stretch). **HRMS** $C_7H_{14}N_2$ expected [M+H] 127.1157 found: 127.2110. **MP** <25 ^{o}C.

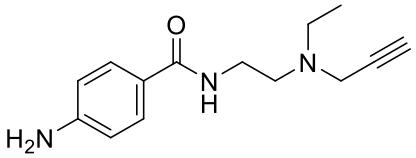
***tert*-Butyl (4-((2-(ethyl(prop-2-yn-1-yl)amino)ethyl)carbamoyl)phenyl)carbamate 43**



Compound **38** (200 mg, 0.59 mmol) was dissolved in DCM (13 mL) and basified using TFA (414 µL, 2.99 mmol 5eq). A solution of **44** (90.5 mg, 0.718 mmol 1.2 eq) in DCM (2 mL) was then added and the resultant solution was stirred overnight at room temperature. Reaction mixture was then concentrated under reduced pressure and purified on a 10 cm silica flash column (Petroleum ether: Ethyl acetate 3:7). After which fractions where concentrated under reduced pressure to reveal a brown Oil **43** which was used without further purification. (275 mg, 0.79 mmol, 75%)

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.72 (d, J = 8.5 Hz, 2H Ar 2 x CH), 7.46 (d, J = 8.5 Hz, 2H Ar 2 x CH), 3.50 (dd, J = 11.0, 5.5 Hz, 2H CH_2), 3.43 (d, J = 2.5 Hz, 2H CH_2), 2.75 (t, J = 6.0 Hz, 2H CH_2), 2.60 (q, J = 7.0 Hz, 2H CH_2), 2.21 (t, J = 2.5 Hz, 1H ≡C*H*), 1.51 (s, 9H Boc), 1.07 (t, J = 7.0 Hz, 3H CH_3).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 166.85(C-**C**O-NH), 152.42 (O-**C**O-NH), 141.40 (HN-**C** Ar), 128.79 (**C**-CO Ar), 128.03 (2 x C-**C**H-CH Ar), 117.74 (2 x CH-**C**H-C Ar), 80.99 (3CH_3-**C**-O) , 78.40(CH_2-**C**≡CH), 73.10 (C≡**C**H), 51.66 (NH-**C**H_2-CH_2), 47.35 (CH_2-**C**H_2-N), 41.30 (N-**C**H_2-CH_3), 37.04 (N-**C**H_2-C), 28.30 (3 x **C**H_3) , 12.78(CH_2-**C**H_3). **HRMS** $C_{19}H_{27}N_3O_3$ expected [M+Na] 368.2052 found: 369.2150

4-amino-N-(2-(ethyl(prop-2-yn-1-yl)amino)ethyl)benzamide (26)

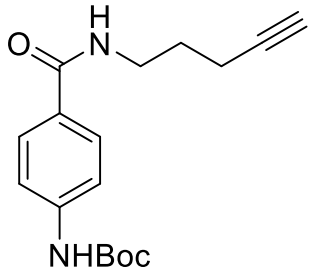


Compound **43** (150 mg, 0.43 mmol) was dissolved in TFA:DCM 1:6 v/v (2 mL)and left to stir at room temperature for 1.5h. Reaction mixture was then concentrated under reduced pressure and resuspended in ACN (10 mL). Amberlite IRA-57 (10 g) was added, and neutralisation was achieved after 2 hours. Amberlite was removed by filtration and solvent was concentrated under reduced pressure to reveal a brown Oil which was resolubilised in mobile phase and purified by flash column

chromatography on a 10 cm silica column run with petroleum ether: ethyl acetate (1:9) to reveal a pale brown oil which later crystallised when chilled **26**. (104 mg, 0.42 mmol, 99%)

^{1}H NMR (400 MHz, CDCl3) δ 7.61 (d, J = 8.5 Hz, 2H Ar 2 x CH), 6.66 (d, J = 8.5 Hz, 2H Ar 2 x CH), 3.49 (dd, J = 11.0, 5.5 Hz, 2H CH_2), 3.43 (d, J = 2.0 Hz, 2H CH_2), 2.74 (t, J = 6.0 Hz, 2H CH_2), 2.60 (q, J = 7.0 Hz, 2H CH_2), 2.20 (t, J = 2.5 Hz, 1H ≡C*H*), 1.08 (t, J = 7.0 Hz, 3H CH_3).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 167.18(C=O), 149.44(C-NH_2 Ar), 128.66 (CH_2 Ar), 124.26(C-CON Ar), 114.12(CH_2 Ar), 78.43(C≡C), 73.03 (C≡C-H), 51.76(NCH_2), 47.28(CH_2-CH_3), 41.26(NCH_2CH_2), 36.91(NCH_2C≡C),12.81 (CH_2CH_3). **IR:** V_{max}/cm^{-1} M 3340 (N-H Stretch), M 3195 (N-H Stretch), M 2974 (C-H Stretch), M 2915 (C-H Stretch) W 2215 (C≡C Stretch), M 1562 (N-H Bend) M 1498 (C=C Ar Bend). **HRMS** $C_{19}H_{27}N_3O_3$ expected [M+H] 246.1528 found: 246.1530 [M+H]. **MP** 170-173 °C

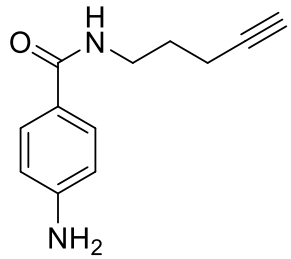
4-(tert-Butoxycarbonylamino)benzoate (51)



Compound **38** (100 mg, 0.3 mmol) was dissolved in DCM (7.5 mL) and stirred at 25 °C followed by the addition of 4-pentyl-1-amine (54 mg 0.66 mmol) and TEA (300 µL). The solution was the stirred overnight. RM was concentrated under reduced pressure and purified by flash column chromatography (EthOAc: Petroleum ether 50:50) Compound (**51**) was obtained as a white solid (80 mg, 0.26 mmol 89% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.71 (d, *J* = 8.5 Hz, 2H, Ar 2 x CH), 7.42 (d, *J* = 8.5 Hz, 2H, Ar 2 x CH), 6.61 (s, 1H, NH), 3.58 (q, *J* = 6.5 Hz, 2H CH_2), 2.32 (td, *J* = 7.0, 2.5 Hz, 3H, C*H*$_2$C≡CH), 2.02 (dd, *J* = 7.5, 5.0 Hz, 1H, , ≡C*H*), 1.86 (p, *J* = 7.0 Hz, 2H, CH_2), 1.52 (s, 9H, Boc). **IR:** V_{max}/cm^{-1} 3300 M (N-H, stretch), 3200 M (C-H, Ar stretch), 1700 S (C=O, Stretch), 1600 M (N-H, Ar bend), 1500 M (C-C, Ar stretch), 1200 V (C-N, stretch), 1100 S (C=O, stretch). **HRMS** $C_{17}H_{22}N_2O_3$ expected [M+H] 303.1630 found: 303.1641 [M+H]. **MP** 98-100 °C.

4-amino-N-(pent-4-yn-1-yl)benzamide (27)



Compound **51** (80 mg 0.26 mmol) was dissolved in TFA:DCM 1:6 v/v (2 mL) and left to stir at room temperature for 1.5 h. Reaction mixture was quenched with Amberlite® IRA-67 (Sigma Aldrich), filtered and purified by flash column chromatography (EtoAC: Petroleum ether 50:50).) Title compound was obtained as a yellow crystalline structure **27** (48 mg 88% yield)

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.59 (d, *J* = 8.4 Hz, 2H, 2xAr H), 6.65 (d, *J* = 8.4 Hz, 2H, 2xAr H), 6.25 (s, 1H, NH), 3.55 (q, *J* = 6.5 Hz, 2H CH_2), 2.30 (td, *J* = 6.9, 2.6 Hz, 2H, CH_2), 2.01 (t, *J* = 2.5 Hz, 1H, , ≡C*H*), 1.84 (p, *J* = 6.8 Hz, 2H, ≡C*H*). **^{13}C NMR (101 MHz, MeOD-4)** δ ppm 170.5 (C=O), 153.0 (C-NH_2 Ar), 129.9 (CH Ar), 123.3 (CH-C=O Ar), 114.7 (2 CH Ar), 84.4 (C≡C), 69.9 (C≡C-H), 40.0 ($NHCH_2$), 29.7 (CH_2CH_2C≡C), 16.8 ($CH_2CH_2CH_2$). **IR:** V_{max}/cm^{-1} 3300 M (N-H, stretch), 3200 NS (C-H, stretch), 3100 M (C-H, Ar stretch), 1700 S (C=O, Stretch), 1500 M (C-C, Ar stretch), 1100 V (C-N stretch), 1200 V (C-N, stretch), 700 SB (N-H, wing) **HRMS** $C_{12}H_{14}N_2O$ expected [M+H] 203.1106 found: 203.1179 [M+H]. **MP** 98-105 ^{o}C

2.7 References

1. I. Turyan, X. Hronowski, Z. Sosic and Y. Lyubarskaya, *Anal. Biochem.*, 2014, **446**, 28-36.
2. K. Xiao, Y. Han, H. Yang, H. Lu and Z. Tian, *Anal. Chim. Acta*, 2019, **1091**, 1-22.
3. T. Keser, T. Pavić, G. Lauc and O. Gornik, *Front. Chem*, 2018, **6**, 324-324.
4. L. R. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A. M. Deelder and M. Wuhrer, *Anal. Bioanal. Chem*, 2010, **397**, 3457-3481.
5. Y. Kita, Y. Miura, J.-i. Furukawa, M. Nakano, Y. Shinohara, M. Ohno, A. Takimoto and S.-I. Nishimura, *Mol. Cell Proteomics*, 2007, **6**, 1437-1445.
6. K.-i. Yoshino, T. Takao, H. Murata and Y. Shimonishi, *Anal. Chem.*, 1995, **67**, 4028-4031.
7. T. Takao, Y. Tambara, A. Nakamura, K. I. Yoshino, H. Fukuda, M. Fukuda and Y. Shimonishi, *Rapid Commun. Mass. Sp.*, 1996, **10**, 637-640.
8. M. Pabst, D. Kolarich, G. Pöltl, T. Dalik, G. Lubec, A. Hofinger and F. Altmann, *Anal. Biochem.*, 2009, **384**, 263-273.
9. S. Klapoetke, J. Zhang, S. Becht, X. Gu and X. Ding, *J Pharm. Biomed. Anal.*, 2010, **53**, 315-324.
10. R. P. Kozak, C. B. Tortosa, D. L. Fernandes and D. I. R. Spencer, *Anal. Biochem.*, 2015, **486**, 38-40.
11. R. A. Kalyan, *Anal. Biochem.*, 2006, **350**, 1-23.
12. F. Higel, U. Demelbauer, A. Seidl, W. Friess and F. Sörgel, *Anal. Bioanal. Chem.*, 2013, **405**, 2481-2493.
13. J. Krenkova, F. Dusa and R. Cmelik, *Electrophoresis*, 2020, **41**, 684-690.
14. C. J. Brown and K. Y. Lonsdale, *P. Roy. Soc. A-Math. Phy.*, 1968, **302**, 185-199.
15. S. Kamoda, R. Ishikawa and K. Kakehi, *J. Chromatogr. A*, 2006, **1133**, 332-339.
16. H. M. Kayili, *Int. J. Mass. Spectrom.*, 2020, **457**, 116412.
17. S. H. Walker, L. M. Lilley, M. F. Enamorado, D. L. Comins and D. C. Muddiman, *J. Am. Soc. Mass. Spectr.*, 2011, **22**, 1309-1317.
18. J. E. Hein and V. V. Fokin, *Chem. Soc. Rev.*, 2010, **39**, 1302-1315.
19. J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, *P. Natl. Acad. Sci. U.S.A,* 2007, **104**, 16793-16797.
20. K. L. Kiick, E. Saxon, D. A. Tirrell and C. R. Bertozzi, *P. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 19-24.
21. H. Wu and N. K. Devaraj, *Top. Curr. Chem.*, 2016, **374**, 3.
22. M. Danese, M. Bon, G. Piccini and D. Passerone, *Phys. Chem. Chem. Phys.*, 2019, **21**, 19281-19287.
23. R. Huisgen, *Angew. Chem. Int. Edit.*, 1963, **2**, 565-598.
24. R. A. Firestone, *J. Org. Chem.*, 1968, **33**, 2285-2290.
25. H. Wang, Y. Ning, Y. Sun, P. Sivaguru and X. Bi, *Org. Let.*, 2020, **22**, 2012-2016.
26. V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem. Int. Edit.*, 2002, **41**, 2596-2599.
27. Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless and M. Finn, *J. Am. Chem. Soc.*, 2003, **125**, 3192-3193.
28. H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem. Int. Edit.*, 2001, **40**, 2004-2021.
29. N. Z. Fantoni, A. H. El-Sagheer and T. Brown, *Chem. Rev.*, 2021, **121**, 7122-7154.
30. C. T. Yuen, Y. Zhou, Q. Z. Wang, J. F. Hou, A. Bristow and J. Z. Wang, *Biologicals*, 2011, **39**, 396-403.
31. M. Allen, Doctor of Philosophy, The Open University, 2022.
32. D. W. Domaille and J. N. Cha, *Chem. Commun.*, 2014, **50**, 3831-3833.
33. *US Pat., US*20150335759, 2015.
34. E. Gelens, L. Smeets, L. A. J. M. Sliedregt, B. J. van Steen, C. G. Kruse, R. Leurs and R. V. A. Orru, *Tetrahedron Lett.*, 2005, **46**, 3751-3754.

35. A. V. Kolotaev, K. R. Matevosyan, V. N. Osipov and D. S. Khachatryan, *Tetrahedron Lett.*, 2019, **60**, 151315.
36. Y.-W. Wang, L. Zheng, F.-C. Jia, Y.-F. Chen and A.-X. Wu, *Tetrahedron Lett.*, 2019, **75**, 1497-1503.
37. J.-M. Huang, H. Chen and R.-Y. Chen, *Synthetic Commun.*, 2002, **32**, 2215-2225.
38. F. Marighetti, K. Steggemann, M. Karbaum and M. Wiese, *ChemMedChem*, 2015, **10**, 742-751.
39. Y. Li, C. Zhan, B. Yang, X. Cao and C. Ma, *Synthetic Commun.*, 2013, **45**, 111-117.
40. S. Rasheed, D. N. Rao, A. S. Reddy, R. Shankar and P. Das, *RSC. Adv.*, 2015, **5**, 10567-10574.
41. R. Baltzly and C. Ferry, *J. Am. Chem. Soc.*, 1942, **64**, 2231-2232.
42. T. Portada, D. Margetić and V. Štrukil, *Molecules*, 2018, **23**.
43. *US Pat., US*7858399B2, 2010.
44. L. T. M. Hoang, L. H. Ngo, H. L. Nguyen, H. T. H. Nguyen, C. K. Nguyen, B. T. Nguyen, Q. T. Ton, H. K. D. Nguyen, K. E. Cordova and T. Truong, *Chem. Commun.*, 2015, **51**, 17132-17135.
45. A. P. Krapcho, M. J. Maresch and J. Lunn, *Synthetic Commun.*, 1993, **23**, 2443-2449.
46. D. W. Lee, H. J. Ha and W. K. Lee, *Synthethic Commun.*, 2007, **37**, 737-742.
47. N. Babu G, W. Belay and T. Endale, *Cogent Chem.*, 2019, **5**, 1708160.
48. A. Nazih and D. Heissler, *Synthesis*, 2002, **2002**, 0203-0206.
49. F. Mu, S. L. Coffing, D. J. Riese, R. L. Geahlen, P. Verdier-Pinard, E. Hamel, J. Johnson and M. Cushman, *J. Med. Chem.*, 2001, **44**, 441-452.

Chapter 3- Quantitative carbohydrate analysis

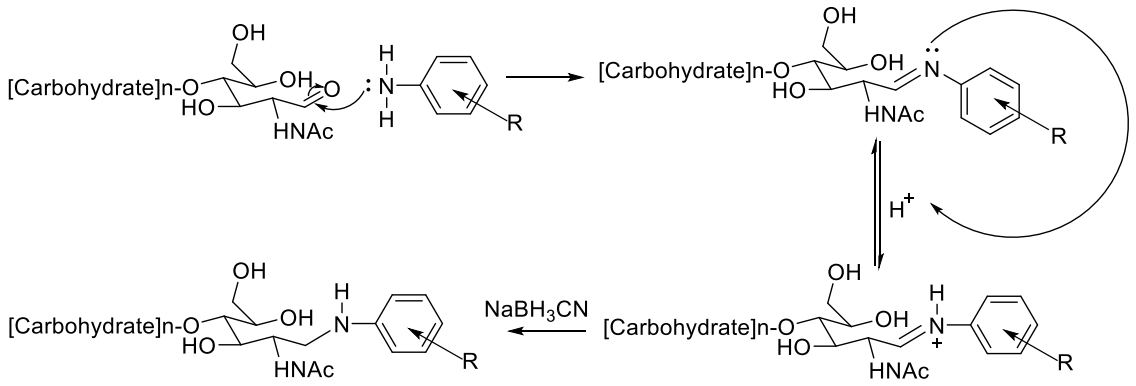
3.1 Introduction

The use of oligosaccharide derivatisation within the field of quantitative oligosaccharide analysis has expanded in recent years with the increase in the effectiveness of biologically derived therapeutic pharmaceuticals as well as biomarker led diagnosis [1-3-5]. While the use of derivatisation agents in qualitative glycan profiling is already well established within scientific literature [6,7], quantitative methods require a significantly higher level of precision in order to be of commercially viable use.

The high polarity of native unmodified oligosaccharides provides a significant analytical challenge, poor separation has been widely reported on reversed phase chromatography while the lack of any fluorescent characteristics makes detection challenging [8,9]. To facilitate reliable separation and detection of oligosaccharides, reducing end modification routinely sees the addition of a fluorophore containing label via a derivatisation agent. The resultant biomolecule is a fluorescent oligosaccharide bearing a region of increased hydrophobicity facilitating both enhanced ionisation and separation [10,11].

Depending upon application the pharmaceutical industry employs a variety of derivatisation agents in order to facilitate the detection of oligosaccharides[12]. These range from single site specific methods such as Michael addition [13] and reductive amination [14] to multi-site conversions such as per-*O*-methylation [10, 15].

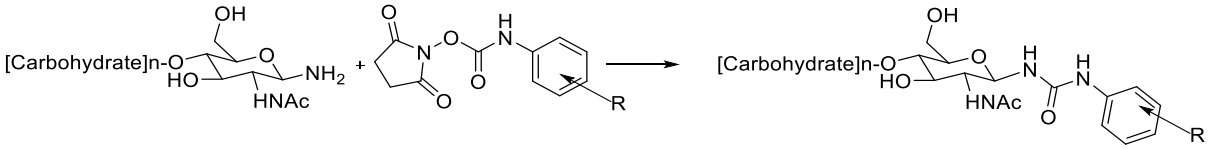
Modern analytical approaches have focused heavily upon single site specific methodologies that produce stable products with efficient and non-reversible product formation. This has seen an acceleration in the research of two branches of coupling chemistry. The first is reductive amination labelling; this process sees the conversion of the primary amine present on the labelling compound to a stable secondary amine between reducing end and label (**Scheme 3.1**) [16 17].



Scheme 3.1-Reductive amination labelling of reducing carbohydrate with a primary amine containing an aromatic label

This takes place via the formation of an unstable imine intermediate which is then reduced, typically by an excess of sodium cyanoborohydride or picoline borane to form a stable secondary amine between both the biologically active carbohydrate and the labelling group [18]. Existing methodologies from both commercial and research settings have carried out derivatisation reactions at between 65 °C and 85 °C at times ranging from 1 to 3 hours depending upon reagent and incubation temperature.

The second classification of coupling labels sees the use of NHS-containing activated carbamate labelling strategies (**Scheme 3.2**). These labels target the glycosylamine generated by the release of oligosaccharides from protein following deglycosylation by PNGase F. These labelling strategies include compounds such as Rapiflour-MS and provide a rapid, mild and effective labelling strategy for released glycan samples [19-21]. These labels are commonly used in pharmaceutical workflows and bring benefits of being driven by speed and are seldom limited by sample amount.



Scheme 3.2 - NHS-carbamate labelling of aminoglycans resulting from PNGase F digestion

While rapid labelling provides a faster rate of derivatisation for some applications, reductive amination labelling is still widely used within biopharmaceutical and research applications. Both labelling strategies form selective single site chemical modification of the target glycan.

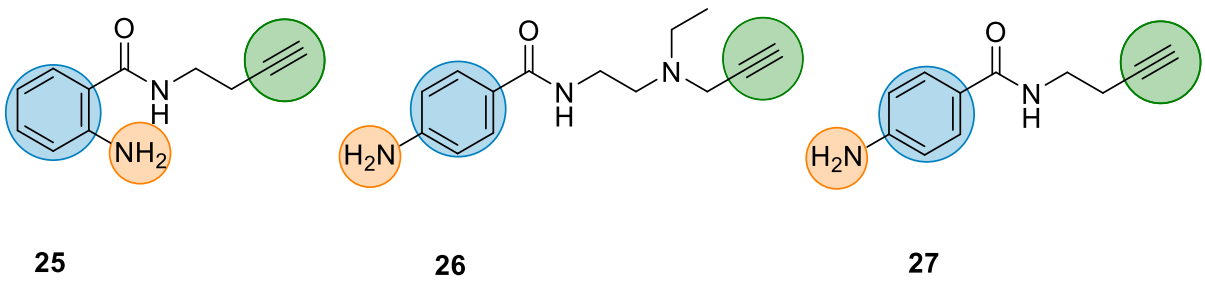


Figure 3.1 -Structures of multifunctional labels **25**, **26** and **27**, highlighting the compounds primary amine in orange, fluorophore in blue and covalently active alkyne in green.

Multifunctional labels **25**, **26** and **27** fall into a widely adopted category of glycan labels utilizing reductive amination as a primary means of permanently reacting with oligosaccharide compounds. This mechanism sees the utilisation of the compound's primary amine highlighted by the orange region in **Figure 3.1**to carry out site specific reductive amination imparting a single fluorophore shown in blue to the reducing end of the released biologically significant oligosaccharides. This reaction mechanism ensures site specific label addition as each glycan structure exhibits only a single reducing end and therefore a single site available for modification. The result is a proportional increase in fluorescence with increasing sample concentration. In addition to selective reactions with released carbohydrates, **25**, **26** and **27** bear an additional alkyne functional group identified in green. This remains active following carbohydrate labelling and therefore facilitates downstream processing, either in surface functionalisation or to aid downstream carbohydrate analysis and improved detection of low abundance species.

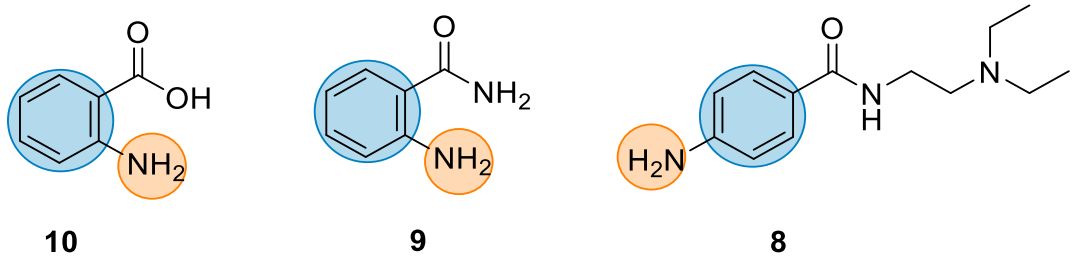


Figure 3.2 - Structures of commercial labels **10**, **9** and **8** highlighting the compounds primary amine in orange, fluorophore in blue.

Existing biopharmaceutical workflows centring around reductive amination labelling are routinely carried out utilising a variety of structurally different labelling strategies. These workflows range from structural determination by oligosaccharide ring number (GU) to deeper structural elucidation by MS fragmentation. 2-Aminobenzamide (2AB) **9** shown in **Figure 3.2** represents the standard for LC/FL based characterisation as it exhibits high levels of stability in solution making it suitable for the

production of retention standards [22]. However its small size, and neutral charge in solution, make it unsuitable for many mass spectrometry (MS) applications [23]. 2AB **9** is best applied in methods where oligosaccharide size and sialylation are already determined and results can be validated against a predetermined standard [24]. This application can be applied to the majority of reductive amination fluorescence labels.

When deeper structural analysis is required or when the sample is of unknown glycan structure, mass spectrometry can be employed. These applications favour less polar labels exhibiting increased hydrophobicity or charged labels in order to improve ionisation efficiency and reduced levels of ion suppression [25]. 2AA **10** exhibits a similar structure to 2AB **9** however its hydroxyl functional group possesses a negative charge in solution resulting in improved analysis by negative mode MS while still facilitating analysis by optical methods. The limitations of 2AA **10** surround its instability in solution when exposed to light. Both prior to, and following, labelling, 2AA **10** undergoes rapid photolysis making products unstable and therefore challenging for the development of retention or quantification standards. Improved compound stability is afforded by Procainamide (ProA) **8** based labelling systems which also provide increased levels of fluorescence in optical analysis. ProA **8** also exhibits increased hydrophobicity generated by the molecule's tertiary amine tail which creates a non polar region helping to aid ionisation efficiency in Electro-spray ionisation (ESI) without the requirement for a permanent charge state.

Both fluorescence and MS analysis of oligosaccharides present effective workflows for the analysis of abundant glycan species. Challenges with glycan ionisation stability mean that structural analysis and quantification of low abundance glycans still presents a significant analytical challenge especially due to the rapid turnaround required for pharmaceutical analysis.

Industrial analysis workflows require derivatisation and clean-up procedures to be high yielding over a short reaction time while maintaining reaction conditions that are sufficiently mild to preserve oligosaccharide structure. Novel labels **25**, **26** and **27** aim to achieve complete sample labelling in comparable times to commercial labels present in kits for both 2AB **9** and ProA **8**, ensuring complete labelling in under three hours at 65 ^{o}C.

Quantification of glycan species has grown with the increasing development of glycoprotein based biotherapeutic drugs, due to the need to ensure that the glycosylation is within a safe range [4, 26]. Quantification of specific glycan structures as biomarkers for underlying mutations in particular enzyme pathways also highlights the need for rapid, accurate and versatile methods [27-29].

3.2 Aims

This work initially aims to generate a series of quantifiable labelled carbohydrate standards. This will allow the study to investigate and optimise the reaction and completion rates for both **10**, **9**, **8** as well as their modified derivatives. Additionally given the proportional relationship between label concentration and fluorescence, this work will also allow for the quantification of released glycan samples from natural sources.

The work in this chapter set out to fulfil three main objectives. Initially the work set out to optimise the conditions for the analysis of novel labels **25**, **26** and **27** in line with existing workflows for glycan profiling. This was initially carried out through investigating their optimal analysis properties by scanning fluorescence to determine excitation and emission energies. These values were then employed to analyse both model and biologically derived carbohydrate labelled with **25**, **26** and **27** inline with liquid chromatography workflows for glycan profiling.

Additionally work will seek to compare the analytical qualities of **25**, **26** and **27** with their commercial derivatives through investigation of limit of detection and labelling rate. This was achieved through synthesis and purification of purified model carbohydrate standards to enable absolute quantification of labelled glycan concentration. Moreover the use of both internal and external standardisation will enable the analysis of labelling rates for both model carbohydrate labelled with both commercial and multifunctional labels.

Finally this investigation sought to understand the effect of time and temperature on the degree of sialylation of a model oligosaccharide derived from natural sources. This was achieved though a series of labelling reactions at varying temperature conditions followed analysis by HPLC-FLD.

These research objectives will enable the effectiveness of novel labels **25**, **26** and **27** to be compared against their commercially available derivatives within the currently adopted pharmaceutical workflow for *N*-glycan analysis. Moreover, maintenance of sialylation forms a significant part of glycan analysis and is a key feature required for downstream analysis.

3.3 Results and Discussion

3.3.1 Fluorescence of multifunctional labels

The fluorescence characteristics of **25**, **26** and **27** will predominantly be analysed following reductive amination of reducing oligosaccharides. It was for that reason that the λ_{ex} and λ_{em} were determined following labelling. This aimed to ensure modifications in fluorescence activity were comparable to analytical conditions. For this trial an α-1,4-linked maltoheptose oligosaccharide was employed as a model carbohydrate for the analysis as it represents a multiring oligosaccharide of comparable size to that of a naturally derived carbohydrate but at a concentration that can be more tightly controlled.

Fluorimetry samples produced from labelled maltoheptose were analysed at 65 ^{o}C over a five hour period, with an excess of sodium cyanoborohydride in borate buffered methanol. Labelled maltopentose was then purified by amide SPE (applied separations Spe-ed 2) cartridges. The purified samples were dried and reconstituted in aqueous 50 mM ammonium formate pH4.4 and ACN (65:35) to replicate analysis conditions.

Samples were diluted to a concentration of 10 µg/ml and 500 µl was transferred to a quartz cuvette. Analysis was then carried out by 3D scanning fluorescence spectroscopy in order to evaluate emission values over a range of excitation wavelengths. Concurrent emission scans were taken between excitation wavelengths ranging from 250-300 nm increasing by 10 nm between each scan at a 5 nm Slit size. Emission wavelengths where then recorded between the values of 300-600 nm. Initial results exhibited levels of emission above the analysable threshold of the instrument so a dilution factor of 2 was applied to the samples reducing them to a final concentration of 5 µg/ml which was then adopted for the remainder of the investigation.

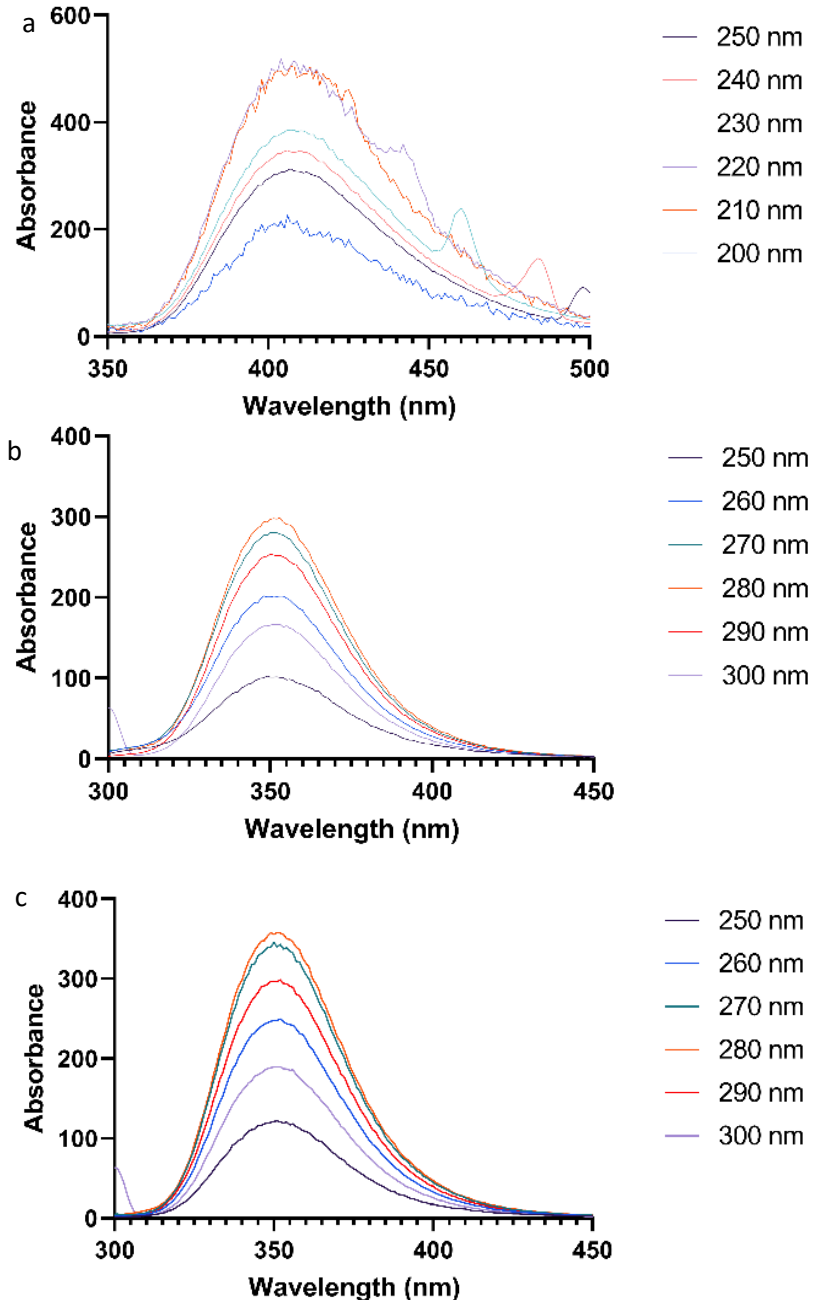


Figure 3.3 -Scanning fluorescence traces for maltoheptose modified with multifunctional labels a) **25** b) **26** c) **27**

Overlaid emission scans for labels **25**, **26** and **27** are outlined in **Figure 3.3** and show emission wavelengths over a range of excitation energies. These figures show the increase and subsequent decrease in emission energy at varying excitation surrounding the value of λ_{em} optima for each label.

Comparing the three modified labels, **26** and **27** display comparable λ_{ex} values between the values of 280-300 nm while the *ortho*-positioned label **25** displays λ_{ex} around 240 nm. These values were then used to determine λ_{ex} and λ_{em} values in single wavelength scans.

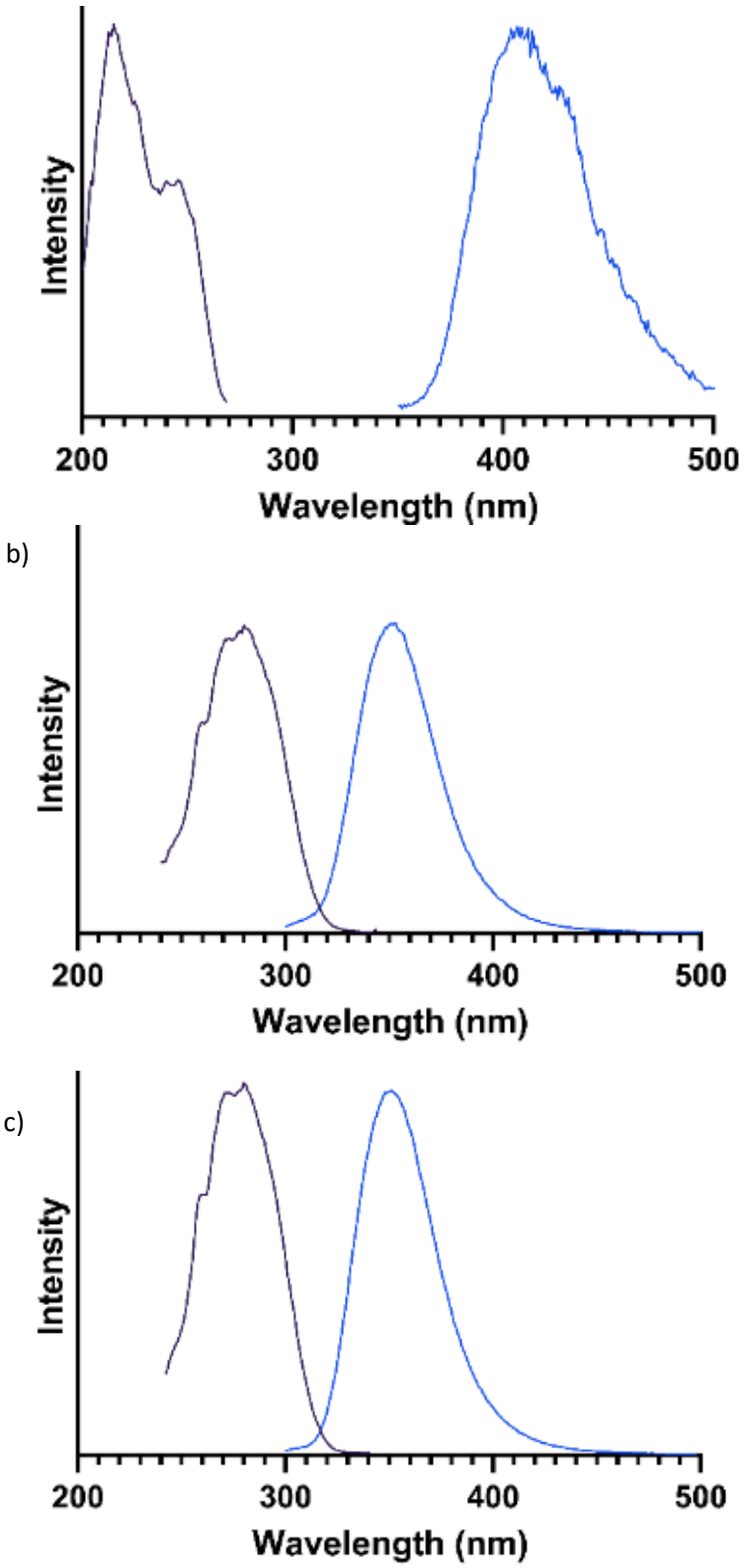


Figure 3.4 -optimal fluorescence conditions following 3D emission scans for GU7 carbohydrate derivatised with a) **25** b) **26** c) **27**

Overlaid single wavelengths scans are outlined in **Figure 3.4**, and present λ_{ex} and λ_{em} optima for **25**, **26** and **27**. These values represent the wavelength at which optimal excitation is achieved and the

values at which subsequent emission takes place. These values therefore dictate the conditions for fluorescence analysis as they represent the values at which the label exhibits the highest levels of fluorescent activity.

Table 3.1 - Fluorescence characteristics for both commercially available and multifunctional labels

Compound	Experimentally established		Literature value		Reference
	Excitation (λ_{ex}, nm)	Emission (λ_{em}, nm)	Excitation (λ_{ex}, nm)	Emission (λ_{em}, nm)	
GU7-2AA **10**	-	-	330	420	[30]
GU7-2AB **9**	240	420	250	428	[31]
GU7-ProA **8**	299	367	310	370	[31]
GU7-**25**	222	405	-	-	
GU7-**26**	289	361	-	-	
GU7-**27**	282	348	-	-	

Table 3.1 outlines the fluorescence conditions for **25**, **26** and **27** as well as their commercially available derivatives. Values for label **27** were found to be λ_{ex} = 282 nm, and λ_{em}= 348 nm, comparable to those of the *para* positioned ProA **8** which were λ_{ex} = 299 nm and λ_{em}= 367 nm [31].

Comparable optima were also observed between the *ortho* positioned label **25**, exhibiting a fluorescence optima of λ_{ex} = 222 nm and λ_{em}= 405 nm although significant differences in label structure may explain variation between the native 2AB **9** label which displays fluorescence optima of be λ_{ex} = 250 nm and λ_{em}= 428 nm [31]

Perhaps the highest level of structural similarity is exhibited between label **26** and its unmodified derivative **8**. Both commercially available and label **26** share close structural similarity with the exception of a single alkyne present on one of there labels' hydrophobic tails. Current literature states excitation and emission of ProA **8** to be λ_{ex} = 310 nm and λ_{em}= 370 nm following the derivatisation of released IgG glycans. Meanwhile, scanning fluorescence spectroscopy found the values for label **26** to be λ_{ex} = 289 nm, and λ_{em}= 361 nm

3.3.2 Calibration of labelled glycan standards

Following the determination of optimal absorbance and emission values, a series of large scale labelling reactions were carried out using lactose as a model carbohydrate. Lactose was chosen as a model carbohydrate as its smaller disaccharide structure facilities fast separation by HILIC mode separation, moreover its increased availability in high purity made it a more achievable donor carbohydrate for the production of large scale carbohydrate standards.

The product of these large scale reactions was a series of pure derivatised carbohydrate standards that facilitated the generation of a series of standard curves. These were recorded at a range of concentrations from 100 nmol of labelled lactose solution down to 10 nmol to represent the range of differing glycan abundances within a sample. Standard curves are often used as an external regression model for quantification in analytical techniques. This allows for the unknown detector responses in both optical techniques as well as MS to be quantified based upon comparison with known standard concentration.

In order to generate quantifiable carbohydrate standard curves, a series of purified carbohydrate standards were generated derivatised with **10**, **9**, **8**, **25**, **26** and **27**. Samples of GU2 (Glucose unit) Lactose were labelled at a 20 mg scale and stirred overnight until reaction completion. Following confirmation of completion by TLC analysis, the crude reaction mixture was concentrated *in vacuo* before being purified by flash column chromatography to remove free label and sodium cyanoborohydride. Fractions containing fluorescent carbohydrate were then pooled and dried under reduced pressure before being freeze dried to remove residual water from the sample. The dried products were then characterised by NMR spectroscopy and mass spectrometry (**Figure 3.5**).

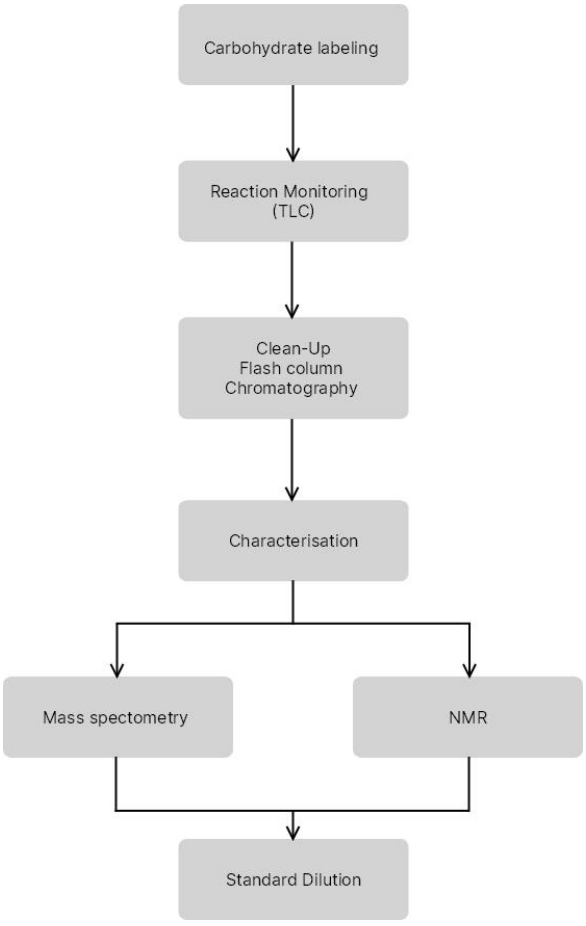


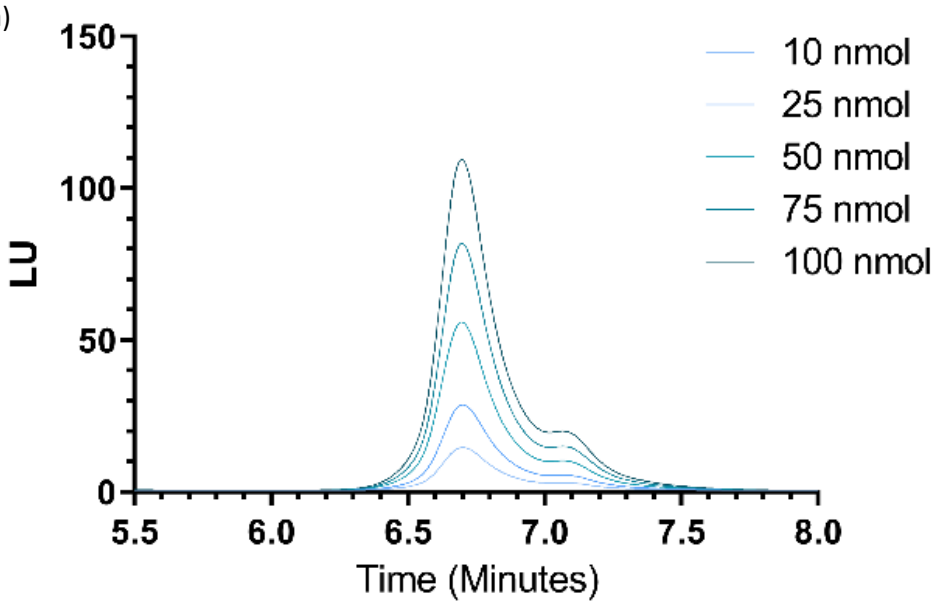
Figure 3.5 - Workflow for the production and analysis for labelled carbohydrate standards

Following this, dried products were weighed and dissolved in H_2O to a concentration of 500 µmolar. Thus resulting in labelled lactose stock solutions equating to 100 nmol of stock solution per each 200 µl sample. From these stock solutions a series of calibration standards were generated and diluted to a final volume of 200 µl for injection (**Table 3.2**). In order to accurately mitigate against instrument injection variations and small recovery losses resulting from sample clean up 10 µl, ProA **8** maltotriose was transferred to each vial via syringe. Vials were then dried at reduced pressure before being resuspended with labelled lactose stock solution and high performance liquid chromatography (HPLC) grade water.

Table 3.2 - Dilution ratios used in the production of calibration samples curve from 100 nmol to 10 nmol per 200 µl sample

Concentration (nM)	Labelled lactose stock solution (µl)	Water (µl)	ProA **8** Maltotriose (µl Dried)
100	200	0	10
75	150	50	10
50	100	100	10
25	50	150	10
10	25	175	10
0	0	100	10

a)



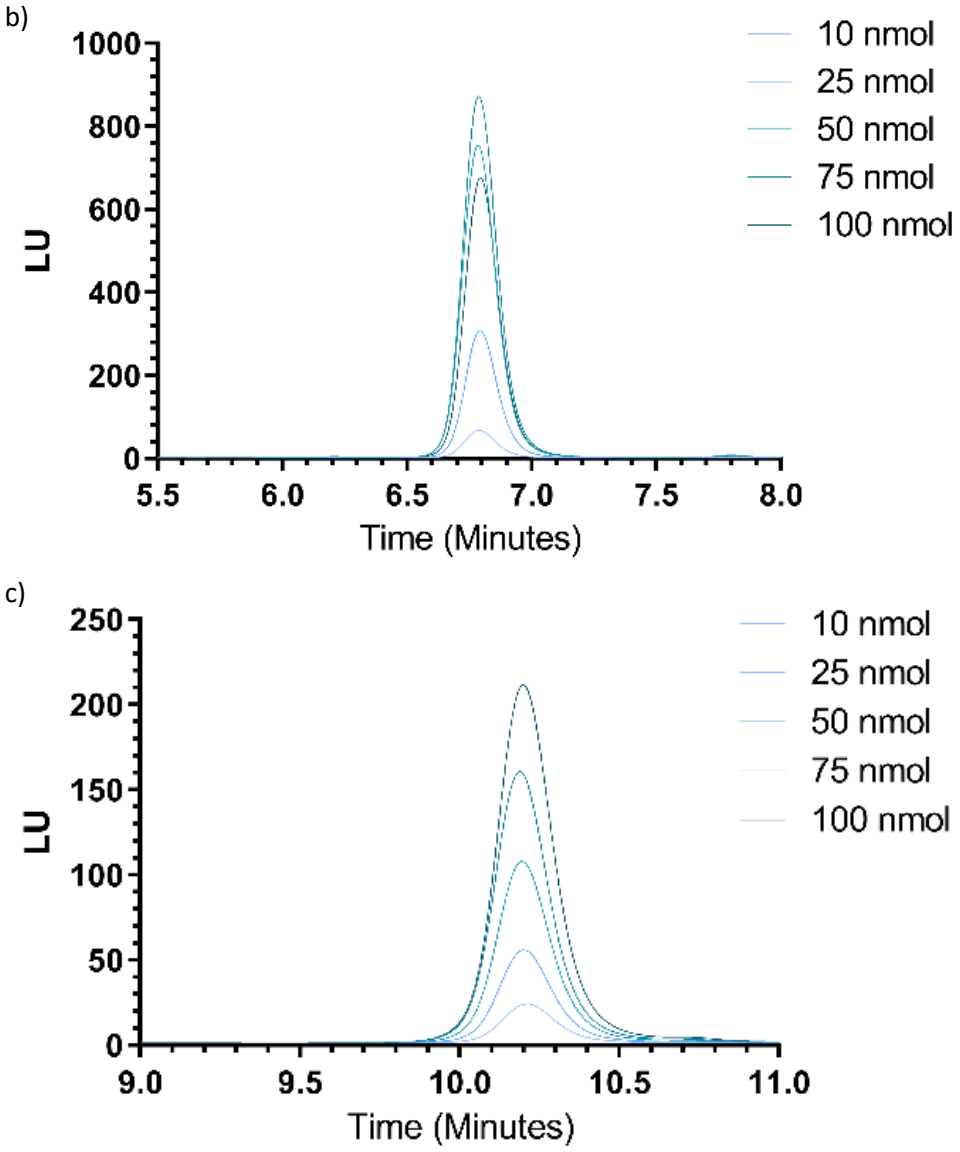


Figure 3.6 -Calibration traces for GU2 carbohydrate derivatised a) **10** b) **9** c) **8**

Overlaid fluorescence traces outlined in **Figure 3.6** show stable hydrophilic interaction chromatography (HILIC) retention was achieved for GU2 derivatised **10**,**9** and **8**. When comparing the three commercial labels it was noticed that the peak shapes between the three labels differed significantly. Thus whilst **9** exhibits a steep narrow peak shape **8** exhibits a shorter peak height but increased peak area due to its shorter and wider peak shape. The increased fluorescence of **8** was also documented by Keser *et al* and Kozak *et al*[31, 32] and in addition it was noticed that **10** exhibited the lowest fluorescence of the commercial labels exhibiting similar characteristics to those observed by Ruhaak *et al*[10].

The repeatability of labelling characteristics in model carbohydrate standards, derivatised with commercially available labels versus biologically obtained carbohydrates aids in the validation of lactose as a model carbohydrate for the testing of multifunctional labels **25**, **26** and **27**.

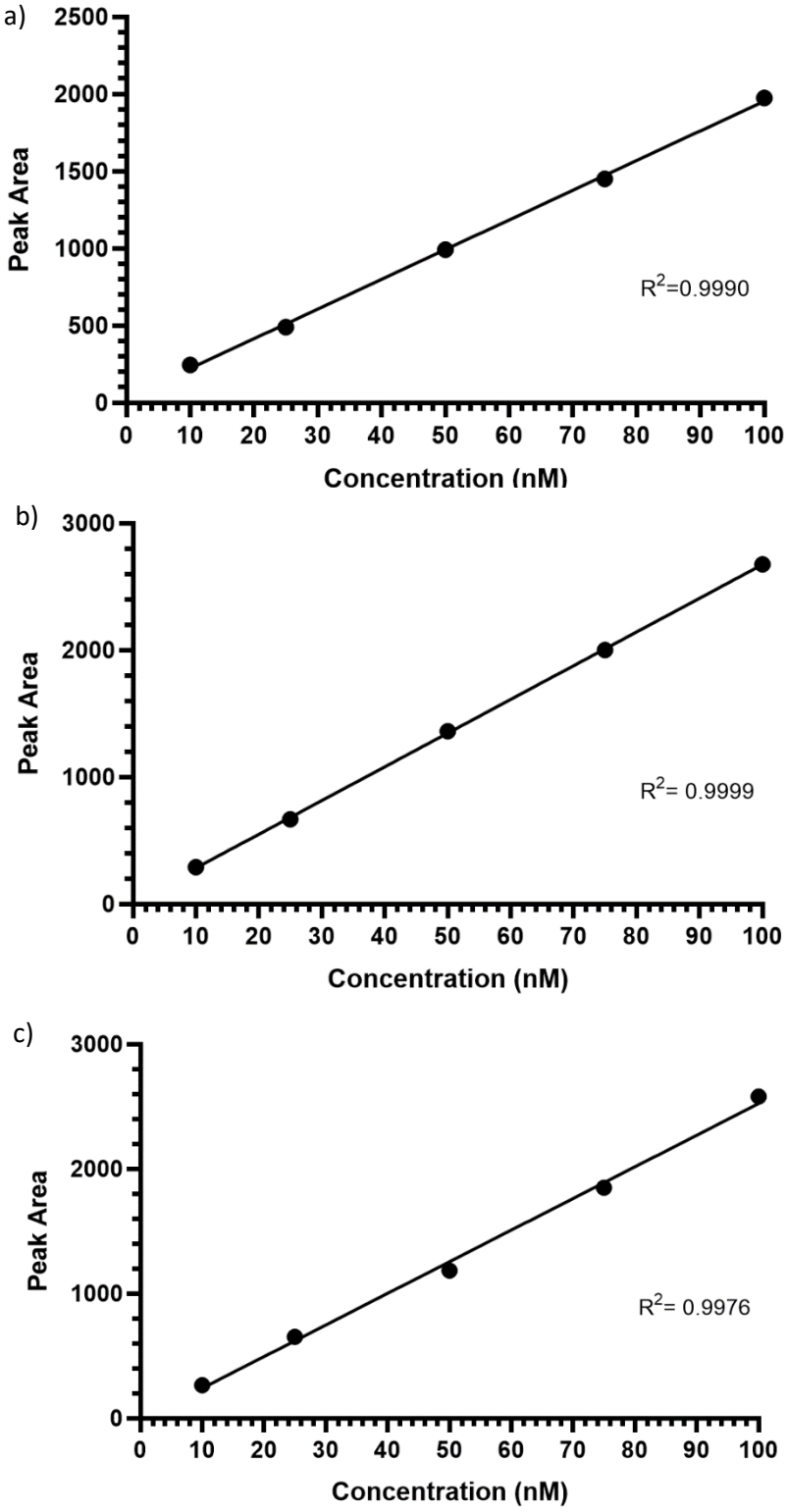
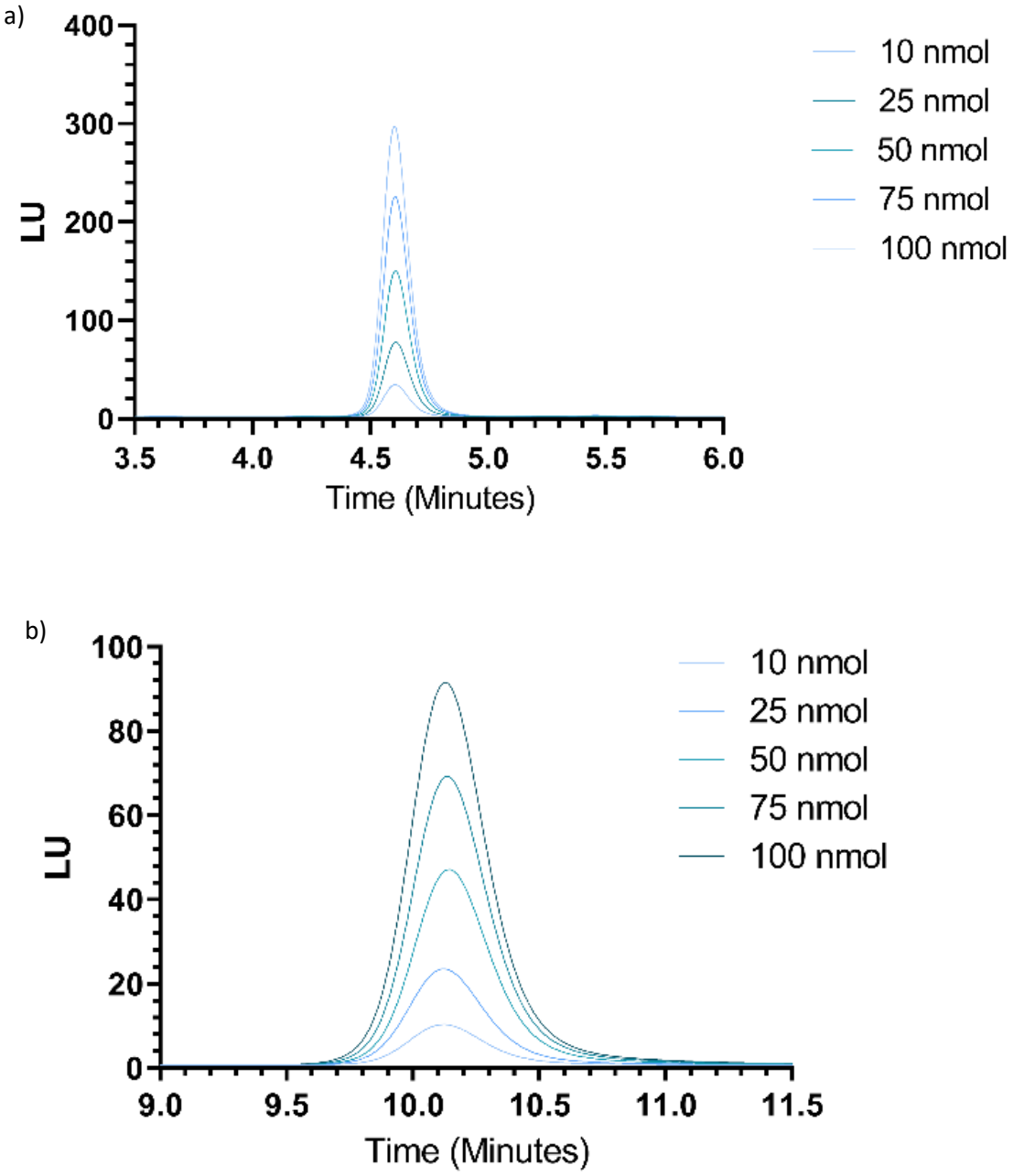


Figure 3.7 -Linearity plots of fluorescently labelled GU2 carbohydrate showing peak area vs standard concentration for the quantitative analysis of carbohydrate derivatised with a) **10** b) **9** c) **8**

A linearity plot of all five standards for each of the three commercial labels revealed a high level of linearity with R^2 values above 0.99 indicating strong positive correlation between increased concentration and increased peak area (**Figure 3.7**). This is explained by the chemistry of reductive amination labelling, where a single label is joined to a single carbohydrate molecule [10]. This provides a linear relationship between analyte concentration and observed detection intensity. This system makes reductive amination labelling an ideal candidate for the further development of quantitative carbohydrate labelling.



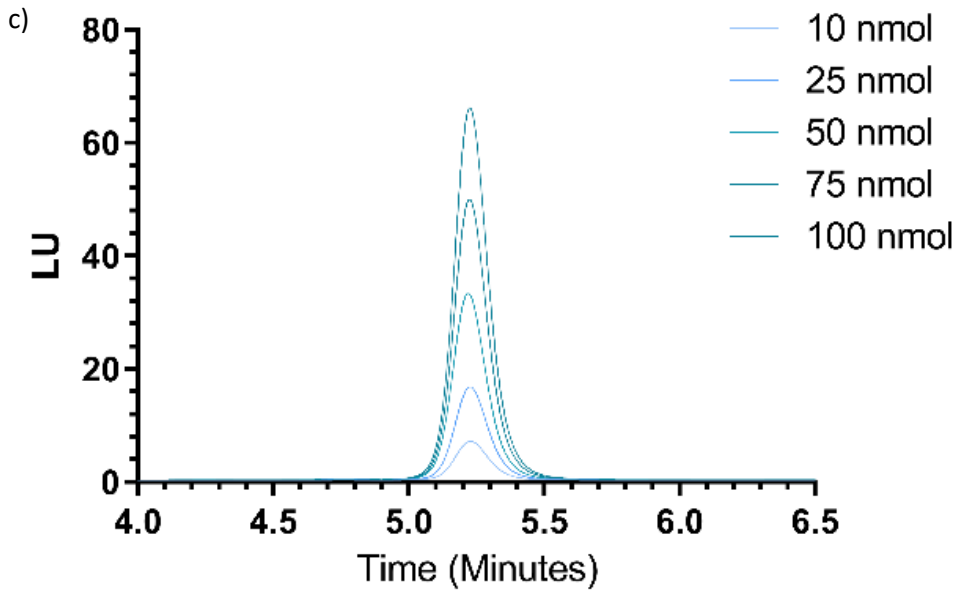


Figure 3.8 -Calibration traces for GU2 carbohydrate derivatised a) **25** b) **26** c) **27**

Like the commercially available labels **10**,**9** and **8,** multifunctional labels **25**, **26**, **27** are all reductive amination labels and therefore react with reducing carbohydrates at a single site [16, 33]. Overlaid traces for labels **25**, **26** and **27** show a high degree of retention time stability between injections (**Figure 3.8**). Although it was observed that **25** and **27** display reduced retention over their commercially available derivatives. Label **27**, exhibiting a retention time of 5.25 minutes while **8** exhibited retention times of 10.15 minutes, in addition **25** gives a reduction in retention of 2 minutes over its derivative **9**. Conversely label **26** appeared to give retention times comparable to that of **8**. Based upon HILIC separation chemistry this is likely to be due to the reduced polarity of the label derivative leading to reduced interactions with the stationary phase. In this case the modifications to **26** involve the lowest change in overall molecule structure leading to similar retention times to that of **8**.

Although there are some variations in retention time between commercially available labels and labels **25**, **26** and **27**, peak shape of all three labels is comparable. Label **25** gives a sharp narrow peak with high detector response while **26** behaves with similar characteristics to **8** exhibiting a broader peak of lower intensity. The larger peak area is likely to be due to the comparatively hydrophobic structure of both **8** and **26** combined with the increased electronegativity of three nitrogen containing groups.

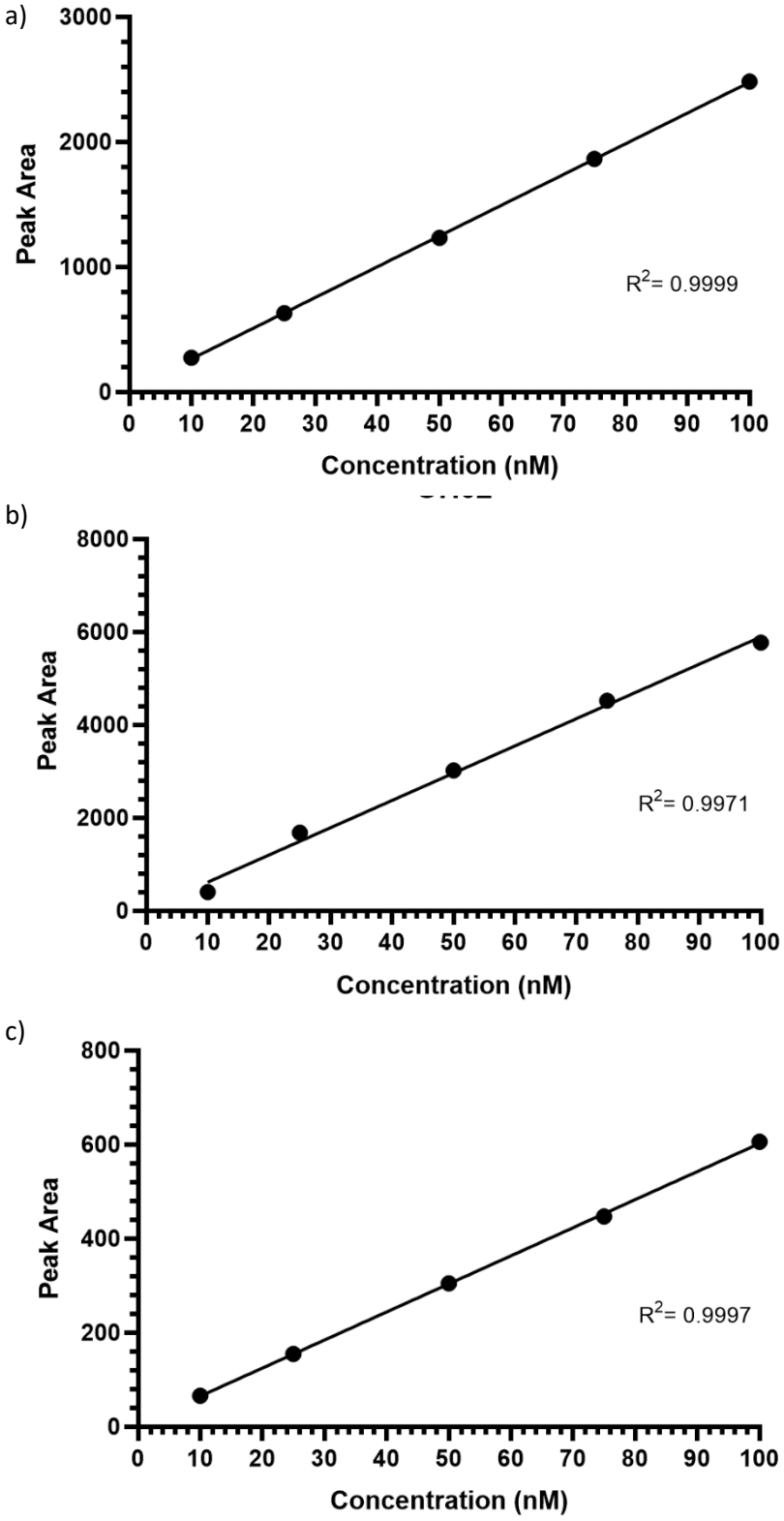


Figure 3.9 -Linearity plots of fluorescently labelled GU2 carbohydrate showing peak area vs standard concentration for the quantitative analysis of carbohydrate derivatised with a) **25** b) **26** c) **27**

Figure 3.9 outlines a plot of sample concentration against peak area. The figures show comparable levels of linearity between all three of the modified carbohydrate labels, when judged by r^2 value with values falling between 0.9999 and 0.9971. These levels fall within comparable values to the unmodified derivatives and align with trends seen in existing literature[32, 34, 35]. These finding help to further aid the validity of the model but also help to justify the use of labels **25**, **26** and **27** as viable derivatisation agents for the quantification of carbohydrate species, following comparable trends to that of commercially available compounds while still maintaining the capability to undergo further post analysis functionalisation to improve further analysis or functionalisation to azide containing biomolecules or surfaces.

3.3.3 Detection limits of multifunctional labels versus commercial labels

Multifunctional labels **25**, **26** and **27** were designed to display comparable analysis characteristics to their commercially available analogues while maintaining the ability to undergo additional post analysis application by use of an alkyne containing 'functional handle'. A fall in detection sensitivity would lead to the label being ineffective as a derivatisation agent for the analysis of low abundance glycan species and limit its viability as a successful glycan label.

The limit of detection (LOD) for labels **25**, **26** and **27** was investigated by HILIC mode HPLC-FLD using labelled model carbohydrate standards. These samples mirrored a similar series of labelling's for **10**, **9** and **8**. Samples were diluted from 500 μmolar stocks of pure labelled lactose to form an incremental series of decreasing concentration of carbohydrate samples as outlined in **Table 3.3**.

Table 3.3 -Volumes of stock solution and diluent required to generate calibration curves for fluorescent GU2 carbohydrate standards.

Concentration pMol/ml	Volume of H_2O (μl)	Volume Standard 500 nmol/ml	Volume Standard 50 nmol/ml
100	800	200	0
50	900	100	0
25	950	50	0
10	800	0	200
5	900	0	100
2.5	950	0	50
0	1000	0	0

2 µl of each solution was injected and separated over 25 minutes by HILIC-HPLC-FLD ranging from low to high concentration in a bracketed sequence with blanks between each series of standards. Separation was achieved using an Agilent 1100 HPLC system fitted with a TSKgel Amide-80 HR 250 x 4.6 mm 5 µm column (TOSOH Bioscience) protected with a TSK® guardgel Amide-80 (15 × 3.2 mm, 5 µm) guard column (TOSOH Bioscience). The separation was performed using a gradient method between ammonium formate buffer (50 mM, pH4.4)(A) and acetonitrile (B). Analyte detection was then carried out on an Agilent 1100 G1321A FLD detector fitted with an 8 µl flow cell at fluorescence characteristics for each label optimised in the previous section.

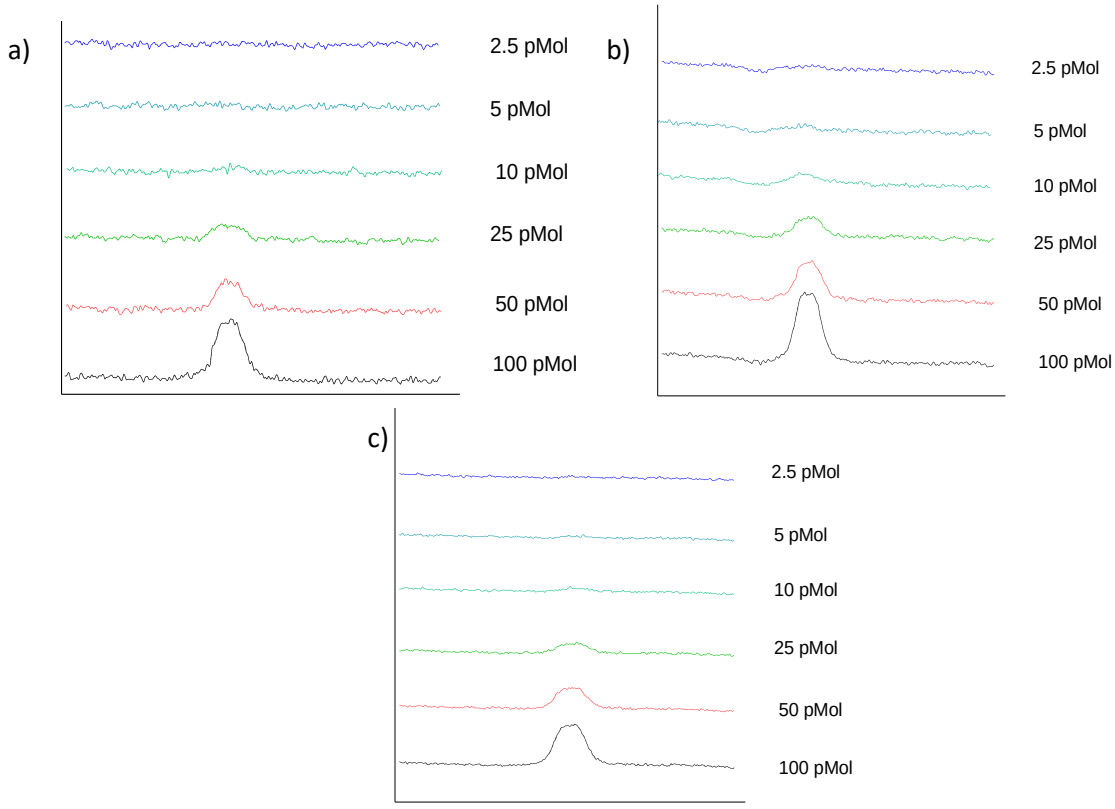


Figure 3.10 -Limit of detection of GU2 carbohydrate labelled with a) **10** b) **9** c) **8**

In previous literature ProA **8** has exhibited a higher level of fluorescence when compared to 2AB **9** [34] and this aligns with the data displayed in **Figure 3.10 a** and **b** which show that **8** samples are no longer discernible from baseline below 10 pMol/ml while **9** was no longer discernible at levels below 10 pmol/ml and would not be easily quantifiable from levels below 25 pmol/ml. 2AA **10** displayed the lowest levels of fluorescence and was not detectable from baseline noise below 25 pmol/ml.

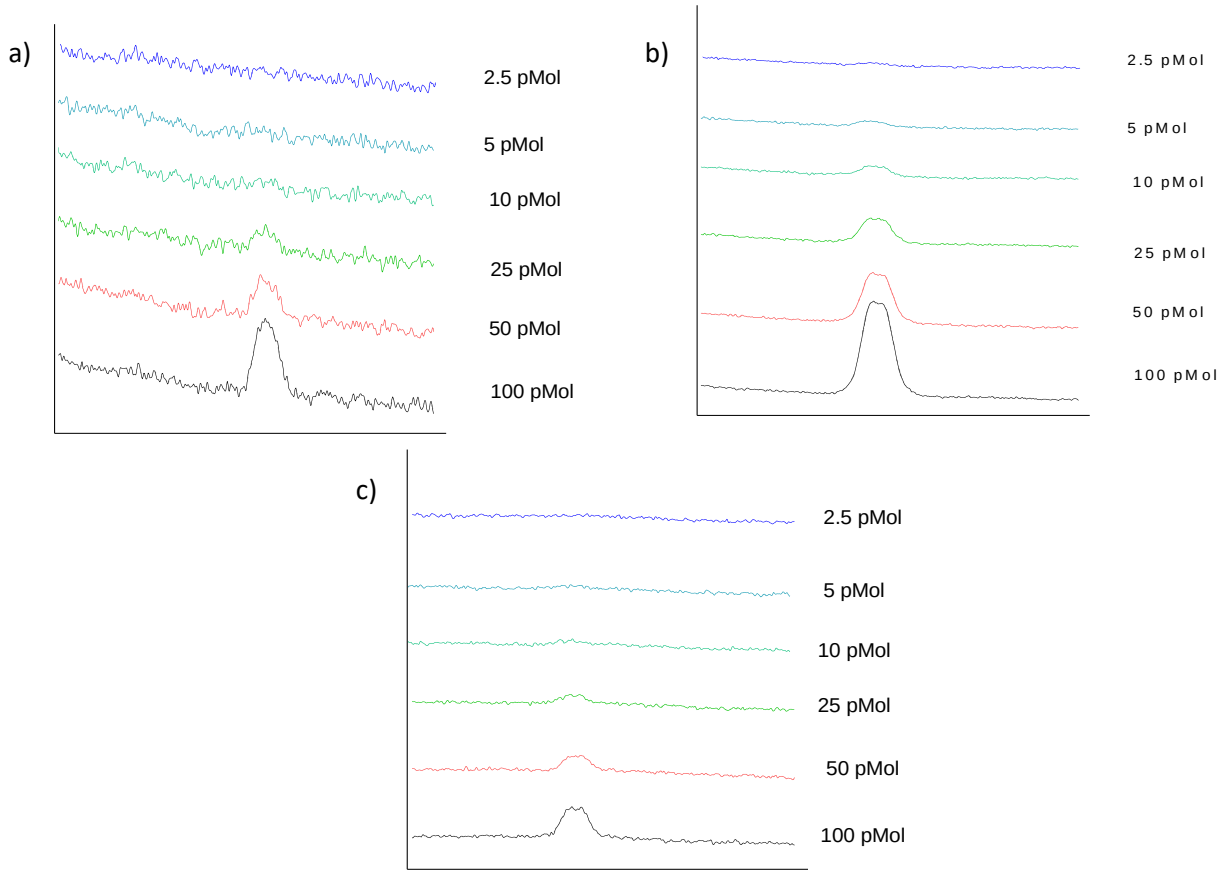


Figure 3.11 -Limit of detection of GU2 carbohydrate labelled with a) **25** b) **26** c) **27**

Modified labels **25**, **26** and **27** also display a pattern of fluorescence similar to that of the commercial labels with **8** derivative, **26** exhibiting higher levels of fluorescence output than **9** derivative **25** (**Figure 3.11**). Carbohydrate labelled **25** was no longer discernible from the baseline below levels of 25 pMol/mL due to poor signal to noise while **26** exhibited enhanced fluorescence and was discernible at levels below 10 pmol/mL although its limit of quantification is likely to be higher. **27** meanwhile exhibited improved signal to nose over **25** but showed reduced fluorescence levels below 50 pmol although it was detected at a 25 pmol/mL level. These detection limits fall within the capabilities of commercially available derivatisation agents demonstrating the ability of multifunctional labels in their unclicked state to give comparable levels of sensitivity while retaining the ability to undergo post analysis functionalisation or derivatisation.

3.4 Optimisation of HILIC mode separation

3.4.1 HILIC Analysis of carbohydrate homopolymer

In order to evaluate the separation and retention characteristics of labels **25**, **26** and **27**, controlled acid hydrolysis of a dextran homopolymer was carried out. Dextran is routinely used in glycomic analysis as a retention standard as following its hydrolysis the large α-1,6-linked glucose structure yields a homopolymer ladder of glucose units ranging in length between 1- 16 glucose units (GU). In

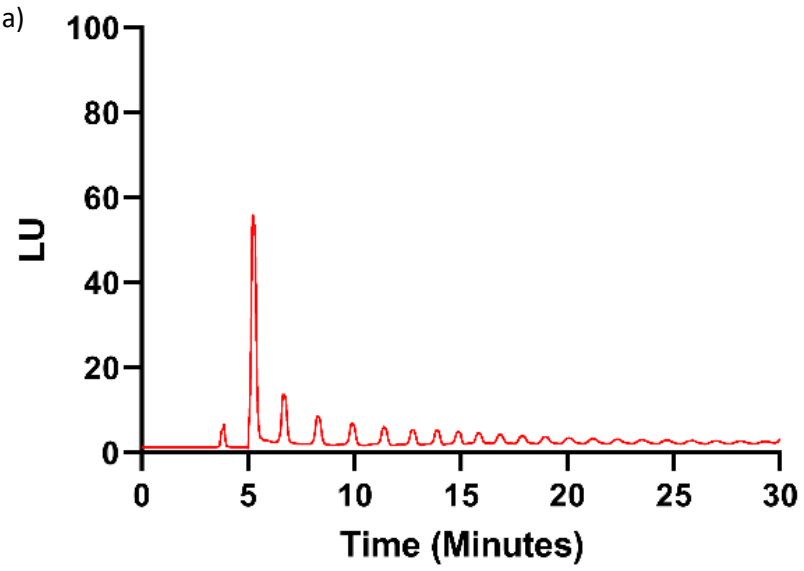
industrial workflows, homopolymer ladders are commonly used as retention standards in HILIC separation for both elucidation of carbohydrate structure as well as for quality control purposes to ensure HILIC method stability [36]. In this investigation the use of a glucose homopolymer will demonstrate whether the separation and retention characteristic of **25**, **26** and **27** align with those seen from commercial labels **10**, **9** and **8**.

Glucose homopolymer samples were generated as a result of the acid hydrolysis of dextran. The predominantly α-1,6-linked glucan which is produced by *Leuconostoc mesenteroides* is first solubilised in dilute hydrochloric acid and heated to 100 °C for 4 hours. The solution is then neutralised and diluted to 100 mg/ml for analysis.

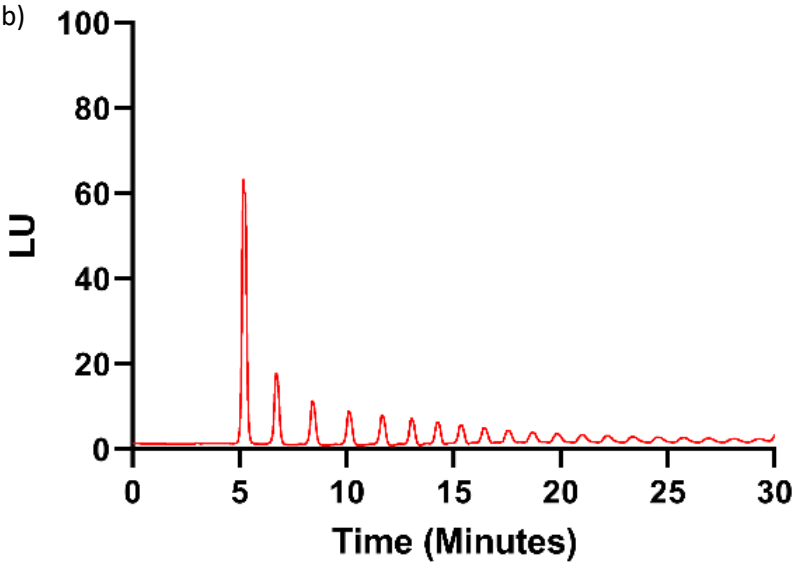
Diluted dextran solution was then labelled with **25**, **26** and **27** as well as **10**, **9** and **8** for comparison. The labelling solution was made up of dissolved label in methanol, and subsequent addition of sodium cyanoborohydride in borate buffered methanol. This solution was then incubated at 65 °C for 3 hours. The samples were then quenched with ACN and applied to spe-ed2 amide HILIC SPE columns. Following purification the samples were eluted in water (600 µl) and dried in a vacuum centrifuge before being resuspended in water (200 µl) for injection.

Sample analysis took place on an Agilent 1100 HPLC system fitted with an Amide-80 HR Column (Tosah bioscience) running a 35 minute gradient method from 65% B 0 minutes, 35% B 20 minutes, 10% B 30 minutes, 65% B 35 minutes with a 5 minute post run equilibration. The separated fluorescent glucose homopolymer was then detected in an Agilent 1100 G1321A FLD with an 8 µl analytical flow cell at the optimum wavelengths found in the work of [31] for commercial labels as well as those previously optimised in the study. Signals were then recorded and processed in Agilent chemstation 4.03b.

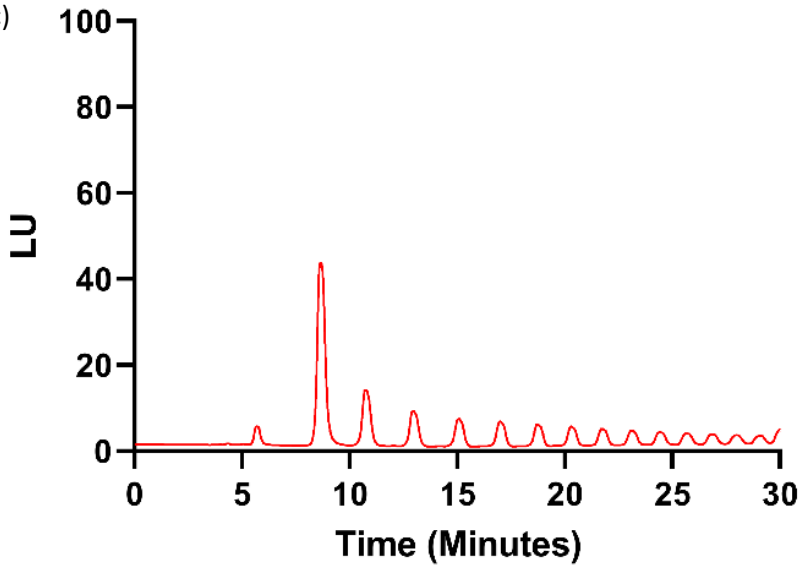
a)



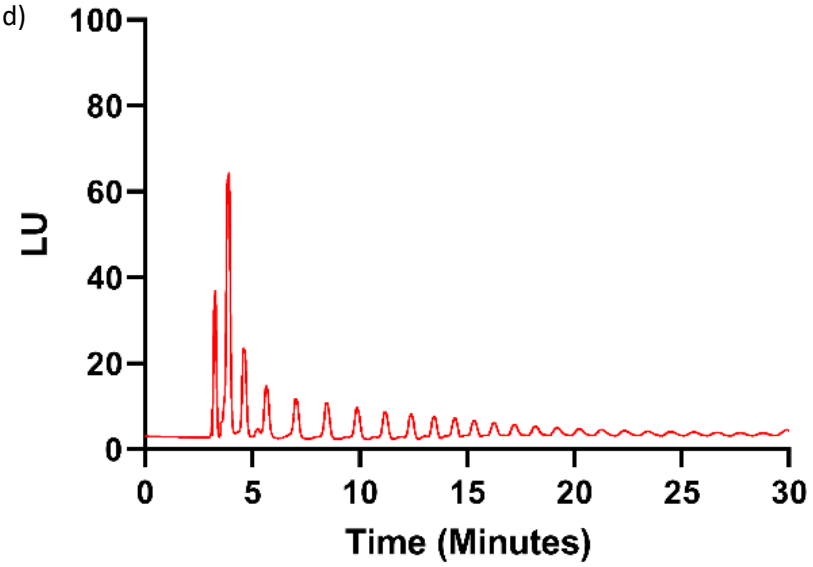
b)


100
80
60
40
20
0
LU
0
5
10
15
20
25
30
Time (Minutes)

c)


100
80
60
40
20
0
LU
0
5
10
15
20
25
30
Time (Minutes)

d)


100
80
60
40
20
0
LU
0
5
10
15
20
25
30
Time (Minutes)

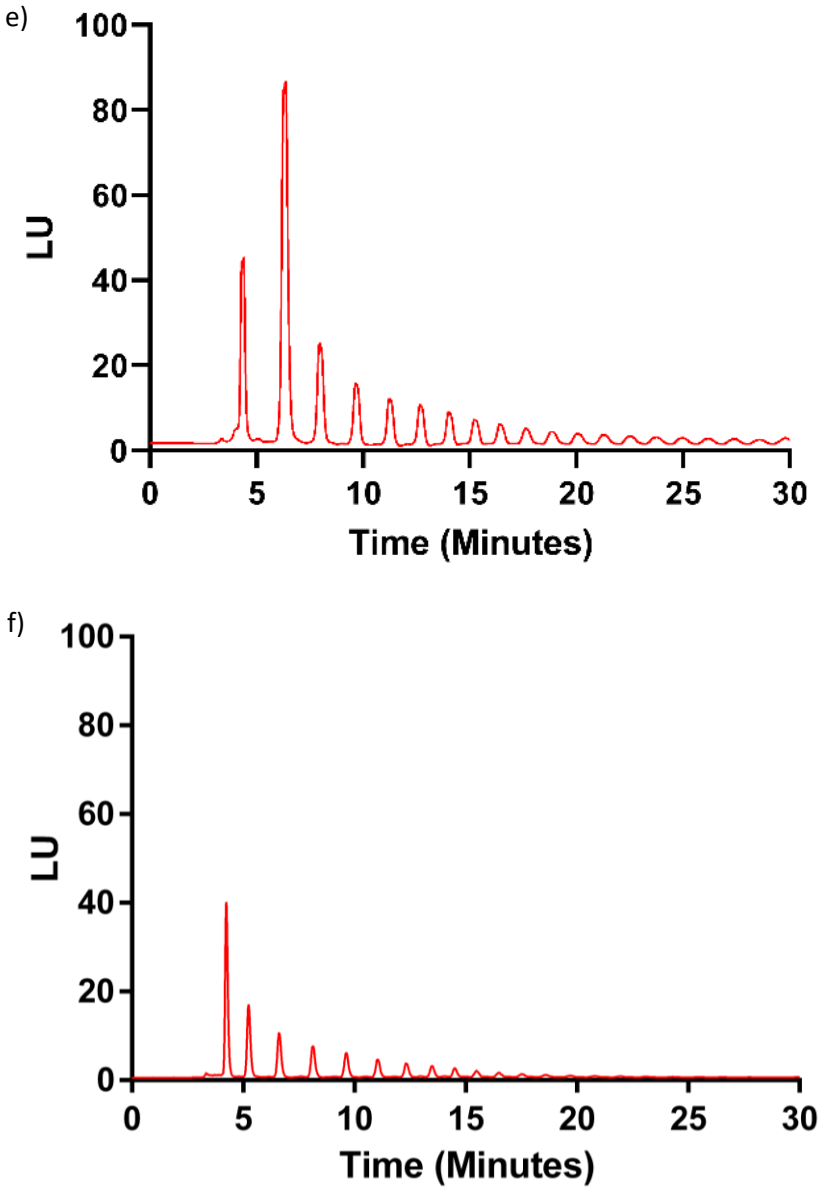
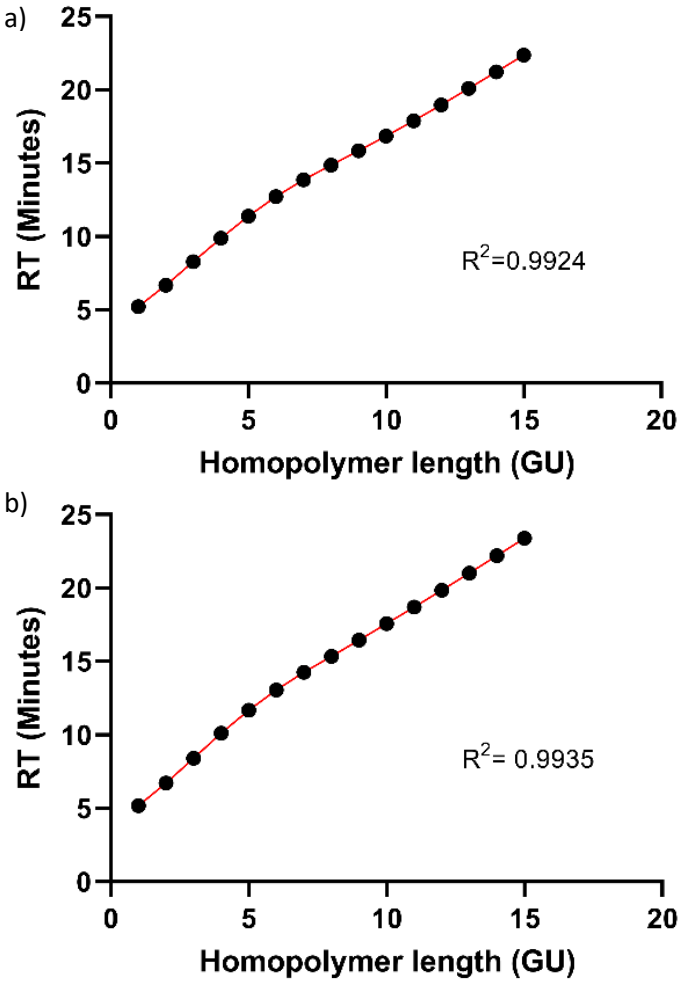


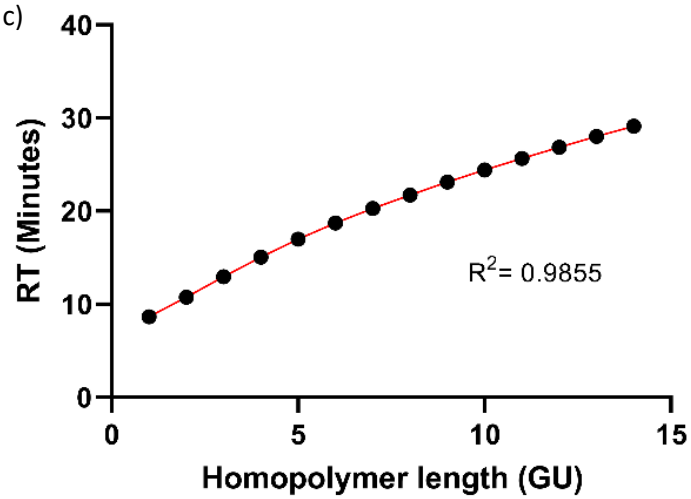
Figure 3.12 -HILIC mode separation of fluorescently labelled glycose homopolymer derivatised with a) **10** b)**9** c)**8** d) **25** e) **26** f) **27**. Fluorescent emission was captured as light units (LU) over time.

Figure 3.12 outlines the chromatograms for dextran homopolymer's labelling with **10**, **9**, **8**,**25**, **26** and **27**. All six of the above chromatograms display strong similarities in separation characteristics and order of elution, with the most abundant and least polar low GU structures eluting first. This is followed by the more polar larger GU species which exhibited increased retention and were therefore eluted later [37]. These trends echo predicted elution orders for HILIC mode separation with increased retention of more polar analytes and therefore elution in order of increasing polarity.

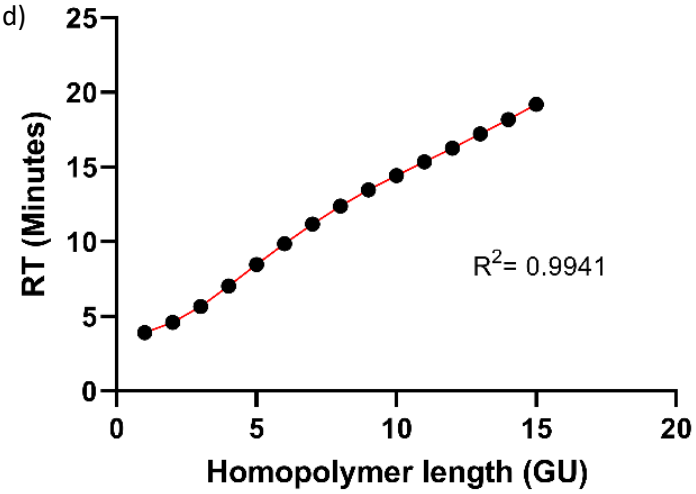
Elution for **10** displays a reduced peak height for GU1 when compared to **9** agreeing with both previous work of [34] and work carried out previously within this study [38]. Separation of **9** labelled dextran samples also revealed increased peak height over GU1 carbohydrate structures labelled with 8. However, the increased peak area of **8** labelled samples gives more than double the detector response in overall peak area, when applied to GU1 structures aligning with the work of [31]. Analysis of **25** labelled dextran samples revealed strong similarities between the levels of fluorescence output by **25** and its commercial derivative **9**. Similarities can also be observed between the fluorescence output of **8** and its modified derivative **26** which bears an additional alkyne functional group, attributing to only a minor decrease in fluorescent activity. In both **25** and **27**, the effect of additional non polar regions leads to a reduced level of retention when compared to the commercially available derivatives although separation was still sufficient to differentiate between structures of differing GU value. Samples **27** exhibited the lowest levels of fluorescence, however adequate separation was still observed. The successful labelling and separation of a glucose homopolymer demonstrates that labels **25**, **26** and **27** display comparable separation characteristics to their commercially available derivatives and may therefore be used to elucidate the structure of unknown glycan samples.



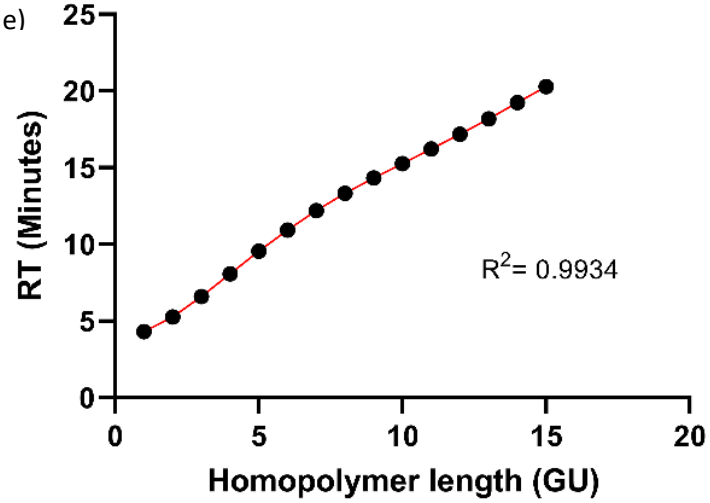
c)


RT (Minutes)
40
30
20
10
0
0
5
10
15
Homopolymer length (GU)
R2= 0.9855

d)


RT (Minutes)
25
20
15
10
5
0
0
5
10
15
20
Homopolymer length (GU)
R2= 0.9941

e)


RT (Minutes)
25
20
15
10
5
0
0
5
10
15
20
Homopolymer length (GU)
R2= 0.9934

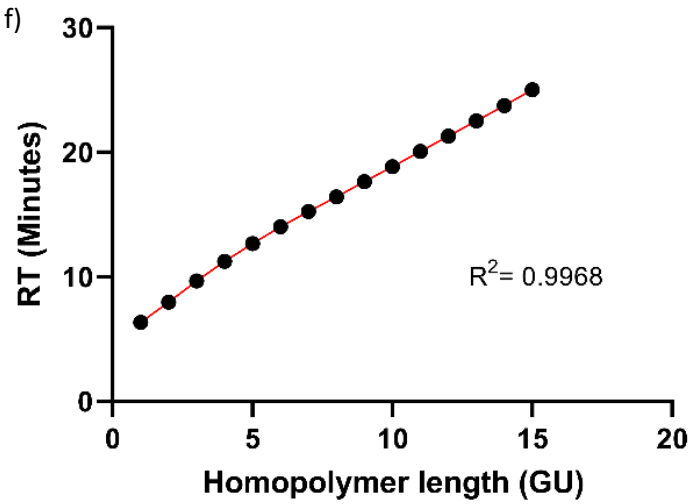


Figure 3.13 -Retention linearity of a dextran derived glucose homopolymer derivatised with a) **10** b) **9** c) **8** d) **25** e) **26** f) **27** separated by HILIC mode LC over a 35minute runtime following a 2 µl sample injection. Fluorescent signals were recorded at optimal wavelength

The figures above (**Figure 3.13**) show plots of analyte retention time against glucose homopolymer length. This section aims to further evaluate the separation characteristics of **25**, **26** and **27** against their commercial derivatives by calculating the r^2 value or the coefficient of determination. This statistic shows the degree of linearity exhibited by the elution of labelled sugars by showing agreement with high levels of positive correlation.

Figure 3.13b shows elution linearity for **10** had an r^2 value of 0.9924 while its amide containing derivative **9** showed slightly improved linearity at 0.9935. This was mirrored by the *ortho* modified label **25** which showed slightly improved linearity of 0.9941. The plot for **8** meanwhile showed the lowest level of linearity of 0.9855 which is likely due to the broader peak shape observed in the lower polarity samples. Improved linearity was observed in both *para*-position modified labels **26** and **27** samples with r^2 values of 0.9934 and 0.9968 respectively. This Illustrates the similarities between multifunctional labels **25**, **26** and **27** with their commercially available derivatives in providing predictable and linear elution.

3.4.2 Analysis of glycan derivatisation time at differing temperatures.

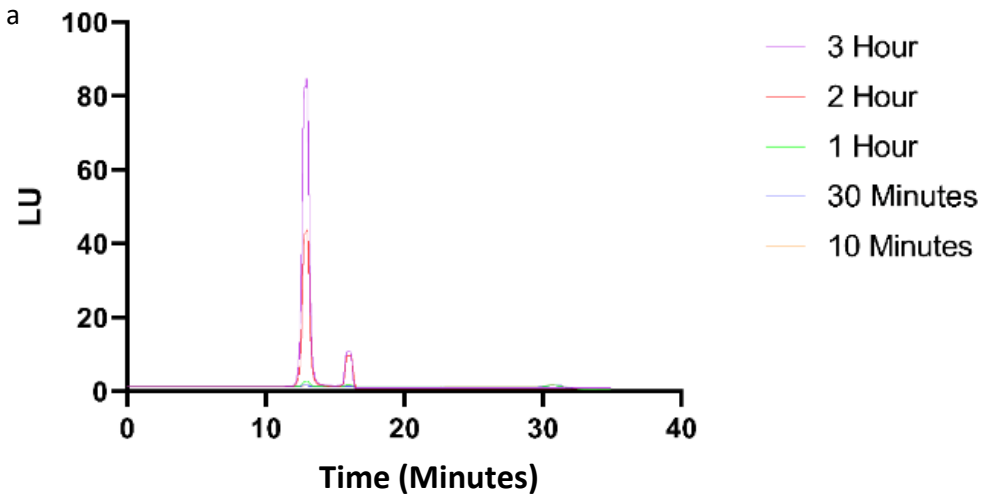
It is commonly accepted that reductive amination labelling is complete within 3 hours at elevated temperatures [16, 18]. Prior research into labelling completion rates has focused on the analysis of free label abundance combined with the decrease in abundance of free sugars in samples analysed by

mass spectroscopy. These samples have typically employed incubation times of longer than 3 hours [32]. This method of reaction monitoring is effective but limited by the poor ionisation efficiency of free carbohydrates. In order to investigate the kinetics of labelling a series 100 nmol GU2 lactose samples were labelled with **10**, **9**, **8**, **25**, **26** and **27** over differing time points at both 65 °C and 85 °C in the presence of a **8** labelled maltotriose internal standard

Lactose was initially added to each reaction tube followed by **8** labelled maltotriose and labelling reagent. Finally, sodium cyanoborohydride in borate buffer was added and the tubes were centrifuged for 10 seconds. Tubes where then placed in the appropriate water bath and heated for 10, 30, 60, 120 or 180 minutes. The 60 µl reaction mixtures were then quenched with ACN up to a total volume of 1 ml forming a cloudy white suspension which was then applied to spe-ed2 amide SPE cartridges and purified with both 1% and 3% water in ACN. The samples were then eluted with water (600 µl) and dried under reduced pressure in a vacuum centrifuge. The dried samples were then resuspended in 200 µl of HPLC pure water before being sonicated, vortexed and transferred to 200 µl vial inserts for injection.

Analysis was then carried out using a 35 minute HPLC-HILIC method, separation was performed on Amide-80 HR gel HILIC (Tosah Biosciences) using a gradient method between ammonium formate buffer (50 mM, pH 4.4)(A) and Acetonitrile (B). Beginning at 0 minutes with 65% B, 25 % 20 minutes 10% 25-27 minutes, 32% 30-35 minutes. Due to the structural similarity between **8** and **26** labelled lactose GU2 and **8** labelled Maltotriose GU3, column temperatures were elevated to 50 °C in order to increase retention of polar analytes as described by Molnarova *et al*. [39]

Analyte detection was then carried out on an Agilent 1100 G1321A FLD detector fitted with an 8 µl flow cell at optimal fluorescence characteristics for each label, both utilising the wavelengths acquired from prior fluorescence analysis and optimal wavelengths from Kozak *et al* [31].



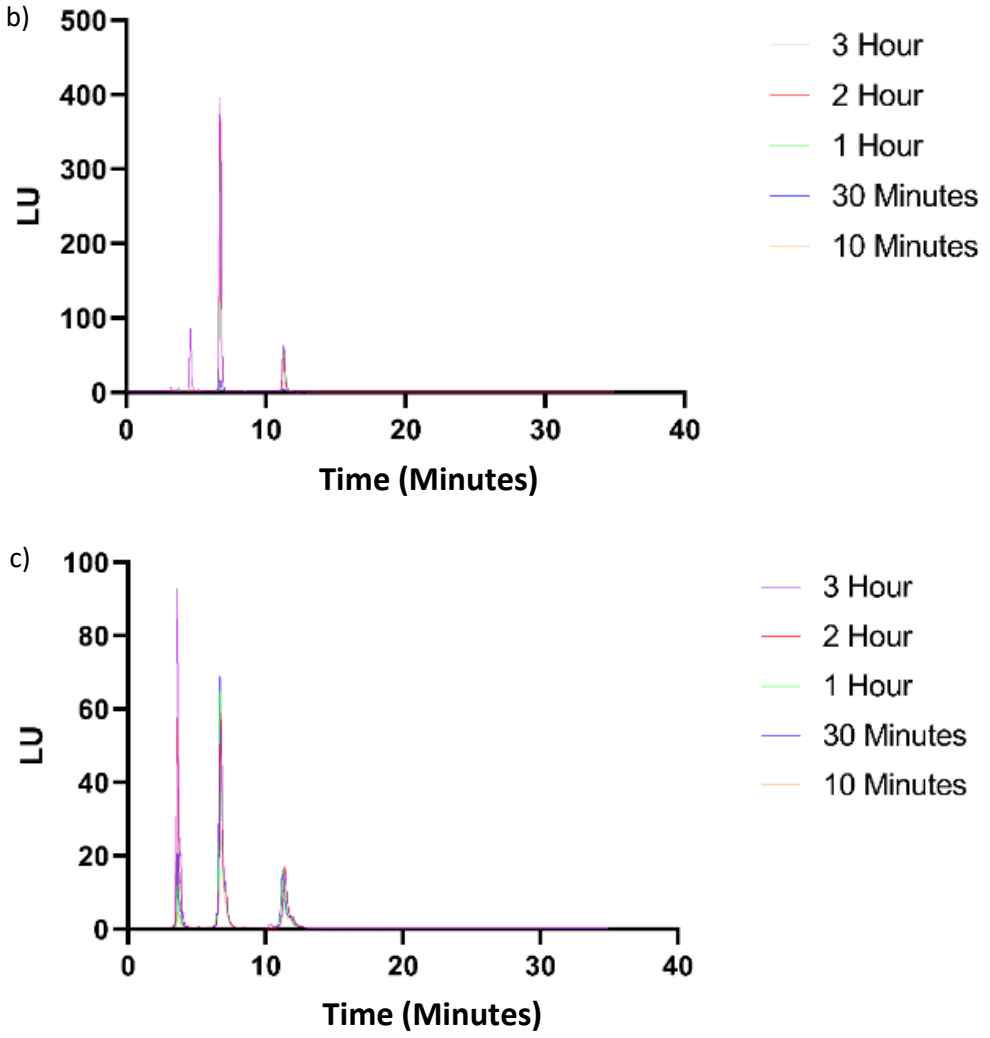


Figure 3.14 -Model carbohydrate labelled with a) **10** b) **9** c) **8** at 65 ^{o}C for 10 - 180 minutes.

The chromatograms outlined above (**Figure 3.14**) all display stable retention for internal standard peaks at 12-13 minutes, with the exception of samples labelled with **8** which was separated using an increased column temperature and displayed equally stable retention between concurrent injections. HILIC retention variation is often heavily dependant upon the size of the injected sample, and although this was not observed in HPLC separation, HILIC mode SPE did lead to some poor recovery in internal standard at lower reaction times making for challenging data processing. However, internal standard concentrations were still well above that of the LOD or LOQ making it possible to infer kinetic data from the samples. In order to further increase the reliability of the findings, samples were extracted in triplicate and concurrent injections averaged to give a mean carbohydrate concentration at each reaction time.

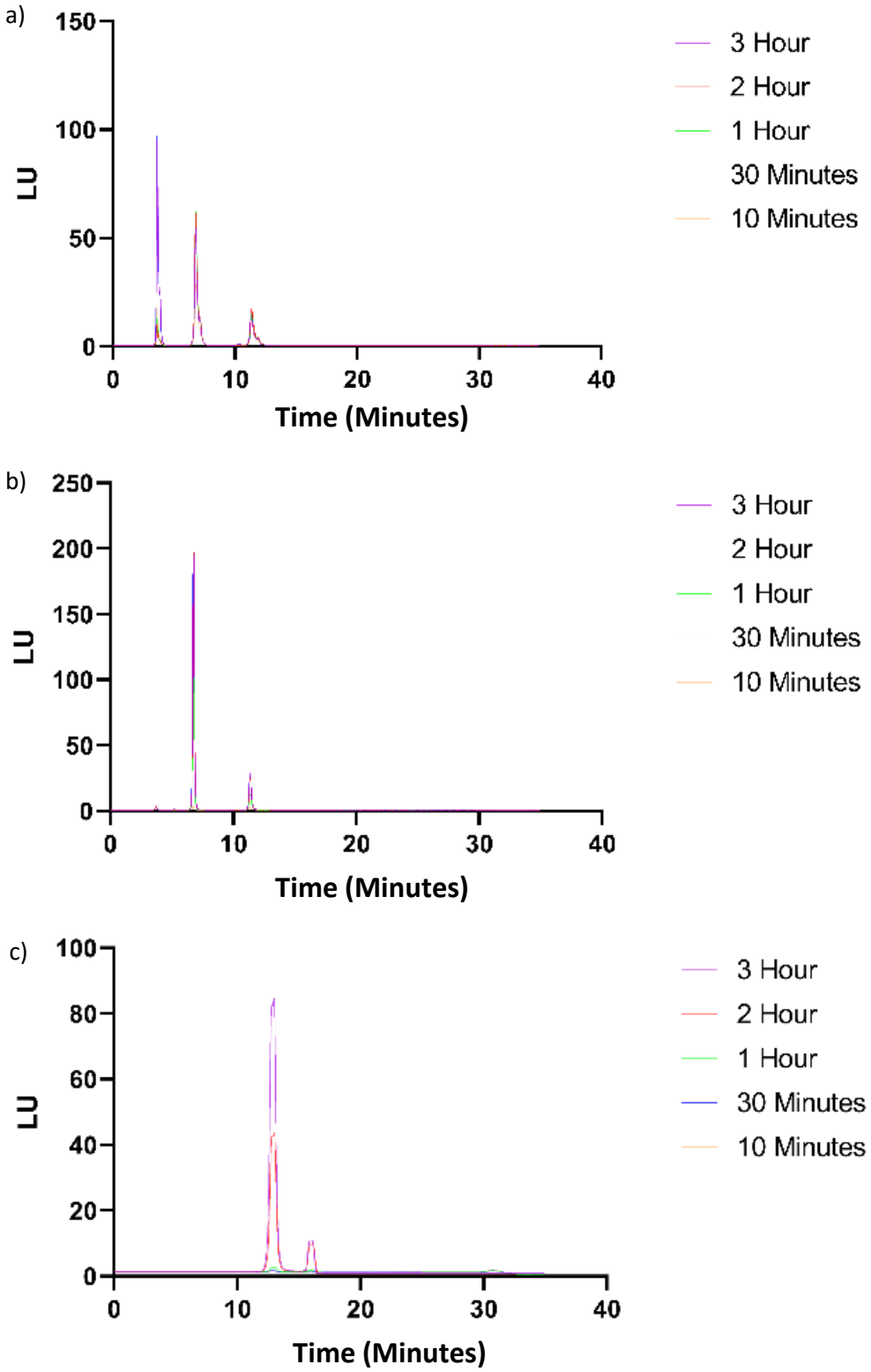


Figure 3.15 -Model carbohydrate labelled with a) **10** b) **9** c) **8** at 85 ^{o}C for 10 - 180 minutes.

The chromatograms in **Figure 3.15** show, increased levels of labelling completion can be observed with increased fluorescence signals at reduced reaction time up to completion, this also reduced the effects of poor SPE HILIC recovery due to the increased concentration of labelled sugar in solution.

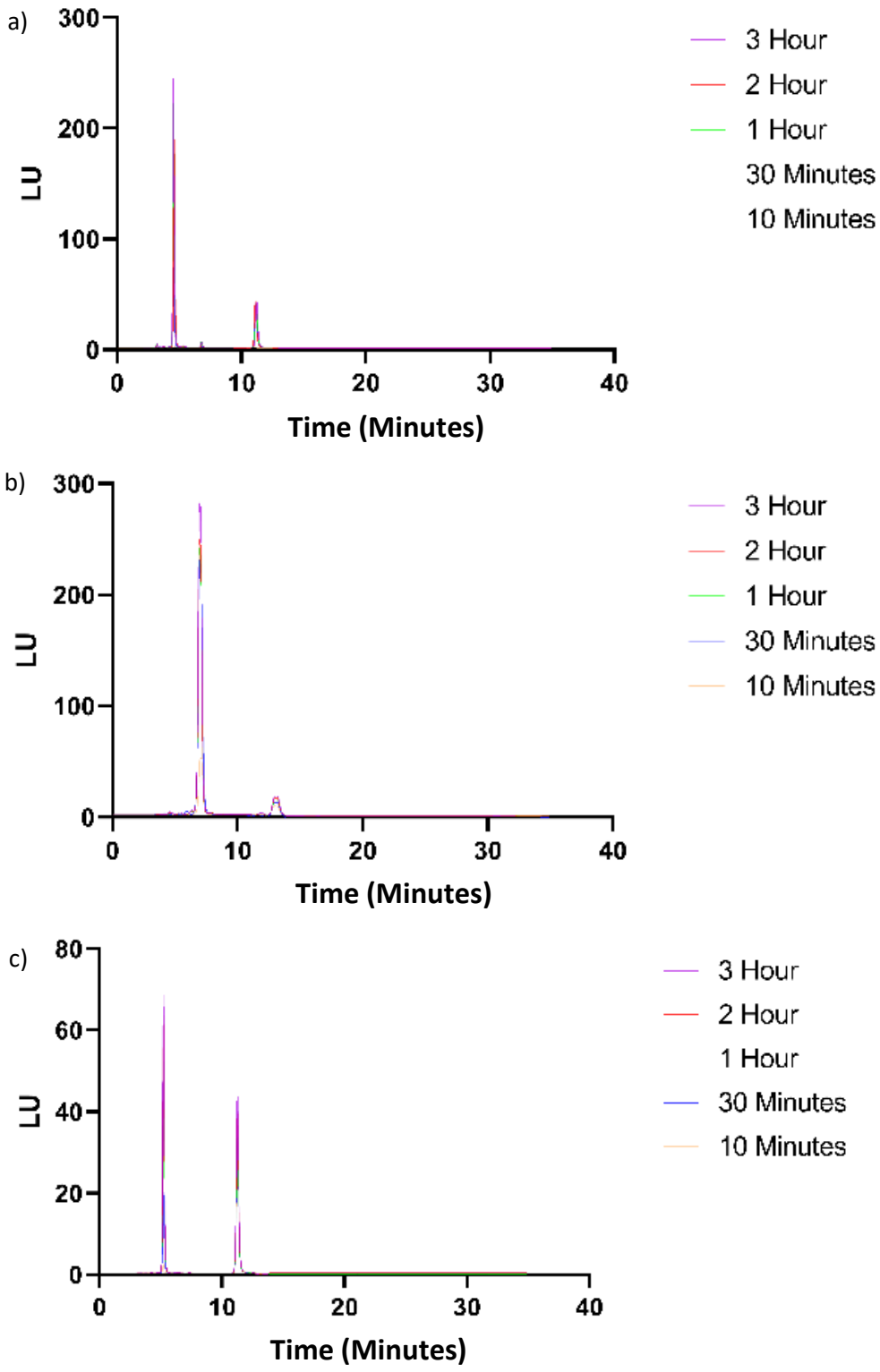
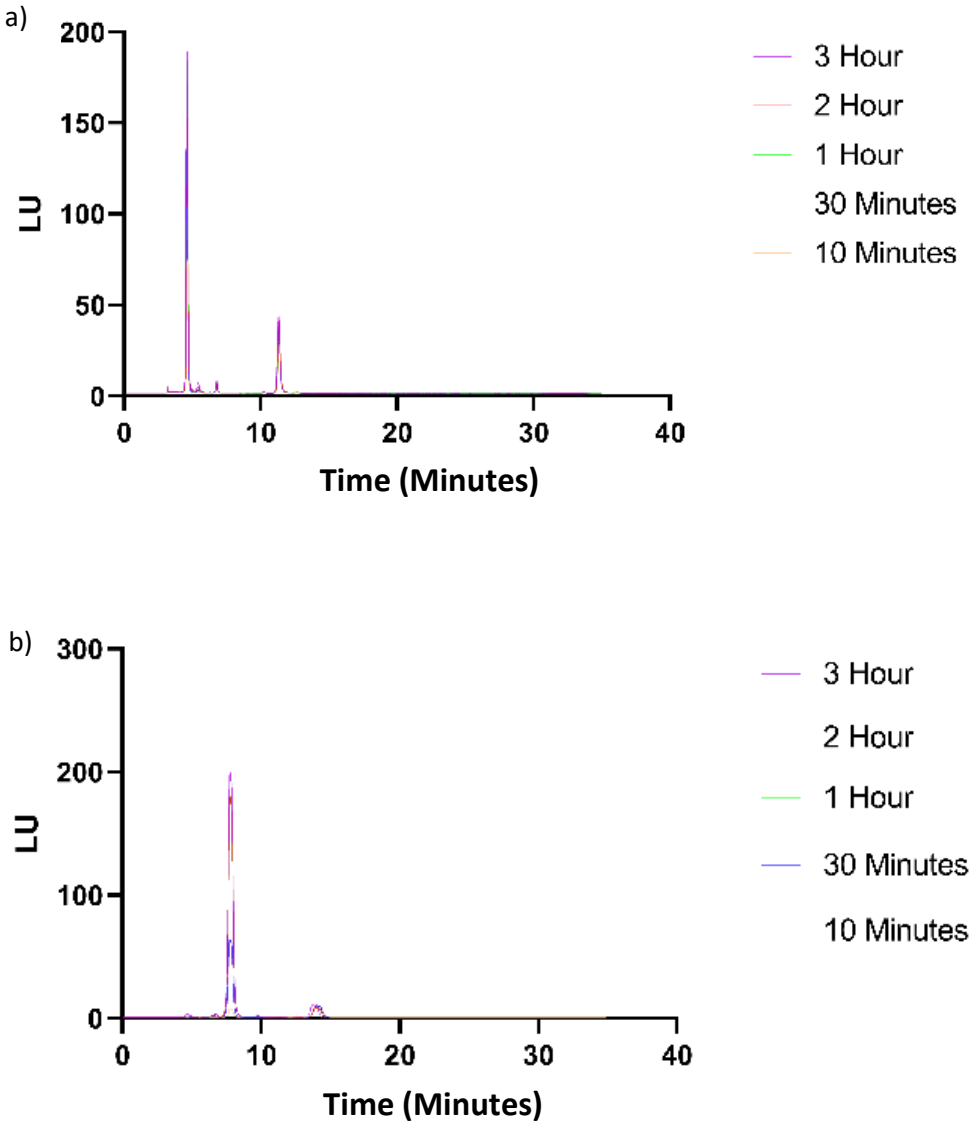


Figure 3.16 -Model carbohydrate labelled with a) **25** b) **26** c) **27** at 65 °C for 10 - 180 minutes.

The chromatographic traces in **Figure 3.16** show overlaid fluorescence traces for **25**, **26** and **27** at 65 °C carrying out labelling reactions with GU2 Lactose. Similar to the commercial samples, stable retention times are observed between concurrent injections both in samples and internal standard peaks. However, increased HILIC SPE retention was observed giving increased sample and internal standard recovery even at short reaction times. Significantly internal standard retention time was comparable and reproducible between **10**, **9**, **25** and **27** all run at 35 °C column temperature. While retention times for **8** and **26** were also comparable following separation at increased column temperature to improve separation between carbohydrates of 1 GU difference over a shorter run time.



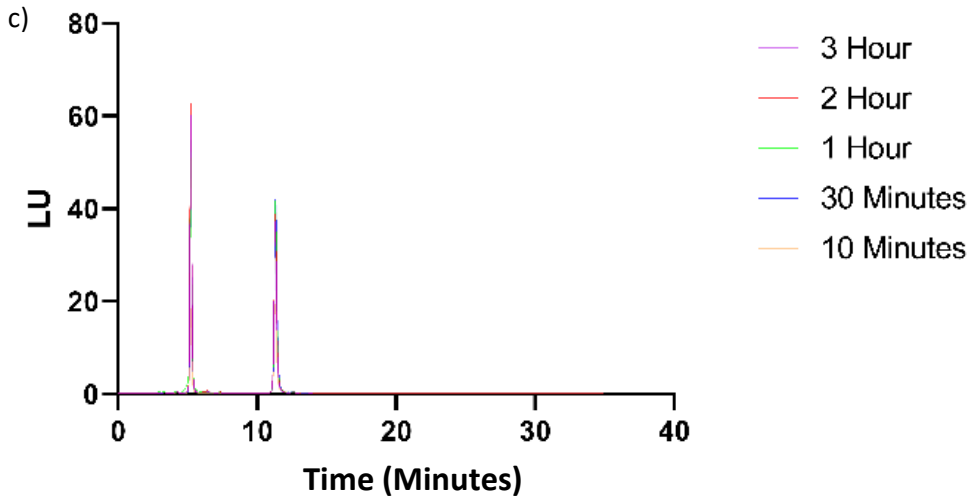


Figure 3.17 -Model carbohydrate labelled with a) **25** b)**26** c)**27** at 85 °C for 10 - 180 minutes.

A 20 °C increase in reaction temperature for labelling GU2 lactose carbohydrate also increased the speed of carbohydrate labelling for all three of the modified carbohydrate labels with no measurable change in retention or peak shape from that of the 65 °C samples. This suggests that they possess comparable stability even at elevated reaction temperatures (**Figure 3.17**).

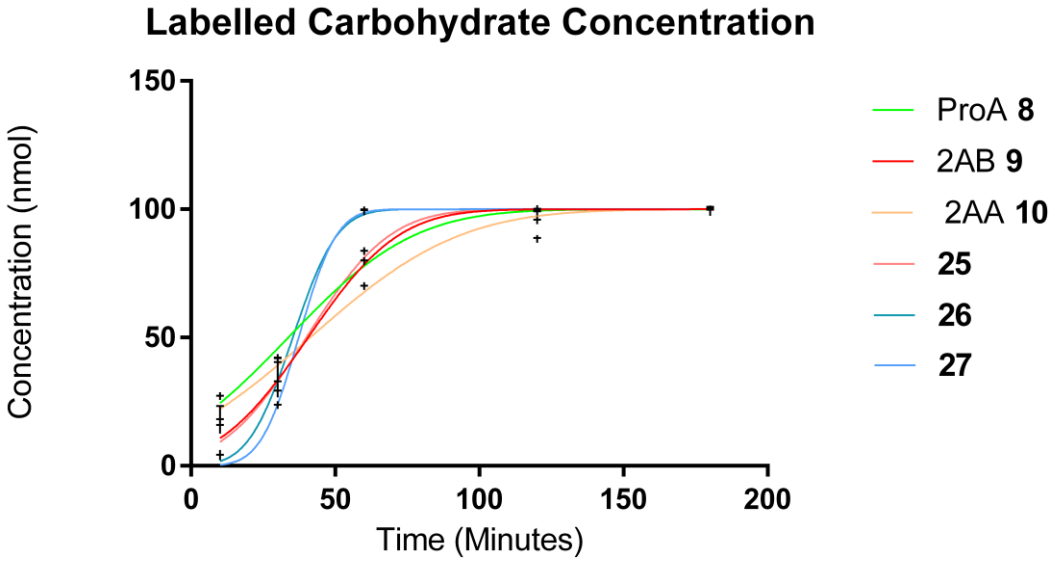


Figure 3.18 - Labelled glycan concentration over time for samples labelled at 65 °C with a) **10** b) **9** c) **8** d) **25** e) **26** f) **27**

Following the analysis of both commercially available and modified carbohydrate labels at both 65 and 85 °C in triplicate, quantified data from each sample was averaged to give a mean concentration of labelled carbohydrate for each time point (**Figure 3.18**). For 65 °C samples **10** and **9** show a linear

increase in labelled carbohydrate concentration up to 60 minutes before climbing to completion at 120 minutes this is mirrored by label **25**, however both *para*-derivatives **26** and **27** follow a slower reaction rate up until 30 minutes at which point both labels reached close to completion by 60 minutes. This rate of reaction is not mirrored by **8** at this temperature as the initial rate of labelling is higher within the first 30 minutes before the rate slows similar to the rates of **10** and **9**.

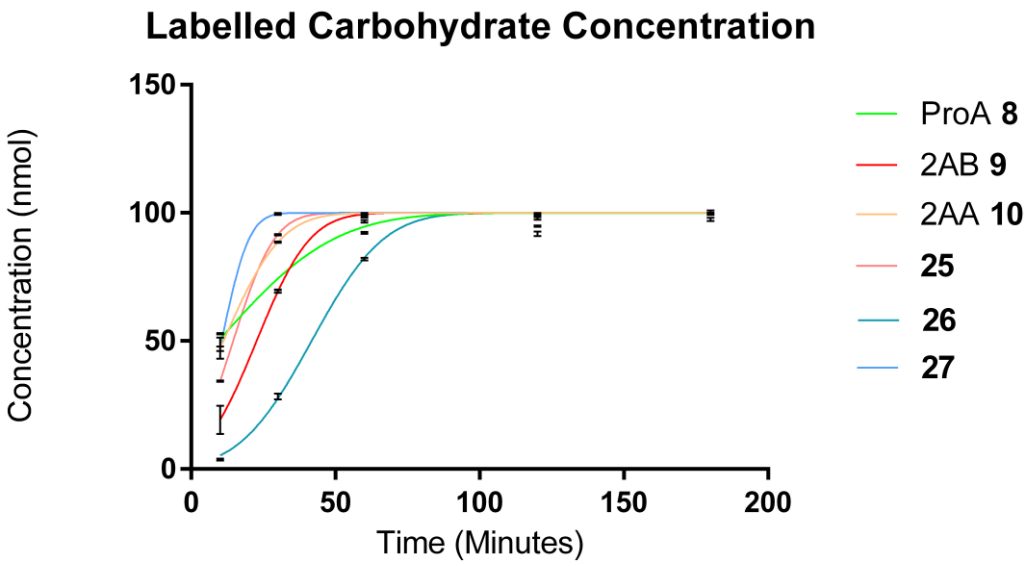


Figure 3.19 -Labelled glycan concentration over time for samples labelled at 85 °C with a) **10** b) **9** c) **8** d) **25** e) **26** f) **27**.

For 85 °C samples shown **Figure 3.19** the increased sample temperature leads to a marked acceleration in completion rate with novel labels **25** and **27** completing between 30 and 60 minutes. This follows the rate of **10** and **9**. Labels **26** and **8** however follow a slower reaction rate initially but **8** completes in 1 hour while **8** reaches completion soon after.

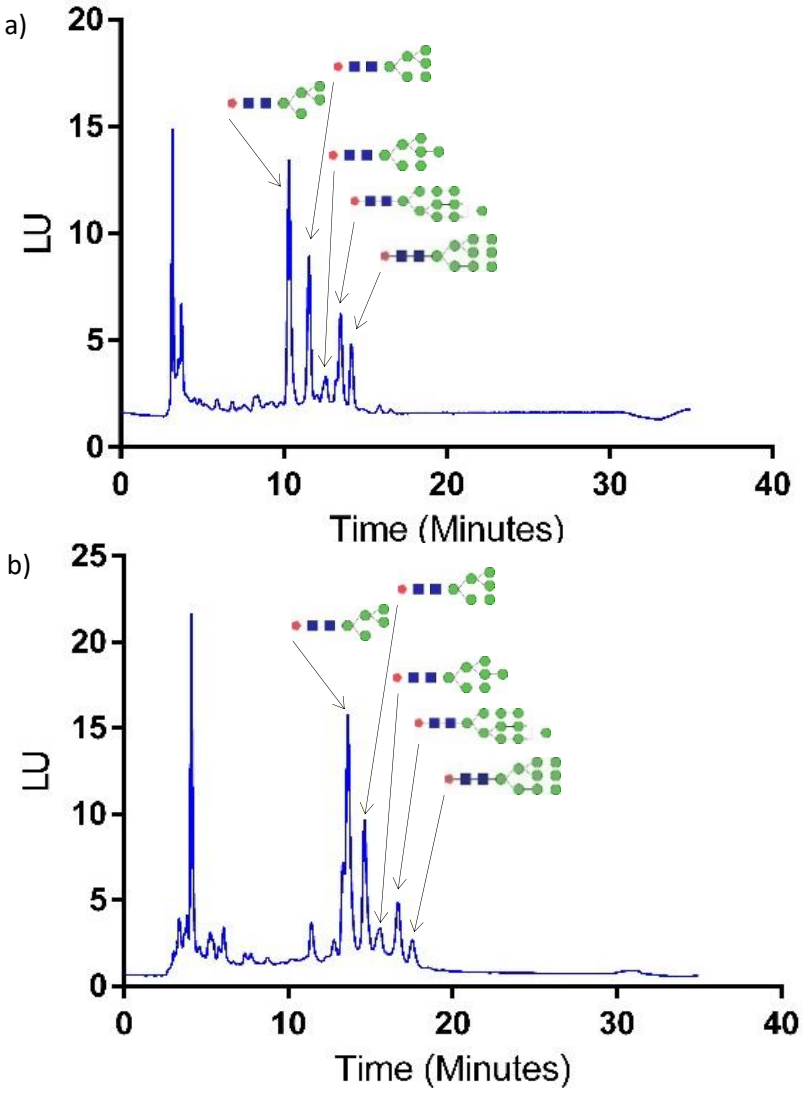
3.4.3 Labelling of released non sialylated glycan species

The successful separation and analysis of model carbohydrates labelled with labels **25**, **26** and **27** provided a proof of concept for multifunctional carbohydrate labels. However further analysis was required to demonstrate that labels **25**, **26** and **27** were capable of labelling glycans from natural sources. Labels are tested commercially against known glycoprotein standards such as Bovine pancreatic RNase B [16, 40]. This glycoprotein exhibits a single site for *N*-linked glycosylation and contains high mannose glycans of either 5-9 mannose subunits. This protein is the most widely

researched glycoprotein and therefore provided an excellent model glycan donor for the modified labels [40].

Enzymatic digestion was performed on denatured RNase B samples using recombinantly expressed PNGase F. Protein denaturation was first carried out by the addition of 40 mM dithiothreitol in a solution of 0.5% SDS at 95 °C for 10 minutes. Denatured samples were then cooled before the addition of recombinant PNGase F. Digestion took place over 24 h and the released glycans were then purified on graphitised carbon SPE with Hypercarb Hypersep PGC SPE cartridges (Thermoscientific).

Purified glycans were then concentrated in a vacuum centrifuge before being resuspended in water for labelling. To this solution borate-acetate buffered methanol was added containing sodium cyanoborohydride and labelling reagent. Samples where then heated to 65 °C in a water bath for 3 hours. Following labelling the reaction was quenched with ACN and free label removed by amide SPE (spe-ed2 amide cartridges applied separations). Purified glycans were then dried and reconstituted in H_2O for injection.



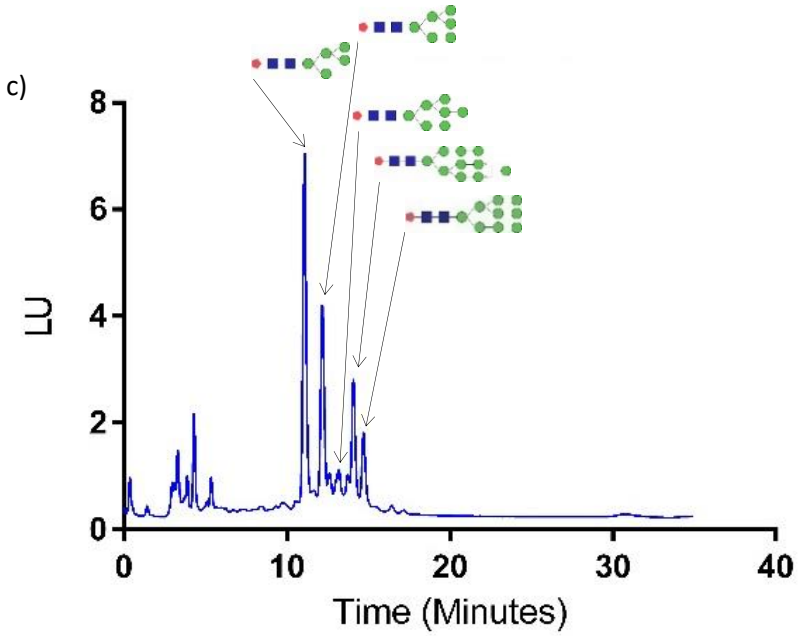


Figure 3.20 - Released RNase B glycans labelled with a) **25** b) **26** c) **27**

The five glycoforms of RNase B can be seen in **Figure 3.20**. The traces for all three modified labels closely resemble that of the commercial labels already published in the literature [16]. Labels **25** and **27** exist as constitutional isomers and display similar levels of retention with Man-5 peaks at 11.03 and 10.33 minutes respectively, while label **26** displays increased retention with an initial peak for Man 5 at 13.69 minutes. These retention differences are also replicated in labelled dextran homopolymer standards for each of the labels with **27** displaying the lowest levels of retention. This data also aligns with both dextran traces, and the literature remarks that HILIC mode glycan separation leads to carbohydrate size dependant elution with lower GU sugars eluting first [32, 41].

Table 3.4 - Chromatographic data showing retention time and peak area (R.T) for RNase B labelled with Multifunctional labels **25**, **26** and **27**

	Man - 5		Man - 6		Man - 7		Man - 8		Man - 9	
Label	R.T (Minutes)	Peak area	R.T (Minutes)	Peak area	R.T (Minutes)	Peak area	R.T (Minutes)	Peak area	R.T (Minutes)	Peak area
25	10.33	177.21	11.55	118.42	12.58	34.73	13.49	85.55	14.16	44.33
26	13.69	338.27	14.64	161.83	15.56	65.39	16.68	88.75	17.54	36.50
27	11.03	101.38	12.19	88.89	13.12	14.85	14.09	44.61	14.71	20.57

Table 3.4 shows collated peak areas and retention times for RNase B glycans labelled with **25**, **26** and **27**. The peak areas for the three modified labels also align with the previous data shown for labelled dextran traces **27** displayed the lowest levels of fluorescence in integrated peak area while only displaying a peak height of 7 LU for Man-5 glycan structures. **26** displayed increased peak height over **25** at 14 LU, combined with the narrower peak exhibited by **25** in previous samples resulted in significantly lower peak areas compared to **26**. It was observed that **26** labelled samples behaved in a similar fashion to its commercial derivative **8** displaying the highest peak areas of the three labels, but with an increased peak width. This could produce some challenges in the separation of highly sialylated glycans when high levels of resolution are required to differentiate between carbohydrates displaying high levels of structural similarity.

3.4.4 Effect of increased temperature on sialic acid stability

After confirming the ability of multifunctional labels **25**, **26** and **27** to derivatise high mannose glycans released from RNase B over reaction times comparable to commercial workflows, the study moved on to investigate the ability of multifunctional labels **25**, **26** and **27** to derivatise sialylated glycans that were enzymatically released from Bovine fetuin samples. Fetuin is a sialylated glycoprotein derived from bovine fetal serum where it functions as a storage and transport protein [42]. Bovine fetuin is commonly used as a standard in both HPLC and MALDI-ToF analysis[17, 43]as it contains mixture of bi- , tri- and tetra- sialylated glycofroms ranging from 9 -12 GU. Fetuin glycans were released enzymatically with recombinantly expressed PNGase F. Deglycosylation was performed following a denaturing step with a solution of 0.5% SDS and 50mM dithiothreitol at 95 ^{o}C after which time the samples were cooled to <37 ^{o}C and 100 µg of PNGase F was added. Samples were incubated at 37 ^{o}C for >12 hours until complete deglycosylation had taken place. Released glycans were purified on Hypercarb hypersep PGC SPE columns (Thermoscientific) and eluted in 50% aqueous ACN. These samples were then dried and pooled to ensure batch reproducibility between samples. The pooled samples were reconstituted in sufficient water so that each sample equated to 100 µg of protein initially digested. To each tube containing released glycan, a labelling solution was added consisting of sodium cyanoborohydride and labelling reagent dissolved in borate-acetate buffered methanol. The resulting samples were gently vortexed before being incubated at either 65 ^{o}C or 85 ^{o}C for 3 hours.

Labelled glycans were then purified on Spe-ed amide SPE cartridges (applied separations) and eluted in H_2O. The purified samples were then dried *in vacuo* before being reconstituted for injection. Analysis of labelled fetuin samples took place over 100 minutes on a HILIC mode amide 80-HR HPLC

column. The mobile phase was run as a gradient between 50 mM ammonium formate buffer (A) with ACN as phase B.

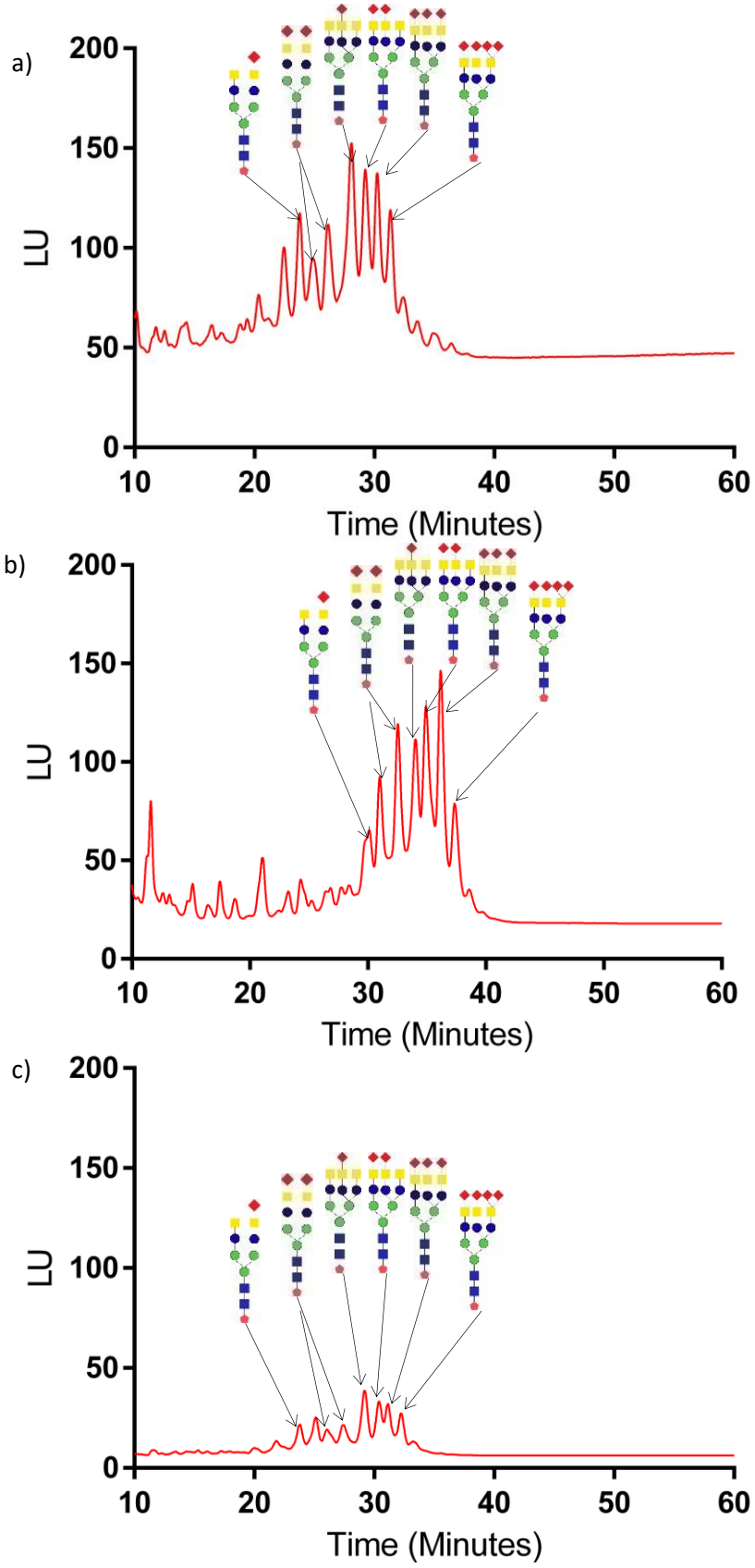
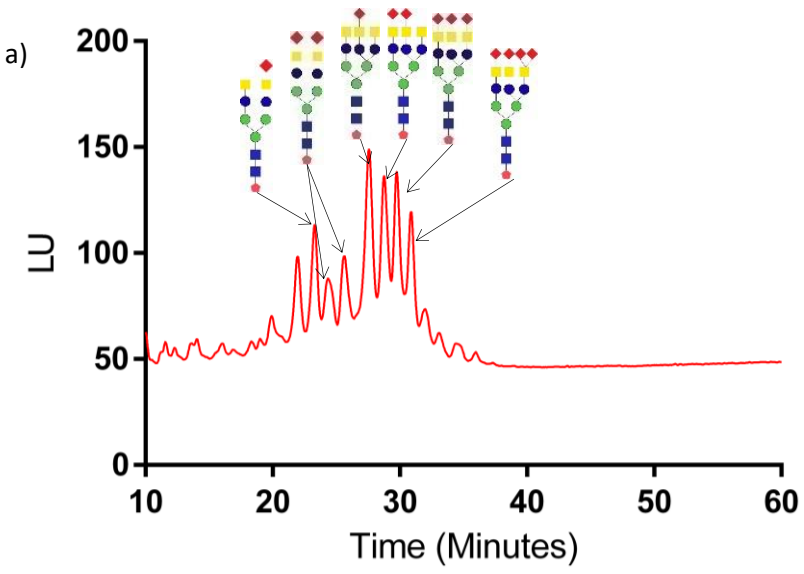


Figure 3.21 - Fetuin Samples labelled with a) **25** b) **26** c) **27** at 65 °C

The chromatographic traces displayed in **Figure 3.21** show separation of released fetuin samples derivatised by multifunctional labels **25**, **26** and **27** at 65 °C for a period of three hours. Multifunctional label **25** displayed the highest peak resolution of the three labels while **26** was not capable of separating all 7 peaks successfully. **27** meanwhile displayed adequate separation but displayed lower levels of fluorescence agreeing with separations seen in RNase B samples. Importantly, the presence of comparable peaks to commercial separations carried out on HILIC HPLC shows that sialylation can be maintained following 65 °C derivatisations with multifunctional labels **25**, **26** and **27**. However, baseline noise in both **25** and **26** samples suggests that minor de-sialylation has occurred [44, 45].

Having established that sialyation could be maintained in 65 °C samples over 3 hours sample extractions were repeated at 85 °C. These samples followed workflows established earlier in the chapter to investigate reaction completion rates where it was established that significantly increased reaction rates were observed at increased temperature. Previous studies have shown links between increased reaction temperatures and sialic acid cleavage from glycans displaying sialylated end capping [46, 47].



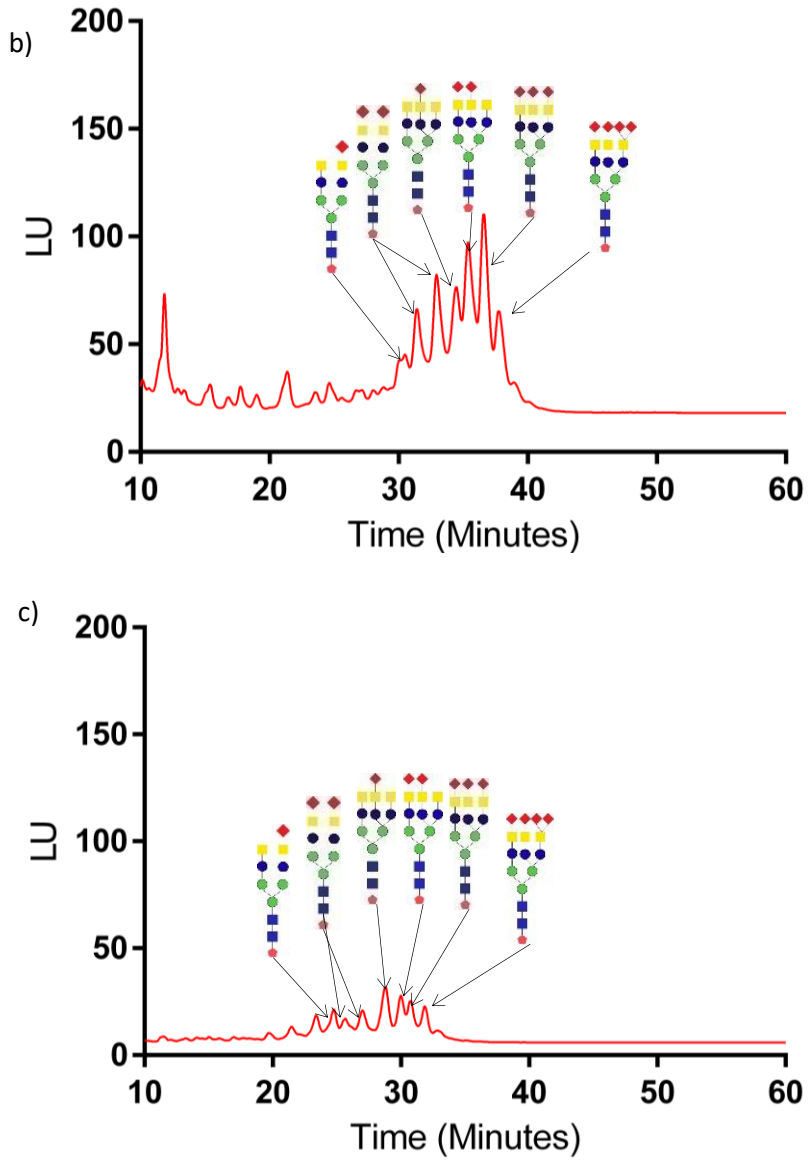


Figure 3.22 -Fetuin samples labelled with a)**25** b) **26** c) **27** at 85 °C

Derivatized fetuin samples were held at 85 °C for three hours and processed following the same clean-up conditions as detailed for the 65 °C samples before being analysed by HILIC HPLC-Fl. The resultant chromatographic traces can be seen in **Figure 3.22**. Multifunctional label **25** once again exhibited the highest resolution in this separation mechanism with **26** exhibiting reduced sample resolution. Interestingly, in all samples, separation and peak size were reduced in samples derivatised at 85 °C and this is most likely due to sialic acid loss and the presence of increased numbers of similar glycoforms within the pool. Label **27** exhibited similar separation characteristics to **25,** however, as the least fluorescently active of the three labels, reduced intensity was seen compared to **25** and **26**.

3.5 Conclusions and future work

The modification of 2AB **9** and ProA **8** yielded three carbohydrate labels displaying comparable detection sensitivity to their commercially available derivatives in HILIC mode HPLC-FLD analysis while possessing the ability to undergo downstream functionalisation with azide containing conjugation partners.

Initially, fluorescent wavelengths were determined by the use of scanning fluorescence spectroscopy. These were then used to obtain the limits of detection for the modified labels as well as their commercially available derivatives. This showed that both commercial and modified labels display LOD values between 10-25 $pMolml^{-1}$. Experimental determination of carbohydrate concentration present in a sample was carried out following the generation of purified carbohydrate standards using a GU2 model carbohydrate. A standard calibration curve was then obtained for each of the modified labelling compounds **25**, **26** and **27** giving comparable results to their commercially available derivatives **10**, **9** and **8** with r^2 values of 0.9997,0.9999 and 0.9971 respectively.

The separation of **25**, **26** and **27** was performed by the use of a labelled dextran homopolymer with labels **25**, **26** and **27** all displaying comparable separation results for the separation of up to 10 GU carbohydrate fragments. Labels **25**, **26** and **27** all displayed comparable sensitivity and resolution to 2AB **9** and ProA **8** as commercially available derivatives. Plotting retention time against polymer length showed positively correlated linear trends for both commercial and modified labelling strategies with r^2 values from 0.9968, 0.9941 and 0.9934 respectively.

Having shown the successful application of **25**, **26** and **27** as carbohydrate labels, their ability to effectively label known concentrations of carbohydrate was analysed by a series of labelling reactions that were quenched at different time points throughout the 3 hour labelling process. Both commercial labelling strategies and the modified derivatives behaved comparably at 65 ^oC with labelling being achieved to a greater than 95% level in 2-3 hours. This was significantly reduced to 1-2 hours with the use of increased temperature in samples labelled at 85 ^oC. However, increased temperature brings the increased risk of sample degradation.

Labels **25**, **26** and **27** were then used to analyse native glycan samples enzymatically released from RNAse B by the action of PNGase F. In these samples the three modified labels displayed comparable separation and resolution both with each other and also to literature results for analysis of RNAse B glycan samples. Following the successful separation of RNaseB, samples of deglycosylated bovine fetuin were labelled with **25**, **26** and **27** as well as their commercial derivatives with the intention to compare the ability of both commercial and modified labels to maintain sialylation following derivatisation. Multifunctional labels provided challenging separation by HPLC-HILIC due to the

similarity of glycoforms released from fetuin samples. However, separation showed intact sialylation for samples labelled at 65 °C. The experiment was then repeated at 85 °C for all samples to investigate the stability of sialic acid at increased temperatures that are associated with increased rates of carbohydrate labelling. This showed partial sialylation and some sialic acid loss resulting in reduced separation resolution.

The aim of this study was to evaluate the optical analysis characteristics of multifunctional labels **25**, **26** and **27** in line with existing glycan analysis workflows for optical profiling with commercially available labelling systems. This was achieved though the successful optimisation and application of multifunctional labels **25**, **26** and **27** to both model carbohydrates as well as from biological sources. Additionally this study was able to develop conditions for reductive amination in under 3 hours while applying existing commercial techniques. Furthermore the generation of multifunctionally labelled biologically derived carbohydrates provides a platform for further modification by CuAAC to facilitate improved analysis by MS applications or for the development of novel glycoconjugates.

3.6 Experimental

3.6.1 General Experimental

All chemicals used were purchased though commercial sources, (Sigma Merck, Fisher Scientific, Alfa Aesar) and used without further purification. Experiments were monitored via thin layer chromatography, TLC plates were aluminium backed 60 F254 silica (Sigma Merck). Visualisation was then carried out under UV light (λ = 254 nm) followed by staining with Ninhydrin (Sigma Merck).

3.6.2 Analytical Equipment

Nuclear Magnetic Resonance

Both proton (δ_H) and carbon (δ_c) Nuclear magnetic resonance, IR spectroscopy, melting point and high resolution mass spectrometry were carried out according to the methods carried out in chapter 2.

Scanning fluorescence spectroscopy

Fluorescence spectroscopy was carried out using a Varian eclipse fluorescence spectrometer fitted with a 10mm pathlength multicell holder.

Protein concentration

Enzyme concentration was determined by either Bradford assay or by use of a NanoDrop Lite spectrometer ay 270nm over a 1mm path length.

High performance liquid chromatography

HILIC mode HPLC was performed on an Agilent 1100 series HPLC system linked to both an Agilent G1321A FLD detector and an G1314A variable wavelength detector. The separation of carbohydrates was performed by amide HILIC on a TSKgel Amide-80 HR column (250 x 4.6 mm 5um particle size) fitted with a TSKgel Amide-80 guard column (15 x 3.2 mm 5 um particle size). Solvent A was composed of aqueous ammonium formate (50 mM pH 4.4) while solvent B consisted of ACN. The flow rate was set at 0.8 ml/min and the column was maintained at 35 ^{o}C for the duration of the run.

Separation of model carbohydrates took place over 35 minutes with a linear gradient beginning with 65% phase B to 30% B over 20 minutes, with a further gradient to 10% at 25 minutes before returning to injection conditions of 65% B between 27-35 minutes.

Siaylated fetuin samples were separated over 100 minutes beginning at 65% phase B with a linear gradient to 50% over 50 minutes. The column was then cleaned with 10% solvent B before being pre-equilibrated prior to the next sample injection.

MALDI-TOF was carried out on a Bruker Ultraflex II system calibrated with Bruker peptide standard II diluted 50% with a saturated solution of sDHB in TA30 (70% $ACN_{(aq)}$ 0.1% TFA). Calibration spots were then pipetted onto a ground steel MALDI plate (1 µl) and dried at reduced pressure.

Derivatized carbohydrate samples (10 µl) were mixed 1:1 with SDHB matrix solution and spotted onto the ground steel target plate. (1 µl) Samples were then dried under reduced pressure before being analysed.

Analysis was then carried out in reflection mode with a mass window of 200-4000 m/z with the laser set to 36% power. Acquisition then took place over 2000 shots in 200 shot increments. The summed spectra were then processed in Bruker Flex analysis.

3.6.3 Solvents and Buffers

2mM Ammonium formate buffer stock solution pH 4.4

Formic acid (184.12 g, 4 Mol) was combined with water (1 L) and cooled to 0 ^{o}C in an ice bath. 25% Ammonia solution (200 ml) was then added in 50 ml increments.. The pH was then raised to 4.4 by the addition of further 25% ammonia solution (5 ml). The stock solution was then diluted to 2 litres in a volumetric flask for a final concentration of 2 Molar. This solution was then diluted to 50 mM and used as a HILIC phase modifier.

Glycan labelling agent stock solution

Glycan labels **25**, **26** and **27** were weighed and diluted in methanol to a final concentration of 0.14 mMol/mL^{-1}

Labelling buffer solution.

Boric acid (3 g, 0.04 Mol) and sodium acetate trihydrate (6 g, 0.045 Mol) were dissolved in MeOH (100 ml)

Sodium cyanoborohydride solution

Sodium cyanoborohydride solution was prepared at 0.42 mMol/mL^{-1} in labelling buffer solution.

Lysogeny broth Medium (LB medium)

25g LB medium powder was dissolved in 1 litre of 18Ω water before being sterilised by autoclave.

Gel electrophoresis

Denaturing SDS page gels were manually cast at 15/4% according to the protocol outlined below.

Running gel: 15% A single SDS-PAGE gel solution was formed by combining 40% acrylamide (3 ml) with H_2O (2.8 ml). This solution was then buffered with 1.5 M Tris pH 8.8 (2 ml). To this solution 10% sodium dodecyl sulphate (80 µl) was added along with 10% ammonium persulphate (80 µl) and tetramethylethylethylenediamine (8 µl). This solution was pipetted into an empty cassette (Biorad mini) until approximately 15 mm from the top. The gel was then topped with isopropyl alcohol and left to set. Following successful gel setting, Isopropyl alcohol (IPA) was removed for the addition of the stacking gel.

Stacking gel: 4% stacking gel solution was formed by the combination of 40% acrylamide (0.75 ml) with H_2O (2.9 ml). This was then buffered with 0.5 mM Tris pH 6.8 (1.25 ml). To this solution, 10% sodium dodecyl sulphate (50 µl) was added along with 10% ammonium persulphate (50 µl) and tetramethylethylethylenediamine (5 µl). This solution was layered on top of the solidified running gel and a 12 well comb added. The gel was then allowed to set for 1 hour before use.

3.6.4 Determining the optical properties of the modified labels.

Generation of fluorescent GU7 maltoheptose

Maltoheptose (10 µg, 0.0086 µMol) was added to a reaction tube, dried under reduced pressure and reconstituted in H_2O (10 µl). Borate buffered methanol containing sodium cyanoborohydride (50 µl) was added along with labelling solution (25 µl). The solution (50 µl) was then heated at 65 ^oC in a water bath for 5 hours. The reaction was quenched with ACN (950 µl) at which point a white precipitate was formed. The suspended sample was then purified on HILIC amide SPE with Spe-ed2

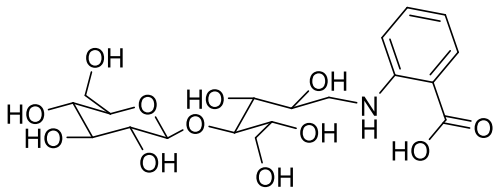
Amide cartridges (25 mg bed volume 1 ml) (Applied separations inc). Columns were washed and equilibrated with MeOH (2 x 1 ml), samples were added in 5% aqueous ACN (1 ml) and allowed to move through the column under gravity alone. Samples were then washed with 1% ACN (aq) (5 x 1 ml) followed by 3% ACN (aq) (5 x 1 ml) to remove free label. Finally, purified labelled GU7 was eluted with H_2O (600 µl) and collected in a clean 1.5 ml centrifuge tube. Purified samples were then dried with vacuum centrifugation at 50 ^oC before being resuspended for analysis.

3D scanning fluorescence spectroscopy of labelled Maltoheptose

Samples of GU7 maltoheptose labelled with modified glycan labels **25**, **26** and **27** were diluted to a concentration 5 µg/mL of GU7 and added to a windowed fused quartz low fluorescence cuvette. Samples were excited at 250 nm increasing by 10 nm per scan up to 500 nm. A detector then monitored the emission between 300-600 nm at a 5 nM optical slit size. Following this, optimal excitation wavelengths were obtained and fixed to obtain optimal emission wavelengths.

3.6.5 Synthesis of GU2 labelled carbohydrate standards

2-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzoic acid, (10 Labelled lactose)

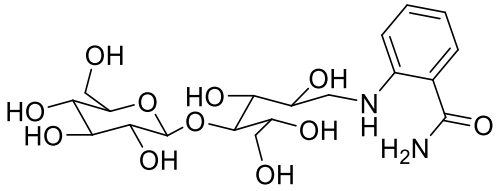


Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) and sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in borate acetate buffered MeOH was added. The solution was heated to 50 ^oC and **10** (40.0 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion. The reaction mixture was dried down under reduced pressure and dry loaded onto silica before being loaded onto a 10 cm silica flash column run with a mobile phase consisting of water, isopropanol and ethyl acetate in the ratio (1:2.5:6). This afforded **10** lablled lactose as a white solid (18 mg, 0.040 mMol, 70% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.72 (d, *J* = 7.5 Hz, 1H Ar CH), 7.33 (t, *J* = 7.0 Hz, 1H Ar CH), 6.83 (d, *J* = 8.0 Hz, 1H Ar CH), 6.73 (t, *J* = 7.5 Hz, 1H Ar CH), 4.39 (d, *J* = 7.5 Hz, 1H CH), 4.19 – 3.06 (m, 16H Carbohydrate).**^{13}C NMR (101 MHz, D_2O)** δ 181.53(C-OOH), 149.29 (C-NHC Ar) , 132.54 (CH Ar),

131.47 (CH Ar), 116.56(CH Ar) , 112.79(CH Ar), 103.09(C-COOH Ar), 79.52 (CH-O), 74.92(CH-**C**H-O), 72.54, 71.23, 71.05, 70.59, 70.20, 68.51, 62.04, 60.71, 45.79 (2CH_2-OH), 23.28(CH_2-NH). **HRMS:** ESI-ve $C_{19}H_{29}NO_{12}$ expected:462.1690 [M-H] found: 462.1685

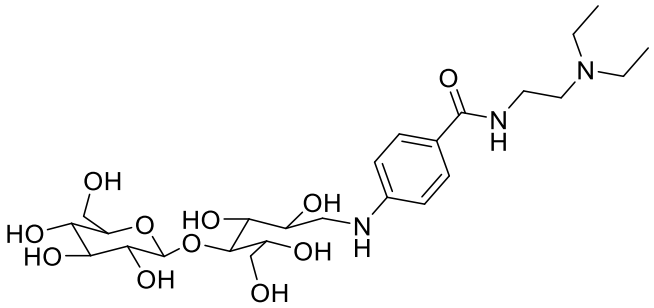
2-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzamide (9 labelled lactose)



Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 ^oC and **9** (39.45 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure and dry loaded onto silica before being loaded onto a 10 cm silica flash column run with mobile phase consisting of water, isopropanol and ethyl acetate in the ratio (1:2.5:6). This afforded **9** labelled lactose as an off white solid (19 mg, 0.041 mMol, 72% yield)

^{1}H NMR (400 MHz, D_2O) δ 7.51 (d, *J* = 7.5 Hz, 1H), 7.39 (t, *J* = 7.5 Hz, 1H), 6.87 (d, *J* = 8.5 Hz, 1H), 6.74 (t, *J* = 7.5 Hz, 1H), 4.43 (d, *J* = 7.5 Hz, 1H), 4.39 – 2.99 (m, 15H). **^{13}C NMR (101 MHz, D_2O)** δ 174.26(C-ONH_2), 148.60(C-NHC Ar), 133.64(CH Ar), 129.11(CH Ar), 116.74(CH Ar), 116.07(C-$CONH_2$ Ar), 113.15 (CH-**C**-2O), 103.03 (O-**C**H-CH), 79.26 (CH-**C**O-CH) , 74.90 (CH-**C**HCH_2-O), 72.52 (CH-**C**HOH-CH), 71.13 (CH-**C**HOH-CH), 71.02 (CH-**C**HOH-CH_2), 70.61 (CH-**C**HOH-CH), 70.09(CH-**C**HOH-CH), 68.44 (CH-**C**HCH_2-OH), 62.01(CHOH-**C**H_2-NH), 60.65 (**C**H_2OH-CH), 45.47(CHOH-**C**H_2-NH). **HRMS:** ESI+ve $C_{19}H_{30}N_2O_{11}$ expected [M+H]:463.1850 [M+H] found: 463.1922

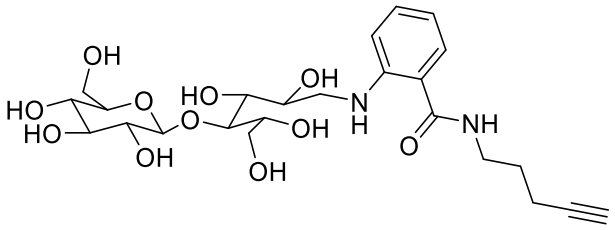
N-(2-(diethylamino)ethyl)-4-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzamide (8 labelled Lactose)



Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 ^oC and **8** (68.2 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure before being loaded onto a glass backed preparative TLC plate and run with mobile phase containing, Chloroform, methanol, water and ammonia in the ratio (8:8:1:2). This afforded **8** labelled lactose as a white solid (21 mg, 0.037 mMol, 65% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.58 (d, *J* = 9.0 Hz, 2H Ar 2 x CH), 6.74 (d, *J* = 8.5 Hz, 2H Ar 2 x CH), 4.32 (d, *J* = 8.0 Hz, 1H CH), 3.98 – 3.01 (m, 22H Carbohydrate), 1.15 (t, *J* = 7.5 Hz, 6H 2 x CH_3).**^{13}C NMR (101 MHz, D_2O)** δ 170.86 (C-ONH), 152.32 (CH-**C**NH-CH), 129.02 (2CH Ar), 120.71 (CH-**C**-CONH-CH), 112.74 (2CH Ar), 110.82 (CHOH-**C**-2O), 79.78 (CHOH-**C**HO-CHOHCH$_2$), 74.79 (CH$_2$-**C**H-O), 72.63 (CHOH-**C**HOH-CHOH), 72.33 (CHOH-**C**HOH-CHO$_2$), 70.79 (CHOH-**C**HOH-CH), 69.29 (CHOH-**C**HOH-CH$_2$), 68.44 (CHO-**C**HOH-CHOH), 64.24(CH$_2$CHO-**C**HOH-CHOH), 63.04 (CHO-**C**HCH$_2$-OH), 60.73 (**C**H$_2$OH-CH), 58.99 (**C**H$_2$OH-CH), 56.95 (CH$_2$-**C**H$_2$-N), 50.86 (2x N-**C**H$_2$-CH$_3$), 47.46 (CHOH-**C**H$_2$-NH), 35.70 (NH-**C**H$_2$- CH$_2$), 9.00(2 x CH$_2$-**C**H$_3$).**HRMS** ESI+ve $C_{25}H_{43}N_3O_{11}$ expected: 562.2898 [M+H] found: 562.2964

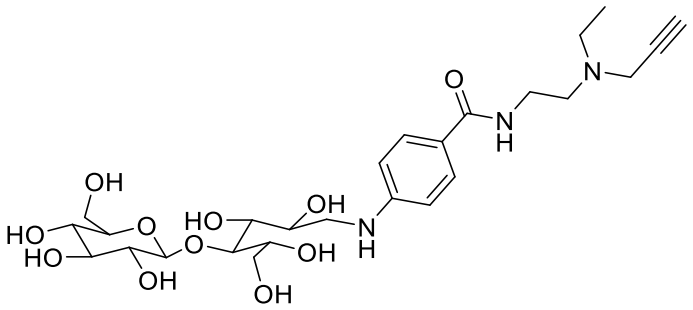
N-(pent-4-yn-1-yl)-2-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzamide (25 labelled lactose)



Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 °C and **25** (58.6 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure and dry loaded onto silica before being loaded onto a 10 cm silica flash column run with mobile phase consisting of water, isopropanol and ethyl acetate in the ratio (1:2.5:6). Following purification **25** labelled lactose was isolated as an off white solid (22 mg, 0.042 mMol, 65% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.36 – 7.22 (m, 2H Ar 2 x CH), 6.77 (t, *J* = 8.5 Hz, 1H Ar CH), 6.70 – 6.62 (m, 1H Ar CH), 4.33 (d, *J* = 8.0 Hz, 1H CH), 4.09 – 3.20 (m, 16H Carbohydrate), 3.02 (dd, *J* = 12.5, 8.4 Hz, 1H C≡CH), 2.17 (t, *J* = 7.0 Hz, 2H CH_2), 1.68 (p, *J* = 7.0 Hz, 2H CH_2).**^{13}C NMR (101 MHz, D_2O)** δ 181.50 (C-ONH), 171.67 (C-NHC Ar), 147.59 (CH Ar), 132.89 (CH Ar), 128.46 (CH Ar), 118.59 (C-CONH Ar), 117.18 (CH Ar), 113.07 (CHOH-**C**-2O), 103.04 (O-**C**HOH-CHOH), 79.22 (CH_2-**C**H-CH), 74.95 (CHOH-**C**HCH_2-O), 72.53 (CHOH-**C**HOH-CHOH), 71.13 (CHOH-**C**HOH-CH), 71.02 (CHOH-**C**HOH-CH_2), 70.65 (CHO-**C**HOH-CHOH), 70.09 (CHOH-**C**HOH-CHOH), 68.46 (CH-**C**H), 62.03 (CHO-**C**HCH_2-OH), 60.73 (**C**H_2OH-CH), 45.60 (**C**H_2OH-CH), 38.56 (CHOH-**C**H_2-NH), 27.26 (NH-**C**H_2-CH_2), 23.25, (CH_2-**C**H_2-CH), 15.27 (CH_2-**C**H_2-CH_2). **HRMS** ESI+ve $C_{24}H_{36}N_2O_{11}$ expected:529.2319 [M+H] found: 529.2392

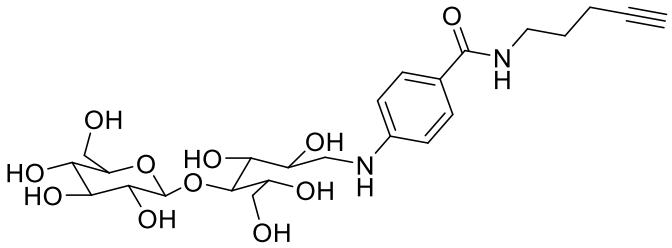
N-(2-(ethyl(prop-2-yn-1-yl)amino)ethyl)-4-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzamide (26 labelled lactose)



Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 ^oC and **26** (71.1 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure before being loaded onto a glass backed preparative TLC plate and run with mobile phase containing, Chloroform, methanol, water and ammonia in the ratio (8:8:1:2). Following purification **26** labelled lactose was isolated as an off white solid (16 mg, 0.029 mMol, 50% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.58 (d, *J* = 8.5 Hz, 2H Ar 2 X CH), 6.75 (dd, *J* = 9.0, 5.0 Hz, 2H Ar 2 X CH), 4.40 – 3.17 (m, 28H Carbohydrate), 2.73 (t, *J* = 6.5 Hz, 2H CH_2), 2.58 (q, *J* = 7.0 Hz, 2H CH_2), 0.98 (t, *J* = 7.0 Hz, 3H CH_3). **^{13}C NMR (101 MHz, D_2O)** δ 215.06 (C-ONH), 195.24 (CH-**C**NH-CH), 194.26(2CH Ar), 128.91 (CH-**C**-CONH-CH), 112.74 (2CH Ar), 103.03 (CHOH-**C**-2O), 80.90 (CHO-**C**HOH-CHOH), 79.46 (CH_2-**C**H-CH), 74.94 (CH_2-**C**H-CH), 72.50 (CHOH-**C**HOH-CHOH), 72.46 (CHOH-**C**HOH-CH), 71.14 (CH-**C**H), 71.02 (CHOH-**C**HOH-CH_2), 70.42 (CHO-**C**HOH-CHOH), 68.44 (CHOH-**C**HOH-CH), 66.62 (CHO-**C**HCH_2-OH), 61.98 (**C**H_2OH-CH), 60.69 (**C**H_2OH-CH), 59.34 (NH-**C**H_2-CH_2), 51.12 (CH_2-**C**H_2-N), 47.32 (N-**C**H_2-CH_3), 41.03(N-**C**H_2-CH), 37.14 (CHOH-**C**H_2-NH), 10.89 (CH_2-CH_3). **HRMS** ESI+ve $C_{26}H_{41}N_3O_{11}$ expected:572.2741 [M+H] found: 572.2823

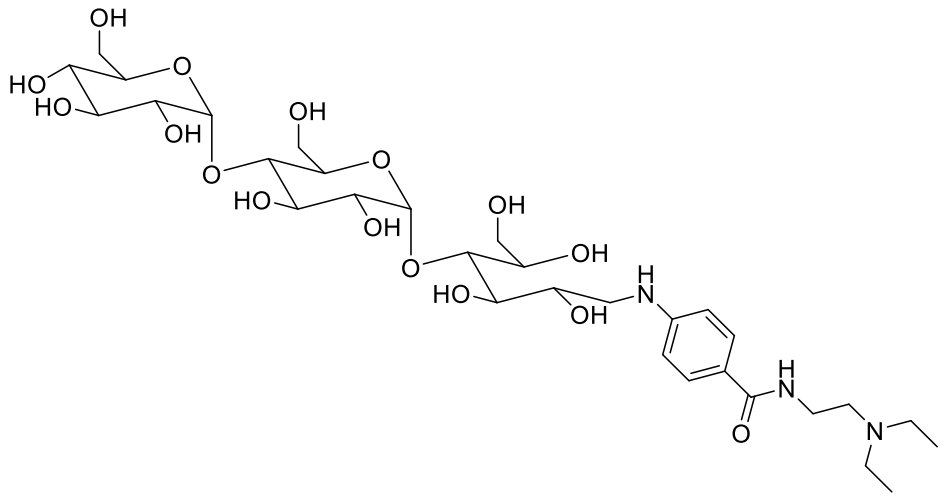
N-(pent-4-yn-1-yl)-4-((2,3,5,6-tetrahydroxy-4-(((2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)hexyl)amino)benzamide (27 labelled lactose)



Lactose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 ^oC and **27** (58.6 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure and dry loaded onto silica before being loaded onto a 10 cm silica flash column run with mobile phase consisting of water, isopropanol and ethyl acetate in the ratio (1:2.5:6). Following purification **27** labelled lactose was isolated as an off white solid (22 mg, 0.042 mMol, 73% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.55 (d, *J* = 8.5 Hz, 2H Ar 2 x CH), 6.72 (d, *J* = 8.5 Hz, 2H Ar 2 x CH), 4.40 (d, *J* = 7.5 Hz, 1H CH), 4.09 – 2.98 (m, 18H Carbohydrate), 2.25 (d, *J* = 2.5 Hz, 1H C≡CH), 2.20 (td, *J* = 6.9, 2.5 Hz, 2H CH_2), 1.78 – 1.65 (m, 2H CH_2). **^{13}C NMR (101 MHz, D_2O)** δ 181.50 (**C**-ONH), 170.46 (CH-**C**NH-CH Ar), 151.70 (2CH Ar), 128.88 (C-CONH Ar), 121.92 (2CH Ar), 112.77 (CH-**C**-2O), 103.05 (CH-**C**O-CH), 79.51 (CH_2-**C**H-CH), 74.96 (CH-**C**HCH_2-O), 72.52 (CH-**C**HOH-CH), 71.18 (CHO-**C**HOH-CH), 71.04 (CH-**C**HOH-CH_2), 70.44(CH-**C**HOH-CH), 70.03 (CH_2-CH), 68.46 (CHOH-CHOH-CH), 62.00 (CHO-CHOH-CH_2), 60.71(**C**H_2OH-CH) , 45.56 (**C**H_2OH-CH), 38.82 (CHOH-**C**H_2-NH), 27.39 (NH-**C**H_2-CH_2), 23.26 (CH_2-**C**H_2-CH), 15.28(CH_2- **C**H_2-CH_2). **HRMS ESI+ve** $C_{24}H_{36}N_2O_{11}$ expected:529.2319 [M+H] found: 529.2390

N-(2-(diethylamino)ethyl)-4-((4-(((2R,3R,4R,5S,6R)-3,4-dihydroxy-6-(hydroxymethyl)-5-(((2R,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl)oxy)tetrahydro-2H-pyran-2-yl)oxy)-2,3,5,6-tetrahydroxyhexyl)amino)benzamide (8 labelled maltotriose)



Maltotriose (20 mg 0.058 mMol) was dissolved in H_2O (200 µl) followed by the addition of Sodium cyanoborohydride (36.72 mg 0.58 mMol 10 eq) in Borate acetate buffered MeOH. The solution was heated to 50 °C and **8** (71.1 mg 0.29 mmol 5 eq) was added in MeOH (1 ml). The solution was stirred overnight and monitored by TLC until completion). The reaction mixture was dried down under reduced pressure before being loaded onto a glass backed preparative TLC plate and run with mobile phase containing, Chloroform, methanol, water and ammonia in the ratio (8:8:1:2). Following purification **8** labelled maltotriose was isolated as an off white solid (15 mg, 0.021 mMol, 37% yield).

^{1}H NMR (400 MHz, D_2O) δ 7.56 (d, *J* = 8.5 Hz, 2H Ar 2 x CH), 6.70 (d, *J* = 8.5 Hz, 2H Ar 2 x CH), 5.26 (s, 1H NH), 4.10 (s, 2H 2 x CH), 3.94 – 3.08 (m, 23H carbohydrate), 3.04 (t, *J* = 6.5 Hz, 2H CH_2), 2.94 (q, *J* = 7.0 Hz, 4H 2 x CH_2), 1.11 (t, *J* = 7.5 Hz, 6H 2 X CH_3).**^{13}C NMR (101 MHz, D_2O)** δ 181.52 (C-ONH), 170.84 (CH-**C**NH-CH), 152.09 (2CH Ar), 129.10 (CH-**C**CONH-CH), 120.55 (2CH Ar), 112.45 (2 CHOH-**C**H-2O), 99.74 (CHOH-**C**HO-CH), 95.87 (CHOH-**C**HO-CH), 87.96 (CHOH-**C**HCH$_2$-O), 72.85 (CHO-**C**HCH$_2$-O), 72.67 (CHOH-**C**HOH-CHOH), 72.36 (CHO-**C**HOH-CHOH), 71.76 (2 CHOH-**C**HOH-CHO$_2$) , 71.57 (CHOH-**C**HOH-CH$_2$), 70.90 (CHO-**C**HOH-CHOH), 69.33 (CH-**C**HOH-CHOH), 68.40 (CHO-**C**HCH$_2$-OH), 63.35

(**C**H_2OH-CH), 60.42 (**C**H_2OH-CH), 50.87 (**C**H_2OH-CH), 47.47 (CH_2-**C**H_2-N), 43.38 (2 N-**C**H_2-CH_3), 35.69 (CHOH-CH_2NH), 23.30 (NH-**C**H_2-CH_2), 9.00 (2 CH_2-**C**H_3). **HRMS** ESI+ve $C_{31}H_{53}N_3O_{16}$ expected:736.3425 [M+Na] found: 736.3427

3.6.6 Expression of PNGase F

Overnight cultures of BL21(DE3) bearing the pOF6 (PNGase F) plasmid (a kind gift from Shaun Lott-addgene plasmid 40315) [48] were grown in LB media containing 100 µg/ml ampicillin at 37 °C. Following incubation overnight 0.8% of the final culture volume was taken and used to inoculate 1 L of fresh LB media containing 100 ug/mL ampicillin. This culture was then shaken (210 rpm) and incubated at 37 °C until an optical density of 0.5-0.7 was reached. The culture was then induced with IPTG (1 mM final concentration) resulting in the expression of pOF6 and the formation of PNGase F. The culture was incubated overnight at 37 °C while being agitated at 210 rpm. Following overnight shaking, the culture was pelleted in a centrifuge (8000 g 30 mins 4 °C).

The Expressed recombinant protein was then purified using a variation of the osmotic shock methodology described by Loo *et al*[48]. The pelleted cell culture was resuspended in a 2% culture volume with chilled sucrose solution (0.5 M), tris-HCL (0.1 M) at pH 8 and incubated for 20 minutes while shaking. The cell lysate was then pelleted at 8000 g 20 mins 4 °C. Following this the supernatant was removed and cell pellet was resuspended in 2% culture volume of chilled 18Ω H_2O and shaken at 4 °C 20 minutes. $MgCl_2$ was then added to the lysate solution to a final concentration of 20 mMol. Tubes containing cell lysate were then incubated for a further 20 minutes. Cell lysate was centrifuged at 8000 g 20 mins 4 °C and the supernatant was filtered through a 0.45 um syringe filter to remove larger cell fragments. Filtered recombinant protein sample was then added to binding buffer composed of f HEPES (20 mM), NaCl (300 mM) and imidazole (20 mM). A HisTrap HP nickel affinity column (5 ml, GE Healthcare) was equilibrated with 5 column volumes (CV) of distilled water followed by 5 CV of binding buffer. The sample in binding buffer (5 ml) was added to the nickel affinity column and flow through recycled three times to ensure optimal protein binding.

The absorbed protein was then washed with 5 CV to remove smaller cell fragments and other proteins before being eluted in 2 mL fractions with a solution containing HEPES (20 mM), NaCl (300 mM) and imidazole (300 mM). Fractions were then analysed for the target protein by SDS-PAGE (15%). Samples were taken from each fraction and were transferred to centrifuge tubes along with denaturing buffer consisting of 0.5% SDS and 40 mM DTT and heated to 95 °C for 10 minutes. Samples were cooled and quenched with 4x running buffer. Samples of each fraction were transferred to each well and the gel was run for 45 minutes at 150V or until the marker reached the bottom of the gel. The gel was then stained with Coomassie stain for 1 hour before being rinsed and

destained in a solution of acetic acid in aqueous methanol. Fractions containing the target protein were then consolidated and dialysed overnight at 4 °C in dialysis buffer (2 L tris-HCL (20 mM) NaCL (20 mM) Na_2EDTA (5 mM) pH 7.5). The dialysed protein was then concentrated by centrifugation at 4000 rpm in a 15 kDa molecular weight cut off centrifugal filter. The concentrated sample was verified for protein content by nanodrop and by Bradford assay and stored at 4°C at a concentration of 1 mg/ml.

3.6.6.1 Activity testing PNGase F

The viability of recombinant PNGase F produced as a result of expression of the pOPH6 plasmid was tested against the deglycosylation of RNaseB. A sample of RNase B (20 µg, 9 µl) was added to denaturing buffer (1 µl 40 mM DTT in 0.5% SDS) and held at 95 °C for 10 minutes. The mixture was cooled and PNGase F was added in a solution of sodium phosphate buffer (50 mM, pH 7) to a final volume of 20 µl. The sample was then incubated at 37 °C for 120 minutes. The digestion was stopped by the addition of 4x loading dye (4 µl) before samples were analysed by SDS-PAGE (15%) at 150V. The gel was then stained with Coomassie stain (2 h) before being rinsed and destained for 6 hours revealing a >1 KDa mass change (**Figure 3.23**).

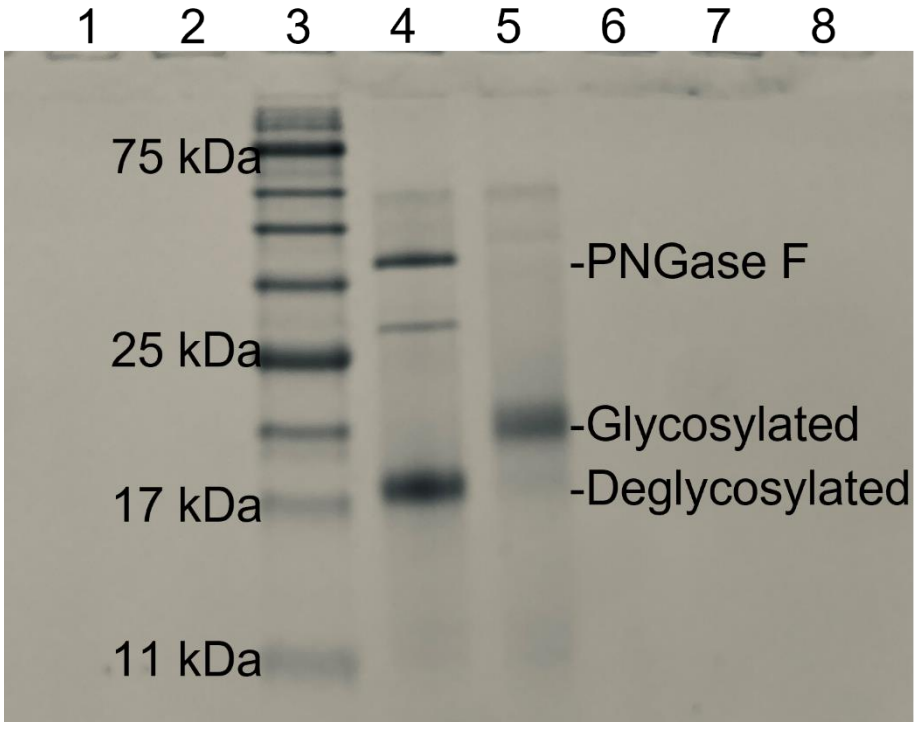


Figure 3.23 - Polyacrylamide gel showing PNGase F activity applied to the deglycosylation of RNase B. Lane 3 contains a ladder of protein molecular mass markers, Lane 4 contains RNase B with PNGase F and Lane 5 contains RNase B.

3.6.7 Preparation of carbohydrates

3.6.7.1 Dextran

Dextran from *Leuconostoc mesenteroides* (Mr ~200,000, 1 g) was dissolved in a solution of 0.1M HCL (10 ml) and heated to 100 °C under reflux conditions. The conditions were maintained for 4 hours while stirring. Following this, the solution was cooled on ice to room temperature and pH neutralized using Amberlite IRA 157 (1.5 g). The solution containing ion exchange resin was stirred until neutrality was achieved, pH 7. The solution was then filtered and collected before being diluted to 0.1 mg/mL in water.

3.6.7.2 Deglycosylation of RNase B

A sample of RNase B (1 mg 270 µl) was added to denaturing buffer (40 mM DTT in 0.5% SDS 30 µl) and was held at 95 °C for 10 minutes. The mixture was cooled and PNGase F (100 ug) was added along with PBS (300 µl). The sample was then incubated at 37 °C 12 hours.

The solution was then made up to 1 ml with 18Ω H_2O centrifuged at 10,000 rpm for 10 minutes. The supernatant was loaded onto a hypersep hypercarb column (Thermo Scientific) previously rinsed stepwise with MeOH (1 ml) sodium hydroxide (1 M, 1 ml), acetic acid (30%, 1 ml) and finally rinsed with H_2O (1 ml). The column was then equilibrated with buffer B (50% ACN, 0.1% $TFA_{(aq)}$ 1 ml) followed by buffer A (5% ACN, 0.1% $TFA_{(aq)}$, 1 ml). The glycan sample was then applied as a crude reaction mixture (1 ml) and allowed to pass through the sorbent by gravity alone. The sample was then washed with H_2O (1 ml) followed by buffer A (1 ml). Glycans were then eluted with buffer B (1 ml) and the collected glycan sample was dried down by centrifugal concentration under reduced pressure. The dried glycan sample was reconstituted in H_2O (100 µl) and stored at -20 °C. The sample recovered allows for 10 labelling reactions to take place.

3.6.7.3 Deglycosylation of Fetuin

A sample of fetuin (100 µg 270 µl) was added to denaturing buffer (40 mM DTT in 0.5% SDS 30 µl) and held at 95 °C for 10 minutes. The mixture was cooled PNGase F (100 µg) added along with PBS (300 µl). The sample was then incubated at 37 °C 12 hours.

The solution was then made up to 1 ml with 18Ω H_2O centrifuged at 10,000 rpm for 10 minutes. The supernatant was loaded onto a hypersep hypercarb column (Thermo Scientific) previously rinsed stepwise with MeOH (1 ml) sodium hydroxide (1 M, 1 ml), acetic acid (30%, 1 ml) and finally rinsed with H_2O (1 ml). The Column was then equilibrated with buffer B (50% ACN, 0.1% $TFA_{(aq)}$ 1 ml) followed by buffer A (5% ACN, 0.1% $TFA_{(aq)}$, 1 ml). The glycan sample was then applied as a crude reaction mixture (1 ml) and allowed to pass through the sorbent by gravity alone. The sample was

then washed with H_2O (1 ml) followed by buffer A (1 ml). Glycans were then eluted with buffer B (1 ml) and the collected glycan sample was dried down by centrifugal concentration under reduced pressure. The dried glycan sample was reconstituted in H_2O (10 µl) for analysis.

3.6.8 General Labelling and purification of model carbohydrates and retention standards for HPLC-FL and MS applications

The carbohydrate sample was applied to a 1.5 ml centrifuge tube (10 µl) and mixed with labelling buffer containing sodium cyanoborohydride (15 µl). To this labelling solution was added MeOH (25 µl). The reaction mixture was then vortexed and centrifuged for 10 seconds before being heated at the required time for a nominal of 3 hrs unless otherwise specified. In the case of kinetic studies, the samples were heated to either 65 ^oC or 85 ^oC for 10, 30, 60,120, 180 minutes respectively.

Following reaction completion, the samples were quenched with the addition of ACN (950 µl) producing a white precipitate. The quenched solution was then transferred to a spe-ed2 amide SPE column pre equilibrated with MeOH (5 ml). The sample was loaded and allowed to drip through under gravity. The sample was then washed with 99% $ACN_{(aq)}$ (5 x 1 ml) followed by 97% $ACN_{(aq)}$ (5 x 1 ml). The purified sample was then eluted with 18Ω H_2O (600 µl), the eluted samples were then dried down and resuspended in 200 µl for injection.

3.6.9 General labelling procedure for released N-Glycans

Tubes containing purified glycan (10 µl) were mixed with labelling buffer containing sodium cyanoborohydride (15 µl). To this labelling solution was added MeOH (25 µl). The reaction mixture was then vortexed and centrifuged for 10 seconds before being heated at 65 ^oC for 3 hours unless otherwise specified.

Following reaction completion, the samples were quenched with the addition of ACN (950 µl) producing a white precipitate. The quenched solution was then transferred to a spe-ed2 amide SPE column pre equilibrated with MeOH (2 ml). The sample was loaded and allowed to drip through under gravity. The sample was then washed with 99% $ACN_{(aq)}$ (5 x 1 ml) followed by 97% $ACN_{(aq)}$ (5 x 1 ml). The purified sample was then eluted with 18Ω H_2O (600 µl), the eluted samples were then dried down and resuspended in 200 µl for injection.

3.7 References

1. A. Pralow, S. Cajic, K. Alagesan, D. Kolarich and E. Rapp, in *Advances in Glycobiotechnology*, eds. E. Rapp and U. Reichl, Springer International Publishing, Cham, 2021, pp. 379-411.
2. R. Zhu, S. Zhou, W. Peng, Y. Huang, P. Mirzaei, K. Donohoo and Y. Mechref, *J. Proteome Res.*, 2018, **17**, 2668-2678.
3. M. Xu, W. Hu, Z. Liu, J. Xia, S. Chen, P. G. Wang and S. Yang, *Bioanalysis*, 2021, **13**, 861-864.
4. B. L. Duivelshof, S. Denorme, K. Sandra, X. Liu, A. Beck, M. A. Lauber, D. Guillarme and V. D'Atri, *Pharmaceutics*, 2021, **13**.
5. Y. Mechref, Y. Hu, J. L. Desantos-Garcia, A. Hussein and H. Tang, *Mol. Cell Proteomics*, 2013, **12**, 874-884.
6. K. Xiao, Y. Han, H. Yang, H. Lu and Z. Tian, *Anal. Chim. Acta*, 2019, **1091**, 1-22.
7. Y. L. Chen, L. Ding and H. X. Ju, *Accounts Chem. Res.*, 2018, **51**, 890-899.
8. C. Grünwald-Gruber, A. Thader, D. Maresch, T. Dalik and F. Altmann, *Anal. Bioanal. Chem.*, 2017, **409**, 2519-2530.
9. L. Han and C. E. Costello, *Biochemistry*, 2013, **78**, 710-720.
10. L. R. Ruhaak, G. Zauner, C. Huhn, C. Bruggink, A. M. Deelder and M. Wuhrer, *Anal. Bioanal Chem.*, 2010, **397**, 3457-3481.
11. H. M. Kayili, *Int. J. Mass Spectrom.*, 2020, **457**, 116412.
12. C. T. Yuen, Y. Zhou, Q. Z. Wang, J. F. Hou, A. Bristow and J. Z. Wang, *Biologicals*, 2011, **39**, 396-403.
13. B. Halfinger, B. Sarg and H. H. Lindner, *Electrophoresis*, 2011, **32**, 3546-3553.
14. W. Morelle and J.-C. Michalski, *Nat. Protoc.*, 2007, **2**, 1585-1602.
15. S. Zhou, K. M. Wooding and Y. Mechref, *Methods Mol. Biol.*, 2017, **1503**, 83-96.
16. S. Klapoetke, J. Zhang, S. Becht, X. Gu and X. Ding, *J. Pharm. Biomed. Anal.*, 2010, **53**, 315-324.
17. W. Morelle, V. Faid, F. Chirat and J.-C. Michalski, in *Glycomics: Methods and Protocols*, eds. N. H. Packer and N. G. Karlsson, Humana Press, Totowa, NJ, 2009, pp. 3-21.
18. L. R. Ruhaak, E. Steenvoorden, C. A. Koeleman, A. M. Deelder and M. Wuhrer, *Proteomics*, 2010, **10**, 2330-2336.
19. A. Kameyama, S. K. Dissanayake and W. W. Thet Tin, *Plos One*, 2018, **13**, e0196800-e0196800.
20. M. A. Lauber, Y.-Q. Yu, D. W. Brousmiche, Z. Hua, S. M. Koza, P. Magnelli, E. Guthrie, C. H. Taron and K. J. Fountain, *Anal. Chem.*, 2015, **87**, 5401-5409.
21. Z. Segu, T. Stone, C. Berdugo, A. Roberts, E. Doud and Y. Li, *MAbs*, 2020, **12**, 1750794.
22. C. Y. Liew, C.-C. Yen, J.-L. Chen, S.-T. Tsai, S. Pawar, C.-Y. Wu and C.-K. Ni, *Commun. Chem.*, 2021, **4**, 92.
23. J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Anal. Biochem.*, 1995, **230**, 229-238.
24. G. R. Guile, P. M. Rudd, D. R. Wing, S. B. Prime and R. A. Dwek, *Anal. Biochem.*, 1996, **240**, 210-226.
25. C. S. Ho, C. W. K. Lam, M. H. M. Chan, R. C. K. Cheung, L. K. Law, L. C. W. Lit, K. F. Ng, M. W. M. Suen and H. L. Tai, *Clin. Biochem. Rev.*, 2003, **24**, 3-12.
26. D. Reusch and M. L. Tejada, *Glycobiology*, 2015, **25**, 1325-1334.
27. A. Corfield, *Histochem. Cell Biol.*, 2017, **147**, 119-147.
28. Y. Y. Zhao, M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume and N. Taniguchi, *Cancer Sci.*, 2008, 99.
29. C. Reily, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, **15**, 346-366.
30. Y. Zhu, X. Liu, Y. Zhang, Z. Wang, Y. Lasanajak and X. Song, *Bioconjugugate Chem.*, 2018, **29**, 3847-3855.
31. R. P. Kozak, C. B. Tortosa, D. L. Fernandes and D. I. R. Spencer, *Anal. Biochem.*, 2015, **486**, 38-40.

32. T. Keser, T. Pavić, G. Lauc and O. Gornik, *Front. Chem.*, 2018, **6**, 324-324.
33. K. R. Anumula, *Anal. Biochem.*, 1994, **220**, 275-283.
34. M. Pabst, D. Kolarich, G. Pöltl, T. Dalik, G. Lubec, A. Hofinger and F. Altmann, *Anal. Biochem.*, 2009, **384**, 263-273.
35. S. Zhou, L. Veillon, X. Dong, Y. Huang and Y. Mechref, *Analyst*, 2017, **142**, 4446-4455.
36. B. Buszewski and S. Noga, *Anal. Bioanal. Chem.*, 2012, **402**, 231-247.
37. Y. Huang, Y. Nie, B. Boyes and R. Orlando, *J. Biomol. Tech.*, 2016, **27**, 98-104.
38. M. Allen, Doctor of Philosophy, The Open University, 2022.
39. K. Molnarova and P. Kozlík, *Molecules*, 2020, **25**.
40. H. C. Joao, I. G. Scragg and R. A. Dwek, *FEBS Lett.*, 1992, **307**, 343-346.
41. M. Melmer, T. Stangler, M. Schiefermeier, W. Brunner, H. Toll, A. Rupprechter, W. Lindner and A. Premstaller, *Anal. Bioanal. Chem.*, 2010, **398**, 905-914.
42. A. M. Sakwe, R. Koumangoye, S. J. Goodwin and J. Ochieng, *J. Biol. Chem.*, 2010, **285**, 41827-41835.
43. J. L. Desantos-Garcia, S. I. Khalil, A. Hussein, Y. Hu and Y. Mechref, *Electrophoresis*, 2011, **32**, 3516-3525.
44. S. Yang and H. Zhang, *Curr. Protoc. Chem. Biol.*, 2014, **6**, 191-208.
45. M. Novotny, W. Alley and B. Mann, *Glycoconjugate j.*, 2012, **30**.
46. R. A. Evangelista, F.-T. A. Chen and A. Guttman, *J. Chromatogr. A*, 1996, **745**, 273-280.
47. A. B. Moran, R. A. Gardner, M. Wuhrer, G. S. M. Lageveen-Kammeijer and D. I. R. Spencer, *Anal. Chem.*, 2022, **94**, 6639-6648.
48. T. Loo, M. L. Patchett, G. E. Norris and J. S. Lott, *Protein. Expres. Purif.*, 2002, **24**, 90-98.

Chapter 4- Optimisation of Mass Spectrometric analysis of low abundance glycan species using CuAAc

4.1 Introduction

The use of derivatisation agents to improve analytical characteristics has been widely used for both optical and mass spectrometric analysis workflows[1-5]. These compounds aim to impart beneficial analytical characteristics to the target molecules, enabling them to be detected at levels below that of the native underivatized molecule. The study of biological oligosaccharide structure is no exception.

Native glycans exhibit high levels of hydrophilicity and lack any sort of basic region capable of protonation [4]. This presents a significant challenge for their analysis by mass spectrometry (MS). Initial methodologies to improve analysis by MS routes sought to employ 'wet-chemistry' steps such as per-*O*-methylation or per-*O*-acetylation. These derivatisation methods seek to add hydrophobicity to the analyte molecule by the conversion of hydroxyl, amino or carboxylic acid groups to their methyl or methyl ester derivatives [6]. These workflows have been shown to be effective in increasing glycan abundance in MS applications and have been shown to produce repeatable fragmentation patterns when analysed by MS/MS. These methodologies, however, show limitations when applied to larger glycan structures because the method presents challenges in reaction completion. This results in a convoluted spectrum, with target compounds possessing unpredictable modifications across multiple m/z.

In contrast, single site modification by reductive amination or carbamate labelling has been shown to successfully derivatise biologically released carbohydrates without glycan size limitations. In optical glycan analysis, labelling reagents such as 2AB **9** and ProA **8** are routinely used to provide a fluorescent region, by which to distinguish carbohydrates from the baseline. In these applications, glycan profiling is possible by direct comparison to a library of known retention standards [7]. This method has proven accurate and reliable for profiling known glycoforms, but is less accurate for unknown glycosylation products. In addition, these methods struggle to differentiate low abundance glycoforms close to the baseline from baseline noise. This presents a problem in the reliability of profiling of low abundance glycan species.

Glycan structural profiling is therefore routinely performed with online MS. This aims to provide an increased level glycan structure though both intact mass and fragmental analysis, allowing both the size and structure of a glycan to be inferred. In MS applications, reductive amination labelling strategies focus around the use of ProA **8**. This is due to its increased hydrophobicity over smaller more polar compounds, such as 2AB **9** [8]. In recent years, the use of carbamate labelling has seen the development of Rapiflour **7**, a fluorescent MS compatible label containing both a quinoline and a hydrophobic tertiary amine containing tail [9]. These labels derivatise ring closed glycosylamine, resulting from the deglycosylation of protein by the action of PNGaseF. The use of carbamate labelling provides a significantly decreased labelling workflow duration when compared to reductive amination, however the non-selective nature of NHS carbamate hydrolysis means that any amines present in solution are available for derivatisation. This provides a particular challenge in the analysis of samples resulting from protein deglycosylation because samples must undergo extensive purification to reduce the formation of by-products [10].

None the less, derivatisation by **26** and ProA **8** have shown between a 10-30 fold increase in the abundance of derivatised oligosaccharides over native carbohydrates. Harvey *et al* also observed that there was a significant increase in intensity of ProA glycans over glycans derivatised by 2AB **9** [11]. It was speculated that the decreased hydrophilicity of ProA **8** resulted in more efficient ionisation. Research carried out by Walker *et al* [12] sought to investigate the effect of increasing hydrophobicity on the detection of low abundance oligosaccharide species. In this study five hydrazine labels were reacted with a model carbohydrate, before being separated by liquid chromatography and analysed by positive mode ESI MS. This scheme of work was designed to investigate the effect of increasing analyte hydrophobicity on ion abundance of carbohydrates at low concentration. This work demonstrated that while hydrophobic derivatisation improved MS detection very high equivalencies (up to 3500 times) of labelling solution were required to fully derivatise the reactant at low concentrations. In similar studies, the effect of charged hydrazine labelling was also used to investigate the effect of charged labelling upon the ion abundance of a model carbohydrate sample in positive mode ESI+MS. These investigations sought to explore claims that detection limits could be reduced by the incorporation of permanent charged MS labelling strategies. Initial investigations by Naven and Harvey [13] and subsequently by Gil *et al* [14] showed that the use of permanent charged hydrazine labels (Girards T reagent) could be used to improve ionisation efficiency in MALDI and ESI-MS by as much as 10 fold. Subsequent research by Bereman *et al* [15] afforded contradicting results, instead showing that charged derivatives decreased ion abundance in positive mode ESI. Later studies Walker *et al* [16] sought to compare the effect of permanent charge by direct comparison of structurally similar charged and uncharged derivatives. These studies saw an approximately 50%

reduction ion abundance when compared to samples labelled with non charged derivatives. Although it has been previously stated that permanent charge derivatisation resulted in confusing fragmentation patterns with partial or total loss of the label molecule itself [17]. Walker *et al* [16] observed no such modification following hydrazine derivatisation with Gerard's T reagent.

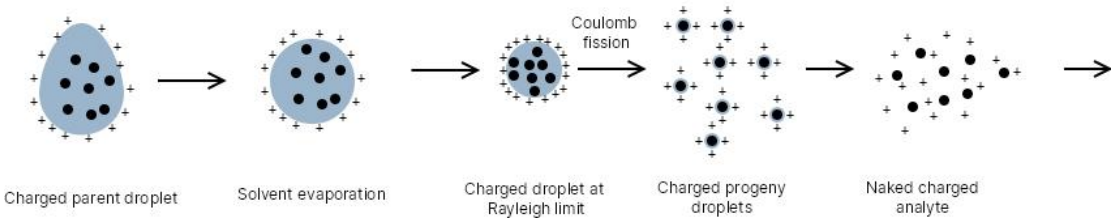


Figure 4.1- A representation of the mechanism behind electrospray ionisation ESI, with progressive desolvation leading to significant increases in surface charge up until fission events.

Electrospray ionisation (ESI) is described as a 'soft' ionisation technique. Similar to MALDI, this is due to charge being passed to the analyte molecules via an intermediate and has been shown to result in less post ionisation degradation (PID) [18]. Unlike MALDI, ESI techniques see sample solutions nebulised by a charged needle, forming charged droplets (**Figure 4.1**). In positive ion mode, these droplets hold a positive charge. Heated drying gas results in desolvation of the droplets and an increase in surface charge. Parent droplets reach their Rayleigh limit resulting in a process of coulomb fission, resulting in the formation of smaller charged droplets [19]. Null *et al* [20] observed that analyte distribution within parent droplets was not homogenous, as previously speculated by Fersht [21]. Instead, more hydrophobic analytes undergo significant enrichment on the outer edges of droplets, while more polar hydrophilic analytes are concentrated towards the centre. This phenomenon is further hampered by the increasing polarity associated with the increased numbers of hydroxyl groups present on glycans of increased molecular weight. It has been speculated that this is due to the higher free energy of solvation present within more hydrophilic analytes, making them less likely to desorb from electrospray progeny droplets into gaseous phase ions[22]. Additionally, the increase in surface charge resulting from coulomb fission events and de-solvation, makes analytes close to the droplet surface inherently more likely to encounter a free proton and be ejected as a gaseous phase ion, than those residing within the centre of the droplet [23].

This results in ESI methods exhibiting a degree of inherent bias towards increased analyte hydrophobicity [24]. This is manifested as an increase in the efficiency of ionisation and increased

abundance of more hydrophobic analytes, when compared to those exhibiting increased hydrophilicity. This bias has been observed in numerous hydrophobic and hydrophilic molecules however its effect in post analysis modification has yet to be studied. In addition the use of permanent charge labels is seldom used within the field of MS analysis due to reduction in ionisation efficiency and challenging fragmentation patterns. Although little work has been carried out the on addition of charged regions in post analysis derivatisation workflows, recent attempts to modify glycan structures with permanent charge labels did not observe any noticeable alterations in fragmentation pattern when fragmentation was carried out by collision induced dissociation (CID)[25].

4.2 Aims

The challenges encountered when detecting glycan species of low abundance by ESI-MS provide a significant barrier to fully understanding the roles of glycan structures. This is due to the hydrophilicity of large glycan structures, resulting in challenges associated with ESI ionisation sources which have been shown to favour more hydrophobic analytes. Initial work carried out by Walker *et al* [12] sought to use hydrazine labelling to investigate both the effect of hydrophobicity and permanent charge upon the ionisation and ion abundance of model carbohydrates. These investigations, while successful, required high molar equivalencies of labelling compound in order to achieve reliable and efficient derivatisation of the target analyte. Subsequent work has sought to develop more rapid labelling strategies by the use of carbamate labelling. While these methods see carbohydrate derivatization achieved in as little as 15 minutes, deglycosylated samples must undergo pre-labelling preparation in order to reduce the impact of amine containing protein fragments present in the derivatisation mixture.

The previous chapters have shown the synthesis and optimisation of a reductive amination labelling strategy containing both a primary amine for reductive amination, as well as an alkyne containing functional handle. This region remains stable and unreacted throughout carbohydrate labelling and separation by HILIC analysis, allowing for post column derivatisation by copper catalysed azide alkyne cycloaddition (CuAAC).

The work in this chapter therefore set out to fulfil three main aims. Initially the investigation sought to isolate oligosaccharide samples at low abundance derivatised with **26**. This was carried out by a series of labelling reactions on maltoheptose with **26** in line with the workflow optimised in chapter 3, followed by the collection of sample peaks from concurrent injections. The samples were then used to optimise selective reaction conditions for CuAAC on a nanomolar scale, with a commercially available benzyl azide. These reactions were carried out at both 30 ^{o}C and 60 ^{o}C over a three hour

period. Finally, the investigation moved to establish a small library of azides in order to investigate the effect of both permanent charge and additional hydrophobicity on the efficiency of ionisation by ESI-MS.

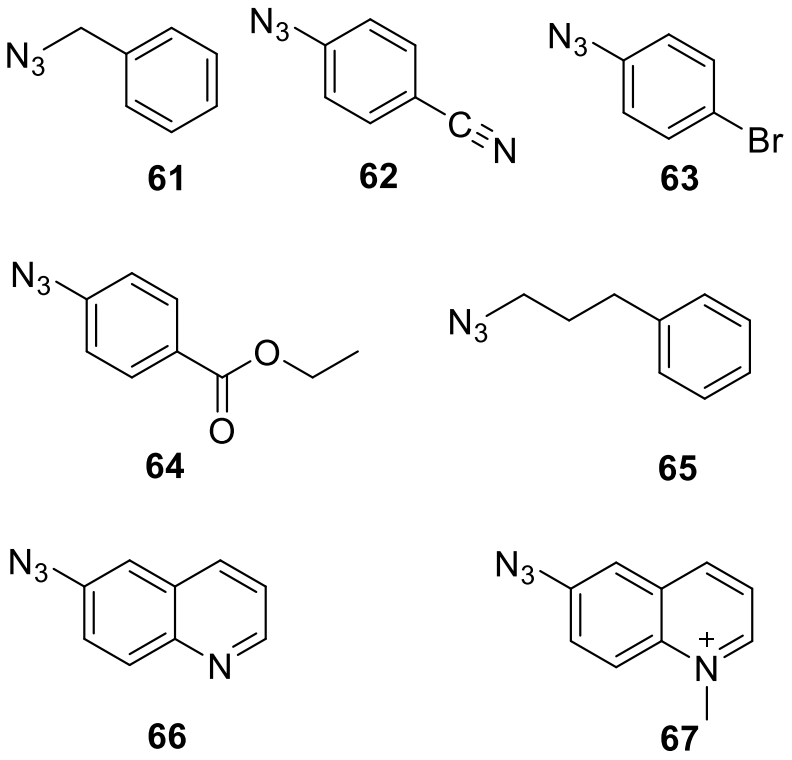


Figure 4.2 - Library of seven neutral and charged azides exhibiting a range of non polar surface areas

Herein, a library of seven azide containing compounds was established, shown in **Figure 4.2**. These were separated into two groups. Initially, five benzyl azide derivatives (two synthesized in house and three available through a collaboration[26]) were selected with the purpose of determining the effect of increasing hydrophobicity upon the ion abundance following the post column derivatisation of a low abundance model carbohydrate following analysis by positive mode ESI-MS. The synthesis of quinoline derivative **66** provided an increased level of hydrophobicity while the synthesis of **67** enabled the comparison of charged and uncharged derivatives in compounds bearing high levels of structural similarity.

4.3 Results and discussion

4.3.1 Derivatization of low abundance carbohydrates

In the previous chapter, the ability of multifunctional label **26** to both derivatise and enhance the detection of derivatised oligosaccharide samples was demonstrated though both investigation of

limit of detection and labelling reactions, on both dextran homopolymers and enzymatically released glycans. Although multifunctional label **26** provided similar detection sensitivity to ProA **8** , the terminal alkyne present on **26** provides opportunities for post column derivatisation by CuAAC. This chapter aims to investigate the effect of post column derivatisation on the ion abundance of multifunctionally labelled carbohydrate. In order to facilitate this, a series of labelling reactions was carried out on a model carbohydrate. Maltoheptose was chosen as a GU7 carbohydrate as it closely reflects the molecular weight of a biologically cleaved *N*-glycan species, such as Man-5. The scale of labelling was designed to reflect quantities of low abundance glycans routinely recovered from biological sources [27]. This typically ranges between 170 pMol and 10 nMol, above that of the LOD, but too small to undergo accurate profiling by fluorescence analysis, and challenging to pinpoint using MS.

Samples of maltoheptose (500 pMol) were transferred to sample tubes and concentrated by vacuum centrifuge. To this, borate buffered methanol was added containing sodium cyanoborohydride as a reducing agent as well as **26**. These samples were then heated at 65 ^{o}C for one hour, in line with the workflow optimised in previous sections. Following incubation, reactions were quenched with ACN and purified by amide HILIC SPE (Applied separations Spe-ed amide 2). The purified oligosaccharides were then concentrated by vacuum centrifuge and resuspended in 100 µl for analysis by HILIC mode HPLC-FLD.

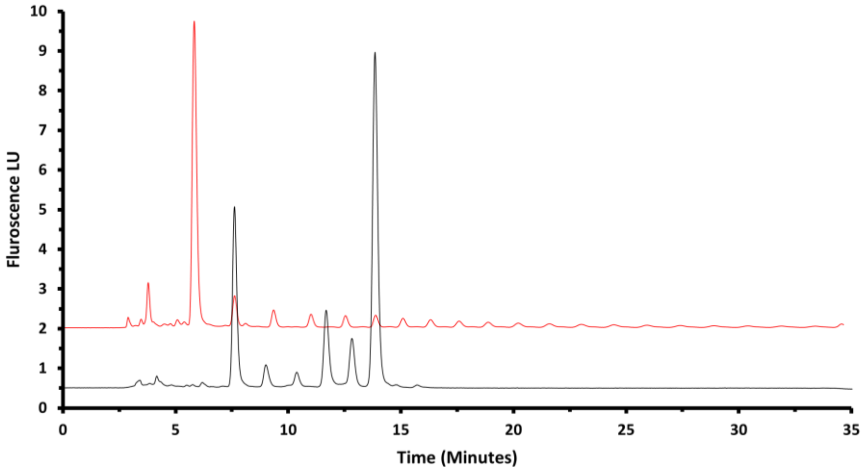


Figure 4.3 - Chromatographic traces for the separation of **26** labelled GU7 maltoheptose (black) compared to a **26** labelled dextran hydroxylate retention standard (red).

The samples were then analysed by HILIC-FLD (**Figure 4.3**) following a 5 µl injection, providing a chromatogram containing five distinct peaks. A homopolymer retention standard was then used to profile the peaks as maltopentose, maltohexose and maltoheptose. The GU7 maltoheptose peak was clearly resolved from interfering impurity peaks and was then collected on an Agilent G1364A

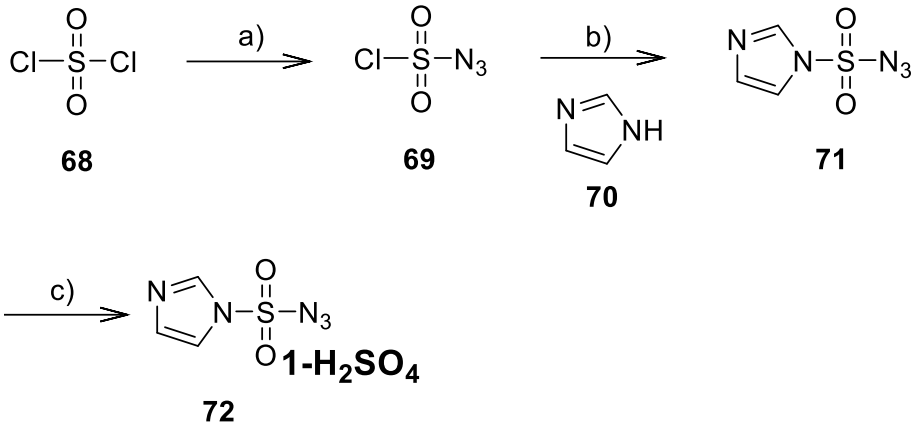
fraction collector. Five concurrent injections were pooled and concentrated to dryness, forming a sample of ~125 pmol of the initial sample for post analysis derivatisation by CuAAc.

4.3.2 Synthesis of hydrophobic and charged azides

In order to generate a library of azide partners to explore the effect of hydrophobicity on the ESI-MS response, a series of diazo transfer reactions were attempted. In previous work, this was carried out by conversion of a primary amine with trifluoromethanesulfonyl azide yielding an azido modified product. Although effective, this reagent is both highly toxic and explosive. In order to overcome this, Goddard and Stick [28] sought to develop an imidazole based alternative. Imidazole-1-sulfonyl azide hydrogen chloride **73** (known as the Stick reagent) provides a more stable alternative to achieving diazotransfer[28]. Further optimisation by Potter *et al* [29] modified this synthetic route developing a method for the synthesis of **72** which omitted the formation of any explosive intermediates. Both the methods of Goddard and Potter were used to synthesise hydrogen chloride and hydrogen sulfate salts of Imidazole-1-sulfonyl azide **71**.

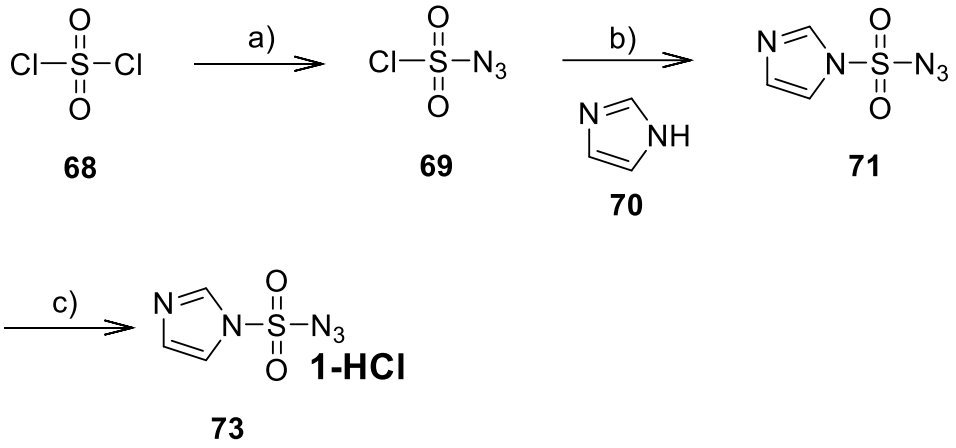
4.3.2.1 Synthesis of Imidazole-1-sulfonyl azide

The synthesis of Imidazole-1-sulfonyl azide hydrogen sulfate **72** was successfully carried out in an 57% yield following the three stage synthesis (**Scheme 4.1**) developed by Potter *et al* [29]. Initial test reactions carried out to generate azido product **72** showed slow reaction times resulting in poor product formation when monitored by TLC. It was hypothesised that this may be due to the increased stability of the hydrogen sulfate derivative.



Scheme 4.1 -Synthetic pathway for the generation of **72** a) NaN_3, ACN, 0 °C b) Imidazole, ACN 0 °C c) **71**, H_2SO_4 0 °C, 57% yield.

Instead the synthesis of Imidazole-1-sulfonyl azide **73** hydrogen chloride was attempted in line with **Scheme 4.2**.

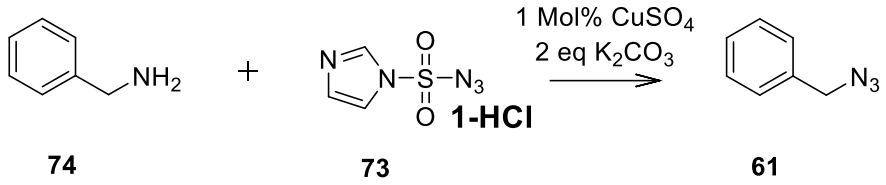


Scheme 4.2 - Synthetic pathway for the generation of **61** a) NaN_3, ACN, 0 °C b) Imidazole, ACN 0 °C c) **71**, HCl, EtOAc 0 °C, 55% yield.

Imidazole-1-sulfonyl azide hydrogen chloride **73** was synthesised over 2 stages as outlined in **Scheme 4.2**. This was initially carried out by dropwise addition of **68** to a cooled suspension of sodium azide in ACN. This suspension was stirred overnight and gradually allowed to warm to room temperature. The formation of **69** was monitored by TLC and visualised by azide staining with PPh_3 and ninhydrin. The reaction mixture was then re-cooled to 0 °C and imidazole **71** was added. The resulting slurry was stirred for a further three hours before being diluted and extracted in ethyl acetate. The solution containing **71** was then precipitated by the addition of HCl in ethanol to yield a white solid. The precipitate was washed and dried to yield **73** in 55% yield following characterisation by ^{1}H NMR spectroscopy and the loss of a singlet integrating to 1 proton indicating formation of a tertiary amine

and the addition of **69**. Diazo-transfer reagent **73** was then used for the synthesis of **61**, **65**, **66** and **67**. This process sees the conversion of the primary amine present on **74**, **75**, and **76** converted to an azide, while also producing a primary amine form of **73**.

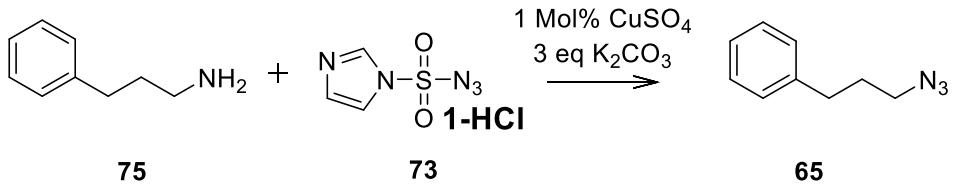
4.3.2.2 Synthesis of Benzyl azide 61



Scheme 4.3 -Diazotransfer reaction conditions for the generation of Benzyl azide **61** in a 58% yield.

The synthesis of **61** outlined in **Scheme 4.3**, was carried out in line with the general procedure set out by Potter et.al[29]. In this reaction, the benzyl amine **74** was stirred with an aqueous solution of 1 mol% $CuSO_4$ and 2 eq K_2CO_3 before **73** was added with MeOH. The solution was then stirred for 6 hours until reaction completion which was determined by TLC and the disappearance of a spot at Rf values comparable to **74**. The reaction mixture was then passed through a sintered funnel to remove any undissolved K_2CO_3 and concentrated by rotary evaporation. Following purification by flash column chromatography, fractions were dried *in vacuo* to yield **61** in 58% yield following characterisation by ^{1}H NMR spectroscopy and HRMS.

4.3.2.3 Synthesis of Phenylpropan-1-azide 65

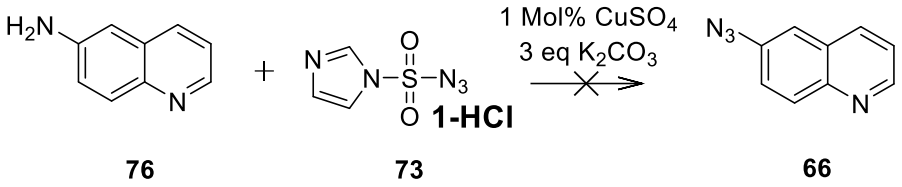


Scheme 4.4 - Diazotransfer reaction conditions for the generation of phenylpropan-1-azide **65** in a 76% yield

The synthesis route (**Scheme 4.4**) for **61** developed by Potter *et al* [29], was also applied to the production of **65**. However the hydrophobicity exhibited by **75** resulted in partitioning occurring

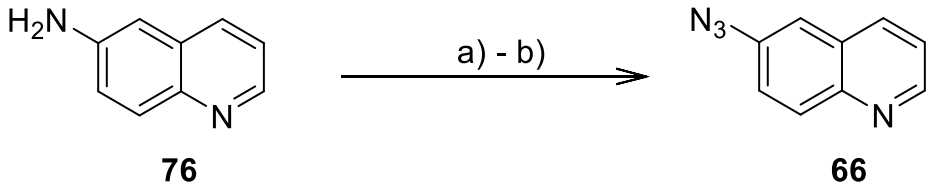
between **75** and the aqueous reaction mixture. Therefore the amine precursor, phenylpropan-1-amine **75** was instead stirred in a 5% aqueous MeOH solution containing 1 mol% $CuSO_4$ and 3 eq K_2CO_3, before 2 eq of **73** was added in 5% aqueous MeOH. The solution was stirred overnight. Completion was monitored by TLC, staining with PPh_3 and ninhydrin to visualise azide product. The reaction mixture was filtered and purified by pTLC, before the isolated product was washed from silica and concentrated *in vacuo* to yield **65** in 76% yield following characterisation by a 26 Da mass shift in HRMS .

4.3.2.4 Synthesis of 6-Azido Quinoline 66



Scheme 4.5 - Attempted synthesis route for the generation of **66** by diazotransfer to amino quinoline **76** with **73**.

Diazotransfer to **76** was attempted by the use of **73**, however after prolonged stirring, (>48 hours) isolation of the product by TLC demonstrated that the product was not afforded at acceptable yields, resulting in the requirement for an alternative synthesis route (**Scheme 4.5**).

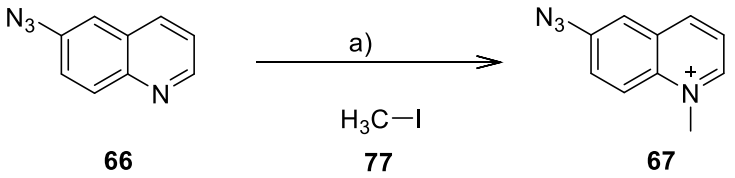


Scheme 4.6 - Alternative diazotransfer to amino quinoline **76** a) $NaNO_3$ 1.5 eq 10% HCl 1h b) NaN_3 2.1 eq 10% yield.

Following the unsuccessful conversion of **76** to **66**, the synthesis route designed by Khan *et al*[30] was adopted (**Scheme 4.6**). In this synthesis, 6-amino quinoline **66** was stirred with sodium nitrate (1.5 eq) in a 10% solution of HCl. Following this, an acidic solution of sodium azide was added and the reaction mixture was stirred for 6 hours. After such time, the reaction's completion was deemed to

be complete by TLC and staining with PPh_3. The reaction mixture was concentrated at reduced pressure before being purified on silica by flash column chromatography. Product **66** was obtained as an orange solid in 60% yield following characterisation by ^1H NMR spectroscopy and HRMS.

4.3.2.5 Synthesis of 6-Azido Quinolinium 67



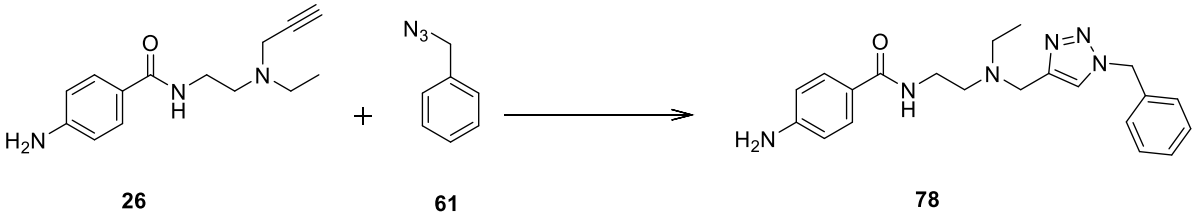
Scheme 4.7 - Quaternary amine generation by the process of methylation a) CH_3I **77**, Dioxane, Reflux

Introduction of a permanent charge was carried out by transformation of the tertiary amine present in **66** to a quaternary amine (**Scheme 4.7**). This was carried out utilising a method developed by Ghandi *et al*[31]. This synthesis saw the azido product formed according to **Scheme 4.7** whereby **66** was heated at reflux in dioxane for one hour with iodomethane. Following purification by flash column chromatography fractions were dried *in vacuo* to yield **67** in 47% yield following characterisation by ^1H NMR spectroscopy and HRMS.

4.3.3 Development of CuAAC conditions for downstream modification of low abundance carbohydrate species.

4.3.3.1 Screening of carbohydrate compatible CuAAc conditions

Following the formation of a small library of azides, the optimum conditions for their conjugation with **26** were investigated. The use of CuAAc has been extensively reported throughout the literature using a wide range of conditions, including for carbohydrate substrates. However, the use of carbohydrates at an analytical scale has not been widely reported, and none of the methods employed a multifunctional labelling reagent.



Scheme 4.8 – CuAAC reaction for the of triazole product **78** from multifunctional label **26** and azide **61** in the presence of $CuSO_4$ catalyst and sodium ascorbate.

The conditions outlined in **Table 4.1** were first applied to a series of test reactions between **26** and **61**, with the aim to produce the triazole product **78** shown in **Scheme 4.8**.

Table 4.1- Reaction conditions screened for the CuAAC conjugation of **26** and **61** over a 5 hour reaction time

Condition	Solvent system	Catalyst	Reducing agent	Temp (°C)	Time (Hours)	pH modifier	Yield	Reference
1	DMSO	Cu(I)Br	-	RT	5	TEA	<5%	[32]
2	DMSO	Cu(I)Br	-	60	5	TEA	<5%	[33]
3	DMF	Cu(I)Br	-	30	5	-	40%	[34].
4	DMF:H_2O (1:1)	$CuSO_4$	Sodium Ascorbate	RT	5	-	98%	[35]
5	THF	CuI	-	RT	5	TBAF	62%	[36]
6	THF:H_2O (1:1)	$CuSO_4$	Sodium Ascorbate	RT	5	-	8%	[37]
7	THF:i-PrOH:H_2O (3:1:1)	$CuSO_4$	Sodium Ascorbate	90°C	5	-	99%	[38]
8	tBuOH:H_2O (1:3)	$CuSO_4$	Sodium Ascorbate	60 °C	5	-	99%	[39]
9	tBuOH:H_2O (1:1)	$CuSO_4$	Sodium Ascorbate	RT	5	-	99%	[37]
10	H_2O	$CuSO_4$	Sodium Ascorbate	RT	5	-	56%	[40]
11	ACN:H_2O (2:1)	$CuSO_4$	Sodium Ascorbate	RT	5	-	98%	[41]
12	MeOH:H_2O (1:1)	$CuSO_4$	Sodium Ascorbate	RT	5	-	99%	[32]

Each of the conditions above were initially trialled as reaction conditions for CuAAC at a 0.1 mg substrate mass, utilising 10 equivalents of azide **61**. Following a 5 hour reaction time, the reactions were then dried down under reduced pressure and reconstituted in 1 ml of H_2O prior to analysis. The samples were then analysed by RP-HPLC over a 30 minute gradient between 0.1% aqueous formic acid and acetonitrile acidified with 0.1% formic acid.

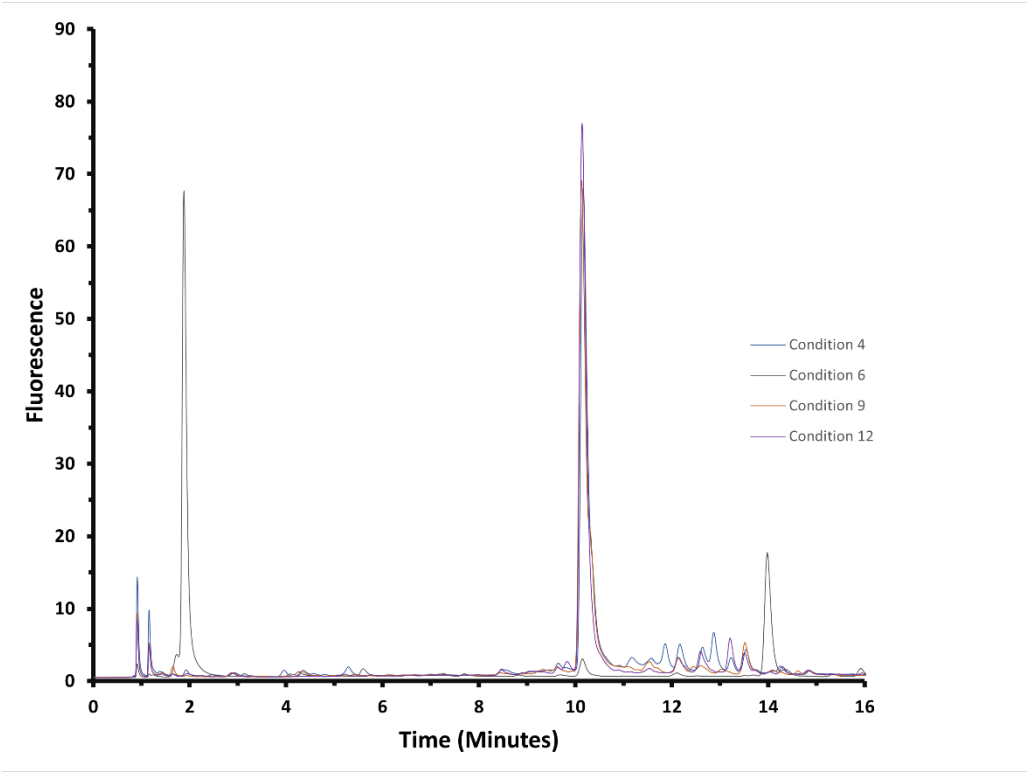


Figure 4.4 - Overlaid chromatographic traces for CuAAC reaction screening samples following incubation in reaction conditions 4, 6, 9 and 12 outlined in **Table 4.1**

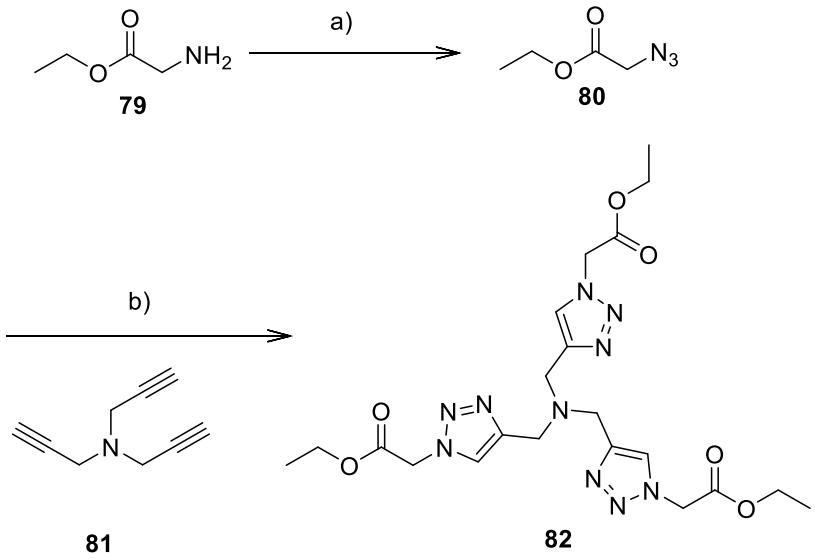
Fluorescence traces for conditions 4, 6 , 9 and 12 are provided in **Figure 4.4** above. The traces show unreacted **26** forms a peak at 1.9 minutes while product formation results in the formation of a new peak at 10.1 minutes which is likely the formation of triazole product **78** later confirmed by MS. In these screening reactions, completion was quantified by the reduction in intensity of the 1.9 minute peak identified as unreacted **26** against the new 10.1 minute product peak. These yields were then collated and are outlined in **Table 4.1** above. Of the twelve conditions analysed in this trial, six provided yields of greater then 70%. Conditions 4, 9 and 12 were then selected for further optimisation. The remaining successful conditions (7, 8, 11) were discounted due to increased reaction temperatures or carbohydrate incompatible solvent combinations.

The subsequent six less successful reaction conditions provided reduced yields of triazole product. This is likely due to incompatibility in the solubility of **26** and **61** in the conditions. It was also observed that in conditions with a reduced proportion of aqueous solvent, that product yield were negatively effected. This is likely due to the reduced solvent polarity limiting solvents ability to dissolve $CuSO_4$. This has been previously linked to a reduction in catalyst efficiency, requiring a phase modifier, such as TEA to ensure sufficient Cu (I) is produced (I) solution[42]. It was also noted that

reactions involving native Cu(I) catalysts such as CuBr and CuI exhibit Cu(I) instability and are commonly stabilised by the use of an amine containing ligand. This trend was also observed by Berg *et al* [43] in Cu (II) catalytic systems and it was also noted that while aqueous conditions improved solubility of copper catalysts, reactions processed at a sluggish pace resulting in increased reaction times. Sharpless and Finn [44, 45] sought to optimise catalyst efficiency by the use of a triazole containing ligand. These were then later applied by Kramer *et al* [46] in the generation of glycoconjugates at high yield.

4.3.3.2 Synthesis of Cu(I) stabilising ligand

The generation of product **78** in the previous section, although successful, required in excess of 5 hours to reach full completion. Due to the requirement for high yielding and efficient generation of triazole products, the synthesis of a triazole containing Cu stabilising ligand was undertaken. This synthesis was carried out in line with the work of Kramer *et al*[46] via a two stage process.



Scheme 4.9 - Synthesis of Triazole containing ligand **82** though a two stage synthetic route, a) 79, DMF NaN_3, b) **81**, CuI, 2,6 lutidine, ACN.

This began with the modification of ethyl chloroacetate **79** to form ethyl azidoacetate **80** (**Scheme 4.9**), this was carried out in a solution of DMF with sodium azide on a 35 mM reaction scale. The reaction mixture was stirred at room temperature overnight and reaction progress was monitored by TLC analysis visualised with PPh_3 and ninhydrin stain. Following removal of the reactant spot the

reaction mixture was extracted with diethyether and concentrated *in vacuo*. This yielded a colourless solution which was taken on to the next stage of the reaction without further characterisation or purification.

The solution of ethyl azidoacetate **80** in diethyl ether was added to a solution of tripropargyl amine **81** and placed under a nitrogen atmosphere (**Scheme 4.9**). To this solution copper iodide and 2,6 lutidine were added and the mixture was stirred at room temperature overnight. Following characterisation of product purity by ^{1}H NMR spectroscopy, product **82** was afforded and identified by the formation of a three proton singlet at 8.02 ppm indicating the formation of three triazole ring structures, as well as a 9 proton triplet at 1.3 ppm indicating the successful addition of **80**. This confirmed the formation of ligand **82** at a 70% yield.

4.3.3.3 Optimisation of aqueous click conditions

Following the synthesis of a Cu(I) stabilising ligand, conditions 7, 8 and 11 were repeated with the addition of triazole ligand **82**. This sought to investigate the effect of increased Cu(I) stability on the yield of triazole product formed between **26** and **61**. These reactions were carried out on a 400 nmol scale and sought to optimise the conditions outlined in the previous section. In addition, the work of Sharpless *et al* [45] suggested that the inclusion of a ligand facilitated the use of increased equivalencies of catalyst to aid in the reduction reaction time and improve product yield without increased risk of oxidative damage to reactants. The reaction conditions outlined in **Table 4.2** were applied to samples containing **26** and azide **61** before incubation for 5 hours at 30 ^{o}C. Following the incubation, the samples were dried *in vacuo* by centrifugal concentration (SpeedVac) and reconstituted in 1 ml of 50% aqueous methanol for analysis.

Table 4.2 - Outlines reaction conditions for the optimisation of triazole product formation with the addition of triazole ligand **82**.

Condition	**26** eq	Solvent	**61** eq	$CuSO_4$ eq	Na Ascorbate eq	**82** eq	Reaction time (Hours)
1	1	tBuOH: H_2O (1:1)	10	5	2.5	1	5
2	1	DMF:H_2O (1:1)	10	5	2.5	1	5
3	1	MeOH: H_2O (1:1)	10	5	2.5	1	5
4	1	MeOH: H_2O (1:1)	10	15	10	10	5
5	1	MeOH: H_2O (1:1)	10	15	25	25	5
6	1	MeOH: H_2O (1:1)	10	15	50	50	5

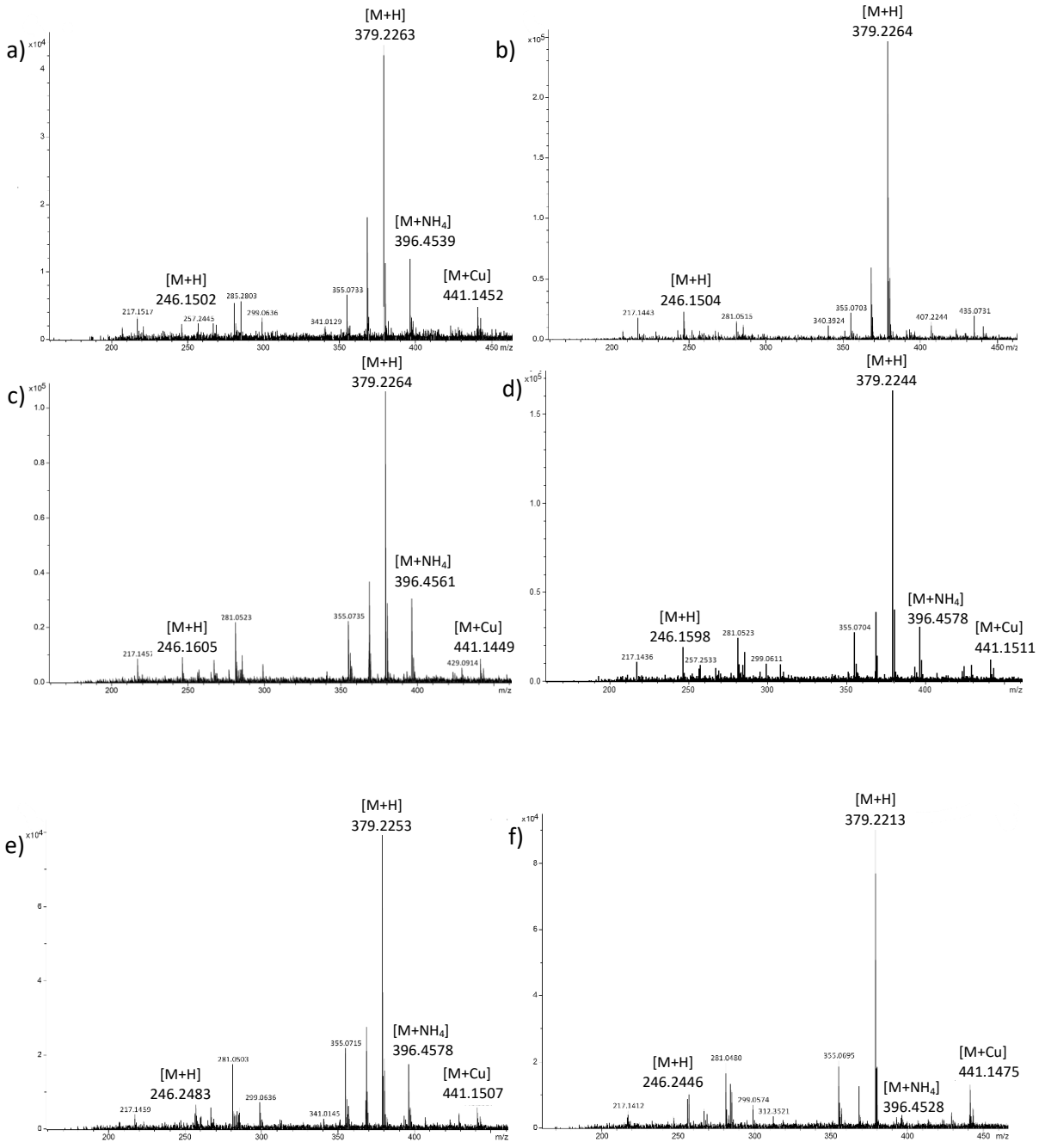


Figure 4.5 - Shows MS spectra generated from the direct infusion of samples extracted in conditions 1-6, a) 1 b)2 c)3 d)4 e)5 f)6. **26** expected [M+H] 246.1528, triazole product [M+H] 379.2168.

The samples were then analysed by direct infusion MS at a rate of 3 µl per minute over a period of 5 minutes. During this time, fragmentation of both reactant product ion peaks were acquired and spectra for the resulting data is provided in **Figure 4.5**. It was anticipated that all spectra would primarily reveal [M+Na] adducts due to the increased concentration of sodium in the reaction mixture. However, in these infusions only [M+H], [M+NH_4] and [M+Cu] adducts were isolated.

Analysis of both aqueous tBuOH and DMF samples shown in **Figure 4.5** A and B demonstrated a mixture of both unreacted **26** and product **78** were present, while the aqueous methanolic conditions formed triazole product in higher yield and thus, unreacted **26** was not detected. In addition to sluggish reaction rates exhibited by condition 2, the conditions relied upon a high proportion of DMF, which has been linked to incompatibility with MS applications. These samples therefore required a lengthy process of concentration due to its high boiling point of 153 ^{o}C at ambient pressure. Additionally, due to the high hydrophilicity of the carbohydrates and the relative hydrophobicity of tBuOH, optimisation of these conditions were not pursued. Instead the investigation sought to optimise aqueous methanolic conditions due to its compatibility with both analytical methods and solubility of highly polar carbohydrates.

4.3.4 Optimisation of aqueous click conditions for the downstream modification of multifunctionally labelled carbohydrates at low abundance.

The successful generation of **78** by aqueous methanol conditions at a 0.1 mg scale, was proof of concept that multifunctional label **26** was capable of undergoing successful CuAAC conjugation with azide containing modifiers such as **61**. The work moved to optimise these conditions at a scale more applicable to that of post column collected carbohydrate samples. These samples where composed of post column collected model carbohydrate, at a scale of 125 pMol. Collected maltoheptose samples were mixed with ligand solution and sodium ascorbate before being dried under reduced pressure. These samples where then resuspended in reaction solvent containing azide and $CuSO_4$. Samples were then incubated at 35 ^{o}C for between one and three hours (**Table 4.3**). In order to analyse the effect of increased azide concentrations on the rate of completion, samples were extracted at both 5 and 10 molar equivalents. Moreover due to the range of hydrophobicity between azide structures, the experiment was repeated for both azide **61** and **66**.

Table 4.3 - Outlines reaction conditions for the modification of **26** labelled GU7 carbohydrate with both azides **61** and **66**.

Sample Number	Time (Hours)	Azide	Catalyst/ ligand	Reducing agent	Solvent composition	Excess azide	Conversion (%)
1	1	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	56
2	1	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	31
3	1	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	63
4	1	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	40
5	2	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	76
6	2	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	60
7	2	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	83
8	2	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	76
9	3	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	85
10	3	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	5	77
11	3	**61**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	91
12	3	**66**	$CuSO_4$(25mol%)/ **82** (50eq)	Sodium Ascorbate	50% MeOH (aq)	10	80

Following analysis by LCMS, reaction efficiency was calculated based upon the peak area of triazole product versus total peak area of product and unreacted carbohydrate. This allowed for comparison between both azide structure and concentration. In contrast to previous direct infusion samples which formed predominantly [M+H] adducts. Analyses by LCMS revealed a mixture of multiply charged [M+Na] and [M+K] adducts. While [M+H] was observed in these analysis it provided the lowest abundance when compared to both sodium and potassium adducts.

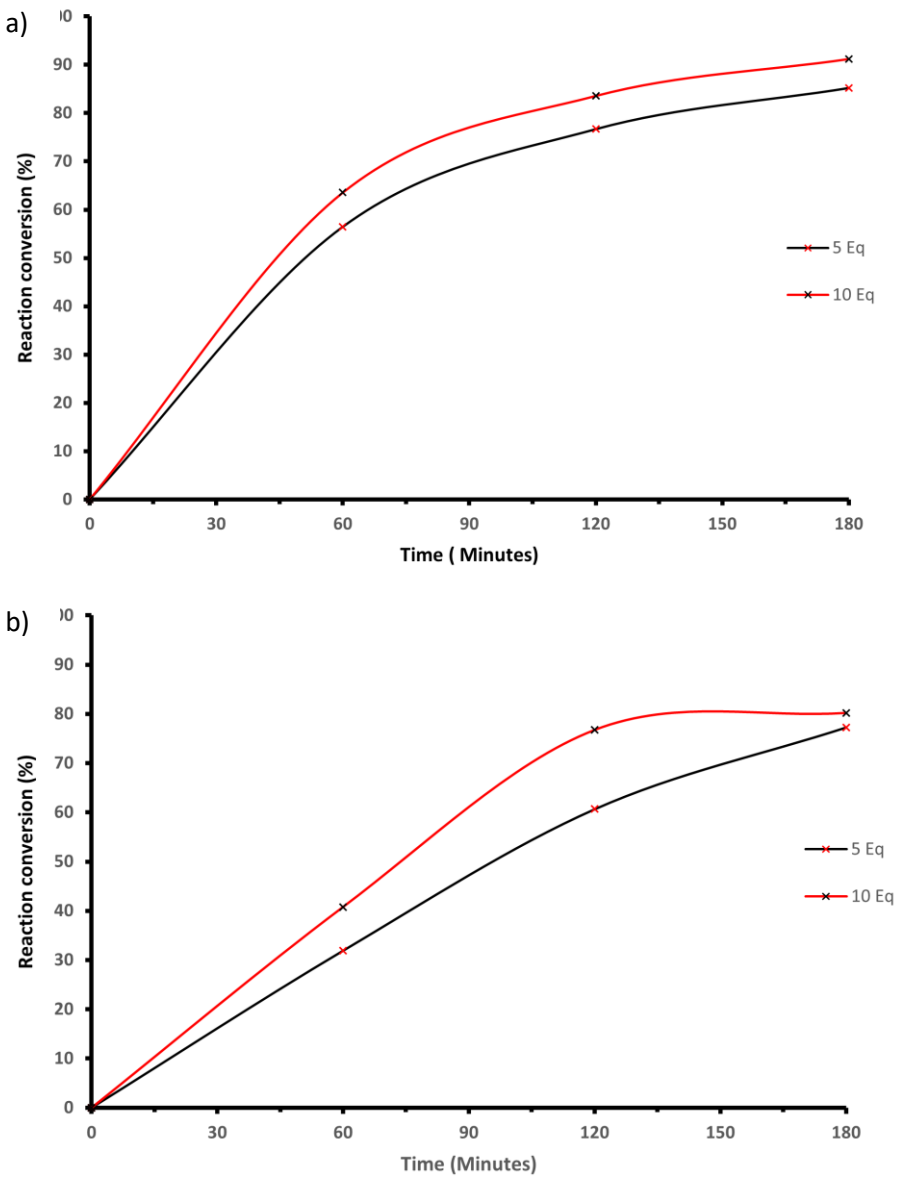


Figure 4.6 Plot of reaction efficiency over time for CuAAC reactions between **26** labelled GU7 carbohydrate and azides a)**61** and b)**66** at both 5 and 10 Molar equivalencies

Figure 4.6 A and B show the increase in relative triazole product concentration over increasing time for both samples at 5 and 10 equivalents. In these samples product yield was judged between peak intensity of unreacted Gu7-**26** and the formation of triazole product. In **Figure 4.6**, maltoheptose labelled with **26** was reacted with azide **61** over a period of three hours. Although the trend showed that product yield increases with time up until three hours, those samples enriched with 10 molar equivalents exhibited a higher reaction efficiency during the initial 2 hours of the reaction. This trend was mirrored in **Figure 4.6**b for the samples reacted with the more hydrophobic azide **66**. Interestingly, while both reaction conditions fell short of >95% product yield, both demonstrated

reduced reaction efficiency between 2-3 hours. This indicated the requirement for an increased proportion of azide in subsequent samples.

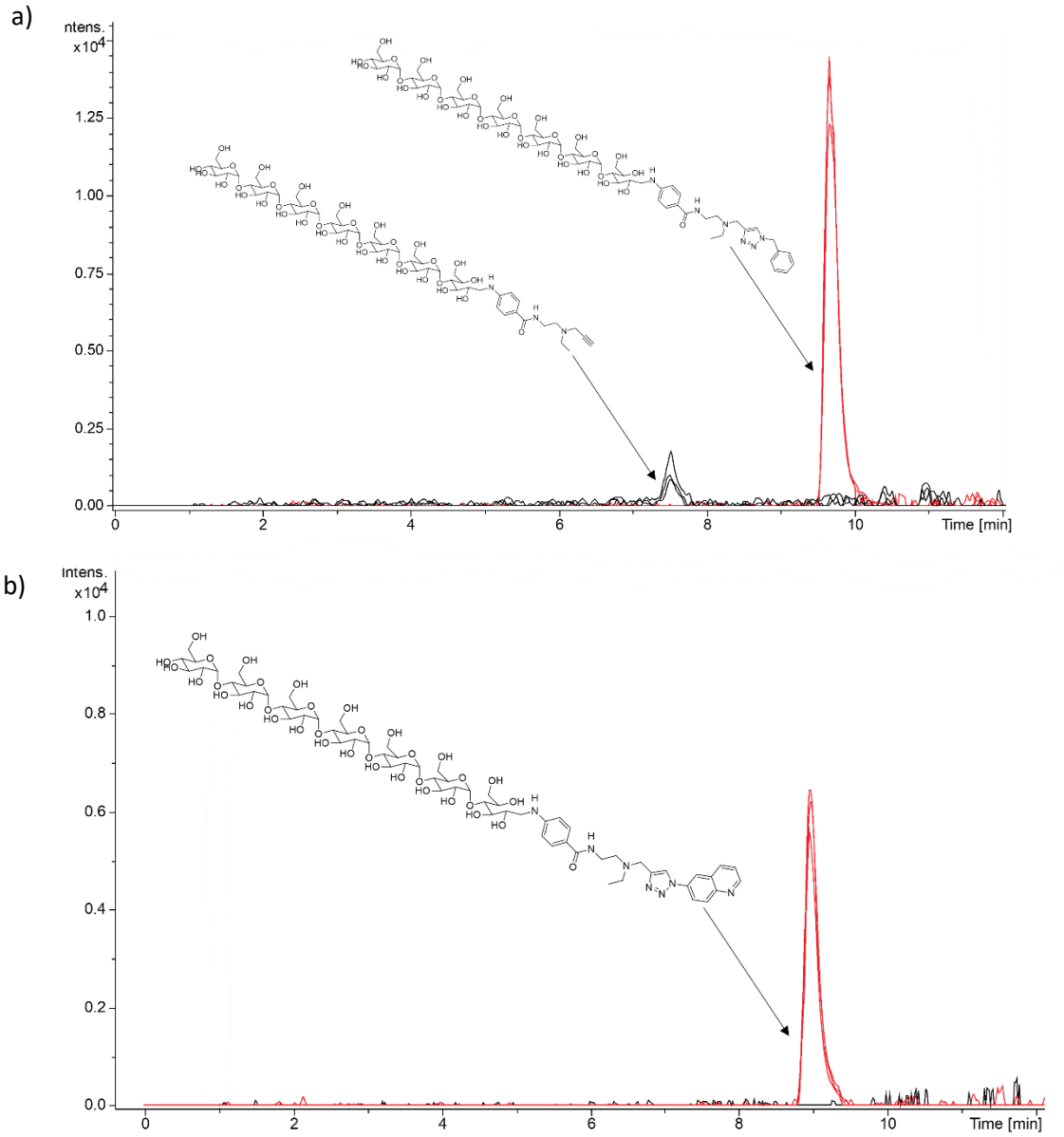


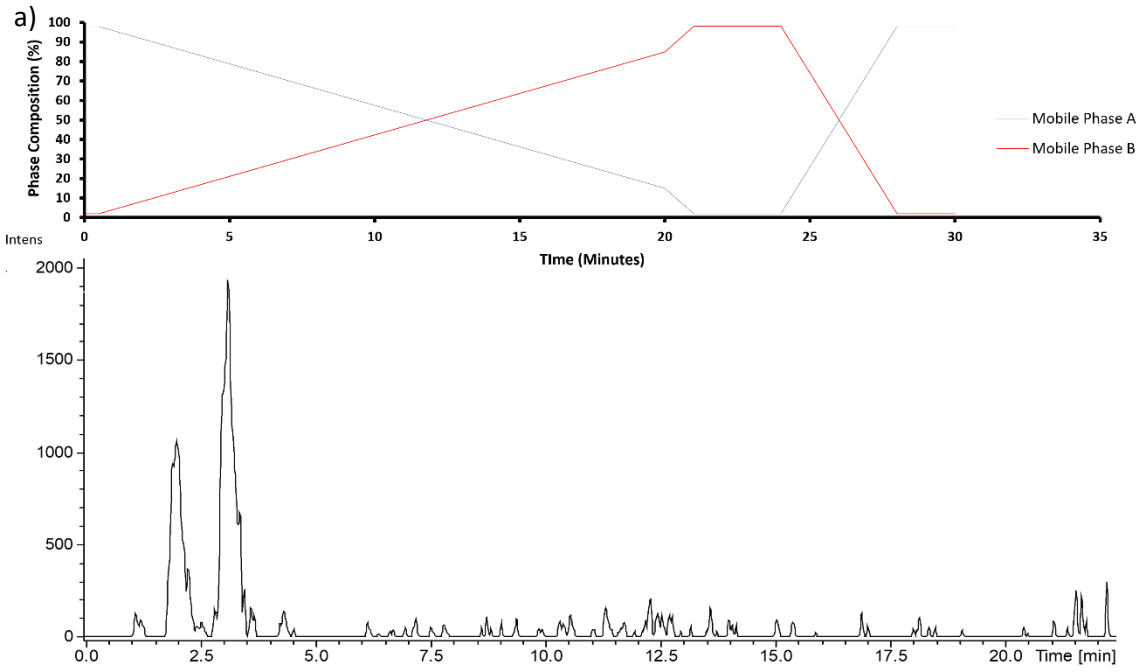
Figure 4.7 - Overlaid EIC for both triazole products (red) and unreacted GU7-26 (black) following 3 hour incubation. A) triazole products of **61** and B) triazole products of **66**.

Having established that triazole product formation could be carried out at a 91% and 80% yield for both azides **61** and **66**, the conditions were further optimised by the addition of a further 10 molar equivalents of azide to the reaction mixture. These samples were extracted in triplicate and analysed by LC-Q-ToF following separation on a Phenomenex aqua C18 5µ column. Following analysis, the reaction yield for both azides **61** and **66** were calculated at 95% and 95% respectively.

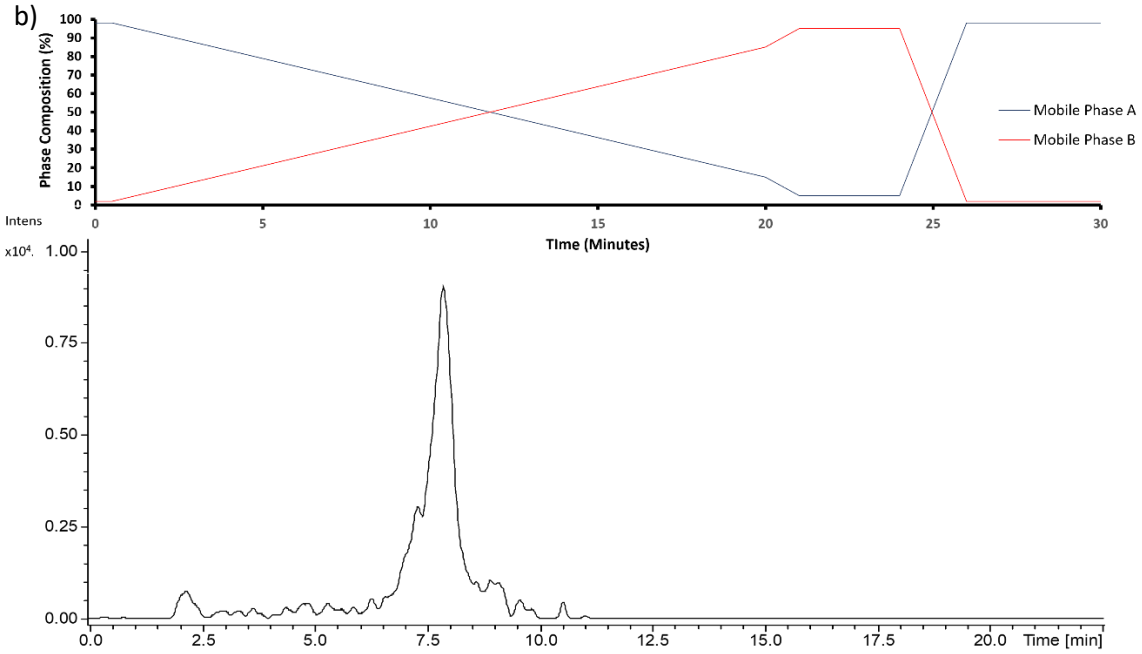
This demonstrated that carbohydrate with **26** could be modified by CuAAc at a picomolar reaction scale. Moreover, azide **61** and **66** represent neutral click conjugation partners, exhibiting both extremes of hydrophobicity. This indicates that the optimised conditions display adequate solvent polarity to dissolve both reductively aminated **26** labelled carbohydrate and the azide conjugation partner. Although these conditions showed significant improvement over the conditions trialled in **Table 4.3**, the chromatographic traces in **Figure 4.7** show that while peak symmetry of the triazole product was adequate, unreacted modified carbohydrate was poorly retained at low concentration. This manifested as sustained period of elution at low abundance. This significantly hindered the calculation of reaction efficiency and led to a process of optimisation for carbohydrate retention on C18.

4.3.5 Optimisation of model carbohydrate separation by reversed phase liquid chromatography.

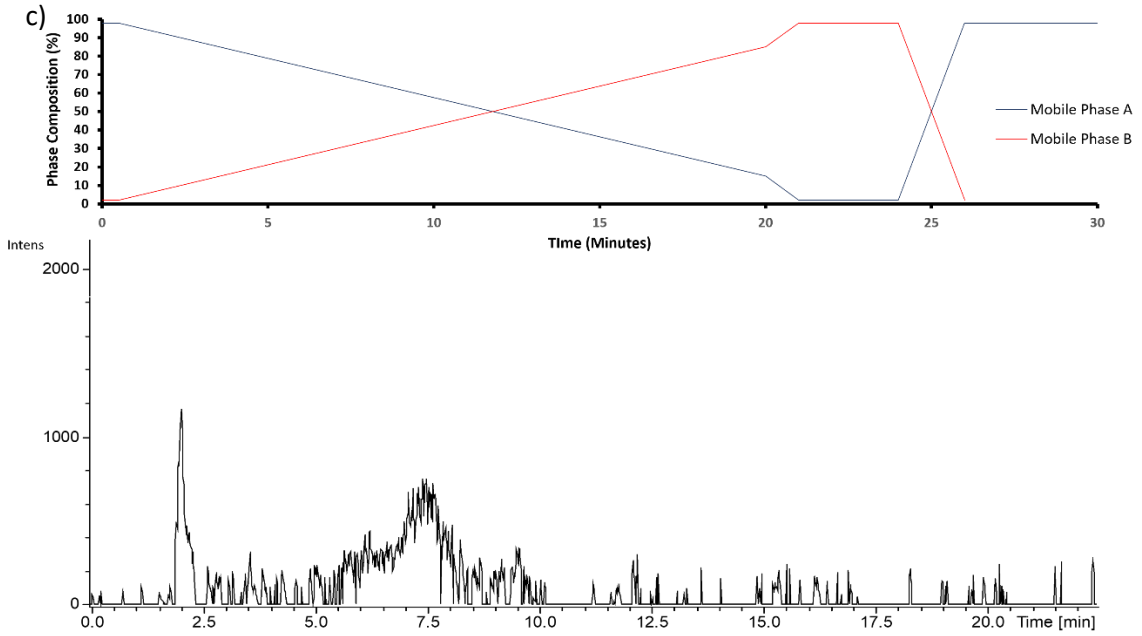
Following the challenges associated with the separation and retention of **26**-GU7 carbohydrate, a series of optimisation conditions were trialled on a pair of polar encapped C_{18} columns. In these samples the GU7 model carbohydrate was used as a sample and separations were carried out over a range of conditions. Previous literature highlights the importance of gradient methods with a low reliance on the proportion of organic mobile phase, emphasising the importance of highly aqueous conditions[6, 47, 48]. This was a significant deciding factor in the choice of column utilised for this application. Following trials of several polar encapped solid phase compositions, it was decided that subsequent separation would be optimised on the Phenomenex synergy hydro.



b)


Phase Composition (%)
Mobile Phase A
Mobile Phase B
TIme (Minutes)
Intens
x10⁴.
Time [min]

c)


Phase Composition (%)
Mobile Phase A
Mobile Phase B
TIme (Minutes)
Intens
Time [min]

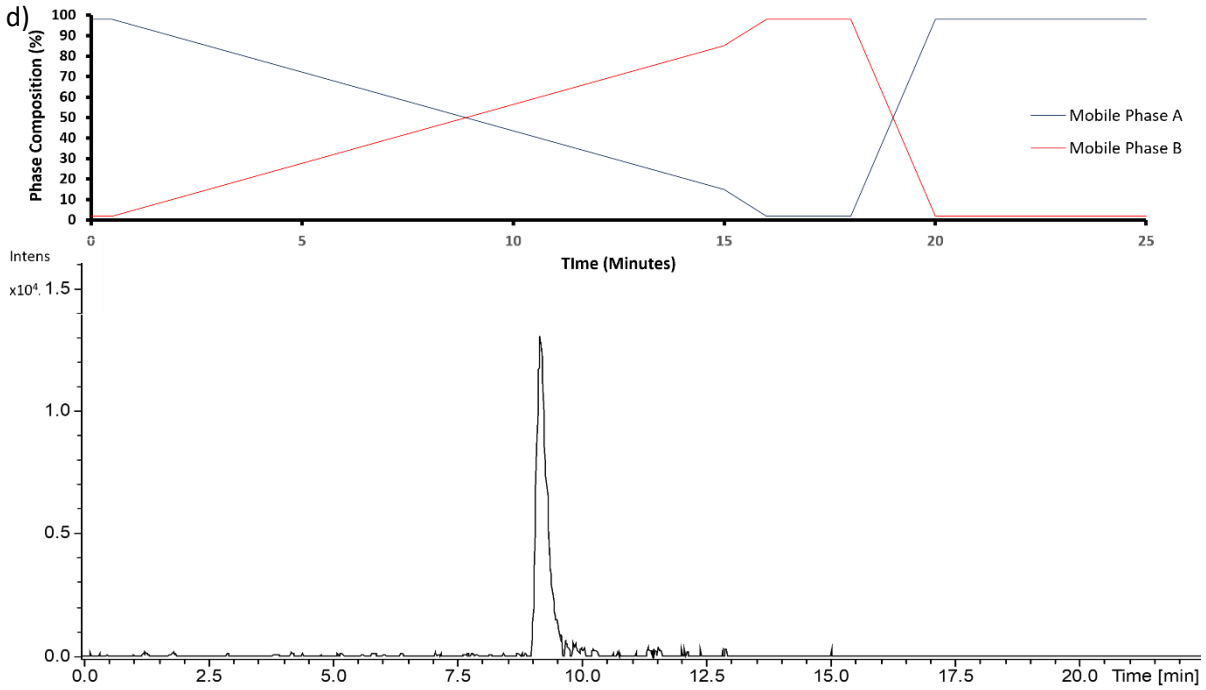


Figure 4.8 - EIC traces for four different gradients a, b, c and d. Aimed at the the separation of **26** labelled GU7 maltoheptose at low concentrations.

Figure 4.8 a, b, c and d display injections using three differing gradients. The retention displayed by GU7-**26**, exhibited in in gradient A provided resolution of the analyte peak, peak splitting and poor retention were observed resulting in a low degree of method control. The elution conditions applied to chromatogram B was a marked improvement over both the previous retention characteristics and those observed in A. However, the elution of GU7-**26** was still prolonged to over 3 minutes and repeated injections exhibited a low degree of control with variation in both retention time and peak symmetry.

Gradients C and D maintained identical separation conditions, with the exception of an increased post run re-equilibration in method D. In comparison to method A, the solvent composition for methods C and D begins with a higher proportion of water with the intention of increasing retention of polar analytes such as GU7-**26**. While this resulted in increased retention and improved peak shape in chromatogram D, a reduction in retention was observed in C. This is likely a result of the increased proportion of aqueous solvent required to generate sufficient retention, and an insufficient equilibration time.

Table 4.4 - Analysis of LCMS peak shape following 5 concurrent injcetions of 26 labelled maltoheptose

Injection	Peak Area	Retention time	Signal/Noise	Peak Asymmetry	Tailing factor
1	162776	9.2	28.5	1.08	1.04
2	162336	9.2	26.7	1.08	1.04
3	175174	9.2	30.8	1.10	1.05
4	175174	9.2	29.2	1.10	1.05
5	169038	9.2	26.4	1.10	1.05
RSD%	3.74	<0.5	6.43	1.00	0.52

Following confirmation of analyte retention in method D, method stability and control was demonstrated though multiple repeat injections of GU7-**26**. These were compared and the relative standard deviation was calculated for peak area, retention time peak asymmetry and tailing factor. The data shown in

Table 4.4 shows variation of less than 5% relative standard deviation for peak area, asymmetry and tailing factor. Increased RSD was observed when comparing signal to noise ratio and this is likely due to operating close to the limit of quantification (LOQ). The low degree of variability observed through repeated injections therefore gives a high level of confidence that repeated injections will be separated uniformly. Demonstrating a high degree of method control for the separation of **26** labelled carbohydrate.

4.3.6 Detection of low abundance carbohydrates bearing hydrophobic derivatisation

Multifunctional label **26** was developed as a tool for the structural analysis of low abundance carbohydrates following separation by HILIC-FLD. The previous sections have seen the conditions for downstream post column click conditions optimised and an improved high aqueous separation method developed. This resulted in a robust method for modification of multifunctionally labelled carbohydrates with azides of increased hydrophobicity. These studies aimed to analyse the effects of increasing hydrophobicity on the efficiency of carbohydrate ionisation by ESI-MS. This aims to provide a tool for both the development of future reductive amination labels, but also as a strategy for the analysis of glycan species at low abundance (**Figure 4.9**).

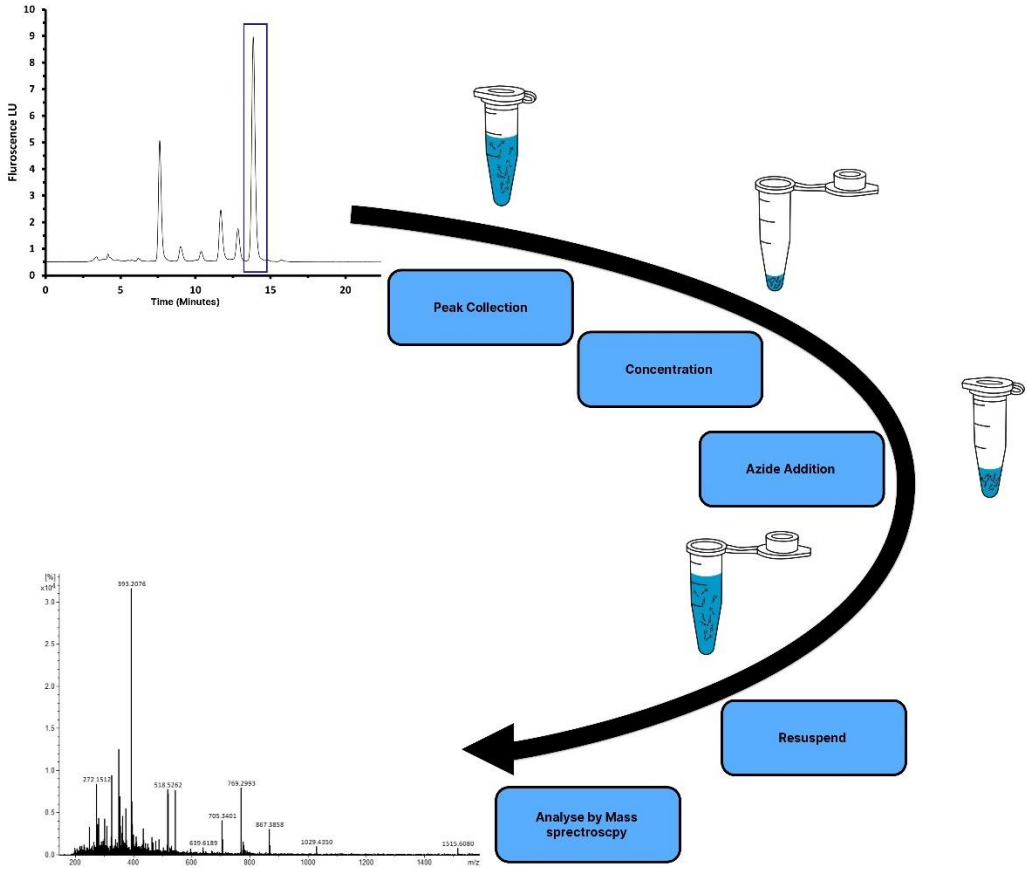


Figure 4.9 - Proposed workflow for the downstream detection of low abundance carbohydrates by LCMS following analysis by HILIC-FLD

Previously, a library of seven azides was created. These compounds predominantly exhibit a neutral charge in solution with the exception of **67**, which possesses a permanent charge modification in the form of a quaternary amine. This allowed the investigation to predominantly focus upon the impact of hydrophobic modification over the effect of permanent charge. The use of neutral compounds facilitates analysis by both positive and negative mode ESI-MS. This has been shown to be advantageous in some MS applications as analysis in negative ion mode has been linked to more informative fragmentation patterns [49, 50]. More importantly, negative mode analysis has been shown to be highly advantageous in the analysis of glycoforms bearing sialic acids [51]. Additionally, the work of Bereman *et al* [15] linked the use of permanent charge derivatives with reduction in ion abundance. As discussed by Walker *et al* [16] this characteristic is of little use in the detection of low abundance samples. Instead, the use of neutrally charged post column derivatisation techniques therefore afford a greater analytical flexibility when analysing structurally diverse low abundance carbohydrates.

Scheme 4.10 - CuAAC modification of 26 labelled maltoheptose GU7 carbohydrate with azide **61**

Multifunctional label **26** was used to reductively aminate a model GU7 maltoheptose carbohydrate at 500 pmol, concurrent with that of a naturally derived carbohydrate at low concentration. As described in the previous sections, labelled carbohydrate samples were then separated by amide HILIC-HPLC. A dextran retention standard was then used to identify Maltoheptose from contaminating GU5 and 6 peaks present in the sample. These peaks were then collected and six concurrent injections were pooled forming an enriched 125 pMol GU7-**26** carbohydrate sample to undergo post analysis derivatisation by CuAAC. **Scheme 4.10** outlines the post analysis derivatisation of **26**-maltoheptose with Azide **61**. These samples were repeated in triplicate to ensure results reliability between collected glycan samples and instrumental analysis conditions. Finally the conditions were applied to azides **61**, **62**, **63**, **64**, **65**, **66** and **76** and 'click' conjugated products are outlined in **Table 4.5**.

Table 4.5 - Outlines the reaction products following modification of GU7 labelled 26 by site selective CuAAC with azide containing conjugation partners along side their corresponding Non-polar surface area (NPSA)

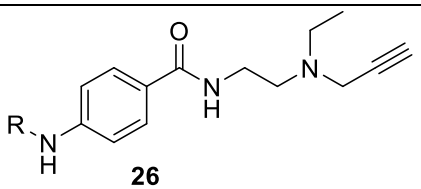
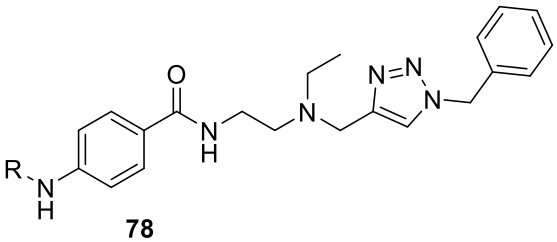
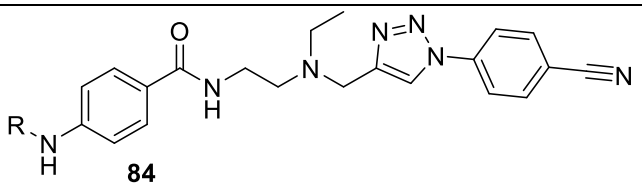
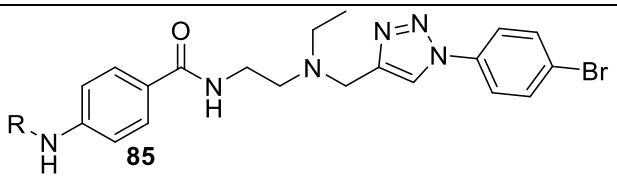
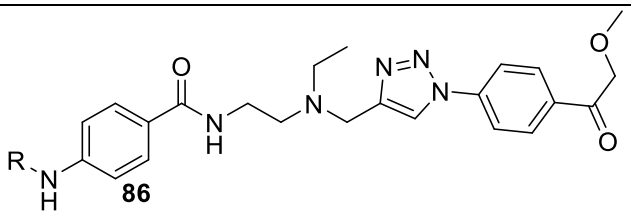
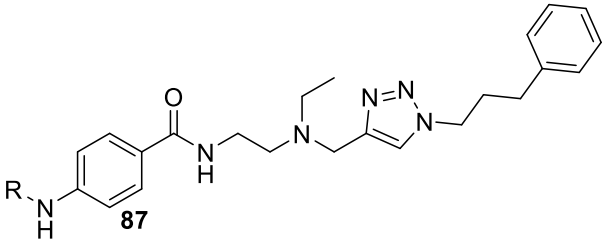
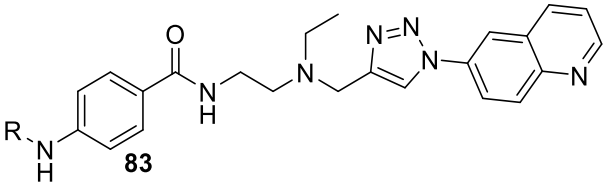
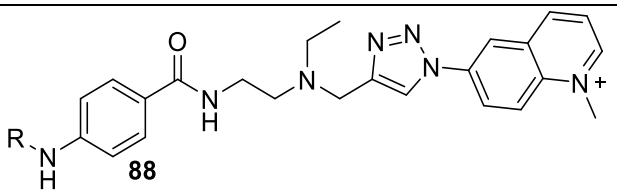
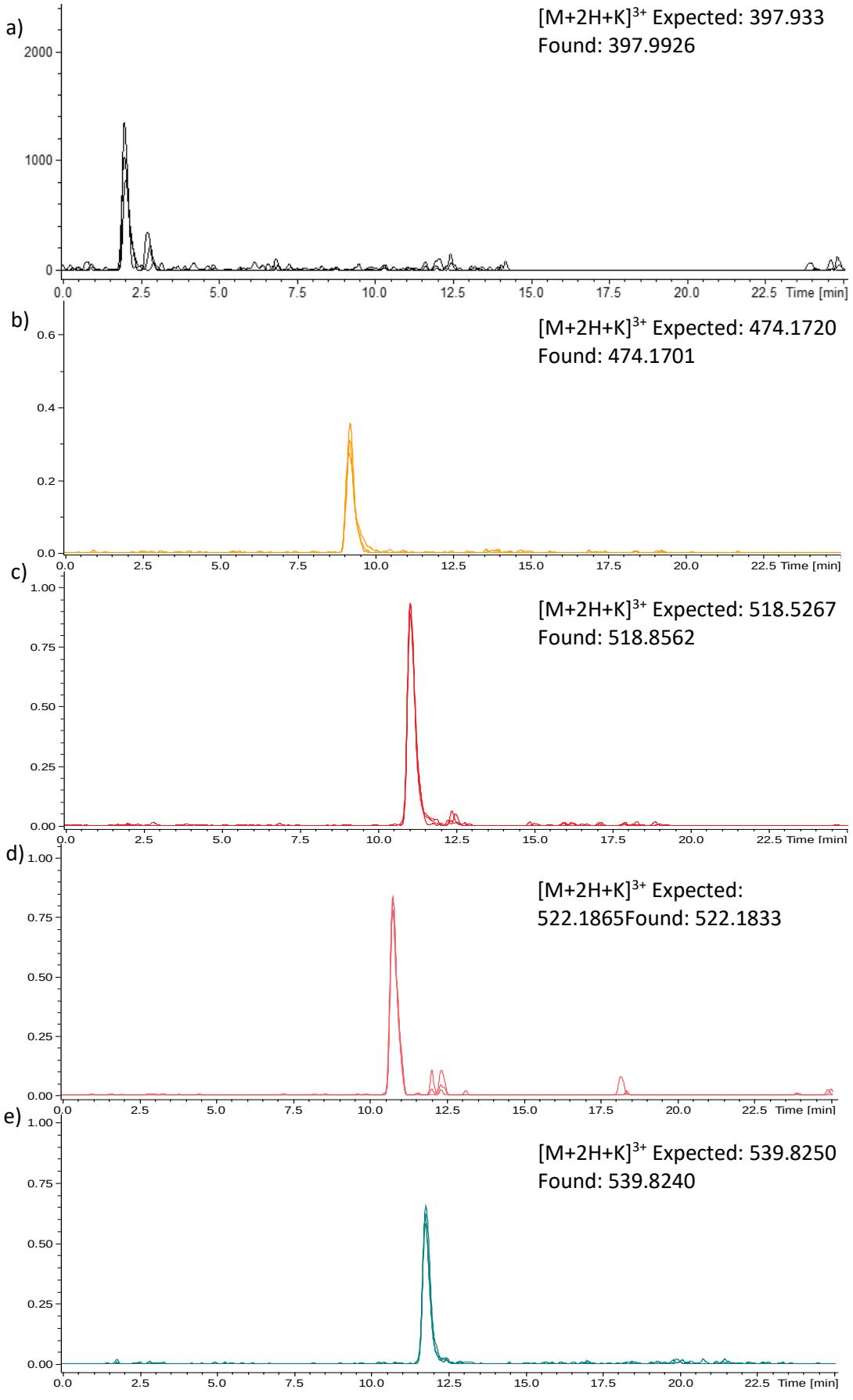
Carbohydrate source	Alkyne No	Azide No	Triazole product structure	NPSA
Maltoheptose GU7	**26**	-	 **26**	184.48
Maltoheptose GU7	**26**	**61**	 **78**	267.15
Maltoheptose GU7	**26**	**62**	 **84**	267.15
Maltoheptose GU7	**26**	**63**	 **85**	243.09
Maltoheptose GU7	**26**	**64**	 **86**	268.63
Maltoheptose GU7	**26**	**65**	 **87**	286.12
Maltoheptose GU7	**26**	**66**	 **83**	305.02
Maltoheptose GU7	**26**	**67**	 **88**	301.04

Table 4.5 outlines the products of CuAAC between both multifunctionally labelled model carbohydrate and an azide conjugation partner. These samples were extracted in line with the optimised click reaction procedure in the previous section over a 3 hour reaction time. Samples were then diluted with a 50% aqueous solution of methanol containing 10 pg reserpine internal standard before being analysed by online LC-MS-QToF. Samples were analysed individually and separated by reversed phase HPLC over a 35 minute runtime inline with the optimised separation gradient covered previously.

Analysis of these samples revealed adequate separation of both triazole, and internal standard peak. Displaying comparable separation characteristics to those experienced in optimisation samples. Additionally although trace amounts of unreacted multifunctionally label carbohydrate remained in the sample, ionic abundance was well below the limit of quantification and inferring a high level of reaction completion of >95% efficiency. Furthermore, this implies that multifunctional label **26** remains stable and intact though both initial and secondary derivatisation workflows. It should also be noted that there were no other carbohydrate containing peaks within the spectra indicating stability of carbohydrate following downstream modification with azide containing conjugation partners.

In order to characterise the effect of downstream hydrophobic click modification, samples bearing triazole modification were analysed by LCMS and observed as a combination of [M+H], $[M+H+Na]^{+2}$, $[M+H+K]^{+2}$ and $[M+2H+K]^{+3}$charges. Additionally, CuAAC samples displayed a variety of [M+Cu] adducts resulting from the presence of $CuSO_4$ catalyst in reaction mixture. These were predominantly observed as $[M+2Cu]^{+4}$ charges but were not observed in native model carbohydrate or **26** derivatised standard. Characterization was therefore carried out on $[M+2H+K]^{+3}$ charge adducts due to its increased intensity over both [M+H] and [M+Na] charge adducts. Additionally, it was observed that the 3+ adduct was more reliably produced than +2 adducts which were rarely observed.


a)
[M+2H+K]3+ Expected: 397.933
Found: 397.9926
b)
[M+2H+K]3+ Expected: 474.1720
Found: 474.1701
c)
[M+2H+K]3+ Expected: 518.5267
Found: 518.8562
d)
[M+2H+K]3+ Expected:
522.1865Found: 522.1833
e)
[M+2H+K]3+ Expected: 539.8250
Found: 539.8240
Time [min]

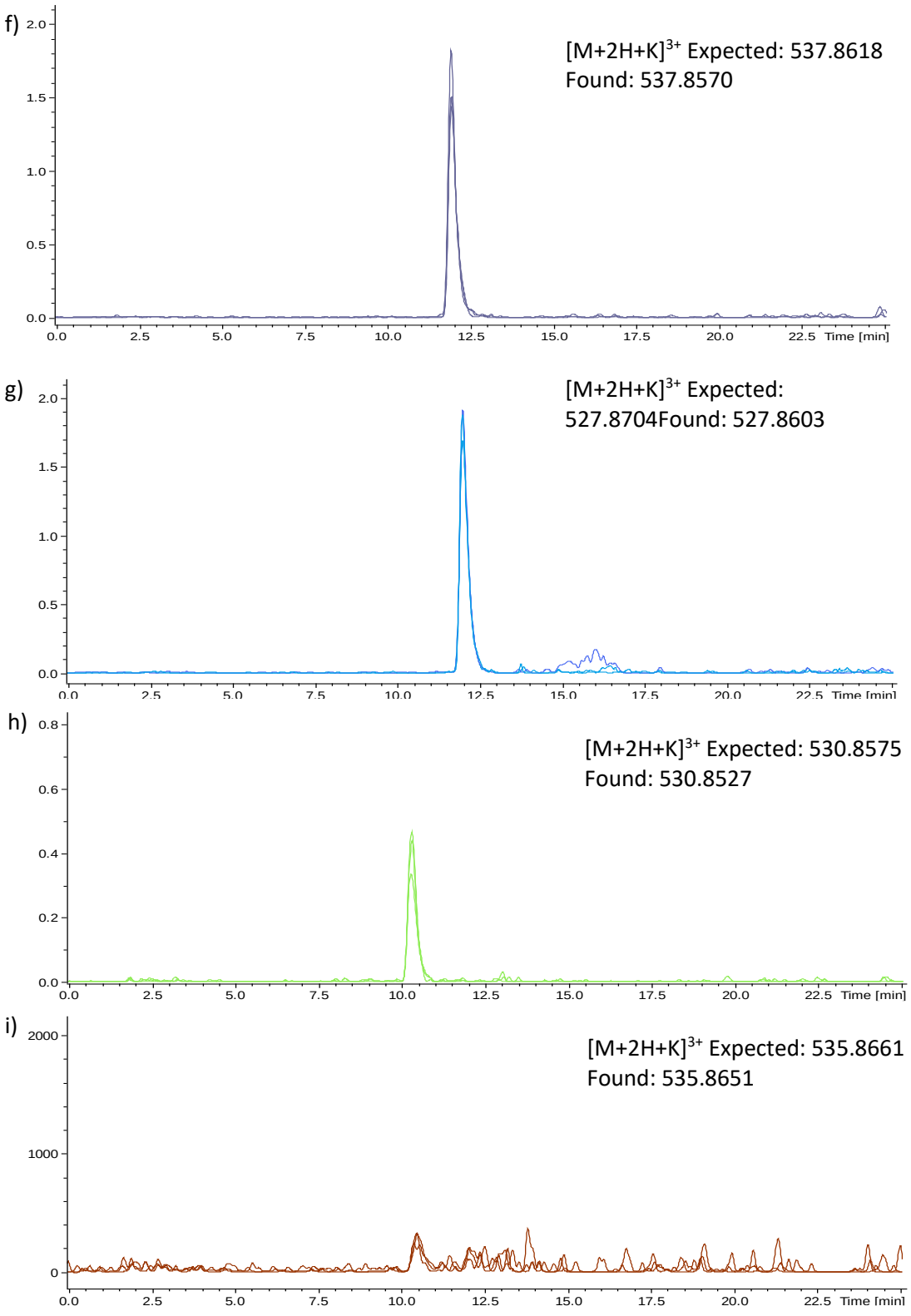


Figure 4.10 - Overlaid EIC for the triazole products resulting from CuAAC modification of **26** labelled carbohydrate. a)GU7 Maltoheptose b) **26** c)**78** d)**84** e)**85** f) **86** g) **87** h)**83** i)**88**

Extracted ion chromatograms (EIC) for each of the reagents and replicates are outlined in **Figure 4.10**. This data shows improved reversed phase sample retention was exhibited by six of the seven modified derivatives when compared to both native model carbohydrate (**Figure 4.10**a) and

derivatised model carbohydrate **Figure 4.10**b. Additionally the data in **Figure 4.10** shows a high degree of repeatability between concurrent extractions of each CuAAC derivative. This trend was also observed between ITSD peaks for each of the replicate injections, representing a high degree of method control between repeat injections over long runtimes outlined in **Table 4.6**below. Conversely EIC for triazole product **88** proved inconclusive due to its proximity to the limit of detection and a high degree of in source fragmentation of the label. This resulted in the formation of ions consistent with the un-charged derivative **83**. It was therefore decided that triazole product would be omitted from any future analysis.

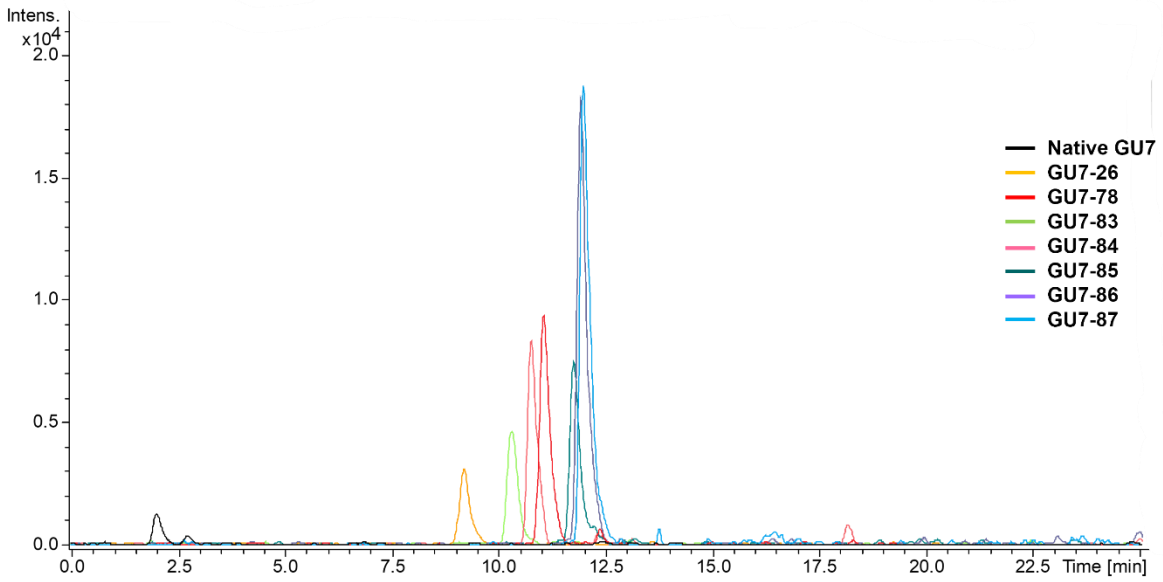


Figure 4.11 - Overlaid EIC chromatographic traces for each of the triazole modifications to **26**

In addition to improved sample retention the remaining six triazole products produced from CuAAC modification of model carbohydrate labelled with **26** resulted in improved peak intensity when conjugated to single aromatic azides when compared to both native and **26** labelled alkyne bearing carbohydrate. **Figure 4.11** shows overlaid EIC for ITSD normalised injections following a 5 µl injection and 25 minute reversed phase separation. In these separations triazole products from modification with **61** and **62** exhibited comparable levels of peak intensity to the **26** derivatised model carbohydrate standard. Conversely conjugation with azides **64** and **65** resulted in substantial increases in ion abundance relative to both native carbohydrate and that derivatised with **26**. The data outlined in **Table 4.6** below also demonstrates that on average a >2 fold increase was observed in CuAAC modified samples when compared to **26** labelled model carbohydrate. Moreover this data also demonstrates up to a 50 fold increase over native model carbohydrate.

Table 4.6 - Mean peak area, retention time and response factor for both Triazole modified carbohydrate and ITSD

Azide	Triazole product formed	Analyte						Internal Standard				Fold Increase over native carbohydrate	Fold Increase over **26** labelled carbohydrate
		Peak Area	RSD %	Retention time	RSD %	Ratio area Analyte/ITSD	RSD %	Peak Area	RSD %	Retention time	RSD %		
-	**26**	59083	3.42	9.21	0.26	0.22	0.80	2626385	3.05	16.49	0.02	9	-
61	**78**	171355	3.62	11.06	0.00	0.16	4.87	10545325	7.90	16.47	0.00	26	3
62	**84**	133552	1.99	10.76	0.38	0.14	2.75	9679952	4.73	16.48	0.01	21	2
63	**85**	108706	3.73	11.78	0.06	0.12	1.82	9370910	4.15	16.46	0.06	17	2
64	**86**	282430	3.05	11.94	0.00	0.33	2	8554580	4.74	16.52	0.29	44	5
65	**87**	322964	2.42	12.01	0.02	0.43	3.38	7563036	3.95	16.49	0.09	50	5
66	**83**	76463	2.23	10.31	0.23	0.08	1.12	9095091	2.14	16.49	0.03	12	1

While CuAAC modification with **61**,**62**, **63**, **64**,**65** and **66** resulted in beneficial impact upon ion abundance resulting in increased peak area over underivatized or alkyne bearing derivatives. Samples modified with Quinoline derivatives **66** and **67** displayed significantly reduced ion abundance with **67** displaying around a 2-fold reduction in peak area. This may be attributed to a degree of in source breakdown or the formation of multiply charged Cu adducts resulting in the reduced abundance of K adducts. Moreover at this scale CuAAc samples bearing azide **66** exhibited an 4 fold reduction in peak intensity when compared to **26** labelled carbohydrate. This data agrees with Bereman *et al*[15] who observed a reduction in detection sensitivity when analysing carbohydrates modified with Girard's T reagent. Interestingly it was also observed that samples modified with **67** displayed a high degree of similar ions with **66** indicating a degree of insource breakdown forming the un-charged derivative, azide **66**.

As previously stated, derivatisation by azides **61**,**62**, **63**, **64** and **65** resulted in a beneficial improvement to ion abundance in when attached to model carbohydrate derivatised with **26**. Of these samples triazole products **78** and **87** outlined in **Table 4.5** show a single phenyl group with varying alkyl chain lengths. This resulted in almost a 2 fold increase in ionisation efficiency and the abundance of triazole product ions when alkyl chain length is increased from one to three. This trend was also exhibited by triazole product **86** resulting from modification with **64** which sees the addition of an aromatic phenyl region combined with a methyl ester and the addition of two alkyl groups. This addition resulted in a comparable increase in ionisation efficiency to that of triazole product **86**. It was therefore hypothesised that the increased number of σ-bonds present following the formation of **78**, **86** and **87** gave an increased level of hydrophobicity resulting in increased ion abundance in ESI ionisation.

This trend was further propagated by modification with azides **62** and **63** resulting in the formation of triazole products **84** and **85**. These samples resulted in similar ion abundance and therefore ionisation efficiency to that of carbohydrate labelled with **26**. In these samples click products resulting from both **62** and **63** do not exhibit the addition of alkyl chain length modification, and instead contain a cyano and covalently bonded bromine molecule respectively. In triazole product **84** the presence of a cyano group provides additional polarity reducing the overall hydrophobicity generated by the addition of an aromatic ring. While in **85** the modification sees the covalent attachment of a bromine molecule, providing significant electron withdrawing power resulting in an exposed area of increased polarity.

4.3.6.1 Effect of non polar surface area on analysis characteristics

Nonpolar surface area was used to estimate the hydrophobic effects of each of the triazole products formed. Previous work by Walker *et al* [12] highlighted the correlation between increasing non polar surface area and increasing ionisation efficiency when applied to hydrazine labelled NA2 glycans. This metric was applied to the triazole products following CuAAc and NPSA are shown in **Table 4.5**. In addition to establishing tends in ionisation efficiency, hydrophobicity is also heavily linked to retention by reversed phase chromatographic separation. In these calculations the effect of carbohydrate surface area was not included in the calculation of NPSA due to the use of model carbohydrate. Instead the NPSA was based upon the structure of the modified label from the site of reductive amination onwards.

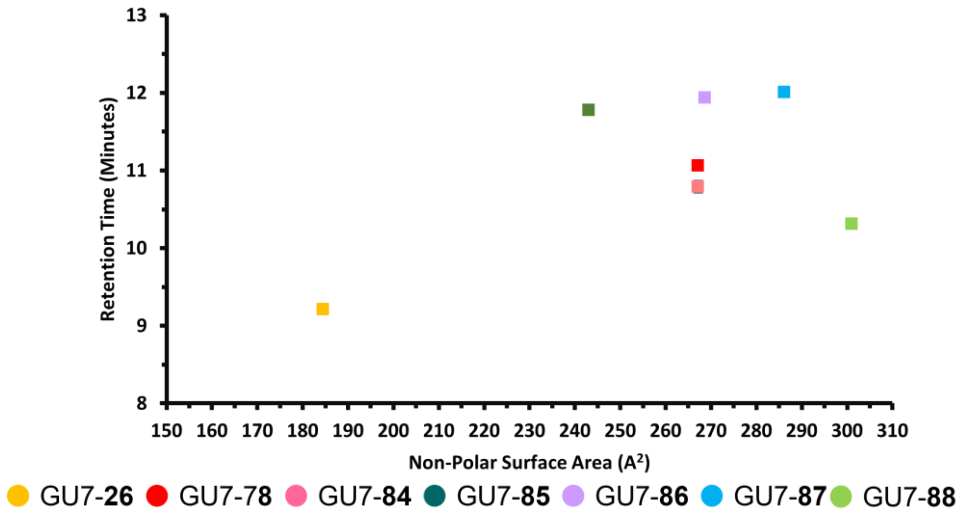


Figure 4.12 - Plot of mean retention time vs NPSA for triazole products formed from conjugation with **26**

The data outlined in **Figure 4.12** shows the relationship between reversed phase retention time and NPSA for each of the triazole products formed from azide conjugation to maltoheptose GU7 labelled with **26**. Unlike normal phase or HILIC based separation methods, reversed phase separation is based upon the interaction of hydrophobicity in the analyte with the C_{18} stationary phase. Separation is therefore heavily influenced by the overall polarity of the analyte. It is therefore expected that increased NPSA would correlate to increased retention and resolution by reversed phase separation. The retention times outlined in **Figure 4.12** broadly align with this trend with the lowest retention times observed for the alkyne bearing **26**. Interestingly, while **78** and **84** exhibited almost identical

NPSA, a 0.3 minute variation was observed in compound retention time. This is likely due to the cyano group present on triazole product **85** which was not considered in the calculation of NPSA. The polarity of the cyano moiety therefore results in a reduced level of retention compared to **78**. This limitation was also observed for **85** which provided a disproportionate increase in retention based upon its NPSA. However, similar to **84,** the hydrophobicity of aryl halides such as **85** was not accounted for in the calculation of NPSA. Meanwhile significant retention enhancement was observed between **86** and **87** although both triazole products formed had an 18 A^2 difference in NPSA. It was therefore hypothesised that the alkyl groups present on both **86** and **87** interacted more heavily with the stationary phase than phenyl groups leading to enhanced retention. Conversely, while CuAAC modification resulted in increased retention of all six triazole products, retention of **88** was significantly less than expected, given its high NPSA. This is likely due to protonation of the tertiary amine present within the pyridine ring resulting in cation interaction with the stationary phase and an altered retention pattern. The broader correlation between calculated NPSA and experimental retention time therefore shows that as hydrophobicity increases analyte retention by reversed phase separation also increases. It was however observed that while there is correlation between retention and NPSA, the increase in retention is not proportional to NPSA increase. This implies that other factors are involved in reversed phase retention on polar encapped solid phase sorbents. The lack of retention offered by **88** also aligns with this observation as it exhibits the largest NPSA of the library but was retained at a lower level than compounds 40 A^2 smaller.

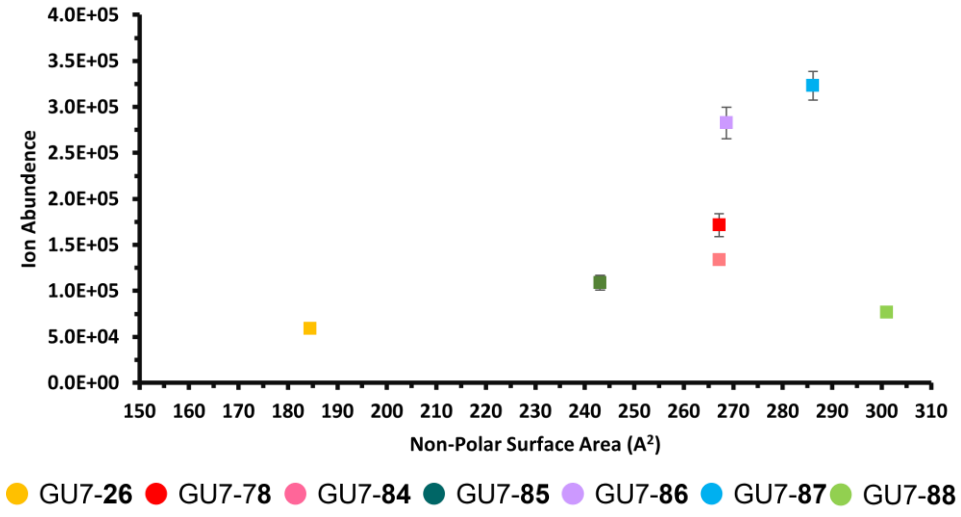


Figure 4.13- Plot of ion abundance vs NPSA for the triazole products formed from CuAAC conjugation with **26**.

The comparison of retention time and NPSA provided an experimental validation for the calculation of hydrophobicity. The data outlined in **Figure 4.13** displays the relationship between NPSA and ion abundance for the products of CuAAC modification of maltoheptose labelled with multifunctional label **26**. Previous research has shown a trend in increased ionisation efficiency of more hydrophobic analytes. It is therefore unsurprising that the data shown in **Figure 4.13** show an increase in ion abundance with increasing NPSA. This data largely aligns with the expected trend with five of the six triazole products exhibiting enhanced ionisation efficiency with NPSA. Interestingly, the increase in abundance for the aryl halide containing **85** was significantly more proportionate to its NPSA unlike its increased retention shown in **Figure 4.12**. Additionally, significantly increased ion abundances were observed for both **86** and **87** indicating a correlation between and increased number of alkyl groups and enhanced ionisation. It should be noted however that the increase in abundance observed was not directly proportional to NPSA indicating that different structural forms have more of an effect on ionisation than merely their overall hydrophobicity. This can be seen for **88** which resulted the lowest increase in ion abundance relative to its NPSA.

4.4 Conclusions

The work outlined in this chapter sought to provide a workflow for the analysis of glycans at low intensity. This was achieved through a process of downstream functionalisation with azide containing conjugation partners. In addition this workflow provides a workflow for the screening of potential glycan labelling compounds though the use of CuAAC.

Initially this investigation sought to apply the workflows optimised in the previous chapter for the derivatisation of carbohydrate with multifunctional label **26** to carry out labellings of a model carbohydrate close to the limit of detection by fluorescence analysis. Sample peaks were then collected and pooled to form a series of samples equivalent to five consecutive injections. In this case this yielded a sample equivalent to ~125 pMol of the initial sample.

Having separated and collected samples resulting from pooled HPLC peaks, a series of CuAAC conjugation reactions was carried out utilising aqueous conditions optimised for multifuctional label **26**. These reactions saw the terminal alkyne present on **26** functionalised with one of seven azides before being analysed by RP-LCMS. In all cases the use of downstream post analysis CuAAc functionalisation resulted in an increased ionisation efficiency over both native underivatized carbohydrate and multifunctional label **26**. It was observed that of all the monophenyl conjugation partners analysed that those containing an addition alkyl chain exhibited increased ionisation efficiency in relation to their non polar surface area.

4.5 Experimental

4.5.1 General Experimental

All chemicals used were purchased through commercial sources (Sigma Merck, Fisher Scientific, Alfa Aeser) and used without further purification. Solvents were purchased commercially at HPLC grade or better and only LCMS grade solvents were used in the production of samples for analysis by mass spectrometry (Fisher Scientific). The maltoheptose model carbohydrate was purchased commercially (carbosynth). All consumables were purchased through Fisher Scientific and Sigma Merck, with the exception of Amide SPE tubes which were purchased from Applied separations. Reaction progress was monitored by TLC. TLC plates were aluminium backed 60 F254 silica (Sigma Merck). Visualisation was then carried out under UV light (λ = 254 nm), azides were visualised by initial staining with triphenylphosphine followed by Ninhydrin (Sigma Merck).

4.5.2 Analytical Equipment

Nuclear Magnetic Resonance

Proton (δ_H) and carbon (δ_c) nuclear magnetic resonance spectroscopy, IR spectroscopy, and high resolution mass spectrometry were carried out according to the methods reported in previous chapters.

High performance liquid chromatography

Separation by HILIC-HPLC was performed on an Agilent 1100 HPLC system equipped with a G1321A FLD detector and an G1314A variable wavelength detector. Labelled carbohydrates were separated on an amide HILIC on a TSKgel Amide-80 HR column (250 x 4.6 mm 5um particle size) fitted with a TSKgel Amide-80 guard column (15 x 3.2 mm 5 µm particle size). Solvent A was composed of aqueous ammonium formate (50 mM pH 4.4) while solvent B consisted of ACN. The flow rate was set at 0.8 mL/min and the column was maintained at 35 ^{o}C for the duration of the run. **26** labelled maltoheptose was treated as a model carbohydrate and separated according to the 35 minute method HILIC method outlined in chapter 3 experimental part 3.6.2.

Separation of Click optimisation samples in **Figure 4.4** was performed by reversed phase liquid chromatography. Analytes were separated on a Phenomenex Aqua C_{18} column (50 x 4.6 mm, 5 µm particle size) fitted with a Phenomenex C_{18} column guard. Solvent A was comprised of 0.1% formic acid in H_2O while Solvent B consisted of 0.1% formic acid in ACN. Separation was carried out at 0.8 ml/minute at a column temperature of 40 ^{o}C following a 2 µL sample injection. Column polarity was modified by a 35 minute linear gradient method beginning with 95% solvent A and gradient to 30% solvent A over 20 minutes before column cleaning in 90% solvent B and equilibration. Analyte

detection was performed on an Agilent G1321A FLD detector set to λ_{ex}=296 nm and λ_{em}=361 nm respectively.

Direct Infusion Mass Spectrometry

Direct infusion mass spectrometry was carried out on a Quadrupole Time of flight mass spectrometer (Bruker Micro-ToF-QII) equipped with an electrospray ionisation ion funnel source. Samples were diluted 100x in a 50% aqueous methanol solution containing 0.1% formic acid. Samples were then infused at a rate of 3 µL per minute and spectra were acquired over a 5-10 minute infusion period.

Electrospray ionisation was carried out at a source temperature of 180 °C and drying gas volume of 4 L/minute. Gas phase ions were introduced to a charge capillary at 4500 V before being analysed at a 3 Hz sampling rate at a mass range between 150-3000 m/z. MS/MS was carried out by collision induced dissociation (CID) at 15 eV in auto MS/MS mode. In this mode up to three MS/MS spectra may be generated per cycle from both a defined inclusion list, as well as the most abundant m/z values in precursor scans. Mass calibration was carried out with Focus mode enabled, according to manufacturer's guidelines and data was acquired in QToF control (Bruker, USA) before being analysed in Compass Data analysis 4.3 (Bruker, USA).

Online Liquid Chromatography Mass Spectrometry

Analysis of hydrophobicity was carried out by reversed phase liquid chromatography coupled to mass spectrometry. Separation was performed on an Agilent 1100 liquid chromatography system coupled to a quadrupole time of flight mass spectrometer. Carbohydrate bearing post column CuAAc derivatisation shown in **Figure 4.10** was separated on a Phenomenex Synergy Hydro column (100 x 2 mm, 2 µm particle size) fitted with a Securguard C_{18} Guard column (Phenomenex). Following a 2 µl sample injection, separation was carried out using a 35 minute gradient method between Solvent A comprising of 0.1% aqueous formic acid and solvent B, 0.1% formic acid in ACN. Chromatographic conditions were altered over the analysis duration, beginning at 98% aqueous phase A reducing to 30% over 20 minutes before the column was cleaned with 90% solvent B over 5 minutes. The gradient then returned to starting buffer composition before the column was re-equilibrated for the final 5 minutes. A 5 minute post-run was also applied to the method to increase retention stability.

Analytes were ionised by Electrospray ionisation and solvated analytes were nebulised at a pressure of 1.2 bar. ESI droplets were desolvated at a source temperature of 198 °C combined with 8.0 L/minute drying gas volume. Gas phase ions were then introduced to the MS via a charged capillary at 4500 V before being analysed at a 0.5 ns sample rate. Fragmentation was carried out by collision

induced dissociation in auto MS/MS mode. In this mode the instrument will select a maximum of three ions from the precursor scan, both taking into account a predetermined inclusion list and high ion abundance. Selected ions were then fragmented at between 10-35 ev. Precursor ions selected were fragmented before being released for the following 60 seconds forcing the fragmentation of a different precursor ion. Data acquisition was then recorded in Qtof control (Bruker, USA) before being analysed in Compass Data analysis 4.3 (Bruker, USA). Quantification was carried out in Quantanalysis 4.3 (Bruker, USA).

4.5.3 Synthesis of a library of azides

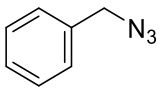
Synthesis of Imidazole-1-sulfonyl azide HCl 73 [28]

Sodium azide (5 g, 77 mol) was added to a flask that was then flushed with nitrogen. Dry ethyl acetate (77 mL) was then added and the resultant suspension was cooled to 0 °C. To this solution, sulfuryl chloride **68** was added dropwise and the reaction mixture was stirred overnight and allowed to warm to room temperature. Imidazole **70** (10 g, 147 mMol, 1.9 eq) was then added under a nitrogen curtain and the resultant slurry was stirred for 3 hours. The reaction mixture was then diluted with EtOAc (100 mL) and then washed with H_2O (4 x 100 mL) and $NaHCO_3$ (2 x 100 mL) before the organic phase was dried on anhydrous $MgSO_4$. The dried organic phase was cooled in an ice bath before a solution of HCl in EtOAc, formed by the dropwise addition of Acetyl chloride (6 mL 115 mmol 1.5 eq), was added. This resulted in the formation of a white precipitate which was filtered and then washed with EtOAc to give **73** as a white crystalline powder (7.32 g, 55% yield).

^{1}H NMR (400 MHz, MeOD) δ 9.81 (t, *J* = 1.3 Hz, 1H CH), 8.24 (t, *J* = 1.9 Hz, 1H CH), 7.84 (dd, *J* = 2.0, 1.3 Hz, 1H CH).

^{13}C NMR (101 MHz, MeOD) δ 138.37(N-**C**-N), 133.96 (N-**C**H-CH), 122.26 (CH-**C**H-N).**HRMS** ESI+ve $C_3H_3N_5O_2S$ expected [M+H] 174.0007. Found [M+H] 174.0077. **MP** Expected 100-102 °C Found 99-102 °C

Synthesis of Benzyl Azide 61[29]

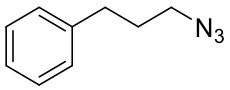


Benzyl amine **74** (50 mg, 0.46 mmol) was added to a 10 ml round bottom flask, to this potassium carbonate was added (128 mg 0.93 mmol 2 eq) and the reaction mixture was solubilized in a solution of imidazole-1-sulfonyl azide HCl **73** in MeOH (97 mg, 0.56 mmol 1.2 eq). The solution was stirred for 5 minutes before $CuSO_4$ was added (5 µL 1% aqueous solution). The solution was stirred overnight before the reaction mixture was filtered and concentrated under reduced pressure and the residue

was purified by flash column chromatography (1:1 EtOAc: Petroleum ether) to afford **61** as a colourless liquid (33 mg, 62% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.43 – 7.29 (m, 5H. 5 x CH), 4.34 (s, 2H, CH_3). **^{13}C NMR (101 MHz, $CDCl_3$)** δ 135.36 (CH-**C**-CH), 128.85 (2x CH-**C**H-CH), 128.32 (C-**C**H-CH), 128.23 (CH-**C**H-CH), 54.81 (CH-**C**H_2-N_3). **HRMS ESI+ve** $C_7H_7N_3$ expected 133.0640 found [M+H] 134.0652

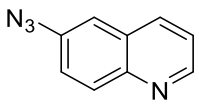
Synthesis of 3-azidobutyl)benzene 65 [29]



A solution of (3-aminobutyl)benzene **75** (200 mg, 1.48 mMol) was added to a 20 mL round bottom flask, Potassium carbonate (410 mg, 2.96 mmol 2 eq) was then added along with Copper sulphate (10 µL, 1% aqueous solution) and MeOH (4 mL). To this solution Imidazole-1-sulfonyl azide HCl **73** (460 mg 1.69 mmol 2 eq) in MeOH (4 mL) was added and the reaction mixture was stirred overnight. Following reaction completion the reaction mixture was filtered and concentrated under reduced pressure. This afforded a colourless liquid which was applied to a glass backed pTLC plate and run with 1:1 EtOAc: Petroleum Ether. Product was then washed from the silica with EtOAc and the solution was concentrated under reduced pressure to yield a colourless Oil which was characterised by the removal of a 2 proton singlet in ^{1}H NMR indicating the modification of the amine present on (3-aminobutyl)benzene **65**. (152 mg, 64% yield)

^{1}H NMR (400 MHz, MeOD) δ 7.21 – 7.10 (m, 2H, 2x CH Ar), 7.10 – 7.00 (m, 3H, 3 x CH Ar), 3.14 (t, *J* = 6.8 Hz, 2H, CH_2), 2.64 – 2.48 (m, 2H, CH_2), 1.82 – 1.66 (m, 2H, CH_2). **^{13}C NMR (101 MHz, MeOD)** δ 141.02 (CH-**C**-CH_2), 128.11(2 x CH Ar), 128.10(2 x CH Ar), 125.70 (CH-**C**H-CH), 50.30(CH_2-**C**H_2-N), 32.36(C-**C**H2-CH2), 30.33(CH_2-**C**H_2- CH_2). **IR** V_{max}/cm^{-1} 1450 M (C-H, bend), 2100 S (azide, stretch), 2800 m (C-H, stretch), 3050 w, b (C-H, aromatic stretch), 3350 w (C-H, aromatic stretch). **HRMS ESI+ve** $C_9H_{11}N_3$ expected [M+H] 162.0953 Found 162.0926

Synthesis of Azido Quinoline 66 [30]

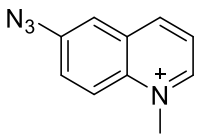


Sodium nitrate (88 mg, 1.04 mmol, 1.5 eq) was added to a round bottom flask and dissolved in 10% aqueous HCl (0.6 mL) before being cooled to 0 °C on ice. 6-Amino quinoline **76** (100 mg, 0.69 mmol) was dissolved in 10% aqueous HCl and added dropwise to the flask. This solution was stirred at a

temperature of <5 °C, after one hour a solution of sodium azide (90 mg, 1.38 mmol, 2 eq) in 10% aqueous HCl (0.4 mL) was added dropwise maintaining reduced temperatures throughout. The reaction mixture was stirred for a further hour and basified by the addition of $NaHCO_3$ (aq)(5 ml). The reaction mixture was transferred to a separating funnel and extracted with petroleum ether (3 x 5 mL). The organic fractions were then dried on anhydrous $MgSO_4$. The orange solvent remaining was filtered and concentrated to yield **66** as an oily orange solid (106 mg, 60% yield)

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.86 (dd, *J* = 4.2, 1.5 Hz, 1H, CH), 8.15 – 8.03 (m, 2H 2 x CH), 7.46 – 7.35 (m, 3H 3 x CH).**^{13}C NMR (101 MHz, $CDCl_3$)** δ 149.87(N-**C**H-CH), 146.09(CH-**C**-N), 138.33(C-**C**H-CH), 135.01(CH-**C**H-C), 131.53(CH-**C**HN_3-CH), 128.91(CHN_3-**C**H-CH), 122.36(CH-**C**-CH), 122.01 (CHN_3-**C**H-H), 115.37 (CH-**C**H-CH). **IR** V_{max}/cm^{-1} 1100 S (C-N stretch), 1350 S (C-H, bend), 1500 S (C-C, stretch), 1590 m (C=C, stretch), 2100 S (azide, stretch), 3100 m, b (C-H, aromatic stretch), 3200 s, b (C-H, aromatic stretch). **HRMS** ESI+ve $C_9H_6N_4$ expected [M+H] 170.0592 found 171.0665

Synthesis of Azido Quinolinium 67 [31]



6-Azidoquinoline **66** (50 mg, 0.29 mmol) was added to a round bottom flask and purged with nitrogen. Dry 1,4 dioxane was added (1 mL) and the solution was stirred until **66** had dissolved. Iodomethane (187 mg, 1.32, mmol 4.5 eq) was then added and the solution was heated at reflux at 120 °C for 1 hour. The reaction was then left to cool and transferred to a separating funnel before being diluted with petroleum ether (5 mL) and washed with $NaHCO_3$ (3 x 5 mL). The organic phase was then dried with anhydrous $MgSO_4$ and concentrated under reduced pressure. This afforded a dark brown product which was characterised by the formation of a three proton singlet at 3.2 ppm ^{1}H NMR as product **67** (25 mg, 47% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.84 (dd, *J* = 4.2, 1.6 Hz, 1H, CH), 8.16 – 7.96 (m, 2H, 2 x CH), 7.47 – 7.29 (m, 3H, 3 x CH), 4.78 (s, 3H, CH_3). **^{13}C NMR (101 MHz, $CDCl_3$)** δ 149.78 (N-**C**H-CH), 145.97 (CH-**C**-N), 138.29 (C-**C**H-CH), 135.01 (CH-**C**H-C), 131.42 (CH-**C**HN_3-CH), 128.86 (CHN_3-**C**H-CH), 122.33 (CH-**C**-CH), 121.97 (CH-**C**-CH), 115.33 (CH-**C**H-CH), 30.32 (+N-**C**H_3). **IR** V_{max}/cm^{-1} 1100 S (C-N stretch), 1350 m (C-H, bend), 1500 S (C-C, stretch), 1550 m (C=C, stretch), 2100 S (azide, stretch), 2900 m, b (C-H, stretch), 3250 s, b (C-H, aromatic stretch). **HRMS** ESI+ve $C_{10}H_9N_4$ expected [M+H] 185.0822 found 185.0822

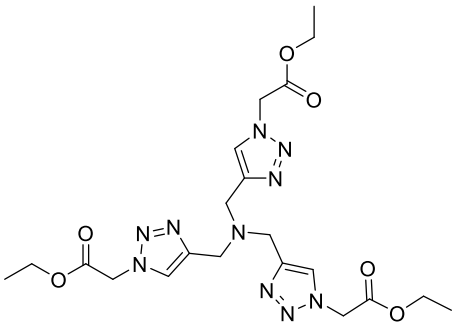
4.5.4 Carbohydrate labelling for low abundance model carbohydrate

Labelling procedure for maltoheptose

Samples of model carbohydrate maltoheptose (10 µl 500 pMol) were labelled according to the general labelling procedure outlined in chapter 3 experimental section 3.6.8. Samples were then incubated at 65 ^{o}C for 3 hours before being quenched in with ACN (950 µL). Excess labelling reagents were then removed by SPE-ed Amide SPE (Applied separations, USA) before being eluted with H_2O (600 µL) and dried under reduced pressure. These samples were then reconstituted in 100 µl of 50% aqueous methanol for analysis.

4.5.5 Post column modification of low abundance carbohydrates bearing multifunctionality.

Synthesis of triethyl 2,2',2''-((nitrilotris(methylene))tris(1H-1,2,3-triazole-4,1-diyl))triacetate 82 [46]



Ethyl chloroacetate **79** (2.9 mL, 27.3 mmol) was added to a round bottom flask containing dry DMF (2.5 mL). The flask was flushed with nitrogen and sodium azide (2.31 g, 35.5 mmol) was added in dry DMF (2.5 mL). The solution was stirred at room temperature overnight after which point the reaction mixture was diluted with H_2O (15 mL) and transferred to a separating funnel. The product was then extracted with diethyl ether (2 x 20 mL) and the combined organic layers were washed with $NaHCO_3$. The organic phase was then filtered and concentrated to ~5 mL and used without further purification.

Tripropargyl amine **81** (0.57 g 4.4 mmol) was added to a round bottom flask and stirred with 2,6 lutidine (0.5 mL, 4.4 mmol). The flask was flushed with nitrogen and dry ACN (5 mL) added. To this solution ethyl azido acetate (3.5 mL, ~20 mmol) was added and the reaction mixture was stirred for 5 minutes. Copper(I) Iodide was then added (17 mg, 0.09 mmol) and the reaction mixture was stirred overnight at room temperature. Following reaction completion by TLC analysis, the reaction mixture was filtered and concentrated under reduced pressure. The yellow oil generated was purified by flash column chromatography yielding a pale yellow oil which crystalised to a white solid upon

cooling to 0 °C. Characterisation by ^{1}H NMR verified presence of a new three proton singlet at 8 ppm indicating the formation of a triazole ring and the formation of **79**. (1.64 g 72% yield).

^{1}H NMR (400 MHz, MeOD) δ 8.02 (s, 3H 3 x CH), 5.35 (s, 6H 3 x CH_2), 4.27 (q, *J* = 7.1 Hz, 6H 3 x CH_2), 3.82 (s, 6H 3 x CH_2), 1.30 (t, *J* = 7.1, 3.3 Hz, 9H 3 x CH_3). **^{13}C NMR (101 MHz, MeOD)** δ 167.08 (3 x O-**C**O-CH_2), 144.19 (3 x CH-**C**-CH_2), 125.66 (3 x N-**C**H-C), 61.79(3 x O-**CH_2**-CH_3), 50.38 (3 x C-CH_2-N), 47.31(3 x N-**C**H_2-CHO), 12.98 (3 x CH_2-CH_3). **HRMS** ESI+ve $C_{21}H_{30}N_{10}O_6$ expected 518.2350 found [M+H] 519.2423.

4.5.6 General click conditions for carbohydrate bearing multifunctionality

Carbohydrate sample collection was achieved through the collection carbohydrate peaks resulting from 6 concurrent separations on HILIC mode HPLC-FLD. Collected samples were dried into centrifuge tubes (~0.125 nmol) and resuspended in sodium ascorbate (6.25 nmol, 50 eq, 10 µL) and ligand **82** (6.25 nMol, 50 eq, 10 µL) was added. Samples were dried under reduced pressure for 10 minutes or until dry before being resuspended in azide solution (2.5 nmol, 20 eq, 5µL) in methanol and copper(II) sulphate (1.8 nmol, 15 eq, 5 µL) in aqueous solution. This solution was then incubated at 30 °C for 3 hours before being centrifuged for 20 seconds and diluted to a volume of 50 µL for injection.

4.5.7 Calculation of Non Polar Surface Area (NPSA)

Calculation of NPSA was carried out computationally using standard Van de Waal's radii and bond length to calculate total surface area as well as polar surface area. These values were used to calculate NPSA. Due to the number of hydrophilic groups present in carbohydrate structures, the NPSA of the carbohydrate itself was deemed negligible. This resulted in NPSA calculation for multifunctional label **26** modification to be calculated from the site of reductive amination downward, omitting the carbohydrate.

4.6 References

1. Y. Zhu, X. Liu, Y. Zhang, Z. Wang, Y. Lasanajak and X. Song, *Bioconjugugate Chem.*, 2018, **29**, 3847-3855.
2. S. Zhou, L. Veillon, X. Dong, Y. Huang and Y. Mechref, *Analyst*, 2017, **142**, 4446-4455.
3. S. Zhou, K. M. Wooding and Y. Mechref, *Methods Mol. Biol.*, 2017, **1503**, 83-96.
4. J. Zaia, *Omics*, 2010, **14**, 401-418.
5. K.-i. Yoshino, T. Takao, H. Murata and Y. Shimonishi, *Anal. Chem.*, 1995, **67**, 4028-4031.
6. S. Kurz, M. O. Sheikh, S. Lu, L. Wells and M. Tiemeyer, *Mol. Cell Proteomics*, 2021, **20**, 100045.
7. J. C. Bigge, T. P. Patel, J. A. Bruce, P. N. Goulding, S. M. Charles and R. B. Parekh, *Anal. Biochem.*, 1995, **230**, 229-238.
8. M. Pabst, D. Kolarich, G. Pöltl, T. Dalik, G. Lubec, A. Hofinger and F. Altmann, *Anal. Biochem.*, 2009, **384**, 263-273.
9. T. Keser, T. Pavić, G. Lauc and O. Gornik, *Front. Chem.*, 2018, **6**, 324-324.
10. Y. Xie, L. M. Mota, A. Bergin, R. O'Flaherty, A. Jones, B. Morgan and M. Butler, *Anal. Biochem.*, 2021, **623**, 114205.
11. D. J. Harvey, *J. Am. Soc. Mass Spectr.*, 2000, **11**, 900-915.
12. S. H. Walker, L. M. Lilley, M. F. Enamorado, D. L. Comins and D. C. Muddiman, *J. Am. Soc. Mass Spectr.*, 2011, **22**, 1309-1317.
13. T. Naven and D. Harvey, *Rapid Commun. Mass Sp.*, 1996, **10**, 829-834.
14. G. C. Gil, Y. G. Kim and B. G. Kim, *Anal. Biochem.*, 2008, **379**, 45-59.
15. M. S. Bereman, D. L. Comins and D. C. Muddiman, *Chem. Commun.*, 2010, **46**, 237-239.
16. S. H. Walker, B. N. Papas, D. L. Comins and D. C. Muddiman, *Anal. Chem.*, 2010, **82**, 6636-6642.
17. M. J. Bowman and J. Zaia, *Anal. Chem.*, 2007, **79**, 5777-5784.
18. D. Reusch, M. Haberger, D. Falck, B. Peter, B. Maier, J. Gassner, M. Hook, K. Wagner, L. Bonnington and P. Bulau, MAbs, 2015,**7**,732-742.
19. M. Wilm, *Mol. Cell Proteomics*, 2011, **10**, M111.009407.
20. A. P. Null, A. I. Nepomuceno and D. C. Muddiman, *Anal. Chem.*, 2003, **75**, 1331-1339.
21. A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, Macmillan, 1999.
22. S. Banerjee and S. Mazumdar, *Int. J. Anal. Chem.*, 2012, **2012**, 282574.
23. L. Konermann, E. Ahadi, A. D. Rodriguez and S. Vahidi, *Anal. Chem.*, 2013, **85**, 2-9.
24. C. M. Shuford and D. C. Muddiman, *Expert Rev. Proteomics*, 2011, **8**, 317-323.
25. G. S. M. Lageveen-Kammeijer, N. de Haan, P. Mohaupt, S. Wagt, M. Filius, J. Nouta, D. Falck and M. Wuhrer, *Nat. Commun.*, 2019, **10**, 2137.
26. C. Smith, unpublished work.
27. F. Higel, A. Seidl, U. Demelbauer, M. Viertlboeck-Schudy, V. Koppenburg, U. Kronthaler, F. Sörgel and W. Friess, *Pharm. Res.*, 2015, **32**, 3649-3659.
28. E. D. Goddard-Borger and R. V. Stick, *Org. Lett.*, 2007, **9**, 3797-3800.
29. G. T. Potter, G. C. Jayson, G. J. Miller and J. M. Gardiner, *J. Org. Chem.*, 2016, **81**, 3443-3446.
30. Z. U. Khan, B. Nay, E. F. V. Scriven and H. Suschitzky, *J. Chem. Soc.*, 1982, 671-672.
31. M. Ghandi and E. Babazadeh, *J. Iran. Chem. Soc.,* 2015, **12**, 379-387.
32. J. Geng, J. Lindqvist, G. Mantovani and D. M. Haddleton, *Angew. Chem. Int. Edit.*, 2008, **47**, 4180-4183.
33. L. Nurmi, J. Lindqvist, R. Randev, J. Syrett and D. M. Haddleton, *Chem. Commun.*, 2009, 2727-2729.
34. X. Ning, W. Rui, D. Fu-Sheng and L. Zi-Chen, *J. Poly. Sc.*, 2009, **47**, 3583-3594.

35. A. J. Cagnoni, J. Kovensky and M. L. Uhrig, *J. Org. Chem.*, 2014, **79**, 6456-6467.
36. H. A. Stefani, N. C. S. Silva, F. Manarin, D. S. Lüdtke, J. Zukerman-Schpector, L. S. Madureira and E. R. T. Tiekink, *Tetrahedron Lett.*, 2012, **53**, 1742-1747.
37. F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless and V. V. Fokin, *J. Am. Chem. Soc.*, 2005, **127**, 210-216.
38. D. S. Ermolat'ev, V. P. Mehta and E. V. Van der Eycken, *Qsar Comb. Sci.*, 2007, **26**, 1266-1273.
39. Y. M. A. Yamada, S. M. Sarkar and Y. Uozumi, *J. Am.Chem. Soc.*, 2012, **134**, 9285-9290.
40. J. A. Shin, Y. G. Lim and K. H. Lee, *J. Org. Chem.*, 2012, **77**, 4117-4122.
41. Y. Jiang, C. Kuang and Q. Yang, *Synlett*, 2009, **2009**, 3163-3166.
42. J. E. Hein and V. V. Fokin, *Chem. Soc. Rev.* , 2010, **39**, 1302-1315.
43. R. Berg and B. F. Straub, *Beilstein J. Org. Chem.*, 2013, **9**, 2715-2750.
44. V. O. Rodionov, S. I. Presolski, D. Díaz Díaz, V. V. Fokin and M. G. Finn, *J. Am. Chem. Soc.*, 2007, **129**, 12705-12712.
45. T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853-2855.
46. H. B. R. Kramer, D.Phil, University of Oxford, 2006.
47. F. Higel, U. Demelbauer, A. Seidl, W. Friess and F. Sörgel, *Anal. Bioanal. Chem.*, 2013, **405**, 2481-2493.
48. G. C. M. Vreeker and M. Wuhrer, *Anal. Bioanal. Chem.,* 2017, **409**, 359-378.
49. D. J. Harvey, *J. Am. Soc. Mass Spectr.*, 2005, **16**, 622-630.
50. A. Hykollari, K. Paschinger and I. B. H. Wilson, *Mass Spectrom. Rev.*, 2022, **41**, 945-963.
51. J. L. Seymour, C. E. Costello and J. Zaia, *J. Am. Soc. Mass Spectr.*, 2006, **17**, 844-854.

Chapter 5- Summary and future proposals

5.1 General discussion

The overarching scope of this thesis was to generate tools to aid and facilitate the analysis of glycan structures, by the use of downstream functionalisation. This was achieved by the development and synthesis of a suite of reductive amination glycan labels bearing additional functionally above that of the primary amine required to react with reducing glycans. In multifunctional labels **25**, **26** and **27**, this was facilitated by the addition of a functional handle bearing a terminal alkyne, which was capable of undergoing selective copper catalysed cycloaddition with azide containing conjugation partners(**Figure 5.1**).

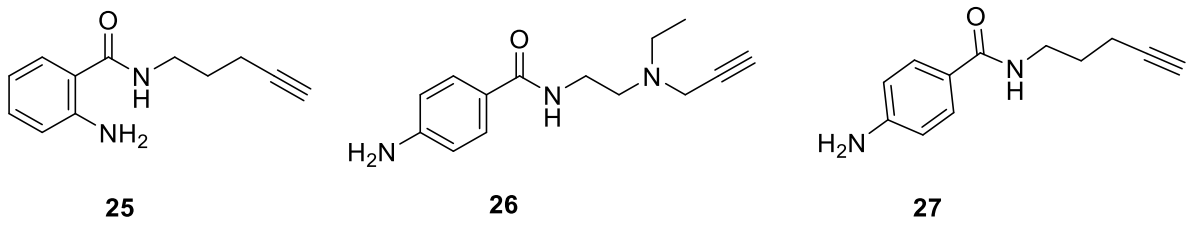


Figure 5.1 - Structures for multifunctional labels **25** ,**26** and **27** developed in Chapter 2

The study then sought to establish the activity, sensitivity and efficiency of the multifunctional labels in comparison with existing reductive amination labelling strategies. This was initially achieved though the generation of quantitative carbohydrate standards based upon a model disaccharide carbohydrate. This facilitated both quantitative evaluation of glycan labelling compilation, as well as a determination of the limit of detection by fluorescence analysis. Following the successful application of the multifunctional labels within existing glycan profiling workflows, multifunctional labelling was also carried out on native glycan species, such as RNase B and fetuin, as well as GHP retention standards for optical profiling.

The project sought to evaluate the capability of **27** to act as a downstream multifunctional tool for the detection of carbohydrates at low abundance. This was initially established though a series of trial reactions on multifunctional label **27**. These trial reactions were then optimised, enabling derivatisation of multifunctionally labelled carbohydrates to occur on a picomolar scale enabling its use on low abundance samples from biological sources.

This allowed the addition of further hydrophobicity to facilitate improved ionisation efficiency in ESI-LCMS. Azide conjugation reactions were then carried out in order to evaluate the impact of additional hydrophobicity upon the detection sensitivity of a known heptasaccharide carbohydrate.

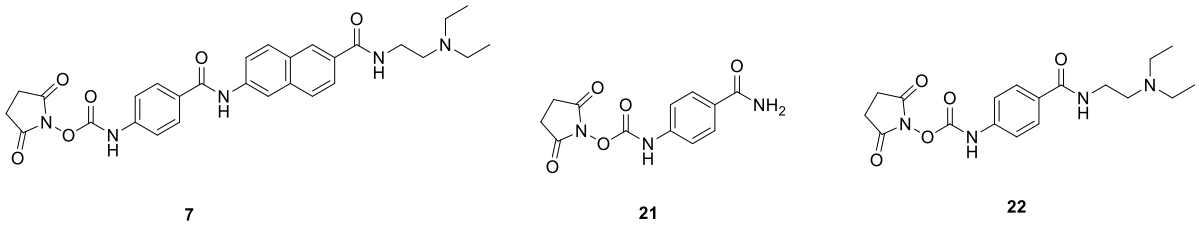
The work outlined in chapter two set out to synthesise a suite of multifunctional glycan labels. These compounds were developed as derivatives of the most widely used commercial labelling systems **9** and **8**. Unlike commercially available labelling systems which possess a single site for reductive amination, multifunctional labels **25**, **26** and **27** possess an additional terminal alkyne capable of undergoing site selective reactions with single azide moieties following the reductive amination of reducing carbohydrates.

The synthesis of **25** was optimised and achieved through a four-step synthesis with an overall yield of 50%. This synthesis followed a similar sequence to that used for the synthesis of **27**, which was optimised in the laboratory and also performed over four steps with an overall yield of 61%. Meanwhile, the synthesis of **26** was previously attempted through a five-step synthetic route, resulting in the formation of several by-products and the formation of impure **26** as a brown oil at <45% [1]. This synthesis was redeveloped and optimised during the work outlined in Chapter 2, resulting in the formation of **26** as a light brown crystalline solid with an overall yield of 54%.

The generation of multifunctional labels **25**, **26** and **27**provides three fluorescent multifunctional labels capable of reacting with reducing carbohydrates in existing optical workflows to aid analysis. In addition, these labelling compounds also possess a downstream capability in the form of a terminal alkyne containing a multifunctional handle. This region is present in all three of the multifunctional labels and represents a site selective region which remains stable and active following the reaction with a reducing carbohydrate.

The future development of these labelling compounds could see their conversion to *N*-hydroxy succinimide containing derivatives. This would involve the conversion of the primary amine to one which is modified with an NHS containing moiety, enabling analysable data to be collected more promptly following the release of the glycan. Reductive amination workflows typically take between two to four hours from label addition to achieve complete labelling. This is then followed by SPE purification and sample concentration prior to resuspension and analysis. This leads to a combined sample turn around time of up to two days in commercial settings [2, 3]. The addition of an NHS containing label has been shown to allow significantly faster labelling, reducing labelling times to between five and thirty minutes [4]. While this labelling system still requires a degree of SPE based purification, reduced labelling times provide an attractive method for reducing sample turnaround time. In addition, the use of NHS carbamate labelling negates such as those shown in **Figure 5.2** the

use of harmful or toxic reducing agents, improving the accessibility of the derivatisation method, because it could be considered bench safe.



7 21 22

Figure 5.2 - Commercially available NHS carbamate labels 7, 21, 22

NHS-Carbamate labels, such as **7**, **21** and **22** shown in **Figure 5.2** are already becoming more widely used in mass spectrometry based workflows. However, the use of NHS carbamate labels is limited by the requirement for released glycans to be in a glycosylamine form, where glycans possess a terminal amine, rather than aldehyde. This enables NHS substitution to take place, removing the NHS structure from the labelling molecule to form a secondary amine with a non reducing carbohydrate[5, 6]. This limits their use exclusively to glycans resulting from enzymatic digestion by PNGase F. Moreover, the reactivity of NHS-carbamates when exposed to primary amines requires addition clean up following glycan release to remove peptide fragments which will be derivatised by NHS labels[7].

The aim of chapter 3 was to compare multifunctional labels **25**, **26** and **27** with existing reductive amination workflows. This began with analysis of their fluorescence characteristics by 3D scanning fluorescence on labelled maltoheptose samples. This enabled the optimum wavelengths for excitation and emission to be established as λ_{ex} =222, λ_{em}=405 and λ_{ex} =296, λ_{em}=361 for multifunctional labels **25** and **26** respectively. These values fell in a similar range to multifunctional label **27**, which was previously determined as λ_{ex} =281, λ_{em}=346 and was re-verified during this investigation as λ_{ex} =282, λ_{em}=348 when attached to maltoheptose.

Following the determination of optical analysis properties, a series of labelled GU2 carbohydrate standards were produced for multifunctional labels **25**, **26** and **27**, as well as their commercial derivatives **10**, **9** and **8**. This aimed to investigate linearity and limits of detection (LOD) for **25**, **26** and **27** when compared to their commercially available derivatives. Due to the stoichiometric addition of label to carbohydrate, both multifunctional and commercial labels gave high levels of linearity (>0.9970) when calibration standards from 10 to 100 nmol were analysed by fluorescence detection following separation by HILIC mode HPLC. The lowest levels of linearity were achieved for both **26** and it's commercial derivative **8** ProA. It was speculated that this was likely due to increased peak

broadening as a result of the increased hydrophobicity present in both **26** and **8** proA. This can be seen in Chapter 3 **Figure 3.6**c and **Figure 3.8**c for both **26** and proA **8** respectively.

The use of purified carbohydrate standards was then translated to the determination of limit of detection for both multifunctional and commercially available glycan labels. This is significant as current industrial benchmarks have focused heavily upon the use of retention standards, rather than quantifiable concentrations, and thus official values for detection limits are not available in existing literature. This investigation established the LOD of **25**, **26** and **27** as 25 pMol, 25 pMol and 25 pMol respectively. This trend was also observed for the commercial labels, with LOD of 25 pMol 25 pMol and 10 pMol for multifunctional labels **10**, **9**, and **8** respectively. This data demonstrated that the addition of a terminal alkyne linker was not prohibitive for the detection sensitivity of the multifunctional labels by optical analysis techniques.

Having established the sensitivity of the multifunctional labels, a series of labelling reactions was carried out on a glycose homopolymer ladder and products were separated by HILIC HPLC over a 35 minute runtime. The resulting traces for **25**, **26** and **27** showed high levels of similarity for both separation order and retention characteristics when compared to their commercially available derivatises. In order to accurately compare between separation characteristics, retention time was plotted against homopolymer length. The linearity of these plots showed high levels of similarity between both multifunctional and commercially available samples, with values of 0.9941, 0.9934, and 0.9968 obtained for multifunctional labels **25**, **26** and **27** respectively. Analysis of **10**, **9** and **8** resulted in R^2 of 0.9924, 0.9935 and 0.9855. This therefore demonstrated the compatibility of a multifunctional glycan label with existing glycan analysis profiling techniques. Both retention and quantification standards were then applied for the quantification of labelling completion at both 65 and 86 ^{o}C. Labelling reactions were carried out in triplicate and halted at time points from 10 minutes to 180 minutes. For samples that were reacted at 65 ^{o}C, all the labels analysed showed high levels of completion by 120 minutes, with the exception of **10**, which required the full three hours to achieve completion. Samples that were reacted at 85 ^{o}C resulted in considerably faster completion times for **25** and **27** as well as **9** and **8** their commercial derivatives, achieving close to completion in 60 minutes. Interestingly, although both **8** and **26** displayed increased reaction rates, both required longer to achieve completion than **10** and **9**. These labelling conditions were then applied to both sialylated and non sialylated *N*-glycan species derived from Bovine Fetuin and RNAse B. These yielded almost identical separation characteristics to literature separations for **9** and **8** [4, 8].This was also demonstrated for the Bovine Fetuin sample. Significantly, this also demonstrated that sialylation was maintained throughout the reductive amination process. This not only helps to demonstrate the commercial potential of multifunctional labels **25**, **26** and **27** within existing glycan analysis

workflows, but also shows their potential to functionalise carbohydrates from cell surfaces, and from wider biomolecules, for the analysis of carbohydrate protein interactions or the generation of novel glycoconjugates.

In addition to this, the work in this chapter could be further developed to gain an increased understanding of the analytical sensitivity of **25**, **26** and **27** as derivatisation agents for intact glycan analysis by mass spectrometry applications. This would enable further structural elucidation to take place, yielding a greater understanding of carbohydrate structure.

The research covered in chapter four aimed to demonstrate and evaluate the use of multifunctional label **26** as a tool for the detection and analysis of carbohydrates released from proteins at low abundance. This section also aimed to develop a set of reliable 'click' conjugation reactions for the post-analysis modification of carbohydrates labelled with **26**. This was completed with hydrophobic azide containing conjugation partners. This provided the opportunity to modify **26** with beneficial hydrophobic or conjugated groups, to facilitate improved detection sensitivity by either FLD or MS detection.

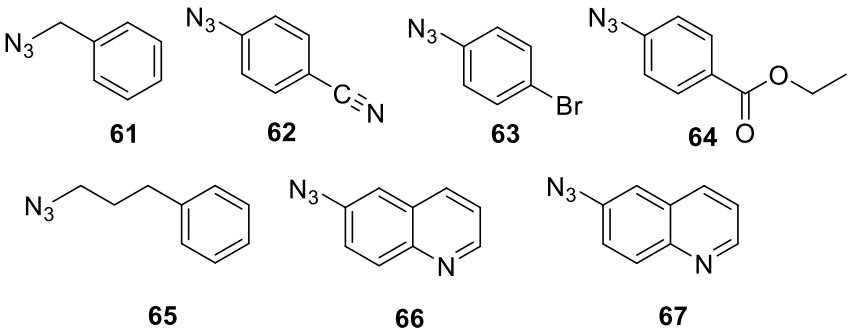


Figure 5.3 - Library of azide containing compounds obtained through synthesis and collaboration.

This was initially achieved though the development of a small library of azide containing compounds (**Figure 5.3**). Three were synthesised while a further four obtained through collaboration were donated. Synthesis of azides **61** and **65** were facilitated by the addition of diazotransfer reagent **73** [9], which was initially synthesised in a yield of 55% following a three step synthesis. The application of **73** stick resulted in the successful modification of **74** and **75** to azides **61** and **65** with yields of 58% and 76% respectively. The conversion of **76** to **66** by the use of diazotransfer reagent **73** was attempted but proved unsuccessful. However, following the application of the method developed by Khan *et al*[10], **66** was successfully generated in a 60% yield. Methylation of **66** was then carried out, leading to the generation of a permanently charged azide species **67** in a 47% yield.

Potential conditions for the coupling of **26** with azide containing conjugation partners were established though a screening process. Twelve conditions were screened for the reaction **26** with Azide **61** following a five hour incubation period. This yielded a series of conditions capable of undergoing CuAAC at high yield while maintaining above 50% aqueous conditions to facilitate carbohydrate solubility. Further optimisation saw the development of aqueous methanolic conditions at a 1:1 ratio due to a combination of high reaction yield and carbohydrate compatible solvent conditions. Prior work by Kramer *et al* [11] [12]and Chan *et al* [13] observed the beneficial effects of triazole containing ligands in stabilising Cu(I) in solution, leading to the synthesis of triazole ligand **82** over a two stage synthesis at 70% yield. The addition of triazole ligand **82** combined with increased equivalencies of sodium ascorbate was shown to achieve increased yields when reactions were carried out at a 400 nMol scale and analysed by offline DI-MS due to the disappearance of reactant masses from the infused spectrum. The revised conditions saw the use of 25 Mol% $CuSO_4$ combined with 50 equivalents of Sodium ascorbate in a 1:1 H_2O:MeOH solvent solution.

These conditions were then applied to the conjugation of **61** and **66** with maltoheptose labelled with **26** on a 125 pMol scale over a range of time periods from one to three hours. Over a three-hour time period, the addition of increased equivalents resulted in significant increases in both product yield and reaction rate. Further samples would therefore incorporate an increased concentration of azide solution equivalent to 20 times that of the anticipated carbohydrate concentration. Subsequent test reactions were then carried out, yielding triazole product yields of 95% for both azide **61** and **66**, representing efficient product generation across the full range of hydrophobicity present in the library.

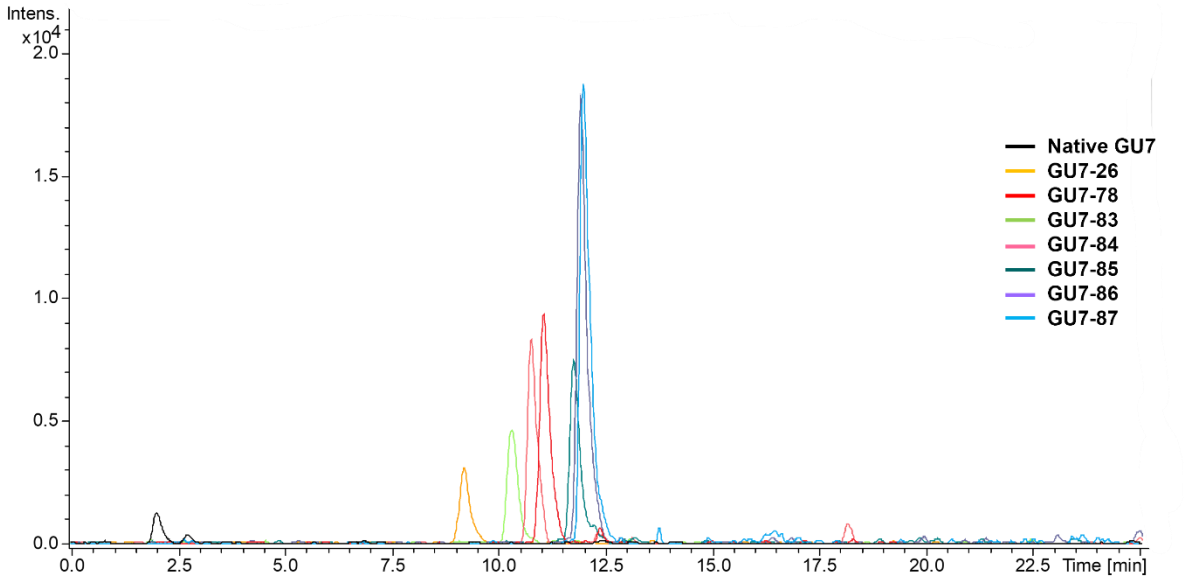


Figure 5.4 - Overlaid EIC for triazole products resulting from GU7 labelled with 26 conjugated to azides **61**, **62**, **63**, **64**, **65**, **66** compared to native underivatized GU7 carbohydrate.

The conditions were therefore applied to the remaining five azides and repeated in triplicate, before being analysed by online LCMS over a 35-minute reversed phase separation method and before analysis by positive mode ESI-QToF operating in full scan mode (**Figure 5.4**). Triazole modified **26** derivatives were then evaluated and compered to their NPSA, which was computationally generated. Much of the data showed a strong positive correlation between calculated NPSA, reversed phase retention, and peak area. The triazole product of **66** resulted in lower then expected retention time and a reduced peak area in relation to its relatively high NPSA of 305.02 A^2. Additionally, the relationship between NPSA and retention, although positively correlated, was not linear, leading to the conclusion that other factors are also involved in hydrophobic retention on polar encapped stationary phase materials. This trend was also observed when comparing ionisation efficiency, leading to the conclusion that the presence of alkyl groups have a more significant effect on in the ionisation efficiency of the triazole products over both cyano- and halo- monophenyl derivatives. This work aligns with the conclusions of Hunter-Walker [14] in the development of hydrophobic hydrazine labels possessing varying degrees of hydrophobicity and alkyl chain length. The data within this chapter therefore concluded that the selective addition of hydrophobicity to derivatised carbohydrates by the use of a multifunctional label increased the detection sensitivity by >50 times over native carbohydrate and up to >2times over **26** labelled carbohydrate prior to CuAAC modification.

This research provides a commercially viable tool for the detection of low abundance carbohydrates, which could be further developed in the future to utilise a wider variety of mono and biphenyl azide

conjugation partners exhibiting differing alkyl chain lengths. Due to the popularity of 2AB **9** in analytical existing workflows, this research could be applied to multifunctional label **25** combined with conjugated fluorescent systems to improve optical properties when analysing low abundance glycans within optical analysis workflows. Furthermore, a multifunctional labelling platform could be developed into a kit for the analysis of released glycans by liquid chromatography. This would enable commercial partners to tailor the analysis to their application by the addition of azide containing groups in order to improve analysis by FLD or MS prior to labelling. Further development could see the advancement of combinatorial soft matter neoglycolipid based array surfaces to study the interaction and movement of glycans over a fluid surface in response to an incoming protein. Glycan arrays have seen numerous advances in recent years including carbohydrates activated onto a solid support in a defined arrangement [15-19]. The development of these systems has provided improvements to both activation chemistry as well as array density and size [20, 21]. This has resulted in the development of microarray surfaces, enabling high throughput elucidation of glycan binding events when exposed to protein samples of interest. Interactions between particular carbohydrates are then used to deduce how interactions will translate into living systems. While array surfaces present a viable and powerful method of probing carbohydrate protein interaction, the application of multifunctional labels would enable the use of biologically significant carbohydrates to be surface activated post analysis, allowing precise glycan interactions to be mapped.

5.2 Future proposals

As part of the investigation into further applications of multifunctional label **25** for the downstream functionalisation of biologically derived carbohydrates. Preliminary work was carried out for the generation of a novel series of neo-glycoproteins. This work would therefore aim to elucidate the effect of glycan size on protein structure and function through the use of 'click' modification with carbohydrates purified though analytical workflows.

5.2.1 Steps towards the generation of novel glycoprotein conjugates by the use of downstream multifunctional glycan labelling techniques.

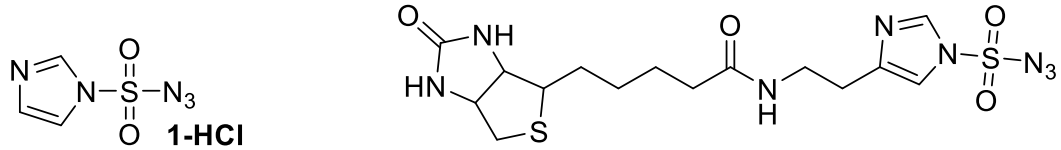
5.2.2 Background

The synthesis of therapeutic protein-based drugs has become increasingly effective in the treatment of a variety of both genetic and environmental disorders, with more than 100 monoclonal therapies currently approved for use within the United Kingdom [22, 23]. As previously stated, glycosylation of these drugs not only affects activity, but also safety, leading to glycosylation being defined as a

critical control attribute. Due to the complexity and size of these biomolecules, synthesis is carried out in a variety of mammalian cell lines. Previous research has shown the ability of prokaryotic cell lines to successfully produce non-glycosylated versions of mammalian proteins under simpler growth conditions at high yield. Primary protein structural synthesis has also been achieved though the use of liquid handling systems [24, 25] or by chemical ligation [26]. However, co-translational modifications, such as glycosylation, present a barrier to the synthesis of glycosylated higher mammalian proteins though chemical or simple cell line approaches.

One such option for the selective addition of modifications to protein structure is though the modification of amine containing amino acids, such as lysine, to azide containing derivatives. Azides have been shown to react in an orthogonal way when introduced to complex biological systems [27, 28]. This has significantly increased the viability of selective modification of protein structure. It is therefore unsurprising that the introduction of azides has facilitated the introduction of fluorophores, phosphates, carbohydrates and small molecule drugs by a range of techniques, including both copper catalysed alkyne-azide cycloadditions (CuAAC) [29, 30] and strain promoted alkyne-azide cycloadditions (SPAAC)[31].

The aforementioned applications have seen the addition of azide moieties into target protein molecules through a range of methods, utilising both chemical and biological techniques. Biologically driven techniques have focused on the introduction of azido modified lysine residues during translation. Initial attempts were carried out by the enrichment of culture media with azido lysine residues. These methods were limited due to the lack of differentiation between native and modified amino acid residues. This resulted in variations in frequency and location of azide modification within the sample generated [32]. Subsequent applications have seen the generation of genetically engineered azido protein conjugates bearing site specific modification though the use of stop codon suppression[33].



73 **89**

Figure 5.5 - Diazotransfer reagents **73** and **89** for the single site diazotransfer to protein structures

Chemical modification techniques have also previously achieved selective chemical modification through the use of protein specific biotin tethered diazotransfer reagent such as **89** (**Figure 5.5**). Subsequent work was also shown to be capable of achieving repeatable azide conversion of

N-terminus lysine residues by pH controlled diazotransfer, utilising imidazole based diazo transfer reagent **73** at 1.75 equivalents at pH 8 [34] (**Figure 5.5**). While these approaches provide a lower level of specificity in the formation of azide moieties, they serve as fast and efficient methods for the modification of proteins. In contrast to biological azido modification methods, the use of **73** negates the requirement for genetic engineering, enabling the formation of novel protein conjugates without the requirement for usage restrictions.

5.2.3 Aims and Objectives

The application of chemical diazotransfer therefore provides an attractive starting point for the generation of novel glycoprotein conjugates. This proposal therefore sets out to achieve four main aims. Initially this section of work of will aim to utilise workflows developed throughout the previous chapters enabling the derivatisation and collection of biologically relevant carbohydrates bearing multifunctional terminal alkyne functionality. This will be achieved though the enzymatic release and subsequent derivatisation of biologically derived carbohydrates with multifunctional label **25**. The project will then seek to optimise a method for the selective diazo transfer of a single azide to target protein molecules, building on the work of Schoffelen *et al*[34]. The study will then aim to optimise and achieve efficient CuAAC modification of azide functionalised protein targets with the anticipation of generating novel glycoconjugates though the conjugation of multifunctionally labelled carbohydrates. The final aim of this work will see activity testing of the novel glycoprotein conjugate against non-glycosylated analogues. This would aim to establish both pH and temperature stability though the use of 96 well plate based optical protein assays.

Herein, steps are presented towards the generation of a novel glycoconjugate. This is achieved through the use of aqueous diazo transfer to bovine pancreatic RNAse A combined with the utilisation of a novel multifunctional glycan label, enabling post analysis functionalisation of labelled carbohydrate structures to azide containing biologically derived structures.

5.2.4 Results and discussion

5.2.4.1 Collection of biologically derived carbohydrates.

Work carried out in the previous chapters saw the development and optimisation of a workflow for the derivatisation, optical analysis and collection of carbohydrate bearing reducing end modification in the form of a multifunctional fluorescent label. These workflows were therefore applied to the collection of **25** labelled GlcNAc following a 35 minute HILIC mode separation shown in **Figure 5.6**.

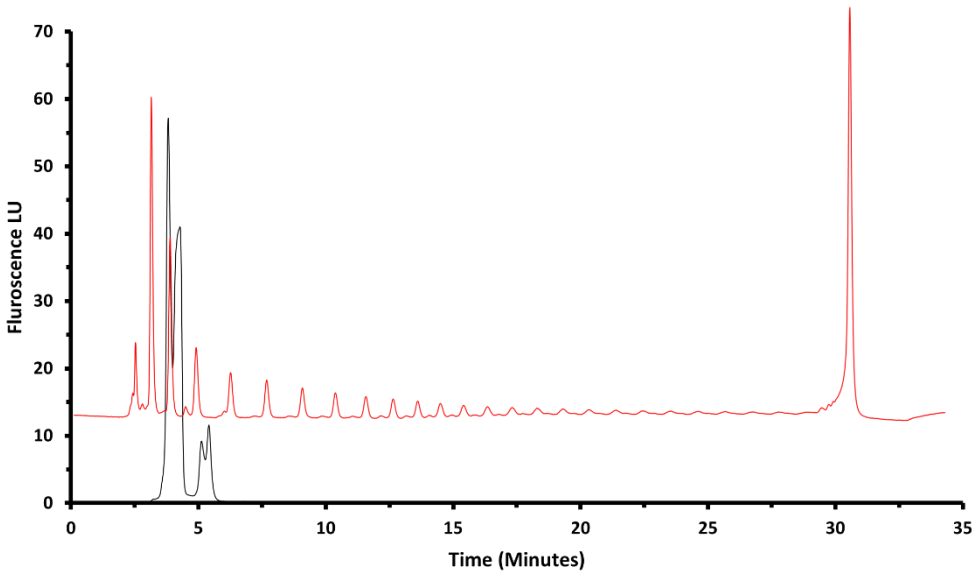


Figure 5.6 - Chromatographic traces for the separation of **25** labelled GU1 GlcNAc (black) compared to a **25** labelled dextran hydroxylate retention standard.

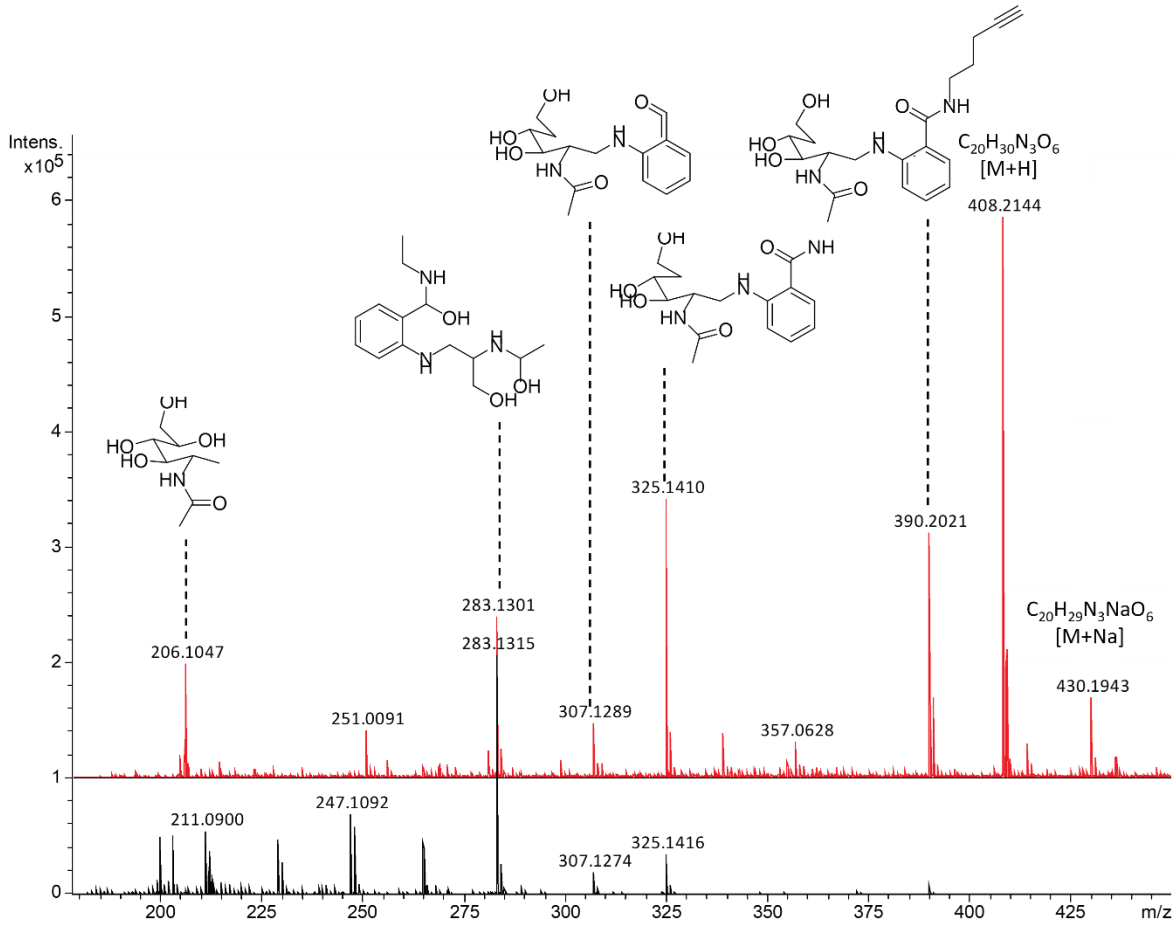


Figure 5.7- Fragmentation data for the 3.4 minute (red) and 4 minute (black) peaks collected from chromatographic separations of GlcNAc labelled with **25** shown in **Figure 5.6**.

The traces outlined in **Figure 5.6** outline the successful separation of GlcNAc derivatised with **25**. In these samples, a second unexpected peak was observed following separation by HILIC- mode LC. The separation was therefore profiled against an **25** labelled GHP retention standard. Minimal retention differences between GlcNAc and a single glucose ring provided challenges in differencing between peaks using a retention based standard. Both the 3.4 and 4 minute peaks were therefore dried before being resuspended and analysed by RP-LCMS-QToF (**Figure 5.7**). Both intact mass and fragmentation of the 3.4 minute peak yielded ions consistent with **25** labelled GlcNAc, with both [M+H] and [M+Na] observed at MS1. Both ions were analysed at MS2 with CID fragmentation at 35 eV, producing ions constant with both intact label and carbohydrate (**Figure 5.7**).

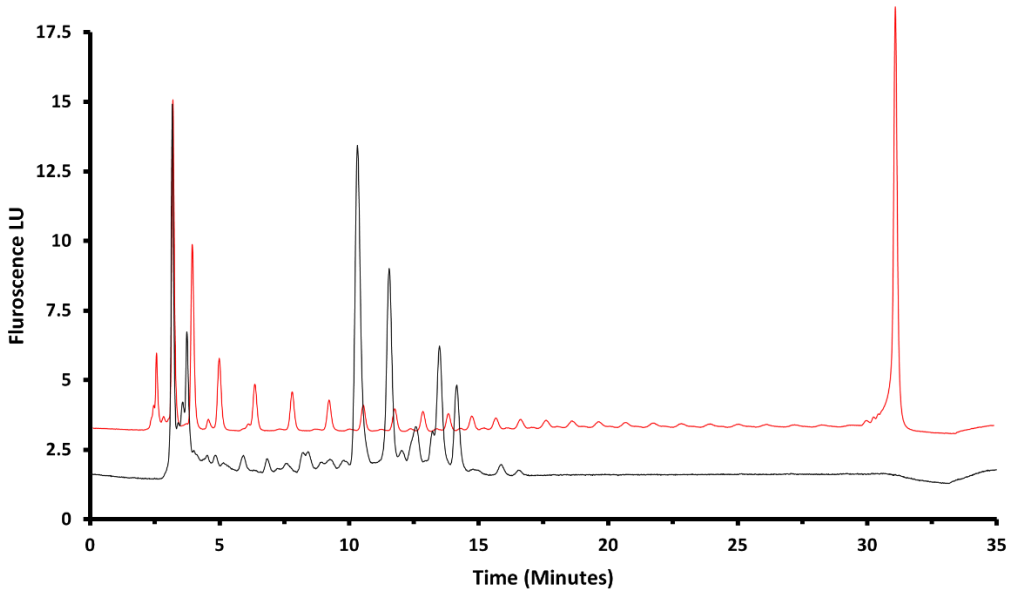


Figure 5.8 -Chromatographic traces for the separation of **25** labelled RNase B glycans (black) compared to a **25** labelled dextran hydroxylate retention standard.

Following the collection of multifunctionally labelled GU1 GlcNAc, the collection method was applied to enzymatically released RNase B glycans. In these samples, 1 mg of RNAse B was glycosylated by PNGase F overnight, before being labelled with **25**. Separations were then carried out by HILIC mode LC in line with methods developed for the separation of non-sialylated glycans in chapter 3. These extractions were then profiled against a GHP retention standard and literature traces, enabling the reliable collection of both Man-5 and Man-6 high mannose glycans bearing reducing end modification with multifunctional label **25** (**Figure 5.8**).

5.2.4.2 Optimisation of single-site diazo transfer to RNase A

Previous work by Schoffelen and Lohse [34, 35] has successfully achieved diazotransfer to multiple protein donors by the use of aqueous diazotransfer using **73**. This work was subsequently developed to increase selectivity using pH modification, resulting in a reduced frequency of azide introduction. As previously stated, it was speculated that the selectivity of pH mediated diazotransfer was likely caused by the specific reactivity of the individual amine. Amine reactivity can therefore be predicted though computational pKa modelling. This process aims to predict the conditions for amine dissociation in solution, resulting in an increased prospect for efficient diazotransfer.

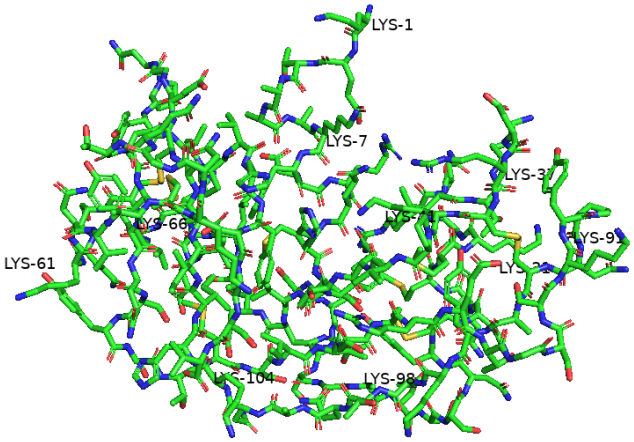


Figure 5.9 - 3D representation of Lysine residues present along the primary structure of Bovine RNase A

To facilitate the development of reliable conditions for the selective addition of a single azide to RNAse A, pKa values and solvent accessible surface area were computationally modelled by the use of PROPKA protein prediction software (**Table 5.1**) and PyMol (**Figure 5.9**) respectively [36].

Table 5.1 - Predicted pKa and solvent accessible surface area for lysine residues present in Bovine RNase A

Residue	pKa	Solvent accessible surface area (A^2)	Relative solvent accessible surface area RSA%, (%)
Lys-1	10.42	242.838	98
Lys-7	10.19	93.659	46
Lys-31	10.35	139.923	70

Lys-37	10.72	134.256	72
Lys-41	8.71	32.475	16
Lys-61	10.53	120.285	61
Lys-66	10.38	169.288	76
Lys-91	10.44	150.157	79
Lys-98	10.39	120.629	62
Lys- 104	10.32	73.735	40

The values in **Table 5.1** show predicted pKa and solvent accessibility values for each of the 10 Lysine residues present in RNase A. While these values suggest that all lysine residues are likely to be protonated at pH 7, Lys-41 exhibited the lowest pKa, indicating the lowest proportion of protonated amines. This would suggest increased amine reactivity of Lys-41 over the adjacent 9 lysine residues present in the proteins structure. However, it should be noted that with an RSA% of <20%, this residue could be considered buried within the protein structure. Significantly higher solvent exposures were seen for lysine residues at the *N*-terminus, making it difficult to selectively react at a specific amine. It was therefore anticipated that the attachment of amine residues to RNase A would primarily see the conversion of *N*-terminus lysine amines, rather than residues displaying lower pKa values.

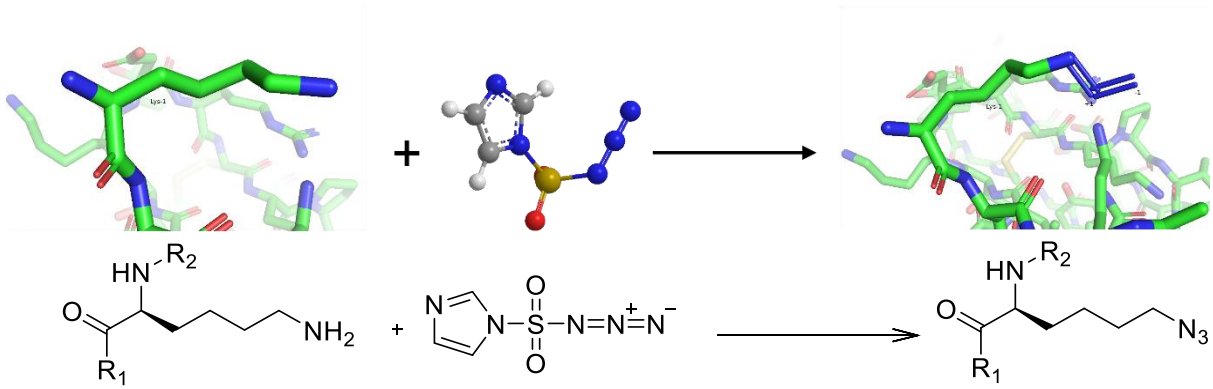


Figure 5.10 - 3D and skeletal diagrams of diazotransfer to Lysine residues present in protein primary structure using **73.**

The modification of individual Lysine residues with additional functionality in the form of a single azide moiety is outlined in **Figure 5.10**. The reaction sees the conversion of amine containing functional groups present on lysine residues to an azide though diazotransfer from **73**. To establish reliable conditions for single azide addition to RNaseA, a series of screening reactions were carried

out on 0.1 mg aliquots of purified RNaseA. These samples were incubated at 25 °C in buffered solutions ranging from pH 4 to pH 11, containing 1.5 eq of **73** per lysine residue. The pH range of these extractions therefore covers the range of primary amine dissociation constants for the 10 residues present in the proteins structure. Following overnight incubation, protein samples were washed with LCMS grade water and concentrated by centrifugal filtration before being diluted to 0.1 mg/mL for analysis. Quantification of diazotransfer was then carried out by RP-LC-QToF in positive ion mode over a 35-minute runtime.

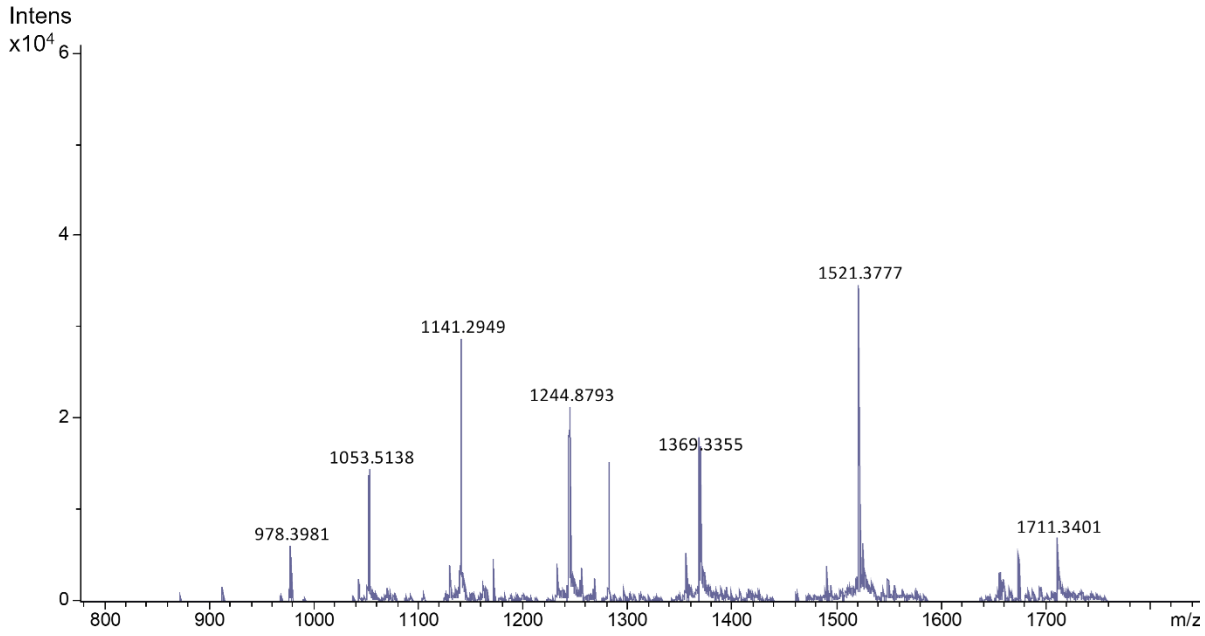


Figure 5.11 - ESI-QToF spectra showing multiple chare adducts following LC separation of native Bovine RNase A

Spectra were then recorded as multiple charge states, as seen in **Figure 5.11**, before being computationally deconvoluted by maximum entropy calculations to establish the intact mass. The recorded masses for samples 1-8 can be seen in **Table 5.2**, along with the corresponding number of azides present in the proteins structure. This was calculated based upon a mass increase of 26 Da over the native protein structure.

Table 5.2 - Reaction conditions for the screening of pH mediated diazotransfer between RNase A and **73**

Sample	pH	Protein	Diazotransfer reagent **73**	Conditions	Unmodified	Mass found	No of Azides
1	4	RNase A	1.75 eq per Amine	40 mM diethanol amine, Acidified with 0.1 M HCl	13,683	13682.8750	0
2	5	RNase A	1.75 eq per Amine	40 mM diethanol amine, Acidified with 0.1 M HCl	13,683	13682.8125 13707.0313	1
3	6	RNase A	1.75 eq per Amine	40 mM diethanol amine, Acidified with 0.1 M HCl	13,683	13682.8125, 13707.5000	1
4	7	RNase A	1.75 eq per Amine	40 mM diethanol amine, Acidified with 0.1 M HCl	13,683	13682.8125, 13707.5000, 13734.2188	2
5	8	RNase A	1.75 eq per Amine	40 mM diethanol amine	13,683	13682.8750, 13707.6250, 13734.0625	2
6	9	RNase A	1.75 eq per Amine	40 mM diethanol amine basified with 0.1 M Na_2CO_3	13,683	13812.8125, 13839.0625, 13865.3125, 13891.5625	9
7	10	RNase A	1.75 eq per Amine	40 mM diethanol amine basified with 0.1 M Na_2CO_3	13,683	13865.3125, 13888.9375, 13916.5000	10
8	11	RNase A	1.75 eq per Amine	40 mM diethanol amine basified with 0.1 M Na_2CO_3	13,683	13865.3125, 13890.2500, 13917.8125	10

The conditions outlined in **Table 5.2** show the relationship between increasingly basic reaction conditions and the number of amine residues converted to azides. This aligns with the previous work of Schoffelen [34] on the use of diazotransfer reagents for azide formation in chicken egg lysozyme. It should be noted however, that future work would be required to ascertain the location of azido

modification in RNaseA. This could be achieved though tryptic digest followed by sequencing by LCMS.

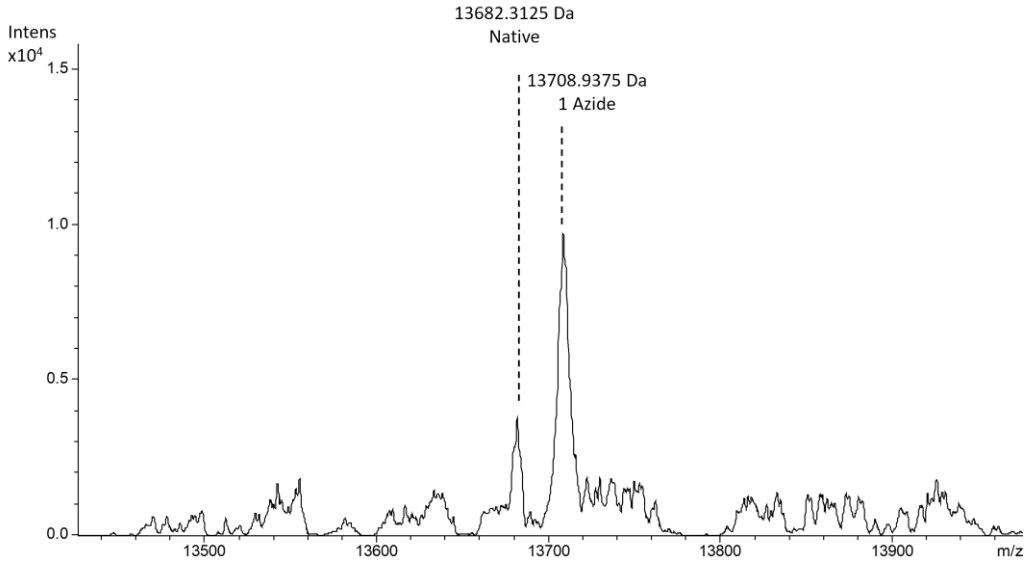


Figure 5.12 - Deconvoluted mass spectra showing single azide addition to RNase A.

The successful addition of a single azide is outlined in **Figure 5.12**. This could be further modified with collected multifunctionally labelled glycan structures through the use of CuAAC. The proposed work would see the generation of novel neo-glycoproteins. Due to the interaction of Cu(I) with biological materials, modification of azido protein would require the use of a copper chelating ligand such as **90**, which was synthesised as part of this work. The generation of these neo-glycoproteins could then be used to probe carbohydrate protein interactions as well as help to understand the impact of glycosylation on protein stability.

5.2.5 Experimental

5.2.5.1 General Experimental

All chemicals used were purchased through commercial sources (Sigma Merck, Fisher Scientific, Alfa Aeser) and used without further purification. Solvents were purchased commercially at HPLC grade or better and only LCMS grade solvents were used in the production of samples for analysis by mass spectrometry (Fisher Scientific). GlcNAc and RNase A and B were purchased commercially (Fisher Scientific).

5.2.5.2 Analytical Equipment

High performance liquid chromatography

Separation by HILIC-HPLC was performed on an Agilent 1100 HPLC system equipped with a G1321A FLD detector and an G1314A variable wavelength detector in accordance with the separation methods outlined in chapter 3.6.2. While collection was carried out in accordance with 4.5.2.

Mass Spectrometry

Both offline DI-MS and online LCMS were carried out in accordance with the protocols outlined in chapter 4.5.2 . Data acquisition was recorded in Qtof control (Bruker, USA) before being analysed in Compass Data analysis 4.3 (Bruker, USA). Maximum entropy spectral deconvolution was used to calculate deconvoluted protein masses using +H ions applied over a mass window of 10-20 KDa.

5.2.5.3 Steps towards the modification of proteins by pH mediated diazo transfer and CuAAC

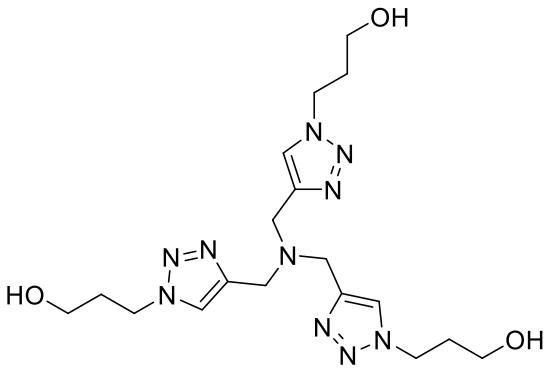
Labelling procedure for carbohydrate derived from model and biological sources

Carbohydrate sample (10 µl) were labelled according to the general labelling procedure outlined in chapter 3 experimental section 3.6.8. Samples were then incubated at 65 ^{o}C for 3 hours before being quenched in with ACN (950 µL). Excess labelling reagents were then removed by SPE-ed Amide SPE (Applied separations, USA) before being eluted with H_2O (600 µL) and dried under reduced pressure. These samples were then reconstituted in 100 µl of 50% aqueous methanol for analysis.

General procedure for diazotransfer to Bovine RNase A

Samples of RNase A (10 µl, 0.1 mg) were added to 1.5 ml centrifuge tubes containing 20 µl of 40 mM diethanol amine buffered to the required pH. Diazotransfer reagent **73** was then added in a solution of aqueous sodium carbonate at pH 7 (20 µl, 0.061 mMol) equivalent to 1.75 eq. Samples were then shaken overnight before being purified by centrifugal filtration. Samples were diluted to a final volume of 500 µl by the addition of 450 µl of HPLC pure water before being transfered to 1.5 ml 3 KDa centrifugal filtration units (Amicon Ultra, MerckMillipore, USA). Samples were then concentrated at 8000 g for 15 minutes before flow through was discarded and the cartridge washed with a further 1 ml of H_2O (2x 500 µl). Purified protein was then removed from the membrane and diluted to 0.1 mg/ml for analysis by LCMS.

Synthesis of 3,3',3''-((nitrilotris(methylene))tris(1H-1,2,3-triazole-4,1-diyl))tris(propan-1-ol) 90 [37]



3-Bromo propanol **91** (2.06 g, 15 mMol) was added to a round bottom flask along containing a solution of 50% aqueous acetonitrile. Sodium azide (1.4 g, 21.8 mM, 1.5 eq) was then added dropwise in a solution of in 50% aqueous acetonitrile. The solution was stirred at room temperature overnight after which point the reaction mixture was diluted with H_2O (5 ml) and transferred to a separating funnel. The product was then extracted with diethyl ether (2 x 20 ml) and the combined organic layers were washed with $NaHCO_3$. The organic phase was then filtered and concentrated to ~2.5 ml and used without further purification.

Tripropargyl amine **81** (0.49 g 3.3 mmol) was added to a round bottom flask and stirred with 2,6 lutidine (0.35 ml, 3.3 mmol). The flask was flushed with nitrogen and dry ACN (5 ml) added. To this solution 3 Azido propanol (2.5 ml, ~12 mmol) was added and the reaction mixture was stirred for 5 minutes. Copper(I) Iodide was then added (17 mg, 0.09 mmol) and the reaction mixture was stirred for 36 hours at room temperature. Following reaction completion by TLC analysis the reaction mixture was filtered and concentrated under reduced pressure. The yellow oil generated was purified by flash column chromatography yielding a pale yellow oil which crystalised to a white solid upon cooling to 0 ^oC. Characterisation by ^1H NMR verified presence of a new three proton singlet at 7.83 ppm indicating the formation of a trizole ring and the formation of **90**. (1.64 g 72% yield).

^1H NMR (400 MHz, D_2O) δ 7.83 (s, 3H 3 x CH), 4.40 (t, *J* = 7.0 Hz, 6H 3 x CH_2), 3.71 (s, *J* = 12 Hz, 6H 3 x CH_2), 3.47 (t, *J* = 6 Hz, 6H 3 x CH_2), 2.02 (q, *J* = 6.5 Hz, 6H 3 x CH_2). **^{13}C NMR (101 MHz, D_2O)** δ 143.22(3 x NCH-**C**-CH_2), 125.22 (3 x N-**C**H-C), 58.11(3 x HO-**C**H_2-CH_2), 47.50(3 x N-**C**H_2-C), 47.17 (3 x N-**C**H_2-CH_2), 31.78 (3 x **C**H-**C**H_2-CH_2). **HRMS** ESI+ve $C_{18}H_{30}N_{10}O_3$ expected[M+H] 519.2423 found 519.2423.

5.2.6 References

1. M. Allen, Doctor of Philosophy, The Open University, 2022.
2. S. Yang and H. Zhang, *Proteomics Clin. Appl.*, 2012, **6**, 596-608.
3. Q. Zhang, H. Li, X. Feng, B.-F. Liu and X. Liu, *Plos One*, 2014, **9**, e94232.
4. S. Zhou, L. Veillon, X. Dong, Y. Huang and Y. Mechref, *Analyst*, 2017, **142**, 4446-4455.
5. M. A. Lauber, Y.-Q. Yu, D. W. Brousmiche, Z. Hua, S. M. Koza, P. Magnelli, E. Guthrie, C. H. Taron and K. J. Fountain, *Anal. Chem.*, 2015, **87**, 5401-5409.
6. T. Keser, T. Pavić, G. Lauc and O. Gornik, *Front. Chem.*, 2018, **6**, 324-324.
7. Z. Segu, T. Stone, C. Berdugo, A. Roberts, E. Doud and Y. Li, *MAbs*, 2020, **12**, 1750794.
8. L. Royle, in *Liquid Chromatography*, Elsevier, 2017, pp. 183-200.
9. E. D. Goddard-Borger and R. V. Stick, *Org. Lett.*, 2007, **9**, 3797-3800.
10. Z. U. Khan, B. Nay, E. F. V. Scriven and H. Suschitzky, *J. Chem. Soc.,* 1982, 671-672.
11. H. B. R. Kramer, D.Phil, University of Oxford, 2006.
12. S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nature*, 2007, **446**, 1105-1109.
13. T. R. Chan, R. Hilgraf, K. B. Sharpless and V. V. Fokin, *Org. Lett.*, 2004, **6**, 2853-2855.
14. S. H. Walker, L. M. Lilley, M. F. Enamorado, D. L. Comins and D. C. Muddiman, *J. Am. Soc. Mass Spectr.*, 2011, **22**, 1309-1317.
15. C. Y. Wu, P. H. Liang and C. H. Wong, *Org. Biomol. Chem.*, 2009, **7**, 2247-2254.
16. C. M. Arthur, R. D. Cummings and S. R. Stowell, *Curr. Opin. Chem. Biol.*, 2014, **18**, 55-61.
17. L. Wang, R. D. Cummings, D. F. Smith, M. Huflejt, C. T. Campbell, J. C. Gildersleeve, J. Q. Gerlach, M. Kilcoyne, L. Joshi, S. Serna, N.-C. Reichardt, N. Parera Pera, R. J. Pieters, W. Eng and L. K. Mahal, *Glycobiology*, 2014, **24**, 507-517.
18. J. G. Briard, H. Jiang, K. W. Moremen, M. S. Macauley and P. Wu, *Nat. Commun.*, 2018, **9**, 1-11.
19. A. Geissner, A. Reinhardt, C. Rademacher, T. Johannssen, J. Monteiro, B. Lepenies, M. Thépaut, F. Fieschi, J. Mrázková, M. Wimmerova, F. Schuhmacher, S. Götze, D. Grünstein, X. Guo, H. S. Hahm, J. Kandasamy, D. Leonori, C. E. Martin, S. G. Parameswarappa, S. Pasari, M. K. Schlegel, H. Tanaka, G. Xiao, Y. Yang, C. L. Pereira, C. Anish and P. H. Seeberger, *P. Natl. Acad. Sci. U.S.A*, 2019, **116**, 1958-1967.
20. A. Shirakawa, Y. Manabe and K. Fukase, *Molecules*, 2021, **26**, 1040.
21. J. P. Merino, S. Serna, A. Criado, A. Centeno, I. Napal, J. Calvo, A. Zurutuza, N. Reichardt and M. Prato, *2D Mater.*, 2020, **7**, 024003.
22. A. Chenoweth and S. Crescioli, The Antibody Society. Therapeutic monoclonal antibodies approved or in regulatory review., https://www.antibodysociety.org/antibody-therapeutics-product-data/, (accessed 27/12/2022, 2022).
23. H. Kaplon, S. Crescioli, A. Chenoweth, J. Visweswaraiah and J. M. Reichert, *mAbs*, 2023, **15**, 2153410.
24. G. B. Fields, *Current Protocols in Protein Science*, 2002, **Chapter 18**, 18.11.11-18.11.19.
25. V. Mäde, S. Els-Heindl and A. G. Beck-Sickinger, *J. Org. Chem.*, 2014, **10**, 1197-1212.
26. S. Wang, Y. A. Thopate, Q. Zhou and P. Wang, *Chinese J. Chem. Phys*, 2019, **37**, 1181-1193.
27. M. L. W. J. Smeenk, J. Agramunt and K. M. Bonger, *Curr. Opin. Chem. Biol.,* 2021, **60**, 79-88.
28. K. Porte, M. Riomet, C. Figliola, D. Audisio and F. Taran, *Chem. Rev.*, 2021, **121**, 6718-6743.
29. S. Presolski, *Methods Mol. Biol.*, 2018, **1798**, 187-193.
30. N. Inoue, A. Onoda and T. Hayashi, *Bioconjugugate Chem.*, 2019, **30**, 2427-2434.
31. J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzi, *P. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 16793-16797.
32. E. Kaya, K. Gutsmiedl, M. Vrabel, M. Müller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858-2861.
33. Z. A. Wang, Y. Kurra, X. Wang, Y. Zeng, Y.-J. Lee, V. Sharma, H. Lin, S. Y. Dai and W. R. Liu, *Angew. Chem. Int. Edit.*, 2017, **56**, 1643-1647.

34. S. Schoffelen, M. B. van Eldijk, B. Rooijakkers, R. Raijmakers, A. J. R. Heck and J. C. M. van Hest, *Chem. Sci.*, 2011, **2**, 701-705.
35. J. Lohse, L. J. Y. M. Swier, R. C. Oudshoorn, G. Médard, B. Kuster, D.-J. Slotboom and M. D. Witte, *Bioconjugate Chem.*, 2017, **28**, 913-917.
36. E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon and N. A. Baker, *Protein Sci.*, 2018, **27**, 112-128.
37. P. Chen, C. Li, D. Liu and Z. Li, *Macromolecules*, 2012, **45**, 9579-9584.