Exploring novel chemical and enzymatic labelling approaches in metabolic oligosaccharide engineering of mammalian cells

Jenny Alex Hayes

Doctor of Philosophy

University of York

Chemistry

September 2021

Abstract

Glycoproteins have many important roles in cell signalling and diseases including cancer, neurodegeneration, and pathogenic infections. Metabolic oligosaccharide engineering has previously been used to incorporate unnatural sugar derivatives into glycoproteins and detect them with bioorthogonal labelling. Additional methods of labelling that work orthogonally to current methods, and are reversible, would allow the simultaneous investigation of multiple glycoproteins and their properties.

Sortase is a transpeptidase found in Gram-positive bacteria that has been previously engineered to perform ligations of glycine- and LPXTG- bearing biomolecules. This project sought to use sortase to specifically label functional groups carried by sugar derivatives incorporated into glycoproteins. Initial incorporation of sugar derivatives bearing glycine-azide showed successful labelling with click chemistry. Subsequent attempts to label incorporated sugar derivatives bearing glycine-amine with the commercial sortase Srt5M failed. An alternative mannosamine-thiazolidine sugar was tolerated in HEK cells up to 300 μM. Live cells that had incorporated this sugar were successfully labelled using the organocatalyst-mediated protein aldol ligation. Labelling of intracellular glycoproteins was also tested but requires further optimisation.

To expand the currently available sortases, 22 novel sortases were cloned and expressed from a metagenome database at the industrial partner Prozomix. Testing for the ability to ligate a range of peptides *in vitro* revealed two functioning enzymes, which were named Srt021 (class C) and Srt025 (class D). The two sortases were shown to function well at physiological temperature and pH, and in a range of buffers including without calcium. This represents the second class C sortase made to function *in vitro*, and the first class D sortase *in vitro*. Further characterisation and directed evolution of these two sortases should result in a wider range of tools for specific labelling of biomolecules.

Contents

List of tables

List of figures

Acknowledgements

First, thank you to Dr Martin Fascione for everything. For accepting me onto this PhD, all the guidance and supervision you have given, the enthusiastic discussions of interesting results and for pushing me to do better when things weren't going well. To Dr Nathalie Signoret for initially agreeing to let me to do some cell culture at your lab and getting so involved with my experiments that you became another supervisor. To Prof Gideon Grogan for your insightful feedback in meetings and reports. To Dr Alison Parkin and Dr Jamie Blaza for useful discussions at TAP meetings.

Thanks are due to all of the MAF-AP-LW lab. To Julia Walton for being a fantastic technician, ensuring that the lab was not overrun with old gels or RB flasks, and for showing me how to use (and fix) so many pieces of lab equipment. To the original Fascione Four for leading the way – Robin for many interesting discussions on science and other topics, Richard for help with the LC-MS demon machine, Harriet for being so kind and supportive at my lowest points, and Emily for showing me the ropes of how to teach undergrads and for entertaining anecdotes from York's nightlife. To Darshita for helping with my initial sugar synthesis troubles, patiently explaining NMR, and baking the best cakes ever. To Mandy for help with making biological probes, Nick for all the jokes and banter, and Tasha for advice on synthesis. To Mohamed for being my M. Chem. student, thank you for all you taught me about my own project through forcing me to explain it clearly to another person. To many other past and present members – Lewis, Mark, Tessa, Tom, Kate, Sol, Lloyd, Angelo, Saeed, Quentin, thank you for being generally wonderful.

To the staff at Chemistry for all their help – Karl Heaton for running mass specs, Heather Fish for running a tight NMR ship and Dr Ed Bergstrom for the many times you have patiently coaxed the LC-MS into working. To the glass-blowing, mechanical and electronics workshops for fixing many broken things and without whom I would be surrounded by rubble.

To the people at Prozomix, thank you for a fantastic 3-month placement. To Simon and Ruth for taking me on and helping the sortase part of the project really shine, and to Carl, Grace, Sam, Jesmine, Darren, Jim and others for making me so welcome.

To my friends Zoe and Joe for many fun nights of dinner, gaming, and cat cuddles. Thank you for keeping me sane during the long road of our studies. To my karate teachers Giacomo, Alice, Dylan, and Sensei Ian for many stress-busting karate sessions. To Tim and Ollie and my other Pokémon Go York raiding buddies for all the virtual adventures when the real world was too much work to deal with.

Finally, thank you to my siblings Heather and Max, for being the best anyone could ask for. Thank you to my parents Chrissy Hayes for teaching me to read and Dave Hayes for being my first science teacher. This thesis is dedicated to you for all your love, support, and belief over the years.

Author's Declaration

I, Jenny Alex Hayes, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. The work submitted for this thesis is my own. The contributions made by others are explicitly stated below. I can confirm that appropriate credit has been given within this thesis where reference has been made to work carried out by others, and I am most grateful to the colleagues and friends below for their assistance.

Chapter 2:

Robin Brabham synthesised the Boc-Thz precursor.

Chapter 3:

Gintare Bucaite synthesised depsispeptide probe.

Amanda Noble synthesised the *N*-terminal precursor for OPAL probe.

Chris Taylor performed MALDI-MS/MS on the protein bands.

Chapter 4:

Mohamed Abd El Bari performed sortase reactions testing Srt021 and Srt025 under varying pH, temperature, and buffers.

Wendy Robinson set up the sortase crystal trays.

Chapter 1 : Introduction

Glycoprotein overview

Glycoproteins are created when proteins are modified by the addition of oligosaccharide chains (glycans). This adds bulky, hydrophilic structures that can assist in folding and function as selective motifs for recognition [1]. There are many monosaccharides found in glycans, and some of the most common appear in all eukaryotic organisms. The standard symbols and abbreviations of important sugars are shown below (Figure 1.1). *N*-acetylglucosamine (GlcNAc) and D-mannose (Man) are found in all *N*-linked glycoproteins in the common core structure. The Asn residue always has a GlcNAc attached directly to it [1]. Other common sugars in glycoproteins are D-glucose (Glc), D-galactose (Gal), L-fucose (Fuc), *N*-acetylneuraminic acid (Neu5Ac, or sialic acid), D-xylose (Xyl) and *N*-acetylgalactosamine (GalNAc) [2]. *N*-acetylmannosamine (ManNAc) is an important sugar in Metabolic Oligosaccharide Engineering (MOE), because it is the first committed step in the biosynthesis of sialic acid [3].

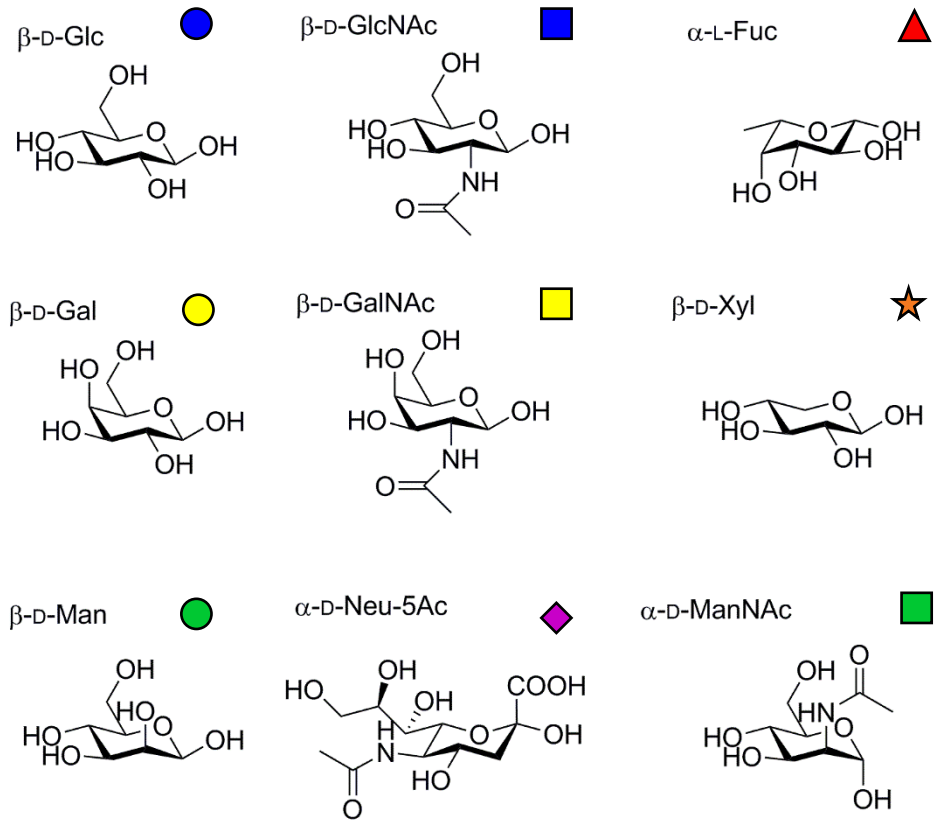


Figure 1.1: Sugars found in glycoproteins. Glc, Gal and Man are hexose sugars. In GlcNAc, GalNAc and ManNAc the 2-hydroxyl group is substituted with an acetylated amino group. Fuc is also a hexose and is present in the L configuration. Xyl is a pentose. Neu5Ac is the most common mammalian sialic acid and is a nine-carbon sugar acid [1].

There are two main types of glycosylation (Figure 1.2), classified according to the type of atom that binds the saccharide [2]. *N*-glycosylation occurs at asparagine residues and requires a consensus sequence of Asn-X-Ser/Thr, with X being any amino acid except Pro due to the restricted secondary structure. The β-hydroxyl group of the Ser/Thr residue hydrogen bonds the side-chain nitrogen of Asn, making it more nucleophilic and allowing attack on a glycan chain [1]. All *N*-linked glycans begin with an *N*-acetylglucosamine residue (GlcNAc) (Figure 1.2a) and also contain mannose (Man) in their core structure [1]. The initial *N*-glycosylation reaction adds a preformed branched oligosaccharyl-lipid, $Glc_3Man_9GlcNAc_2$-PP-dolichol, to a nascent protein by an oligosaccharyltransferase (OSTase) complex [4]. In fungi, the three glucose residues are cleaved by glucosidase enzymes and the chain extended to yield high-mannose glycans [5]. In mammals, mannosidase enzymes cleave mannose residues to leave the first five

attached sugars, $Man_3GlcNAc_2$, which are a conserved core in complex *N*-linked glycans [6]. *O*-glycosylation usually occurs at the hydroxyl group of Ser and Thr residues, though also occurs at hydroxylysine in collagen [6] and hydroxyproline in plant cells [7]. *O*-glycosylation usually adds a shorter and simpler oligosaccharide to the protein [2]. Unlike *N*-glycans, *O*-glycans do not have a consensus core sequence [1]. A frequent modification in vertebrates is the addition of an *O*-linked *N*-acetylgalactosamine (GalNAc) to Ser/Thr (Figure 1.2b) [6].

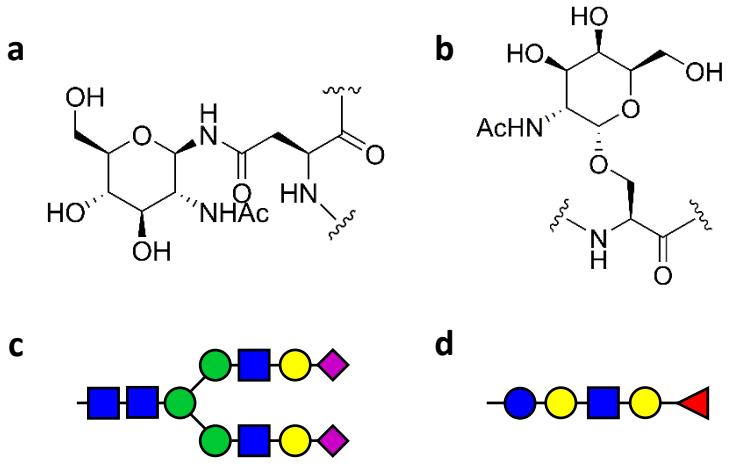


Figure 1.2: Examples of *N*-linked and *O*-linked glycans. (a) GlcNAc linked to asparagine in *N*-linked glycosylation. (b) GalNAc linked to serine in *O*-linked glycosylation. (c) An example of a complex *N*-linked glycan. The first two GlcNAc and three Man residues make up the core structure common to all *N*-glycans. (d) An example of an *O*-linked glycan, the blood group O antigen.

Proteins on the cell surface can be studied by simply incorporating a protein tag into their genetic sequence, such as the gene for green fluorescent protein (GFP) [8], or SpyTag, which covalently binds its partner SpyCatcher [9]. However, glycans are indirectly coded into the genome and hence cannot be labelled in this manner [10]. The glycans present in a cell depend on the sugar substrates available, the expression of glycosyltransferase enzymes that perform biosynthesis, and the expression of glycosidase enzymes, which cleave glycosidic bonds [11]. The glycosidic linkages between monosaccharides exist in multiple configurations, as there are multiple carbons able to form the bond and each position can take α or β configuration [10]. A

different enzyme is therefore required for each type of glycosidic bond, and multiple bonds are possible so unlike peptide chains, oligosaccharide chains can be branched [1]. The expression levels of different enzymes may vary, the enzymes may compete for substrates, and biosynthesis products may be exported from the Golgi apparatus where the enzymes are located [10]. Therefore, the glycan products are micro-heterogeneous, varying in response to the environment even on the same genetically encoded polypeptide chain [6]. Studying glycans is also made more difficult by the chemical similarity of the sugars [1].

Glycoproteins in cancer

The labelling and study of glycans is important because glycoproteins are involved in cell adhesion and signalling, and are therefore important in areas such as immunology and cancer metastasis [11] [12]. Cancer cells have abnormal metabolism and show altered expression of surface glycoproteins [13]. The Warren-Glick phenomenon describes how the presence of larger *N*-linked sugar chains on the surface of cells correlates with the ability of the cells to metastasize [14]. Overexpression of hypersialylated glycoproteins in particular is associated with metastasis, with a fluorinated sialic acid analogue that inhibits sialylation causing a decrease in growth and migration of tumour cells [15]. Sialyl Lewis^X is a tetrasaccharide that binds selectins on endothelial cells and regulates the extravasation of leukocytes during inflammation [16]. However, expression on cancer cells enables cell migration and metastasis [17] (Figure 1.3). Increased sialyl Lewis^X is associated with poor prognosis in prostate cancer [18], and is thought to be caused by a combination of impaired sulfation and hypoxic conditions [19].

Tumour cells also express various unusual sialic acids. The onco-foetal sialyl-Tn antigen is expressed in over 80% of cancers and is linked to increased growth in some cancer types [20]. Neu5Gc is a hydroxylated derivative of Neu5Ac that is abundant in mammalian cells, but absent in humans due to a deletion in the cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH) gene [21]. It is hypothesised this mutation was advantageous because it prevented the binding of malaria parasites to erythrocytes [22]. Neu5Gc is still present in humans in small amounts due to incorporation from dietary sources, mainly red meat [23]. It is incorporated more readily into hypersialylated cancer cells, and this causes inflammation and immune activation [24]. Tumour-specific glycans are being

investigated as targets for monoclonal antibody therapy [25]. MOE-based labelling methods able to differentiate between hypersialylated glycans and normal glycans could potentially aid cancer diagnosis.

Sialidase enzymes have also been suggested as therapeutic targets, with at least three types known in mammalian cells that have different effects on tumour progression [26]. Overexpression of the lysosomal-type sialidase Neu1, in murine melanoma cells, causes an increase in desialylated glycoproteins and results in decreased growth and metastasis [27]. In contrast, expression of the plasma-membrane associated sialidase Neu3 is increased in human colon cancer cells and this inhibits apoptosis [28].

GlcNAc transferase V increases branching on glycans by addition of β1,6-linked GlcNAc, which allows the formation of tri- and tetra-antennary glycans [29]. GlcNAc transferase V gives a transformed phenotype when overexpressed in cell culture [30]. It is suggested the branched glycans decrease cell-cell interactions and allow cells to detach from the primary tumour and enter the circulation [31], as mice deficient in this enzyme have suppressed tumour growth [29]. Conversely, overexpression of GlcNAc transferase III, which competes for the same acceptor, decreases metastasis in melanoma cells [32]. Cancer cells also undergo mutations to evade the immune system, for example in overexpression of siglec ligands, which are recognised by the sialic acid-binding immunoglobulin-type lectins on natural killer cells as “self” and prevent immune attack [23]. Core2 *O*-glycan expression also improves immune evasion and correlates with metastasis, whereas core3 and core4 *O*-glycans are downregulated in migrating tumour cells [33].

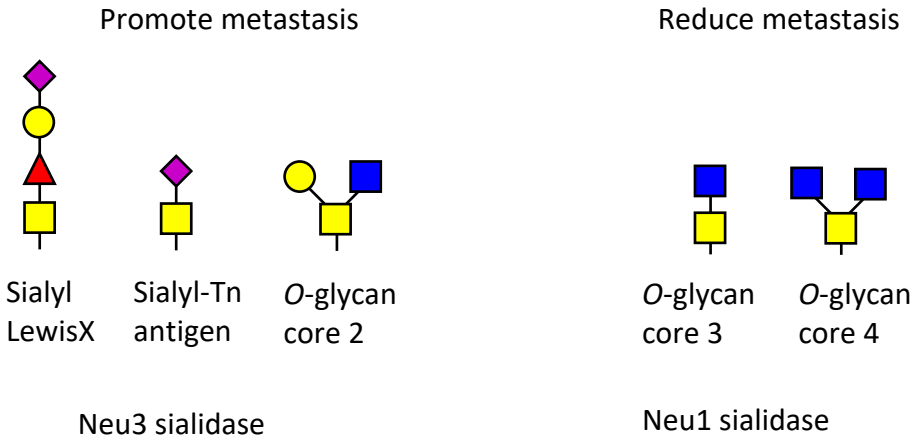


Figure 1.3: A summary of glycans and associated enzymes that are associated with the promotion or reduction of metastasis in tumour cells.

Glycoproteins in other diseases

Glycans have been linked to other diseases. *O*-GlcNAcylation is a dynamic protein modification in the nucleus and cytoplasm, where a single *O*-GlcNAc is added to Ser/Thr sites by *O*-GlcNAc transferase and removed by *O*-GlcNAcase [34]. This is similar to the activation or inhibition of proteins by phosphorylation, and can involve the same Ser/Thr sites [35]. This mechanism may be involved in neurodegenerative diseases such as Alzheimer's, where hyperphosphorylated Tau protein tangles form in the neurons [36]. Tau is also extensively *O*-glycosylated, so it is possible that a lack of *O*-glycosylation results in the hyperphosphorylation of Tau and subsequent formation of Tau tangles associated with the disease [37].

Glycoproteins are also involved in interactions between pathogens and the immune system. Bacteria coat their surface in sugars to hide from the immune system [38], such as the use by *Staphylococcus aureus* of wall teichoic acids to prevent antibody binding to cell wall epitopes [39]. *Streptococcus pyogenes* is coated in a hyaluronic acid capsule, which is required for the bacterium to change to a hyperinvasive phenotype [40]. *Neisseria meningitidis* synthesises Neu5Ac and conjugates it to surface components in order to mimic host cells and evade the immune system [41]. *Haemophilus influenza* likewise uses Neu5Ac to mimic host cells, but sources its Neu5Ac by cleaving it from host cells [42]. This mimicry technique is also used by non-pathogenic bacteria. *Bacteroides fragilis* is an intestinal symbiont that cleaves fucose from host cells and incorporates it into surface glycoproteins to prevent attack by the immune system [43]. As different species can favour different sugars for immune evasion, a method for specifically recognising individual types of sugar could have applications in diagnosing bacterial infections. Genomic profiling would also be able to identify the species, but it is still advantageous to know which sugars are currently being expressed by the cell to evade the immune system. In the case of bacteria that incorporate sugars from host cells, it would be more difficult to use genomics to deduce which sugars the bacteria are using for immune evasion.

Bacterial sugars on the cell surface can help cells control the immune system, for example *S. pyogenes* can bind complement factor H to sialic acids on its surface, therefore protecting the bacterial cells from the complement system [44]. Bacteria can also respond to host glycans. *Streptococcus pneumoniae* upregulates the sialidase NanA in response to Neu5Ac compared with Neu5Gc, therefore detecting if the host is human, and switches to a more virulent phenotype [45]. *S. aureus* expresses Staphylococcal Protein A, which binds the Fc portion of antibodies and prevents opsonisation [46].

Influenza viruses are classified according to two surface proteins that are both clinical targets. Hemagglutinin is a glycoprotein that binds sialic acid on human cells in an α-2-6 link with galactose, allowing the virus to attach to cells and insert its genome. Avian flu binds sialic acid in an α-2-3 link with galactose, which is why it does not easily transmit to humans [47]. Umifenovir is a drug that inhibits this binding [48]. Neuraminidase cleaves sialic acid from an infected cell, releasing viral progeny into circulation. This is inhibited by oseltamivir [49]. Heparin is another clinically important glycan. It binds antithrombin and enhances its inhibition of the coagulation cascade, and is therefore used as a blood thinning agent [50].

Metabolic Oligosaccharide Engineering

Incorporation of sugars

Metabolic oligosaccharide engineering (MOE) is a process whereby unnatural sugar precursors are fed to cells in order to alter cell surface glycoproteins [51]. The aim is to incorporate small noncanonical functional groups that can be recognized by a chemical probe into glycoproteins [52]. This does not require genetic modification, unlike the incorporation of unnatural amino acids (UAAs). UAAs typically require cells to be transfected with plasmids containing tRNA and an evolved aminoacyl-tRNA synthetase, as the naturally-occurring aminoacyl-tRNA synthetase enzymes are very stringent [53]. The sugars are modified with functional groups suitable for bioorthogonal reactions. The functional group is not found in living systems, and must not react with the other components of the cell, but reacts specifically with a chemical probe under physiological conditions [54].

For successful MOE, the monosaccharides must be imported into the cell, be enzymatically converted to the nucleotide sugar donor, then added to glycoproteins by

glycosyltransferases. At each step, the proteins involved must tolerate the functional group that has been added to the structure of the usual substrate. In eukaryotes, there are 2 classes of sugar transporter proteins: Glucose transporters (GLUTs) and sodium-glucose symporters (SGLTs). GLUTs are transporter proteins that facilitate diffusion across the membrane gradient. These enzymes have varying K_m values for glucose uptake. GLUT1 constantly transports glucose into cells, as its K_m of 1 mM is less than normal serum glucose level (4-8 mM). GLUT2, present in the liver and pancreas, has a K_m between 15-20 mM, so only transports glucose when the levels in the blood are higher than normal. This allows the pancreas to sense high glucose levels and release insulin, and the liver to take up and store excess glucose [55]. GLUT2 is known to also transport glucosamine efficiently [56], although the specificity of most GLUT transporters has not been fully characterized. SGLTs are energy-dependent transporters. SGLT-1 transports glucose from the gut lumen, and has a K_m less than 1 mM [55]. The SGLT proteins have previously been shown to tolerate unnatural moieties such as fluorine. The isoforms SGLT-2 and SGLT-3 also transport galactose, although have a higher affinity for glucose, and the SGLT-4 isoform accepts mannose [57]. It is thought that mannose is normally taken up through the glucose transporter [58]. The tolerance of some SGLT and GLUT proteins for different sugars means they likely also accept modified sugars during MOE. However, feeding unnatural sugar precursors can inhibit cell surface glycosylation, so incorporation is not always successful [51]. In addition, the modified sugars are usually added at concentrations around 100 μM, an order of magnitude lower than the K_m of many transporter proteins, so uptake is not rapid.

Once the sugars have been transported into the cell, they must be converted to the nucleotide sugar donors. *N*-acetyl mannosamine (ManNAc) phosphorylation is the first committed step in the biosynthesis of sialic acid [59]. When unnatural ManNAc analogues such as *N*-levulinoyl mannosamine (ManLev) are added to mammalian cells the downstream enzymes of the Roseman-Warren pathway are sufficiently promiscuous to accept the analogues as substrates (Figure 1.4) [60]. These enzymes include *N*-acetylmannosamine kinase (GNE), *N*-acylneuraminate-9-phosphate synthase (NANS), *N*-acylneuraminate-9-phosphatase (NANP), and *N*-acylneuraminate cytidylyltransferase (CMAS). It is possible that these enzymes do not require a high

degree of selectivity because sialic acid is the only nine-carbon sugar in mammalian cells [11]. The levulinoyl group therefore ends up on the C-5 of Neu5Ac residues at the terminus of *N*-linked glycans, and can be detected by chemoselective ligation of a hydrazide fluorescent marker to the ketone-coated cells [60]. Certain mannosamine analogues cannot be incorporated, such as replacing the entire *N*-acetamido group with an azide directly attached to C-2, which prevents the enzymes of the sialic acid biosynthesis pathway accepting the ManNAc analogue and converting it to sialic acid [61].

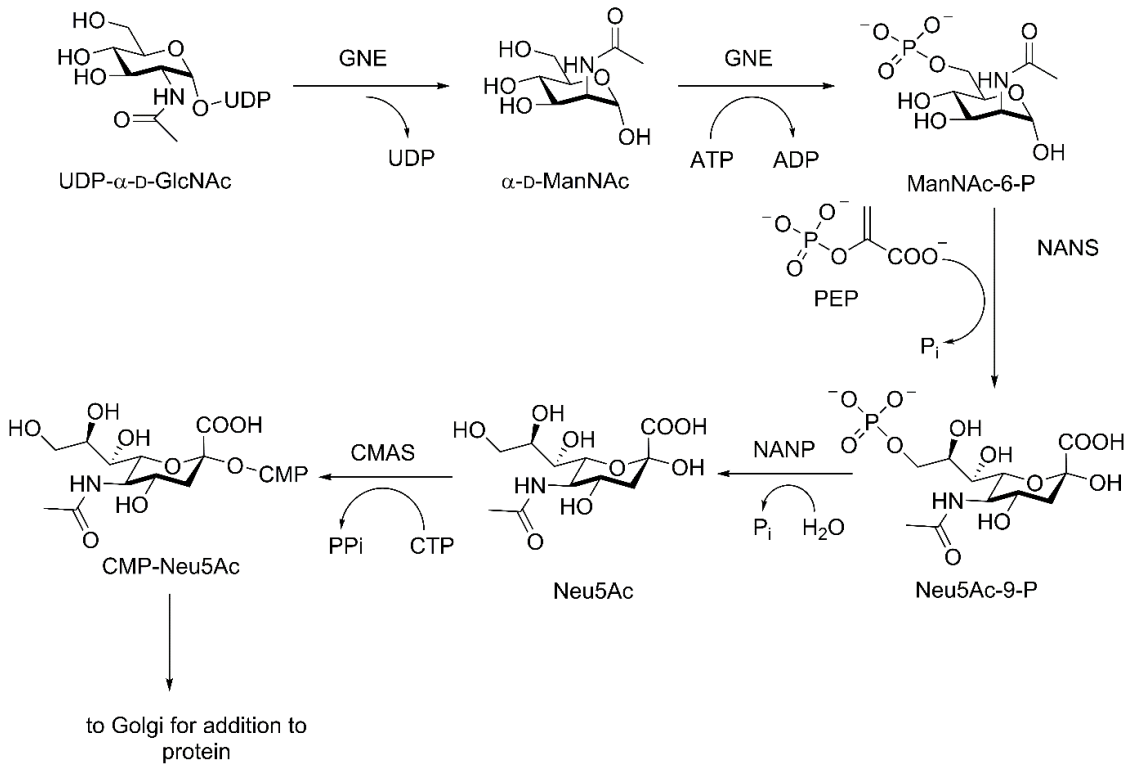


Figure 1.4: The Roseman-Warren biosynthetic pathway for sialic acid inside the lumen of the endoplasmic reticulum [62] [63]. The enzymes *N*-acetylmannosamine kinase (GNE), *N*-acylneuraminate-9-phosphate synthase (NANS), *N*-acylneuraminate-9-phosphatase (NANP), and *N*-acylneuraminate cytidylyltransferase (CMAS) can all accept unnatural substrates.

For the sugars that can be incorporated, inhibition of viral uptake suggested 18-55% incorporation of *N*-acyl derivatives of mannosamine depending on the cell type [64]. Cell lines have been developed specifically to improve uptake of unnatural sugars [65]. BJA-B K20 and HL-60 myeloid strains lack the bifunctional GNE enzyme and are therefore

hyposialylated. The lack of naturally-synthesised ManNAc renders the cell lines dependent on ManNAc being provided in the cell medium, and therefore unnatural ManNAc derivatives are more easily incorporated and can constitute up to 85% of sialic acids on the cell [66].

GlcNAc analogues can similarly be accepted by enzymes including GlcNAc kinase (GNK), phospho-*N*-acetylglucosamine mutase (AGM1) and UPD-GlcNAc pyrophosphorylase (AGX1) in the hexosamine synthesis pathway and be incorporated in place of GlcNAc residues in *N*-linked and *O*-linked glycans (Figure 1.5) [67].

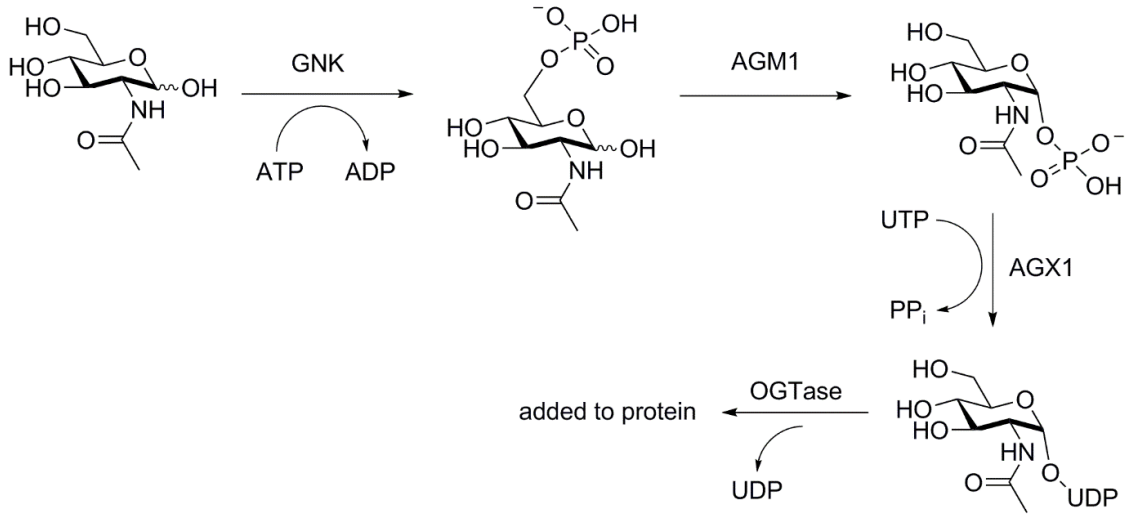


Figure 1.5: The hexosamine salvage pathway [67]. The enzymes GlcNAc kinase (GNK), phospho-*N*-acetylglucosamine mutase (AGM1), UDP-GlcNAc pyrophosphorylase (AGX1) and *O*-GlcNAc transferase can also accept unnatural substrates.

The nucleotide sugar donors must then cross another membrane to reach the glycoprotein they are added to. Protein transporters move sugar donors such as CMP-Sia and UPD-GlcNAc into the Golgi apparatus. UPD-GlcNAc is also transported into the endoplasmic reticulum [55]. Several of these transporters recognize multiple substrates, for example, the bifunctional transporter for UDP-Xyl and UDP-GlcNAc, which again suggests that modified GlcNAc sugars can be tolerated [10].

ManNAc and GlcNAc Analogues

Many different ManNAc analogues have been developed to improve the rate of incorporation into glycans or add new functional groups [62] [68] [69]. For example, it was found that radiolabelled *N*-acyl derivatives of GlcNAc, such as *N*-propanoyl-, *N*-butanoyl- and *N*-pentanoyl-D-glucosamine could be incorporated into homogenised mouse cells. The *N*-pentanoyl-D-glucosamine molecule displayed the highest rate of incorporation, particularly with intestinal cells [70]. The most rapidly metabolised derivative carried a hexanoyl group. After a 30 minute incubation, 46% of the hexanoyl derivative was phosphorylated and 16.7% converted to the nucleotide sugar donor. This suggests that the biosynthesis enzymes tolerate unnatural sugars very well, and the limiting step is likely to be the initial transport of sugar into the cell.

Subsequent studies used sugars such as ManLev **1** (Figure 1.6) and its acetylated form Ac_4ManLev **2**, which incorporate ketone functionality that is able to ligate to chemical reporters such as a hydrazide probe linked to biotin [71]. Biotin is tightly bound by avidin and streptavidin, and this interaction is frequently used in biochemical assays such as pull-down purification [72]. After the ketone reacts with hydrazide, the biotin binds fluorescein isothiocyanate (FITC)-coupled avidin, allowing detection of the glycan by fluorescence microscopy [71].

N-azidoacetylmannosamine **3** (ManNAz) (Figure 1.6) incorporates an azide functional group into cell surface glycans, which can then undergo labelling by a phosphine probe in a modified Staudinger ligation [73].

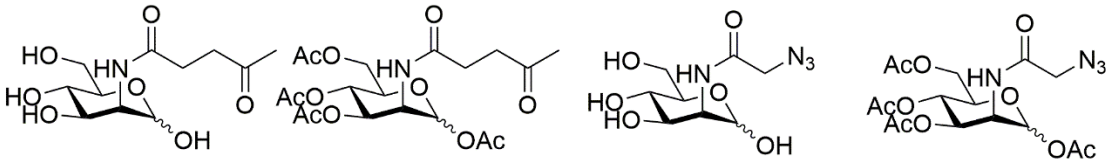


Figure 1.6: ManNAc derivatives that have been incorporated into *N*-linked cell surface glycans and bioorthogonally labelled. The acetylated Ac_4ManNAz is the most popular due to more efficient incorporation [52].

The azide group can also undergo “click” chemistry with an alkyne bearing a fluorescent or purification tag [11] [74]. This very specific and high-yielding reaction is a cycloaddition between an azide and an alkyne, which gives a 1,2,3-triazole product (Figure 1.7). Typically it is catalysed by copper (I), however a copper-free method using a strain-promoted cyclooctyne in a destabilised ground state due to ring strain [75] has also been developed.

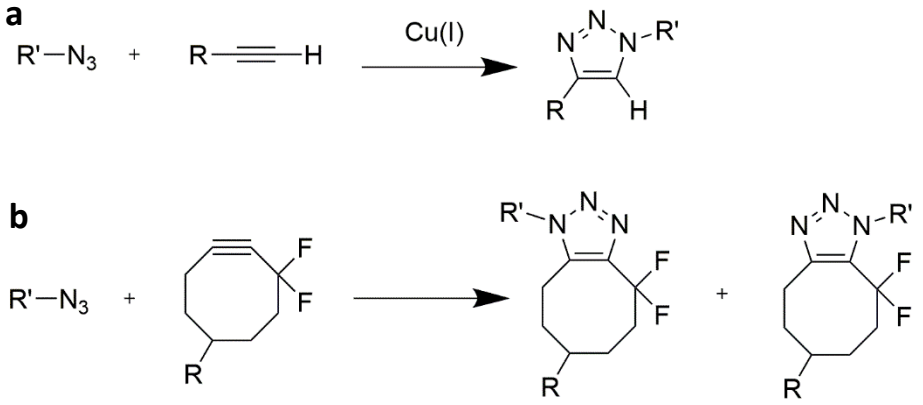
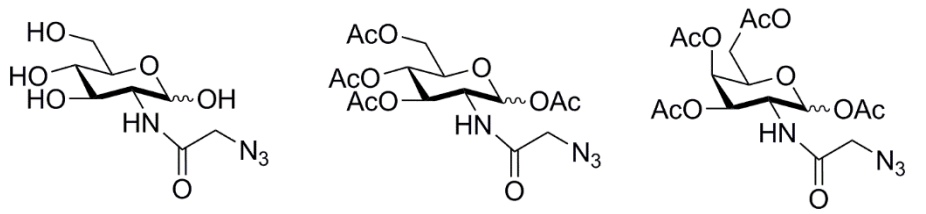


Figure 1.7: “Click” chemistry reaction between an azide and an alkyne catalysed by copper (a) or without copper using a strained alkyne (b).

Incorporation of sugar analogues including ManNAz **3** is improved up to 600-fold by acetylation of the four hydroxyl groups, suggested by Yarema to be due to improved cellular uptake or increased metabolic flux through biosynthetic pathways [76]. It is thought the acetyl groups are removed by nonspecific esterases [77] [62]. The acetylated Ac_4ManNAz **4** (Figure 1.6) has been incorporated into *N*-linked glycans in a multitude of studies [52] [78]. A comparison of Jurkat cells treated with Ac_4ManNAz **4** and biotin probe with biotinylated beads of known biotin density showed up to 4.5 million biotin moieties on the cell surface. Imaging uses have included live chondrocytes [79], zebrafish embryos [80] and the heavily sialylated tumour cells in mice implanted with lung carcinoma cells [81]. The mice were injected with Ac_4ManNAz **4**, which after incorporation over 7 days was reacted with a biotinylated phosphine probe followed by a fluorescent avidin conjugate. Ac_4ManNAz **4** has been used to identify sialylated *N*-linked sites in prostate cancer cells [82]. Ac_4ManNAz **4** has also been used to track the movement of sialylated glycans on the surface of live cells [83]. Controlled bleaching after the fluorescent probe was added allowed single-molecule tracking over time,

which showed *N*-linked Neu5Ac diffused more slowly in metastatic cells, possibly as the cell surface is more sterically hindered by sialylated glycans. Similarly Ac_4ManNAz **4** has been used to track the movement of neural cell-adhesion molecule (NCAM), which is modified with polysialic acid, over the surface of neurons [84].

N-azidoacetylglucosamine (GlcNAz **5**) (Figure 1.8), a GlcNAc analogue, is readily incorporated into intracellular glycoproteins in place of GlcNAc, or in place of GalNAc due to conversion to GalNAz by an epimerase [85]. The alkynyl-modified analogue, GlcNAlk, is not interconverted to the Gal epimer and so is only incorporated in place of GlcNAc [86]. Studies on *N*-levulinoyl glucosamine (GlcLev) showed that it is incorporated into the sialic acid pathway because of epimerisation, but the number of ketones reaching the cell surface glycans is 15-fold lower than for ManLev **1** [61]. Instead, GlcNAc derivatives such as Ac_4GlcNAz **6** (Figure 1.8) are mainly incorporated into cytosolic and nuclear proteins [67]. Ac_4GalNAz **7** is epimerised at a later step than Ac_4GlcNAz **6** in the biosynthetic pathway, from UDP-GalNAz to UDP-GlcNAz by UDP-galactose 4'-epimerase, so it allows specific labelling of mucin-type *O*-linked glycoproteins and has improved flux into *O*-GlcNAc residues compared with Ac_4GlcNAz [85]. Another GalNAc derivative, 2-keto-GalNAc, is incorporated into CHO cells via the salvage pathway and then reacted with a hydrazide probe [87]. It was detected in both *O*-linked glycoproteins and chondroitin sulfate, a cell surface glycosaminoglycan (GAG).



GlcNAz **5** Ac_4GlcNAz **6** Ac_4GalNAz **7**

Figure 1.8: GlcNAc derivatives that have been incorporated into *O*-linked glycans.

Some derivatives are incorporated into unusual cell components. Mannosamine has also been modified at the C-4 position, to give C-7 modified sialic acids when incorporated *in vivo*, with Ac_3-4-azido-ManNAc shown to be incorporated into 50% of

sialic acid residues in *GNE*-deficient cells [88]. Unlike the C-2 modified ManNAz, treatment with PNGase F to cleave *N*-linked glycans revealed that 4-azido-ManNAc was not incorporated into *N*-glycans, but into *O*-glycans such as mucin-1. Another sugar analogue, 1-deoxy-N-pentynyl glucosamine (1-deoxy-GlcNAlk), is curiously not incorporated into glycans at all, but is incorporated onto proteins that undergo lysine acetylation, such as histones [89]. This compound could be used to probe the epigenetic regulation of gene expression.

Azides also react with alkyne in a copper-catalysed azide-alkyne cycloaddition (CuAAC) to form a 1,2,3-triazole product [90]. This covalent bond allows the detection of azide sugars by labelling with alkyne-biotin followed by streptavidin-fluorophore [62]. As copper is cytotoxic, a copper-free strain-promoted azide-alkyne cycloaddition (SPAAC) has been developed [68]. The dibenzocyclooctyne (DBCO) molecule has a ring strain of 18 kcal/mol and this allows the reaction between azide and alkyne to proceed spontaneously. Using a DBCO-fluorophore conjugate therefore allows visualisation of glycans by fluorescence microscopy or flow cytometry (Figure 1.9).

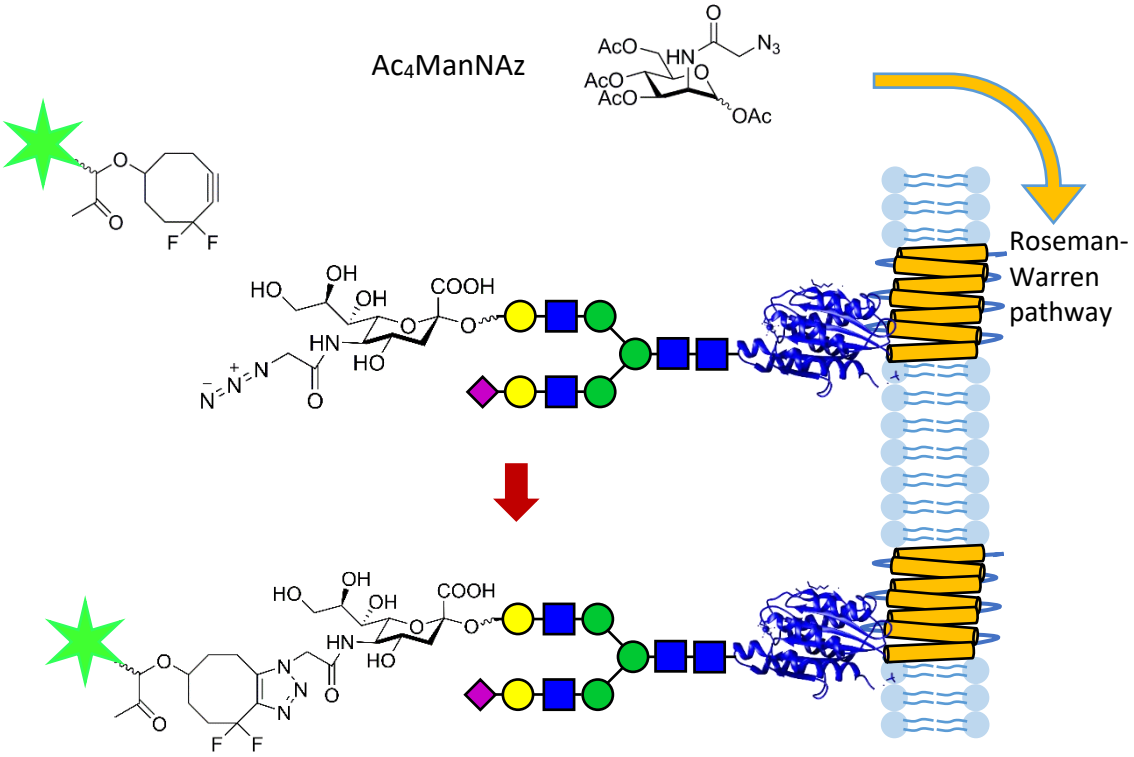


Figure 1.9: Incorporation of Ac_4ManNAz into cells and labelling. Ac_4ManNAz enters the pathway for sialic acid biosynthesis and is incorporated into cell surface proteins such as integrin. The azide group then undergoes click chemistry with a strained alkyne

fluorophore to form a 1,2,3-triazole product, attaching the fluorophore to the cell surface.

Azide can also react with DBCO linked to a purification tag [75]. Using biotin allows azide-containing sialylated glycans to be collected by streptavidin beads, cleaved by formic acid, and the glycoproteins to be digested by trypsin and characterised by tandem mass spectrometry. Work by the Bertozzi group has developed this technique, termed IsoTaG, in order to profile the glycome of various cell types [91]. Glycoproteins overexpressed in prostate cancer cells were identified by culturing normal and cancerous prostate cells in Ac_4ManNAz **4** and using IsoTaG to enrich for and analyse sialylated glycans, and comparing the two populations [92]. This technique has similarly been used to identify overexpressed glycoproteins such as ecto-5'-nucleotidase in pancreatic cancer cells. In this case, the DBCO-biotin tag was improved by the addition of a thiol link between the DBCO and biotin components [13]. The thiol is cleaved under relatively mild conditions to release the DBCO-azide-glycoproteins from the streptavidin beads, allowing detection of glycoproteins present at lower concentrations by tandem mass spectrometry.

This purification technique can also be used on bacterial sugars. Cell surface glycans are important in allowing many pathogenic bacteria such as *Toxoplasma gondii* to bind and infect host cells. Incorporation of Ac_4GlcNAz **6** followed by a click reaction with phosphine-biotin, purification on avidin beads and tandem mass spec identified glycoproteins known to be *O*-GlcNAcylated in other species, but also new proteins that may be specific to how the parasite infects host cells [93].

Additional Bioorthogonal Probes

The diazo derivative, Ac_4GalNDiaz **8** has a smaller functional group than the azide sugar and can also labelled with DBCO-biotin probe **9** [94]. However, to enable the labelling of multiple sugars on a cell at the same time, other bioorthogonal sugar-probe pairs have been developed (Figure 1.10). Terminal alkenes undergo the Diels-Alder reaction with inverse electron demand with tetrazines, and this reaction has been shown to be compatible with physiological conditions. ManNAc derivatives carrying a carbamate side-chain such as Ac_4ManBeoc **10**, Ac_4ManPeoc and Ac_4ManHeoc, have been

incorporated into HEK 293T cells and subsequently labelled with a tetrazine-biotin probe **11**. Ac_4ManBeoc **10**, having the shortest side chain, was the most efficiently incorporated and labelled [69]. The cyclopropene derivative Ac_4ManNCp **12** also reacts with tetrazine probe **13**, and has been shown to be incorporated into cell lines such as HeLa more efficiently than Ac_4ManNAz **4** [78]. Another bioorthogonal pair has been developed with ITag-Man **14**, a mannosamine derivative with an imidazolium group that after being incorporated can be non-covalently labelled by fluorescent-*N*-nitrilotriacetate **15**. This also allows purification of glycoproteins by a nickel-*N*-nitrilotriacetate column, the non-covalent interaction allowing the glycoproteins to be purified under mild conditions before analysis [95].

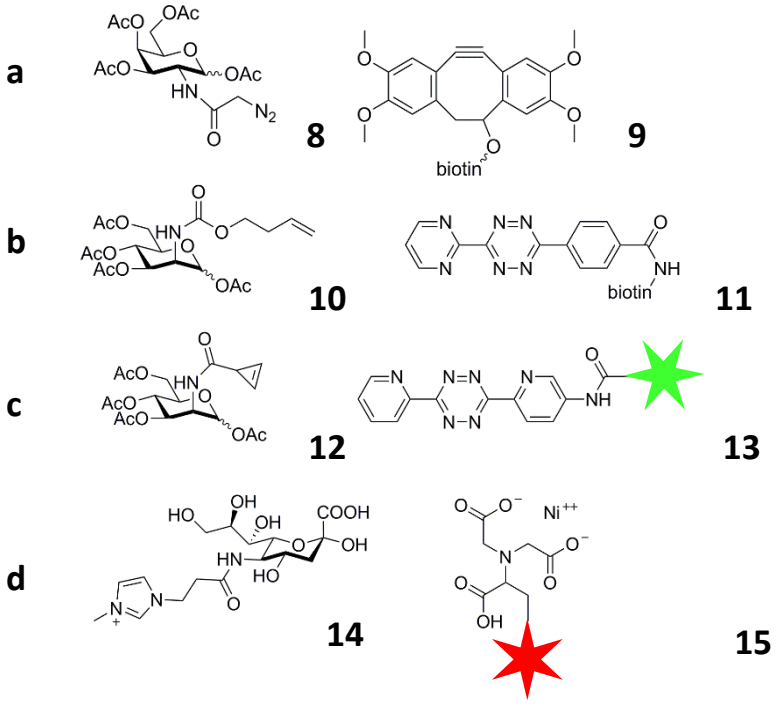


Figure 1.10: Other bioorthogonal sugar derivative-probe pairs. (a) Diazo sugar and DBCO-biotin probe. (b) Carbamate sugar and tetrazine-biotin probe. (c) Cyclopropene sugar and tetrazine-fluorophore. (d) Imidazolium sugar and *N*-nitrilotriacetate-fluorophore.

Labelling of multiple sugars in the same experiment can allow visualisation of how different glycans interact and is possible if mutually orthogonal pairs of bioorthogonal sugars and probes are combined. For example, the simultaneous incorporation of Ac_4ManLev **2** and Ac_4ManNAz **4** allowed dual labelling with biotin hydrazide and

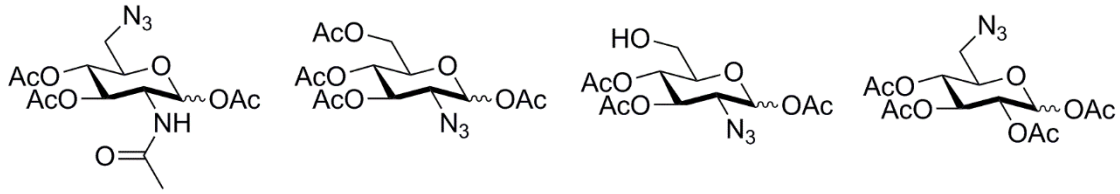
phosphine-FLAG respectively, and then detection with phycoerythrin-avidin (red) and FITC-anti-FLAG (green) [61]. This required Ac_4ManLev **2** to be added at 4 times the concentration of Ac_4ManNAz **4** in order to have comparable green and red fluorescence, as Ac_4ManNAz **4** is more efficiently incorporated. It is also possible to label different types of glycans at the same time. Addition of Ac_4GalNAz **7** and Ac_4ManLev **2** to Jurkat cells allows the labelling of *O*- and *N*- linked glycans respectively with different fluorophores so they can be visualised simultaneously [96]. Ac_4GalNAz **7** has also been simultaneously incorporated with Ac_4ManNCp **12**, into colon cancer cells, allowing simultaneous labelling with DBCO probe and tetrazine probe respectively [97]. Finally, Ac_4GalNAz **7** and Ac_4ManNPtl (a pentenoyl derivative that reacts with tetrazine) have been used in HeLa cells to label cell membrane glycans [98].

It is also possible to give multiple noncanonical functional groups to the same sugar. 9-azido sialic acid (9AzSia) can be incorporated into cells without going through the Roseman-Warren pathway – instead it is taken up by endocytosis and released from the lysosome [99]. A bifunctional unnatural sialic acid has been synthesised with an azide at C-9 and an alkyne on the *N*-acyl side chain [100]. It is incorporated into various cell lines such as HeLa as efficiently as 9AzSia. First, to prevent azide and alkyne groups incorporated on the same cell from reacting with each other, copper-free DBCO-Alexa Fluor 488 was reacted with the exposed azides. This was followed by the detection of exposed alkynes using copper-catalysed cycloaddition with azide-Alexa Fluor 647. Another bifunctional sialic acid contained diazirine on the *N*-acyl side chain. Once incorporated, UV light was used to crosslink sialic acid glycans *via* the diazirine group. Then an alkyne-biotin affinity tag was covalently bound to the azide group, allowing purification of multiple proteins by streptavidin beads.

O-GlcNAcylation

Different sugars are incorporated with varying rates of efficiency depending on the structure and cell line [101]. A comparison of Ac_4GlcNAz **6**, Ac_4GlcNAlk, Ac_4GalNAz **7** and Ac_4GalNAlk across cell lines showed that Ac_4GalNAz **7** was particularly well incorporated into many cell lines, followed by Ac_4GlcNAz **6** and then Ac_4GlcNAlk, while Ac_4GalNAlk showed very poor incorporation. These are all incorporated into *O*-glycans, so are not specific to the *O*-GlcNAcylation dynamic switching of proteins by addition and removal of a single *O*-GlcNAc to Ser/Thr sites. For this purpose, the analogue 6-azido-6-deoxy-N-

acetyl-glucosamine ($Ac_3$6AzGlcNAc **16**) was employed [102] (Figure 1.11). It cannot be phosphorylated at the 6-hydroxyl, so cannot be converted to the UDP-sugar donor and is therefore a specific marker for *O*-GlcNAcylation. 2-azido-2-deoxyglucose ($Ac_4$2AzGlc **17**) is also selectively incorporated in place of intracellular *O*-GlcNAc [103] (Figure 1.11). It is not removed from proteins by *O*-GlcNAcase, as treatment with the *O*-GlcNAcase inhibitor Thiamet-G does not prevent its removal over time. This allows $Ac_4$2AzGlc **17** to persist at the *O*-GlcNAc site long enough for purification and analysis of glycoproteins by reacting with DBCO-biotin followed by enrichment with streptavidin beads. Further derivatives include $Ac_3$2AzGlc **18** (Figure 1.11), which lacks the 6-acetyl group and was found to be less toxic than $Ac_4$2AzGlc **17** [103], and 6-azido-6-deoxy-glucose **19** (Figure 1.11), which is also accepted by *O*-GlcNAc transferase and may be used to investigate *O*-glucose modification [104].



$Ac_3$6AzGlcNAc **16** $Ac_4$2AzGlc **17** $Ac_3$2AzGlc **18** $Ac_4$6AzGlc **19**

Figure 1.11: GlcNAc and Glc analogues that are specific to dynamic *O*-GlcNAcylation.

Cell-selective labelling

Liposomes – vesicles composed of phospholipids – have been used to label specific cells with unnatural sugars. The phospholipid layer contains embedded ligands that are recognised by certain cell receptors [105]. 9-azido sialic acid (9AzSia) was encapsulated in folate-targeted liposomes and these were tested on HeLa cells overexpressing folate receptor [106]. The liposomes bound to the cells and 9AzSia diffused directly into the cell and was more efficiently incorporated than 9AzSia alone. This allowed fluorescent labelling of cells by alkyne-biotin followed by Alexa Fluor488-streptavidin at 9AzSia concentrations of 10 µM, instead of the 100 µM typically used to incorporated 9AzSia. The liposomes were later used to incorporate 9AzSia or Ac_4ManNAz **4** into the sialylated glycans of tumour cell xenografts in mice, allowing the detection of tumour cells by

fluorescent DBCO-Cy5 dye [107]. This also allowed the glycoproteins to be purified with alkyne-biotin and streptavidin beads.

Liposomes containing 9AzSia were also successfully incorporated into brain sialoglycans in mice. Ac_4ManNAz was not incorporated, presumably due to the blood-brain barrier [108]. This allowed imaging of the neurons with a fluorescent probe. A pulse-chase experiment showed sialoglycan turnover in the hippocampus to be much slower than in the rest of the brain. Finally, the liposomes have been used to label multiple types of cells specifically. B-lymphocyte K20 cells were co-cultured with HeLa cells expressing folate receptor. Liposomes containing alkyne-sialic acid (SiaNAl) and carrying surface ^{BPC}NeuAc, a ligand for the B-cell receptor CD22, were added along with folate-targeted liposomes containing 9AzSia. These were then reacted with DBCO-carboxyrhodamine to conjugate dye to exposed azides in the cell surface glycans, followed by azide-Alexa Fluor 647 to label exposed alkynes. This specifically labelled the K20 B cells with Alexa Fluor 647 and the HeLa cells with carboxyrhodamine [105].

Specific cells can also be labelled with aptamers – oligonucleotides that bind a target. Incorporated Ac_4ManNAz **4** in Jurkat or Ramos cells was detected using a DBCO-aptamer-fluorophore probe that binds to protein tyrosine kinase-7 (PTK7) or μ-IgM [109]. In this method, the aptamer binds specifically to the receptor on one type of cell, then the DBCO group reacts with the azide and creates a covalent bond. This combines the specificity of the aptamer for a particular glycoprotein with the bioorthogonal click chemistry labelling. Labelling is therefore very specific and efficient. DBCO-aptamer-biotin probe and streptavidin purification was used to analyse the proteins, revealing a novel glycoform of PTK7.

Another method of studying specific cell types is to modify rarer sugars than ManNAc or GlcNAc. Fucose is a component of the cancer associated tetrasaccharide sialyl Lewis^X. Testing derivatives of fucose has revealed that 6-azido fucose (6AzFuc) is incorporated by the fucose salvage pathway, but the 2- and 4- azido derivatives are not. 6AzFuc needs to be added at 100 μM or higher for incorporation, but causes toxic effects at this concentration, either because it inhibits enzymes or alters protein function [110]. The 6-alkynyl fucose derivative is efficiently incorporated, and is less toxic, being tolerated by cells up to 200 μM [111].

Incorporation of sugars in bacteria

Bacteria use a range of non-mammalian sugars that are ideal targets for species-specific labelling (Figure 1.12). 3-deoxy-D-mannosoctulosonic acid (Kdo **20**) is an essential component of the inner core of lipopolysaccharides [112]. The derivative 8-azido-8-deoxy-Kdo (8AzKdo) has been shown to allow labelling of the membranes of Gram-negative bacteria [113]. This derivative has also been used in plants, in which Kdo is only found in the cell wall pectin rhamnogalacturonan-II. Adding azide-Kdo and alkyne-fluorescent probe specifically labelled the *Arabidopsis thaliana* plant cell wall, allowing imaging of pectin as the cell wall elongated during root growth [74]. 8AzKdo has also been injected into the mouse gut microbiota and labelled with alkyne-tetramethylrhodamine (TAMRA) dye, concurrent with the addition of a fluorescent vancomycin-derivative. This allowed two colour imaging of the distribution of Gram-negative bacteria and Gram-positive bacteria respectively [112]. A similar technique was used for gut commensal species including the symbiotic *B. fragilis*, which efficiently incorporates Ac_4GalNAz **7** into Polysaccharide A. Ac_4GalNAz **7** was then labelled with fluorescent-DBCO, allowing observation of the distribution of the bacteria when they were injected into the intestine [114].

Other cell wall components include muramic acid (MurNAc **21**), an amino sugar [80]. UDP-MurNAc derivatives linked to FITC have been incorporated into the cell wall of Gram-positive bacteria [115]. These derivatives have also been incorporated into Gram-negative bacteria, although it requires calcium chelation to increase the outer membrane permeability to the large FITC molecule. A more conventional method is used to label trehalose monomycolate (TMM **22**), a component of the mycomembrane that is uniquely found in mycobacteria. *O*-AlkTMM has a terminal alkyne and can be incorporated into the membrane and clicked with an azide fluorophore, while *N*-AlkTMM is similarly incorporated into trehalose-containing glycolipids [116].

Another rare sugar is pseudaminic acid **23** (Pse), found in *Campylobacter jejuni* [117]. The bacterial flagella are typically heavily glycosylated with this, and this is required for motility. Addition of an azide-pseudaminic acid precursor restored flagella assembly and motility in a non-motile strain deficient in Pse biosynthesis, proving that the downstream PseI and PseF enzymes can accept unnatural substrates [118].

Legionella pneumophila serogroup 1 has an *O*-antigen containing 5-*N*-acetimidoyl-7-*N*-acetyl-legionaminic acid **24** (Leg5Am7Ac). An azide analogue of this was successfully incorporated into four strains of *L. pneumophila* and labelled with alkyne-biotin probe followed by Alexa Fluor 488-IgG anti-biotin antibody conjugate [119]. This labelling was specific to the serogroup 1 strains, as other bacteria including *Escherichia coli*, *Pseudomonas aeruginosa*, and also other Legionella species did not incorporate the azide analogue.

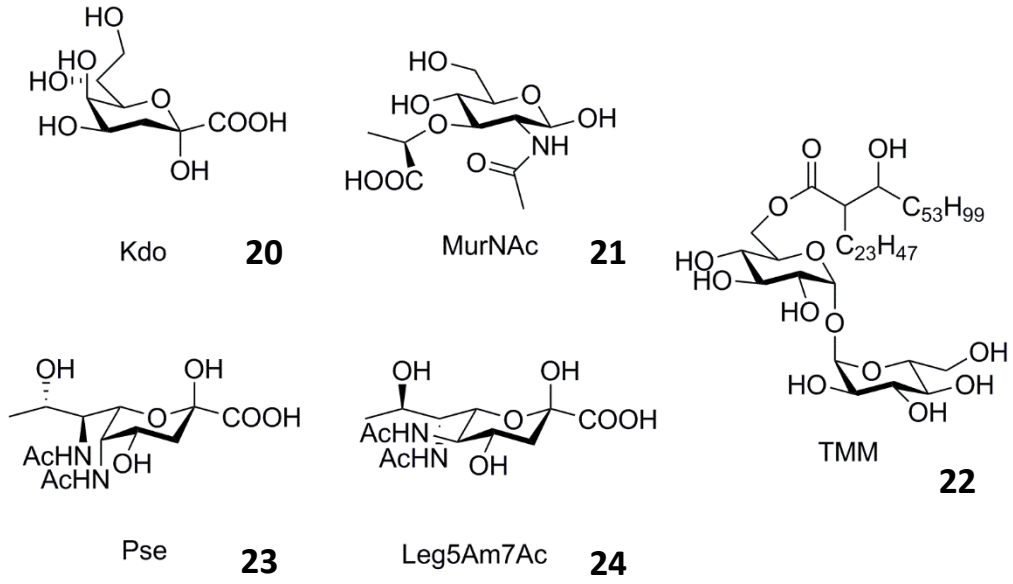


Figure 1.12: Rare bacterial sugars that have been targeted by MOE.

Biological and medical uses of MOE

While bioorthogonal labelling by definition aims to incorporate functional groups that will not interfere with biological functions, it is also possible to deliberately introduce groups that have a biological effect, including several that block the binding of native proteins. The incorporation of *N*-glycolylmannosamine (ManNGc) results in non-human Neu5Gc expression on the surface of neuroblastoma-glioma hybrid cells and prevents the binding of myelin-associated glycoprotein, a Neu5Ac-binding lectin that stabilises myelin and has been linked to inhibition of nerve regeneration [120]. Incorporation of the alkyl derivatives ManNProp, ManNBut and ManNPent has been shown to decrease virus binding and subsequent infection, including polyomavirus infection of mouse fibroblasts [121] and influenza A infection of dog kidney epithelium cells [122]. The addition of a bioorthogonal probe can also be utilised to disrupt binding. Incorporation

of GlcNAz into *S. aureus* followed by alkyne dye reduced adherence to bladder carcinoma cells by 48% as binding carbohydrates were blocked. This could potentially be used to treat *S. aureus* bladder infections [123].

Incorporation of sugars can be improved by butanoylation, which improves flux through the Roseman-Warren pathway. It also results in different toxic effects on cells, as the 3,4,6-*O*-butanoylated version of ManNAz causes apoptosis at concentrations of just 20 μM, whereas the 1,3,4-*O*-butanoylated version is well tolerated up to 400 μM and more efficiently labels sialoglycans than Ac_4ManNAz **4** [124]. Halogenated sugars have also been investigated for their toxic effects on tumour cells. GlcNAc derivatives carrying a trifluoromethanesulfonyl group in place of the C-1 hydroxyl have been found to have the most toxic effect out of a group of GlcNAc analogues tested on mouse mammary adenocarcinoma cells [125]. Similarly, 3-fluoro sugars reduce the growth of murine leukaemia cells when added at 30 μM [126].

The fluorescent probes used to detect incorporated azides are best suited to use in transparent organisms such as zebrafish, as their use in other organisms such as mice is restricted to the organism's surface. Probes are being developed to allow non-invasive scanning of glycans by using Magnetic Resonance Imaging (MRI). Ac_4ManNAz **4** can be labelled with DBCO-cryptophane-A, a xenon-binding MRI contrast agent. After the click reaction, hyperpolarized Xe is administered to the cells and allows detection of sialylated glycans by hypersensitive MRI [127]. Ac_4GalNAz **7** has been tested on adenocarcinoma tumours injected into mice, by clicking with a DBCO-gadolinium contrast agent, followed by scanning with MRI. This labelled the tumour, although other organs – kidney, gut and liver also showed significant incorporation of the MRI contrast agent [128].

Unnatural sugars can also be used to synthesise glycoproteins with novel properties. Interferon-β is a multiple sclerosis treatment that has a single sialylated N-linked glycan at Asn80 [129]. Ac_4ManNAz **4** can be incorporated into CHO cells transfected with interferon-β, such that the secreted interferon-β has NeuNAz at 29% of its sialic acid sites. Incorporating azide in this way would enable the covalent ligation of other groups to improve the pharmacokinetic properties of interferon [130]. It is also possible to combine unnatural sugars with genetically-modified yeast to make glycoproteins with new functional groups. For example, a GlcNAc auxotrophic yeast strain has been created

by removal of the gene *GNA1* to disable de novo GlcNAc synthesis, and addition of the genes *NGT1* (GlcNAc transporter) from *Candida albicans* and human *NAGK* (GlcNAc kinase) to create a GlcNAc salvage pathway. This enabled good incorporation of GlcNAc analogues such as GlcNAz and GlcNAlk at the conserved chitobiose core of *N*-linked glycans, as the analogues do not need to compete with the native GlcNAc substrate [131].

MOE can also be used to allow virus-mediated gene transfer in cells that lack natural receptors. ManLev **1** has been incorporated into NIH-3T3, a fibroblast cell line lacking adenovirus receptors, and labelled by a hydrazide-biotin probe, followed by an avidin-anti-adenovirus antibody conjugate. This allowed adenovirus to bind and improved gene transfer to the cells [132].

Finally, MOE can be used to distinguish between samples of healthy and diseased cells. SiaNAl has been shown to be poorly incorporated into fibroblasts from patients with glycosyltransferase deficiencies. The cells therefore do not bind azide-linked fluorophores and can be distinguished from control cells from healthy people by microscopic examination [133]. This could allow diagnosis of congenital glycosylation disorders. In another application, caged Ac_4ManNAz linked to a protease substrate was released into CHO cells when prostate-specific antigen (PSA), a protease produced in excessive amounts by prostate cancer cells, was added. Cleaved Ac_4ManNAz **4** was incorporated into surface glycoproteins and fluorescently detected, whereas control cells had only a low level of background labelling due to hydrolysis of the bond [134]. In future this could be used to image prostate tumour cells from patients to test PSA levels.

Sortase

Sortase mechanism

Many of the discussed methods for detecting incorporated sugars rely on chemical ligation to permanently attach a fluorescent group. This project aimed to use MOE to incorporate sugar substrates carrying peptide sequences into cells, and then detect the peptide by enzymatically ligating a probe using sortase. This would allow reversible labelling with multiple tags in the same experiment, as the sortase forms an equilibrium between the ligated product and hydrolysed tag. For example, a purification tag could be used followed by a fluorescent tag for microscopy. It would also be desirable to

engineer the sortase to only recognise, for example, an amine functional group incorporated into a sialic acid residue, and not incorporated into any other sugar residue. This would have applications in, for example, detecting unusually high amounts of sialic acid on cancer cells that would indicate an aggressive tumour with a high risk of metastasis. It would potentially allow the detection of sialic acid within specific glycans such as sialyl-Tn antigen, as chemical labelling methods such as click chemistry will label all azide-carrying sialic acid residues and are not specific to sialic acid residue attached to another specific sugar residue. As sortase is a biological enzyme, it could in theory be evolved to specifically bind and label only amine borne by sialyl-Tn antigen.

Sortase is a transpeptidase enzyme found in some Gram-positive bacteria that binds virulence factors including fibronectin-binding proteins to the cell wall [135]. Class A sortases are known as “housekeeping” sortases and recognise an LPXTG motif [136]. Class B sortases recognise the NPQTN motif and anchor iron transporter proteins to the cell wall [137]. Class C sortases recognise LPXTG, but unlike class A the nucleophile is the ε-amine group from a lysine residue on a pilin protein [138]. Class D sortases recognises an LPNTA motif, and are involved in spore formation in bacteria [139]. Class E sortases are a housekeeping sortase found in some bacterial species that do not have class A, and recognise the LAXTG motif instead [140]. Class F sortases are the least well-characterised class of sortase, but also appear to be a housekeeping class of sortase that accepts the LPXTG motif [141].

Sortase A from *Staphylococcus aureus* recognises an LPXTG motif [142]. The enzyme breaks the T-G peptide bond using a cysteine protease mechanism, creating a thioacyl intermediate, which is attacked by the nucleophilic N-terminal glycine of peptidoglycan, ligating peptidoglycan in place of the C-terminal glycine [143] (Figure 1.13). Sortase has previously been used to label cell membrane proteins expressing the LPXTG motif at the C-terminus with a fluorescent oligoglycine probe [142]. Another sortase has been used to label unnatural amino acids bearing a glycyl-glycyl-lysine motif with a modified ubiquitin bearing an LPLTG tag [144]. Therefore, it was thought that unnatural sugars with a G or GG motif may be accepted by sortase and ligated to a probe bearing the LPXTG motif. The covalent bond formed between sortase motif and sugar should allow efficient labelling and ensure that sugars can be detected by fluorescence at comparable concentrations to chemical ligation methods.

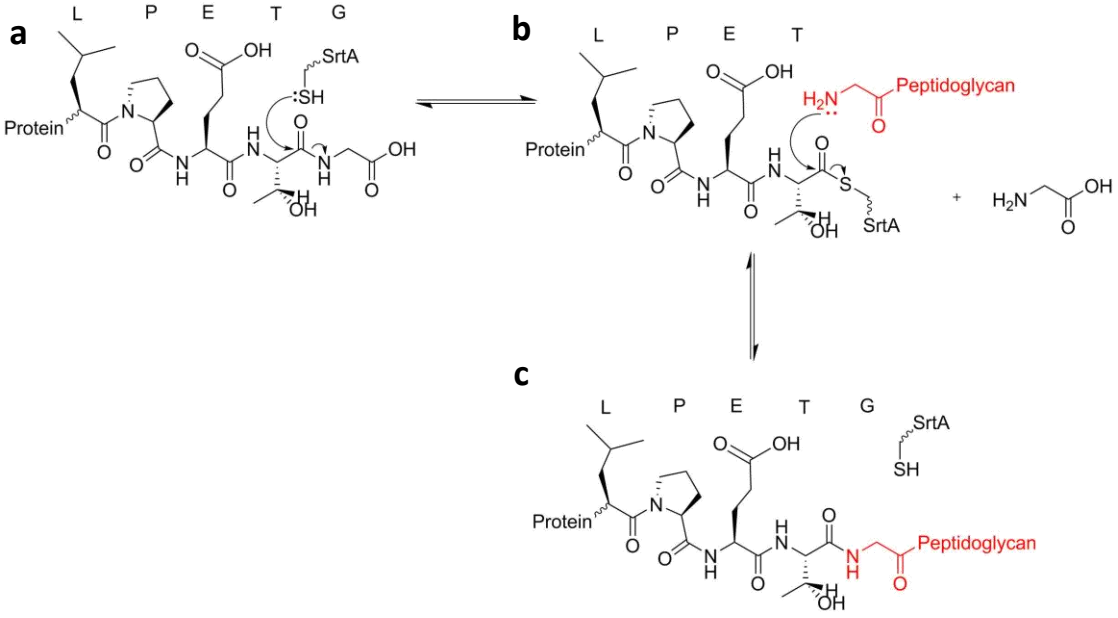
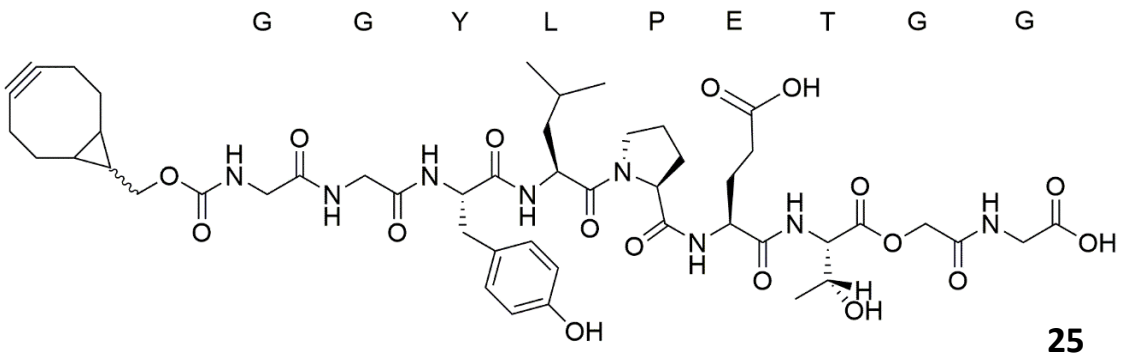


Figure 1.13: Mechanism of SrtA-catalysed ligation between pilin proteins carrying an LPXTG motif and peptidoglycan. The enzyme breaks the T-G peptide bond (a) in a cysteine protease mechanism. The *N*-terminus of peptidoglycan attacks the thioacyl intermediate (b) and ligates peptidoglycan to the pilin protein (c).

Sortase ligation is a reversible reaction, although it can be made irreversible by using depsipeptide probe **24** containing LPXTG (Figure 1.14) [145], in which an ester linkage replaces the native amide bond between T-G residues. The irreversible reaction leads to higher conversions to ligated product with lower sortase amounts [146]. Mannose-binding protein (ManBP) and galactose-binding protein (gMBP) have previously been shown to be suitable for ligating to alkyne depsipeptide **25** [147].



25

Figure 1.14: Alkyne depsipeptide containing GGYLPET-o-GG motif that improves conversion of the reaction with sortase.

The catalytic residue is Cys^{184}, with Arg^{197} and His^{120} being conserved in sortase from Gram-positive bacteria [148]. Calcium binds to the β3/β4 loop at the residues Glu^{105}, Glu^{108} and Glu^{112} and facilitates substrate binding [149].

Sortase variants

The class A sortase, Sau SrtA, has previously been modified using a directed evolution method [150]. Yeast display was used to express a library of Sau SrtA fused to Aga2p cell surface mating factor. Random mutations were introduced, at an average of two mutations into each member of a library of 7.8×10^7 sortase genes, using PCR and mutagenic dNTP analogues [151]. Biotinylated LPETG probe was added and conjugated to surface-immobilised triglycine peptide by the sortase enzymes, and cells were stained with streptavidin-PE and fluorescence-activated cell sorting used to pick cells with high sortase activity. Four rounds of enriching sortase activity were performed. Each time the selected sortase variants were subjected to decreased LPETG substrate concentration, shorter reaction times, and selecting smaller percentages of fluorescent cells to carry through to the next round. This process resulted in the creation of the pentamutant, SrtA 5M, which has the 5 mutations P94R, D160N, D165A, K190E, and K196T (Figure 1.15a) [150]. SrtA 5M has a 140-fold increase in LPETG-coupling activity and has been used to label membrane proteins [152]. SrtA 5M also has a 20-fold increase in catalytic conversion of –GGG substrate compared with the wild-type enzyme. This would allow LPXTG probe covalently bonded to G sugar to be cleaved by the use of sortase and a GGG peptide. This would remove the large fluorophore and allow subsequent labelling with a different probe, for example an affinity tag for purification.

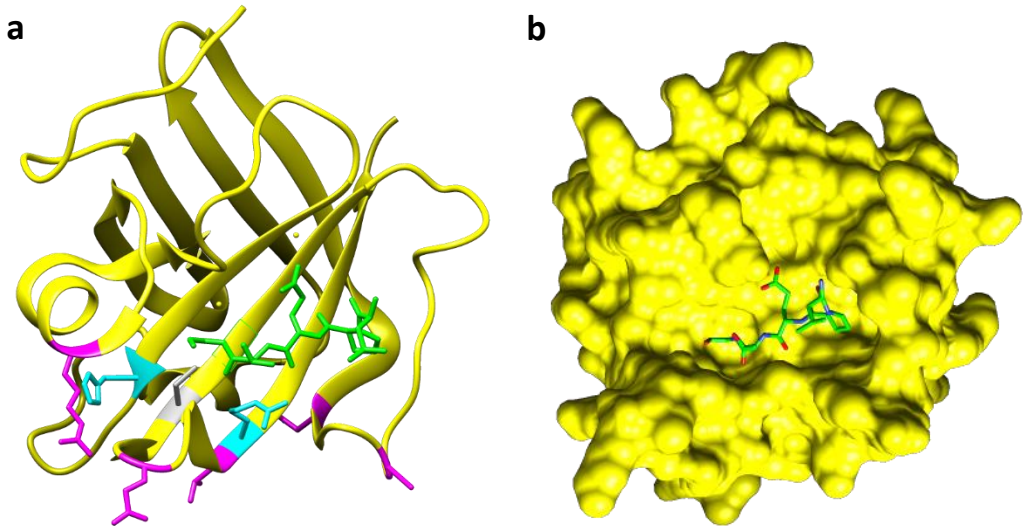


Figure 1.15: Structure of sortase. (a) Structure of SrtA 5M showing the mutations at P94R, D160N, D165A, K190E, and K196T (magenta). The LPXTG peptide (green) is shown interacting with the conserved H120 and R197 residues (cyan) and the catalytic C184 (grey). (b) The surface of sortase shows a large cleft at the LPXTG binding site that is suitable for mutagenesis.

Sortase has a large cleft at the LPXTG binding site (Figure 1.15b) [148], and other mutant sortases have been previously engineered [153]. Further mutations to the SrtA 5M enzyme created variants with improved ligation to anti-HER2 antibody with a sortase recognition sequence, which would allow efficient manufacture of antibody-drug conjugates [154]. Directed evolution has been used to create a calcium-independent sortase [155] by mutating E105K and E108A to alter the structure to a conformation closer to the *S. pyogenes* sortase, which can function without calcium. This double mutant was then combined with the pentamutant to make a fast-acting calcium-independent heptamutant [156]. Another mutated sortase has an improved yield due to the insertion of a β-hairpin that inhibits binding of the product [157]. Cyclization of sortase has been shown to increase thermostability and tolerance of denaturing agents such as DMSO [158]. A triple mutant, D165Q/D186G/K196V also has improved stability in organic solvents [159].

It has also been possible to create sortase variants that accept different motifs, for example, an orthogonal pair of evolved sortases that recognise LAXTG and LPXSG [160]. Another screening, focused on mutations in the β6/β7 loop, yielded mutants that

recognised APXTG or FPXTG [161]. Sortase is therefore amenable to mutations to alter the specificity and efficiency of the enzyme. The use of variant sortases, either created by directed evolution or harvested from different Gram-positive bacteria, could allow different substrates to be covalently labelled. If multiple sortases each specific to their own single substrate were used simultaneously, this would expand the bioorthogonal labelling methods available.

It would also be possible to create useful sortase variants by looking beyond class A sortase. Spy SrtA is a variant class A sortase from *S. pyogenes* that recognises an LPXTA motif [162], and should therefore be able to ligate probes to alanine-bearing sugars in a reaction orthogonal to the LPXTG SrtA. Class B sortases recognise the NPQTN motif [137]. Class C sortase recognises LPXTG, but unlike class A the nucleophile is the ε-amine group from a lysine residue on a pilin protein [138]. A class C sortase has previously been altered to function *in vitro* by creating a double mutant [163]. The crystal structure of the *Corynebacterium diphtheriae* sortase showed a "lid" partially blocking the active site. This lid contained a conserved DPW sequence, and mutating this sequence to GPG allowed the enzyme to polymerise pilin proteins *in vitro*. Incorporation of a lysine sugar would also allow labelling by LPXTG probe, and specific labelling would be possible if the SrtA- and SrtC- catalysed reactions were performed concurrently with LPXTG probes linked to different fluorophores. Class D sortase recognises an LPNTA motif, and is involved in spore formation in bacteria [139]. Class E sortase recognises the LAXTG motif [164]. Class F sortase, the least well studied, also accepts LPXTG [141].

Aims

This project aimed to use sortase to specifically label distinct glycans on the surface of live cells. Current chemical methods can detect, for example, an azide incorporated into sialic acid. As sortase has a large binding site and has been shown to be amenable to directed evolution to accept different substrates, it should be possible to create a sortase that binds not just a glycine motif on incorporated sugars, but glycine incorporated into sialic acid and no other residue. This would allow detection of a particular glycan, for example a hypersialylated phenotype on a cancer cell that diagnoses it as aggressive and likely to metastasize. It may also be possible to evolve the

sortase so it only recognizes glycine-sialic acid bound to another specific sugar – something that current chemical methods cannot do. This would allow easier detection of other cancer-associated glycans such as sialyl-Tn antigen. It would also be possible to do multiple experiments on the same cell sample with a reversible sortase-based labelling system – for example, adding a purification tag, removing it, and adding a fluorescent tag for microscopy, which cannot be done with chemical probes that permanently bond their recognized functional group. The use of orthogonal sortase enzymes to recognize different glycans would also allow labelling of multiple sugars in the same experiment. For example, a glycine-bearing mannosamine derivative would be incorporated into the cell through the Roseman-Warren pathway simultaneous with an alanine-bearing glucosamine derivative being incorporated through the hexosamine synthesis pathway, and these could be labelled with two different fluorophores ligated by Sau sortase and Spy sortase respectively (Figure 1.16).

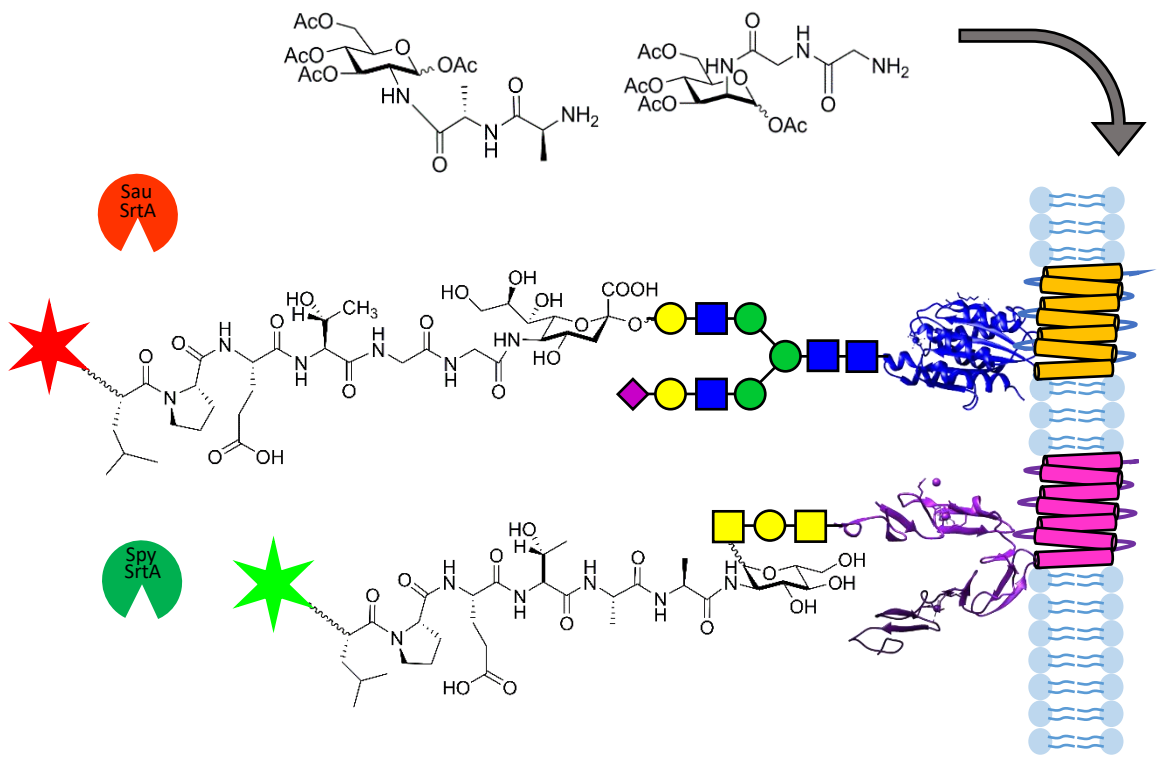


Figure 1.16: Labelling of two incorporated sugars by two orthogonal probe-sortase pairs.

The different sortase variants may be found by mining of metagenomes or by mutation of existing sortase using the combinatorial active-site saturation (CASTing) method

[165]. This is a method for expanding the substrates accepted by an enzyme. Libraries of enzyme mutants are generated by simultaneously mutating pairs of amino acids located close to the active site. Each pair tested gives up to 400 different mutants that are screened and may yield variants with improved binding to unnatural substrates such as the peptide sugars.

The main aims were:

I. Synthesise sugars with glycine and alanine side chains bearing azide groups, feed to mammalian cells and test for incorporation using click chemistry to fluorescently label the cells.
II. Reduce sugars that have been established as being incorporated to the amine sugar and test sortase labelling of cells with LPXTG/A probes.
III. Use the GRASP-cloning technology and metagenome database at the industrial partner Prozomix to expand the library of available sortases.
IV. Use the CASTing method to evolve new sortase variants that specifically detect different sugar-Gly/Ala derivatives, for example a mutant that preferentially modifies GlcNAc-Gly over Neu5Ac-Gly.
V. Characterise the kinetics of the new enzymes by testing on defined glycans followed by testing in live cells using flow cytometry and microscopy.

Chapter 2 : Sugar incorporation and click labelling

Incorporation of azide sugars

The initial approach was to feed sugars carrying suitable sortase motifs to cells so that they would be incorporated into glycoproteins and it would then be possible to label these with sortase. First, it was decided to synthesise azide sugars as some sugars with the azide group have already been shown to incorporate into cells [11] and it would be possible to check the incorporation of novel sugars using strain-promoted azide-alkyne cycloaddition (SPAAC) to react with a dibenzocyclooctyne (DBCO) dye [75]. It would then be possible to reduce the azide to an amine, which would allow recognition by sortase.

The sugars Ac_4ManNAz **4**, henceforth called AcManGN$_3$ and Ac_4GlcNAz **6**, henceforth called AcGlcGN$_3$ were chosen as the first sugars to be synthesised and tested, as these sugars have previously been incorporated into the *N*-linked and *O*-linked glycans of mammalian cells and the reduction of azide to amine would expose a glycine suitable for sortase-catalysed ligation to an LPXTG probe. The derivatives AcManGN$_3$ **4,** AcManGGN$_3$ **26,** AcManThz **27**, AcGlcGN$_3$ **6**, and AcGlcAN$_3$ **28** were synthesised following a protocol from the Bertozzi group (Figure 2.1) [68]. AcManThz **27** used a Boc-Thz precursor kindly synthesised by Robin Brabham.

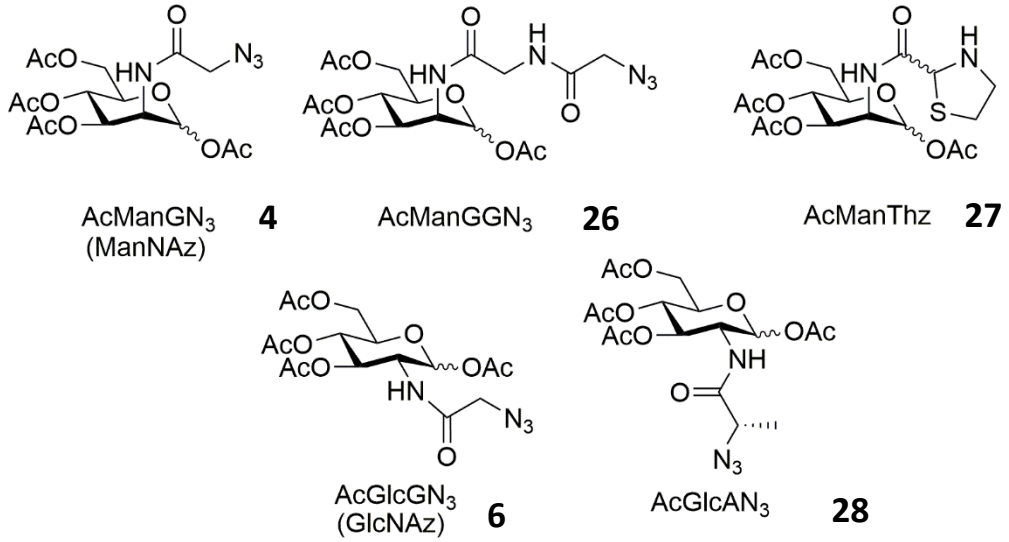


Figure 2.1: Mannosamine and glucosamine derivatives that have been synthesised for cell viability studies in HEK293.

Four of the precursors had azide groups to allow detection by “click” chemistry [75] (Figure 2.2). The AcManThz **27** precursor has a thiazolidine group that can undergo a

decaging reaction catalysed by allylpalladium (II) chloride dimer to give a glyoxyl aldehyde [166]. This can react with a fluorescent aminooxy probe in the Organocatalyst-mediated Protein Aldol Ligation (OPAL) method [167].

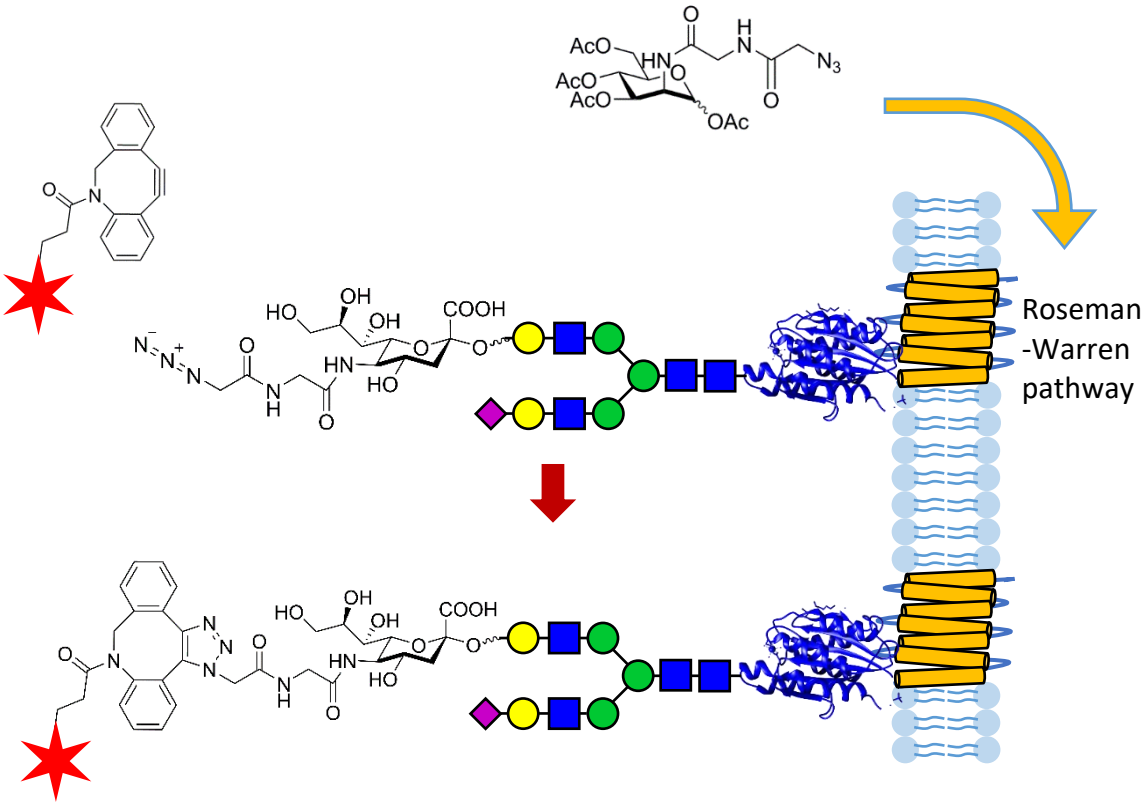


Figure 2.2: MOE incorporation of AcManGGN$_3$ similarly to previous incorporation of AcManGN$_3$ [68], and detection by Cy5-DBCO labelling.

Cell Viability

Previous work on sugar incorporation has shown large differences in the toxicity of sugars [103]. AcManGN$_3$ **4** and AcGlcGN$_3$ **6** are known to have a toxic effect on cells at high concentrations above 100 µM [68]. Many sugars are acetylated to improve cellular uptake [76]. These acetyl groups can have different effects on toxicity depending on their number and placement, hence less toxic sugars are sought after as they can be added to cell growth media at higher concentrations to improve sugar incorporation. For example, acetylated 2-azido-2-deoxy-glucose has been used to label *O*-GlcNAc modifications, with removal of the acetyl group at the 6-position affording a partially protected sugar that is tolerated by mammalian cells up to 200 µM [103]. Similarly, the 1,3,4-*O*-butanoylated ManNAz does not increase cell death at concentrations up to 400

µM, whereas the 3,4,6-*O*-butanoylated ManNAz significantly increases apoptosis at just 50 µM [124].

To test the toxicity of the sugars, HEK293 cells were initially cultured on 10 cm plates for 72 hours, in media containing up to 100 µM sugar. Examples of typical cells are shown in Figure 2.3.

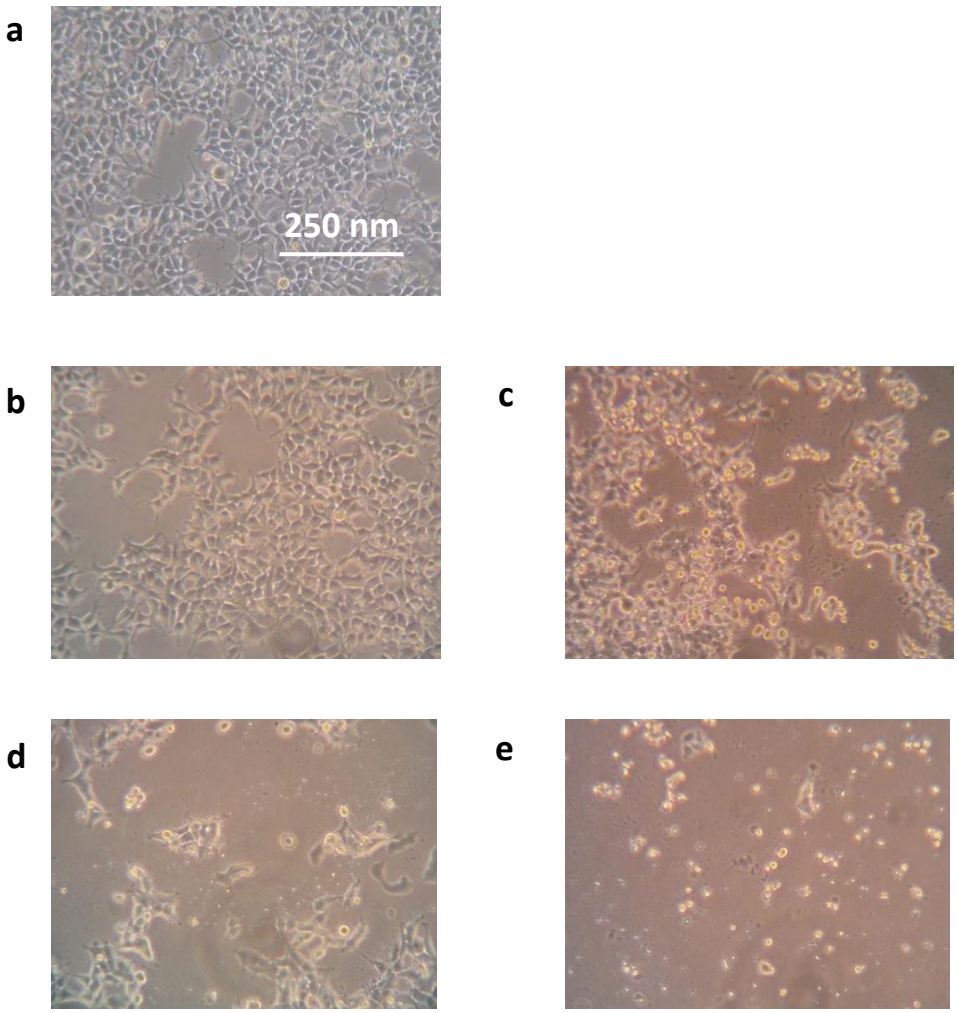


Figure 2.3: Comparison of cell growth in mannosamine derivatives. (a) With the addition of only ethanol (control) cells grew to 80% confluence in 72 hours from an initial seeding of 2 x 10^6 cells. (b) In AcManGN$_3$ (100 µM) cells grew to an average of 80% confluence, and (d) (same plate, different area) patches of dead cells and debris appear. (c) In AcManGGN$_3$ (100 µM) cells grew to 50% confluence and have a more granulated appearance, and (e) (same plate, different area) patches of debris appear.

The cells exposed to 100 µM $AcManGN_3$ **4** grew to about 80% confluence in 72 h, while cells in 100 µM $AcManGGN_3$ **26** grew to about 50% confluence in 72 h. In addition, there appeared to be more fragments of dead cells at 100 µM sugar compared with the control cells, suggesting a higher death rate.

Cells cultured on plates pre-coated with sugar to a concentration of 100 µM $AcGlcGN_3$ **6** grew to about 60% confluence, and cells in 100 µM $AcGlcAN_3$ **28** to about 50% confluence in 72 h (Figure 2.4). Microscopic examination again showed an increase in fragments of dead cells at higher concentrations.

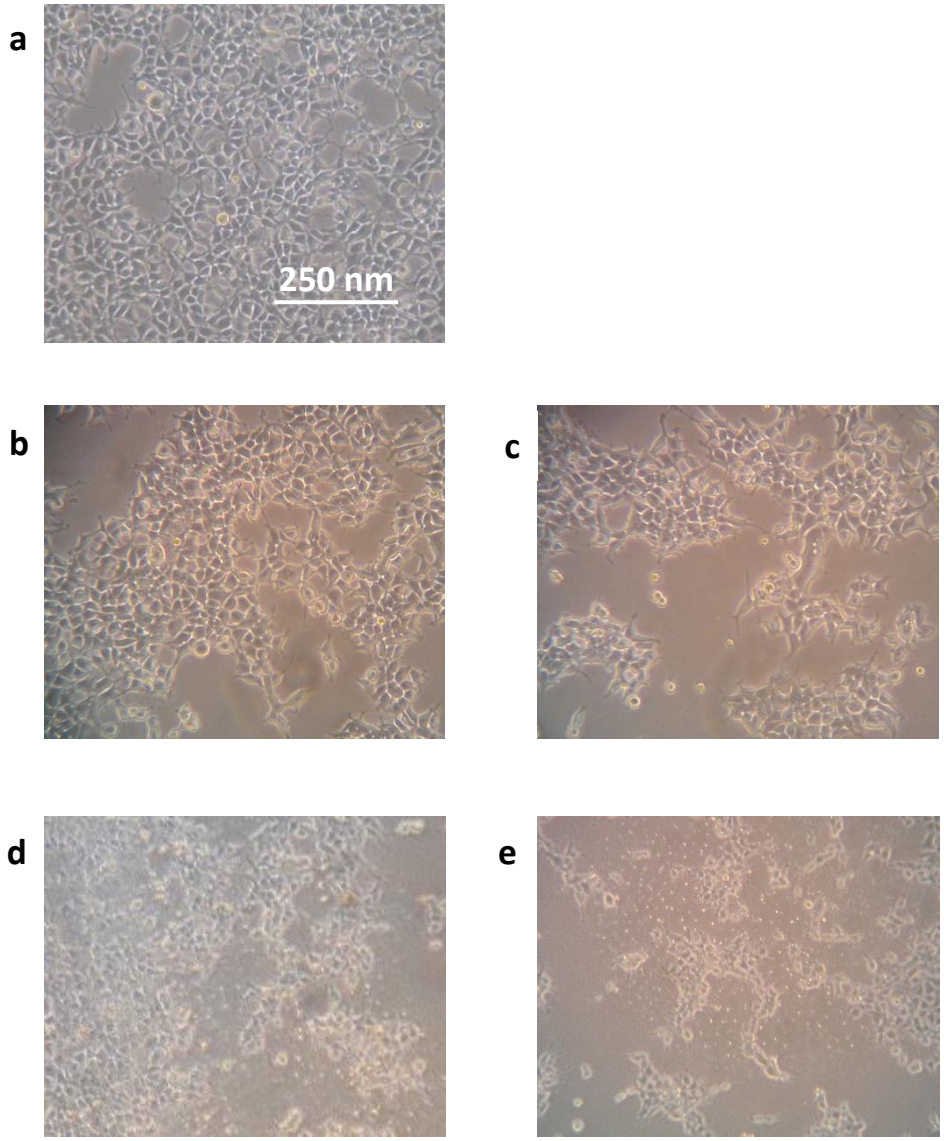


Figure 2.4: Comparison of cell growth in glucosamine derivatives. (a) In ethanol (control) cells grew to about 80% confluence in 72 hours from an initial seeding of 2 x 10^6 cells. (b) In $AcGlcGN_3$ (100 µM) cells grew to 60% confluence. (c) In $AcGlcAN_3$ (100 µM) cells grew to 50% confluence. Patches of dead cells and debris appear with both $AcGlcGN_3$ (d) – same plate as (b), and $AcGlcAN_3$ (e) – same plate as (c) at 100 µM.

Trypan Blue Viability Count

To quantify cell death, HEK293 cells were cultured in triplicate in 6 well plates and live and dead cells were counted using Trypan blue exclusion [76]. This showed that treatment with mannosamine derivatives decreased the number of viable cells per mL

as the sugar concentration increases to 100 µM (Figure 2.5). The larger $AcManGGN_3$ **26** may have a slightly larger effect on viability.

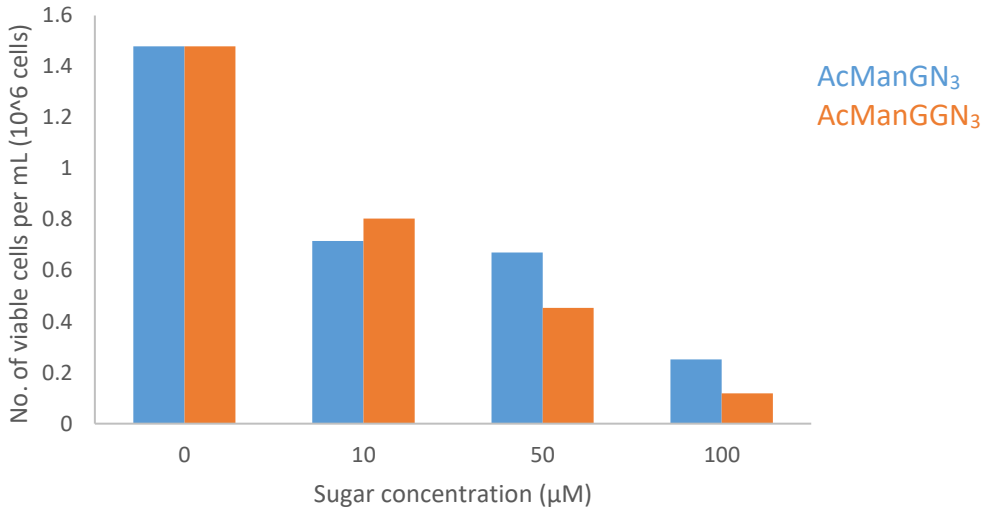


Figure 2.5: Effect of mannosamine derivatives on HEK293 cell viability quantified by Trypan blue exclusion. Number of viable cells per mL is shown with increasing sugar concentrations. Cells were treated with the indicated concentrations of sugars for 72 h and their viability measured by counting live and dead cells in a haemocytometer by Trypan blue exclusion.

The testing of glucosamine derivatives in HEK293 cells also showed a decrease in the number of viable cells per mL as sugar concentration increases to 100 µM (Figure 2.6). Curiously, here the larger $AcGlcAN_3$ **28** may be slightly less toxic.

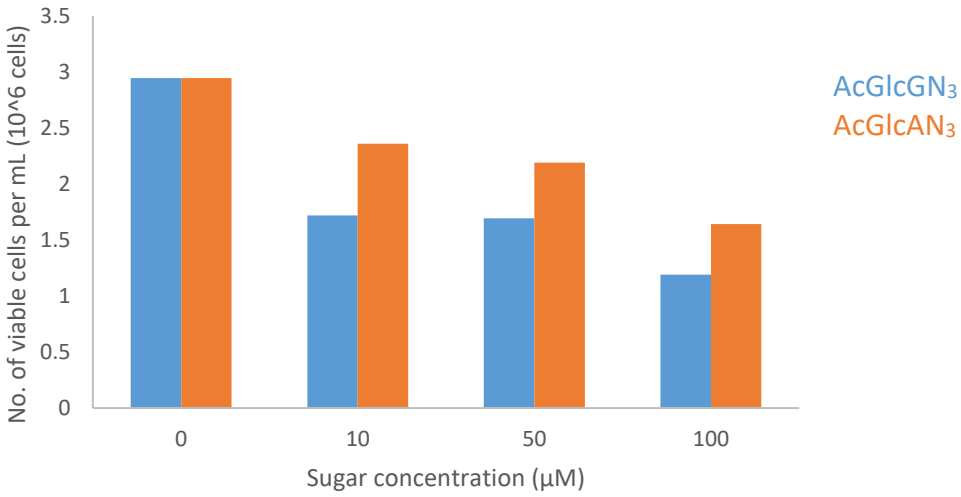


Figure 2.6: Effect of glucosamine derivatives on HEK293 cell viability quantified by Trypan blue exclusion. Number of viable cells per mL is shown with increasing sugar concentrations. Cells were treated with the indicated concentrations of sugars for 72 h and their viability measured by counting live and dead cells in a haemocytometer by Trypan blue exclusion.

AcManThz **27** was also tested for toxicity. It showed no effect on the number of viable cells per mL up to 100 µM (Figure 2.7).

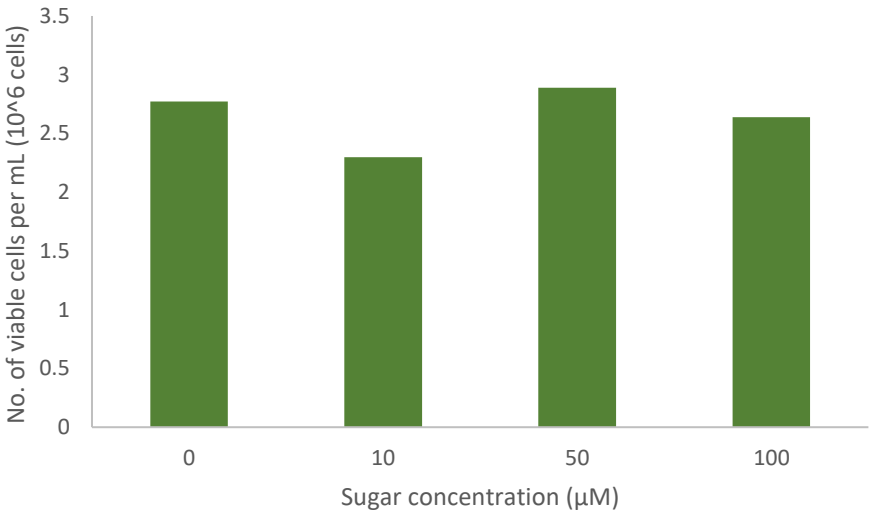


Figure 2.7: Effect of AcManThz on HEK293 cell viability quantified by Trypan blue exclusion. Number of viable cells per mL is shown with increasing sugar concentrations. Cells were treated with the indicated concentrations of sugars for 72

h and their viability measured by counting live and dead cells in a haemocytometer by Trypan blue exclusion.

The difference between the large decrease in the number of live cells and the much smaller decrease in percentage viability could be explained by the sugars inhibiting cell growth, rather than increasing cell death. Another possibility to explain the number of live cells decreasing faster than percentage viability is that the cell counting method removed relatively more dead cells than live ones, giving a higher number for percentage viability than the true value. The process required the media to be aspirated and cells washed with phosphate-buffered saline (PBS), then detached from the well plates by incubation with trypsin-ethylenediaminetetraacetic acid (EDTA). Dead cells may be removed during the aspiration and washing steps if they are not as adherent as live cells. To test this, cells were cultured in media containing 100 µM $AcManGN_3$ **4** or ethanol (control) for 72 h as before and the media and PBS was collected and centrifuged at 200 g, 4°C, for 3 min as in the protocol for labelling cells with fluorescent probe. The supernatant was aspirated and the cell pellet resuspended in 100 µl media, then mixed 1:1 with Trypan blue and cells counted under the haemocytometer (Figure 2.8).

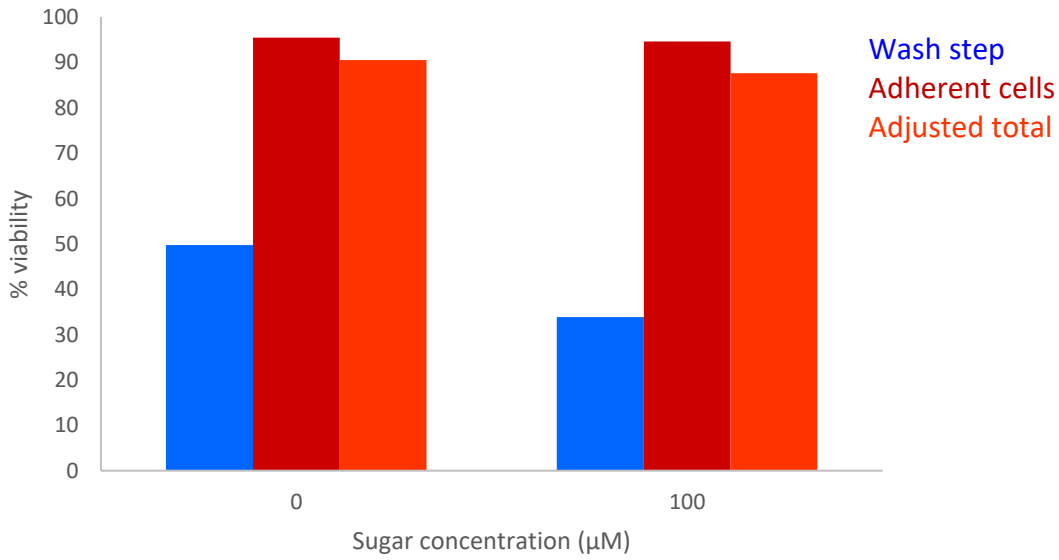


Figure 2.8: Comparison of viability of cells removed at wash step and cells removed by incubation with trypsin-EDTA. The cells removed at the wash step have a much lower percentage of live cells, though the much smaller total number of cells results in only

a small decrease in the adjusted viability. Cells were treated with the indicated concentrations of sugars for 72 h and their viability measured by counting live and dead cells in a haemocytometer by Trypan blue exclusion.

The cells removed at the wash step after incubation in 100 μM $AcManGN_3$ **4** were 33.9% viable compared with 94.6% viability in the adherent cells, a large decrease. Therefore, the dead cells were indeed much less adherent than the live cells, and a disproportionate amount of dead cells was removed by the wash step. This could potentially have caused cell viability to appear higher than the true value. However, the total number of cells in the wash step was much smaller, with the ethanol (control) cells averaging 64 cells counted compared with 372 cells counted in the adherent cell sample. The volume of PBS removed at the wash step (0.5 mL) was also smaller than the volume of adherent cell sample (1 mL). Taking this into account, the larger percentage of dead cells in the wash step meant the cell viability was adjusted downwards by 3.5% in the ethanol (control) cells and 7% in the 100 μM sugar cells as shown above (Figure 2.8). This is not a significant change.

Although the Trypan blue method of counting live cells has been previously used to establish the toxicity of unnatural sugars [76] [168] and nanoparticles [169], it had produced variable results and it was therefore decided to use an MTS assay as an independent method of measuring viability [170].

MTS Assay

The MTS assay uses 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2- (4-sulfophenyl)-2H-tetrazolium inner salt (MTS), which is reduced by the NAD(P)H-dependent oxidoreductase enzymes in living cells to a coloured formazan product [170]. It therefore measures how many viable cells are metabolising in a sample. The MTS assay is frequently used to measure the toxicity of unnatural sugar compounds [103].

The MTS assay suggests that sugars did not have a significant effect on cell viability at low concentrations (Figure 2.9). Most sugars caused cell viability to decrease as the sugar concentration increases. $AcManGN_3$ **4** and $AcManGGN_3$ **26** generally resulted in similar viabilities, with decreases to 40% and 35% viability respectively at 300 μM, whilst $AcGlcGN_3$ **6** and $AcGlcAN_3$ **28** also had similar viabilities, decreasing to 66% and 57%

respectively at 300 µM. AcManThz **27** gave an unexpected result. Viability at 100 µM appeared to be 113%, suggesting the cell may be growing and metabolising more rapidly in the sugar. However, it is most likely AcManThz **27** simply does not have a toxic effect at any concentration tested.

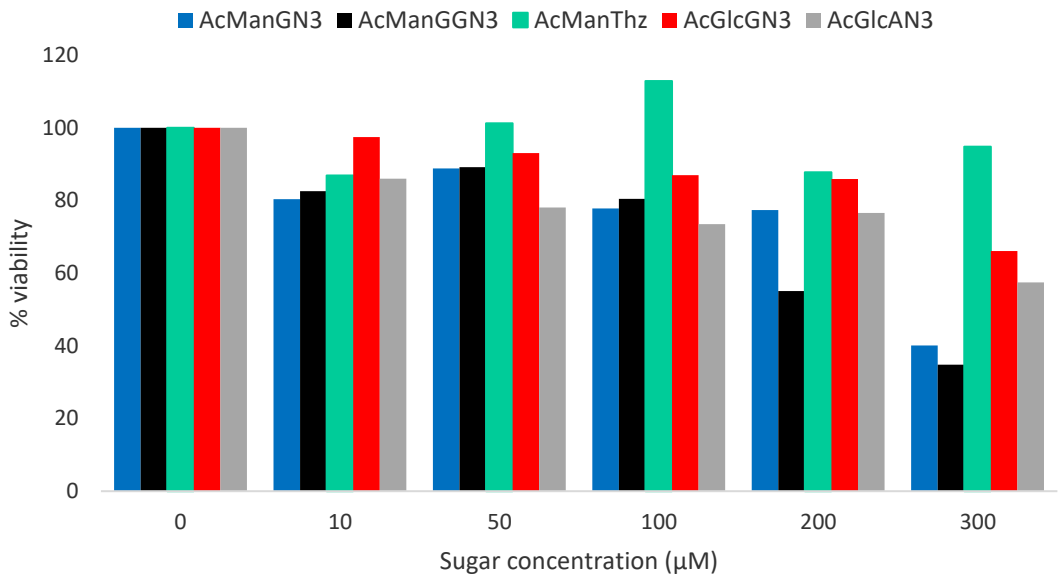


Figure 2.9: MTS viability assay of HEK293 cells treated with up to 300 µM azidopeptide sugars for 72 h and A_{490} read after 1 h incubation. Most sugars show a decrease in viability that becomes larger as the concentration increases.

Comparing MTS results with Trypan blue results for all sugars (Figure 2.10) showed broad agreement between the two assay methods up to 100 µM sugar concentration. However, the MTS viability for AcManThz **27** was unusually high at 100 µM, compared with the Trypan blue viability at the same concentration. Without unequivocal confirmation that AcManThz **27** incorporated into the cell surface, as this stage it was not possible to discount that cells are not taking the sugar up and therefore are not affected by toxicity. The MTS and Trypan blue values for $AcGlcAN_3$ **28** also show significant differences. On balance the MTS results are likely more reliable as the Trypan blue method requires multiple steps of washing and detaching cells, followed by manual counting that may be subject to operator error. The MTS assay requires one step of

reagent addition followed by scanning by a plate reader, so is less susceptible to operator error.

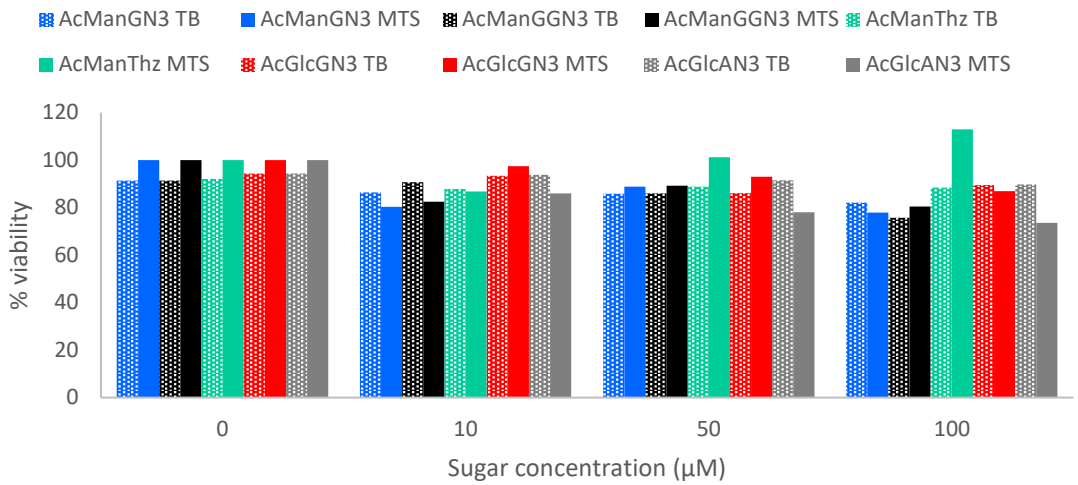


Figure 2.10: Viability of HEK293 cells treated with azidopeptide sugars for 72 h. Dotted bars are for cell viability measured by Trypan blue exclusion, solid bars are for cell viability measured by MTS assay.

Incorporation of mannosamine derivatives

The incorporation of AcManGN$_3$ **4** and AcManGGN$_3$ **26** into HEK293 cells was tested by labelling the cells with Cy5-DBCO (DBCO) dye. The reaction between azide and alkyne can be catalysed by the addition of copper (I) [90], however as copper is toxic to cells this would prevent labelling of live cells. Instead, a strain-promoted reaction with cyclooctyne was used as this does not require copper [171]. AcManGN$_3$ **4** and AcManGGN$_3$ **26** were added to HEK293 cells at 100 µM and the cells incubated for 72 h. Cells were detached by incubating with 10 mM EDTA for 5 min, then the EDTA was diluted with PBS. The cells were transferred to flow cytometry buffer and incubated in 10 µM Cy5-DBCO dye for 1 h.

The fluorescent cells were detected using fluorescence-activated cell sorting (FACS) with a 670/14 filter. As the cells pass single file through a capillary tube, they are illuminated by a laser beam and the light detected at various angles [172]. FSC-A is the amount of forward scatter and measures light diffraction around the cell, and hence the cell size. SSC-A is the amount of side scatter and measures light at a 90° angle to the laser, which

is scattered by the internal cell components, and hence measures the cell granularity. Combining these measurements allows the live cell population to be selected. The live population of 61.3% of events is shown to the lower left of the graph (Figure 2.11). Cells shift up and to the right as they begin to die as they first become larger and more granulated. The dying cells then shift left as they become smaller, and finally end up in the lowest left corner as debris, shown by the 24.9% gate.

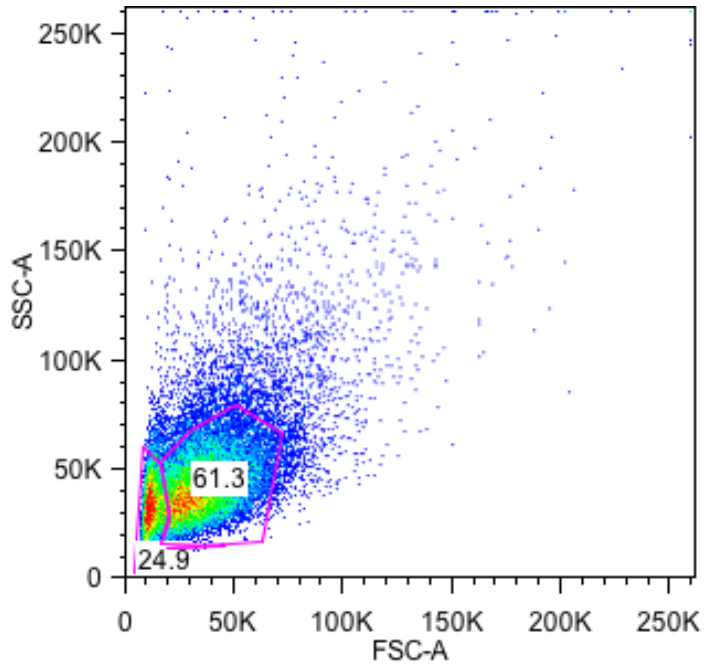


Figure 2.11: FACS FSC-A – SSC-A scatter plot showing population of 10000 events. Cells were incubated in AcManGN$_3$ and AcManGGN$_3$ for 72 h and were labelled by SPAAC with Cy5-DBCO dye. There are 61.3% live intact cells based on FSC/SSC. The 24.9% gate to the left shows dead cells and debris.

Fluorescent intensity analysis was carried out on the 61.3% live gated cells. Extracting the mean fluorescence intensity (MFI) suggested a small difference between AcManGN$_3$ **4** and the control cells, and a slightly larger difference between AcManGGN$_3$ **26** and the control cells. This suggested that AcManGN$_3$ **4** was being incorporated to a smaller extent than AcManGGN$_3$ **26**. However, the ethanol (control) tube showed the background labelling was very high. The Cyanine5 dye appeared to non-specifically bind the cell surface. Previously, sugars such as AcManGN$_3$ **4** have been incorporated into cells and fluorescently labelled at levels at least 100-fold higher than the background

fluorescence [79]. Several modifications were therefore made to the cell labelling protocol to try to reduce the high background fluorescence shown by the control cells and establish if AcManGN$_3$ **4** was also being incorporated.

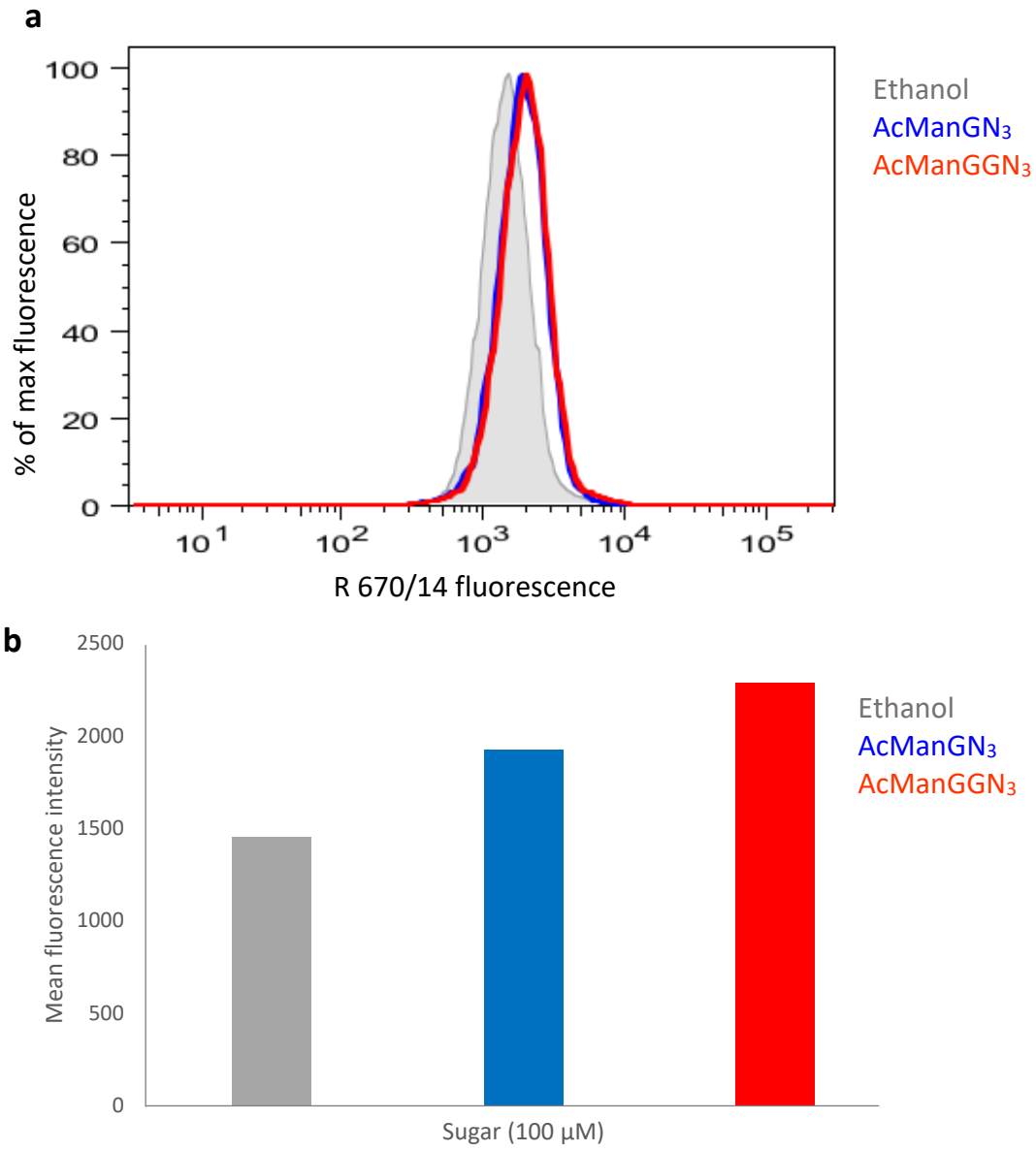


Figure 2.12: (a) Fluorescence shift suggesting incorporation of both ManNAc derivatives. The background labelling is high. (b) Averaged mean fluorescence intensity (MFI), suggesting possible incorporation of sugars.

The second method followed the previous protocol, but afterwards the cells were left in FACS fixing buffer at 4°C in the dark overnight to try to reduce nonspecific binding by

Cy5-DBCO dye. Overnight fixation of the samples did not improve the ratio between fluorescent labelling and background (Figure 2.13).

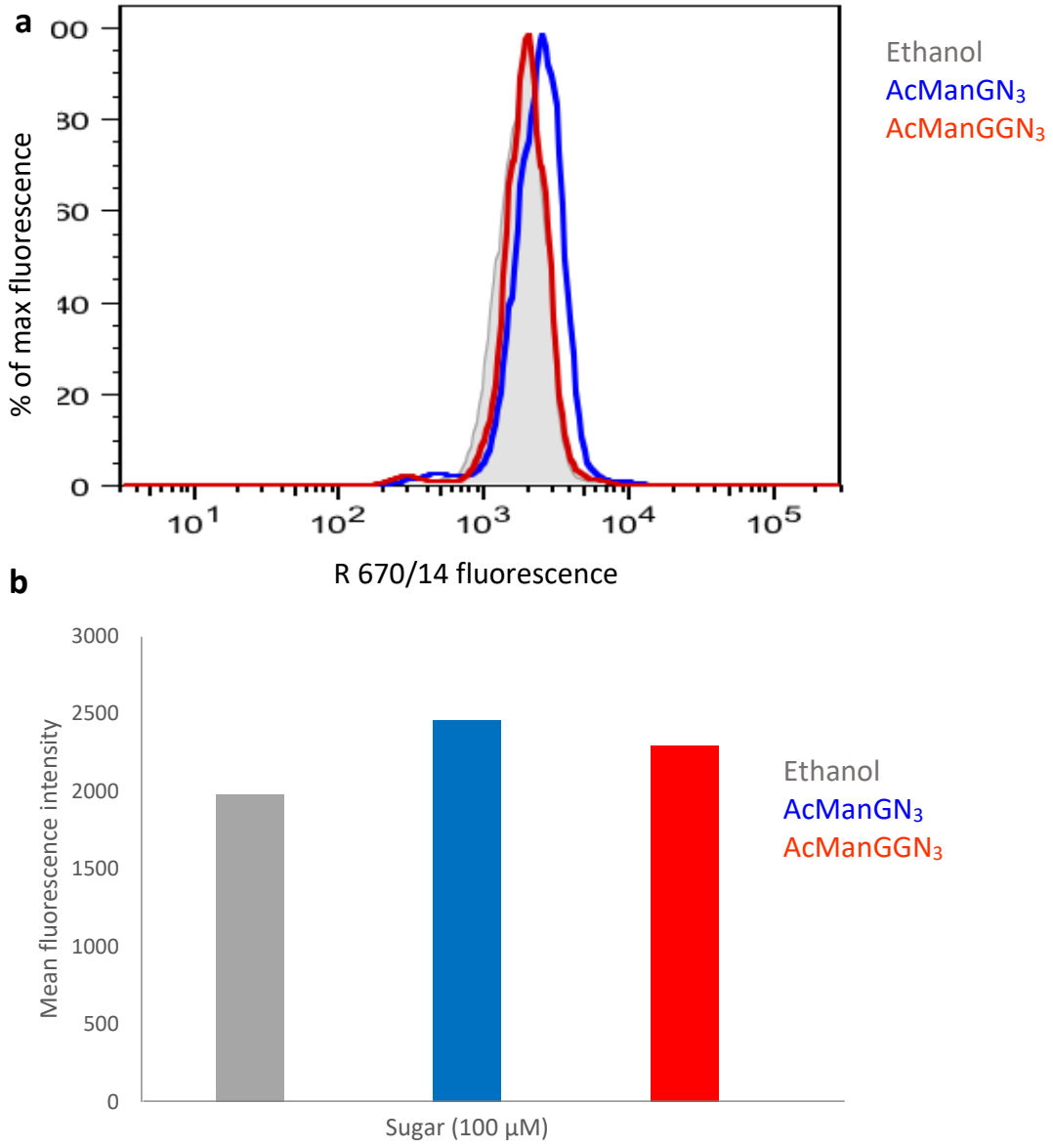


Figure 2.13: (a) Fluorescence shift due to ManNAc derivative incorporation after overnight fixation. (b) MFI shows no apparent difference between background labelling and fluorescence due to $AcManGN_3$ or $AcManGGN_3$.

A third method was designed to be closer to a published protocol for labelling chondrocytes with DBCO-650 dye [79]. The cells were incubated in PBS with 10 µM Cy5-DBCO in the dark for 1 hour before detaching with EDTA and transferring to flow cytometry buffer. Altering the order of labelling and detaching cells also did not improve

the ratio between fluorescent labelling and background (Figure 2.14). There was also more variation in how fluorescent the cells are, as a smaller population of poorly fluorescent cells appeared to the left of the main peak. The presence of the media in this experiment may interfere with the cell uptake of the dye.

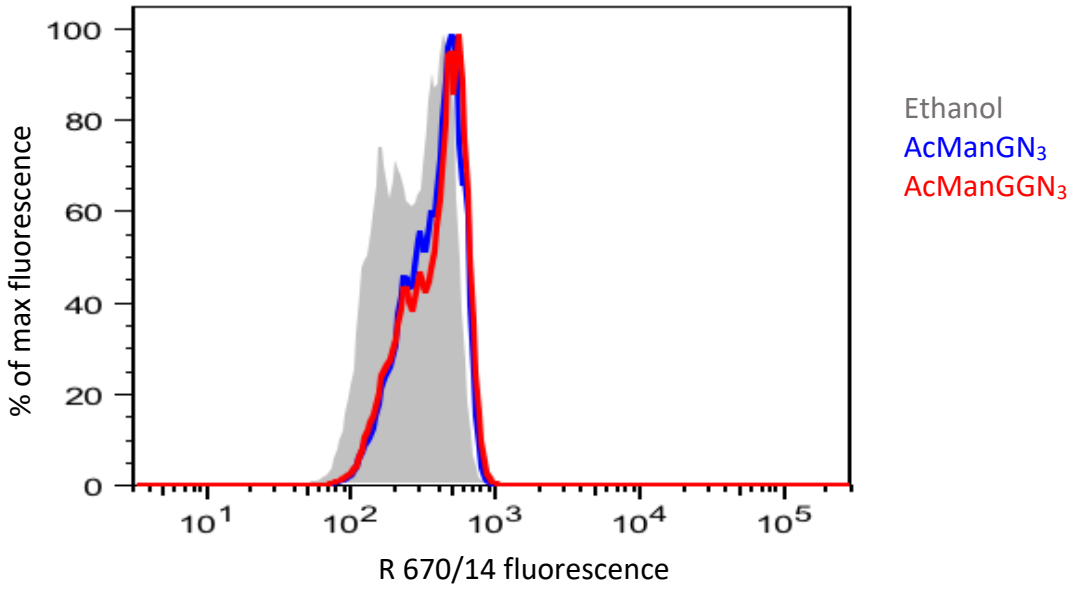


Figure 2.14: Fluorescence shift due to incorporation of ManNAc derivatives with labelling carried out before cells were detached using EDTA. The fluorescent signal is worse and there is no difference between the background labelling, AcManGN$_3$, and AcManGGN$_3$.

It was also found that using polystyrene tubes instead of polypropylene tubes during the 1 h incubation on ice gave results where there was no difference between control cells and cells in sugar (Figure 2.15). The Cy5 dye is positively charged and may stick to the polystyrene.

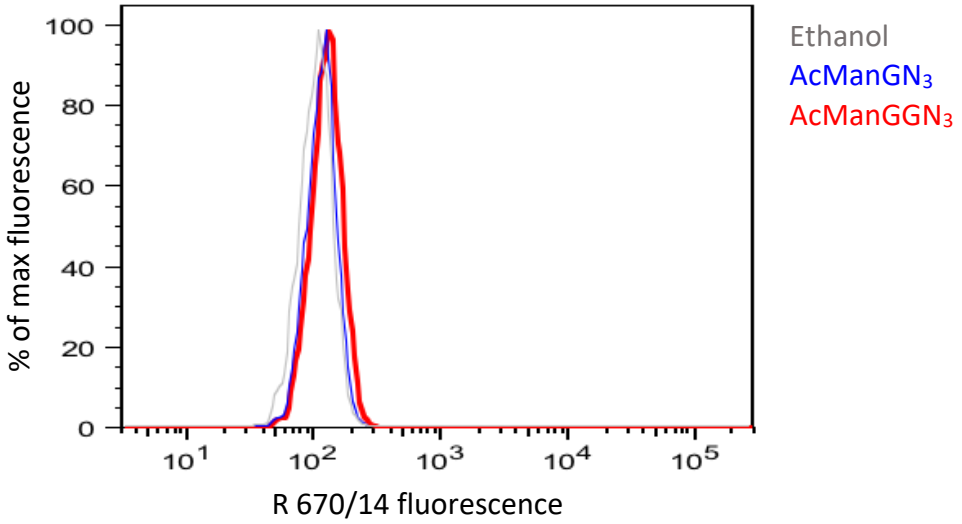


Figure 2.15: Fluorescence shift due to incorporation of ManNAc derivatives and labelling in polystyrene FACs tubes. The only fluorescent signal is the background from non-specific labelling.

Another protocol change tested rocking the cells at 4°C for 1 h instead of keeping stationary tubes on ice for 1 h. This also did not improve the labelling, and the percentage of live cells gated dropped to between 5-17%, which is very low (Figure 2.16). The HEK cells are probably too fragile to survive this treatment.

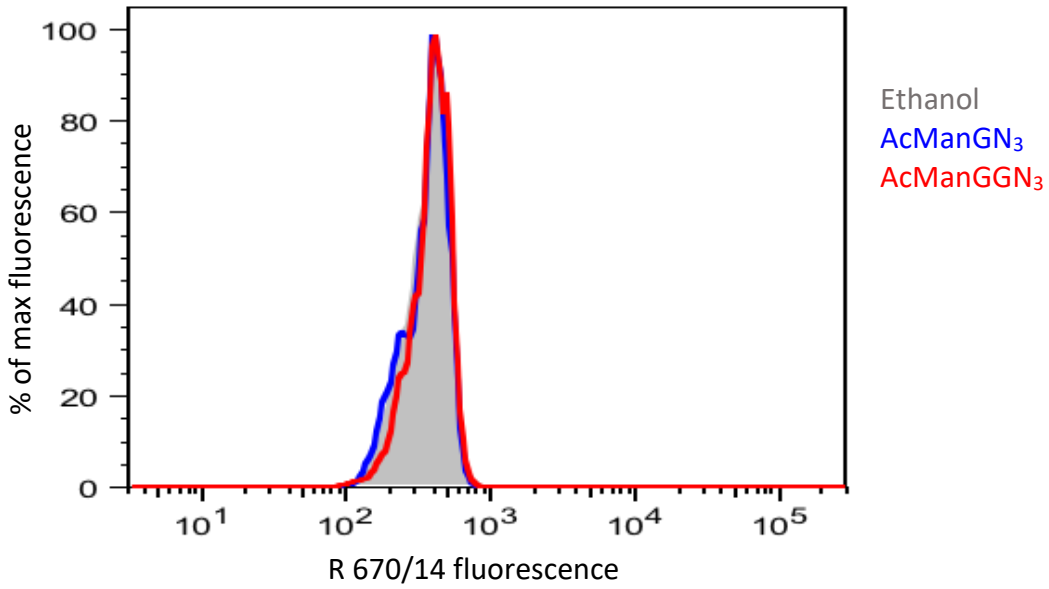


Figure 2.16: Fluorescence shift to due incorporation of ManNAc derivatives after labelling on a rocker at 4°C. The only fluorescent signal is the background from non-specific labelling.

A further protocol change did however show more promise. Cells were removed by EDTA as before, but after 5 min incubation this was neutralised with media instead of diluted with PBS to more rapidly stop the effect it has on the cell membranes. This was followed by an extra wash step in PBS to remove the media-EDTA mixture before transferring to FACS buffer. The FACS results suggested a small difference between the cells that have incorporated ManNAc derivatives and the control cells (Figure 2.17a).

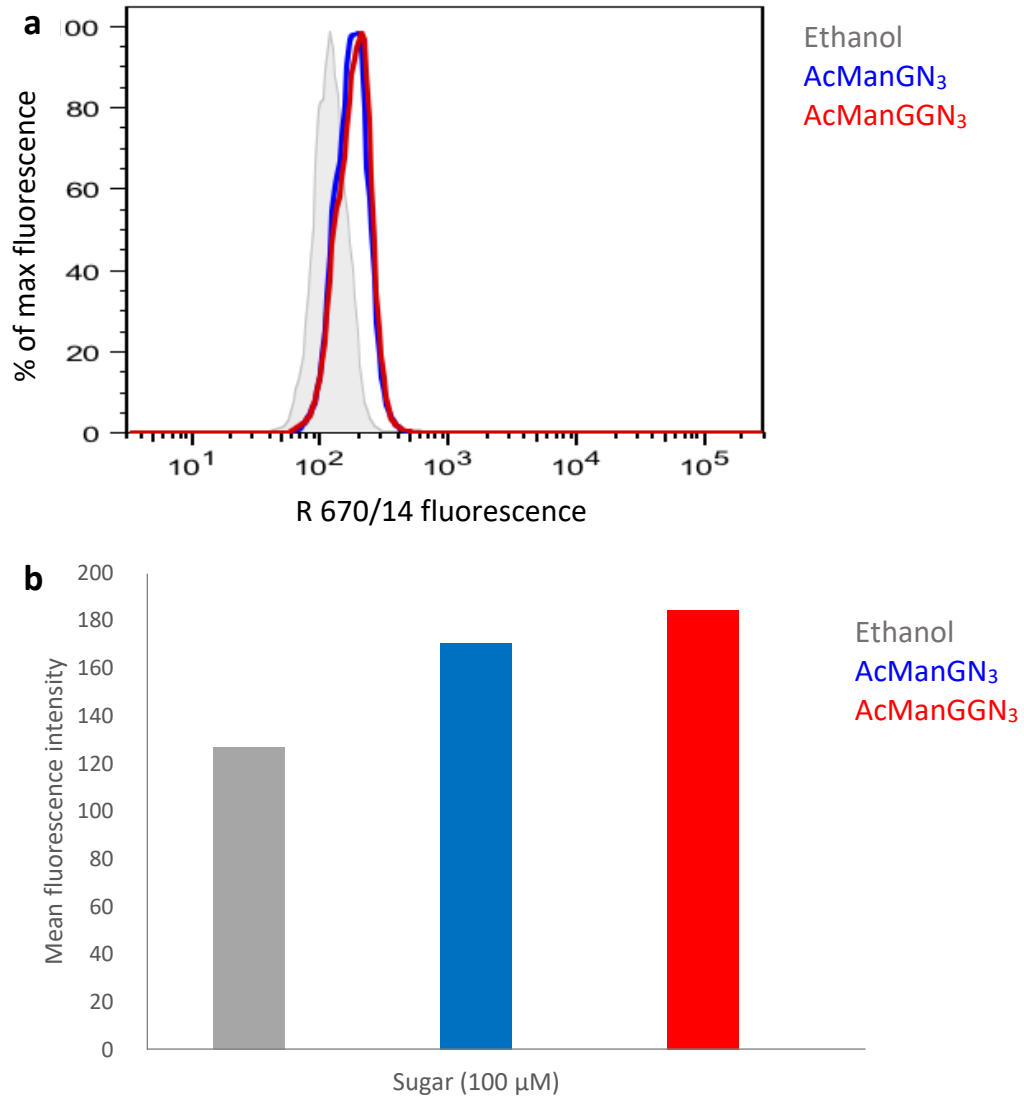


Figure 2.17: (a) Fluorescence shift due to incorporation of ManNAc derivatives showing no apparent difference between $AcManGN_3$ and $AcManGGN_3$. (b) Averaged MFI. There appears to be a small increase in fluorescence.

Pooling results from multiple experiments confirmed the incorporation of both mannosamine-based sugars (Figure 2.18). The MFI for each sample was divided by the smallest control MFI from that experiment to compare fluorescence ratios, and a t-test performed on 3 experiments each containing 3 replicates of each sugar and the control (n=3). The P values for this were 0.012 for $AcManGN_3$ **4**, 0.052 for $AcManGGN_3$ **26** and 0.554 when comparing the two sugars. As the value is significant for $AcManGN_3$ **4** and just outside of significance for $AcManGGN_3$ **26**, it is likely both sugars are incorporated. However, the level of fluorescent increase was below previous work incorporating $AcManGN_3$ **4** [79]. The low level of labelling may be due to side reactions between alkyne and thiol [173].

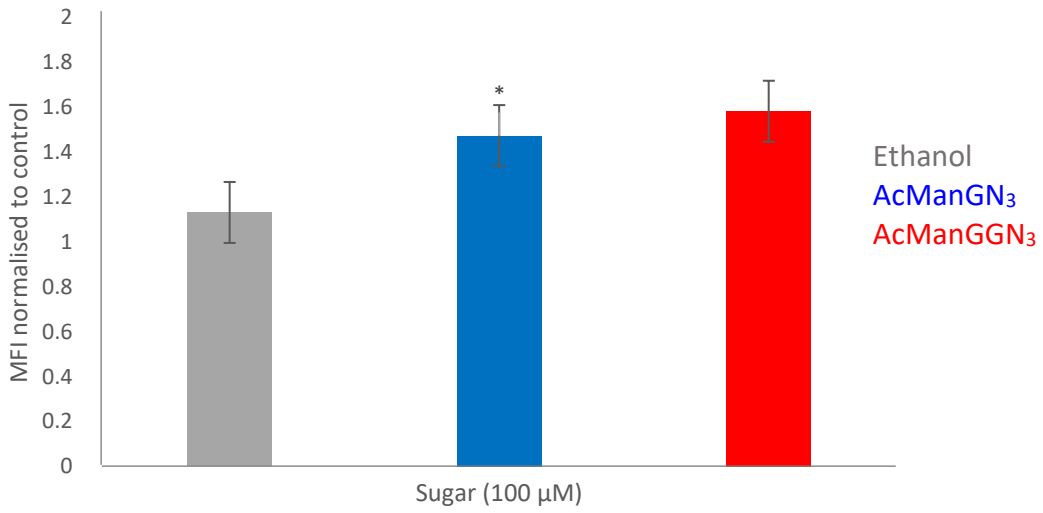


Figure 2.18: Incorporation of the mannosamine-based sugars into HEK293 cells. The average MFI for a pooled set of 9 total replicates for each sugar across 3 experiments. $AcManGN_3$ shows a significant increase in fluorescence (P = 0.012), for $AcManGGN_3$ P = 0.052. Error bars represent mean ± S.E.M. of biological triplicates (n=3). * = P < 0.05.

Next, different concentrations of sugars were tested. Although a higher concentration of sugars should lead to better incorporation, if the sugars have a toxic effect on the cells it may lead to fewer cells being labelled overall. The sugar incorporation and Cy5-DBCO labelling was therefore tested at a range of concentrations to find the optimum.

Cy5-DBCO labelling of $AcManGN_3$ **4** shows the highest increase in fluorescence at 100 µM (Figure 2.19).

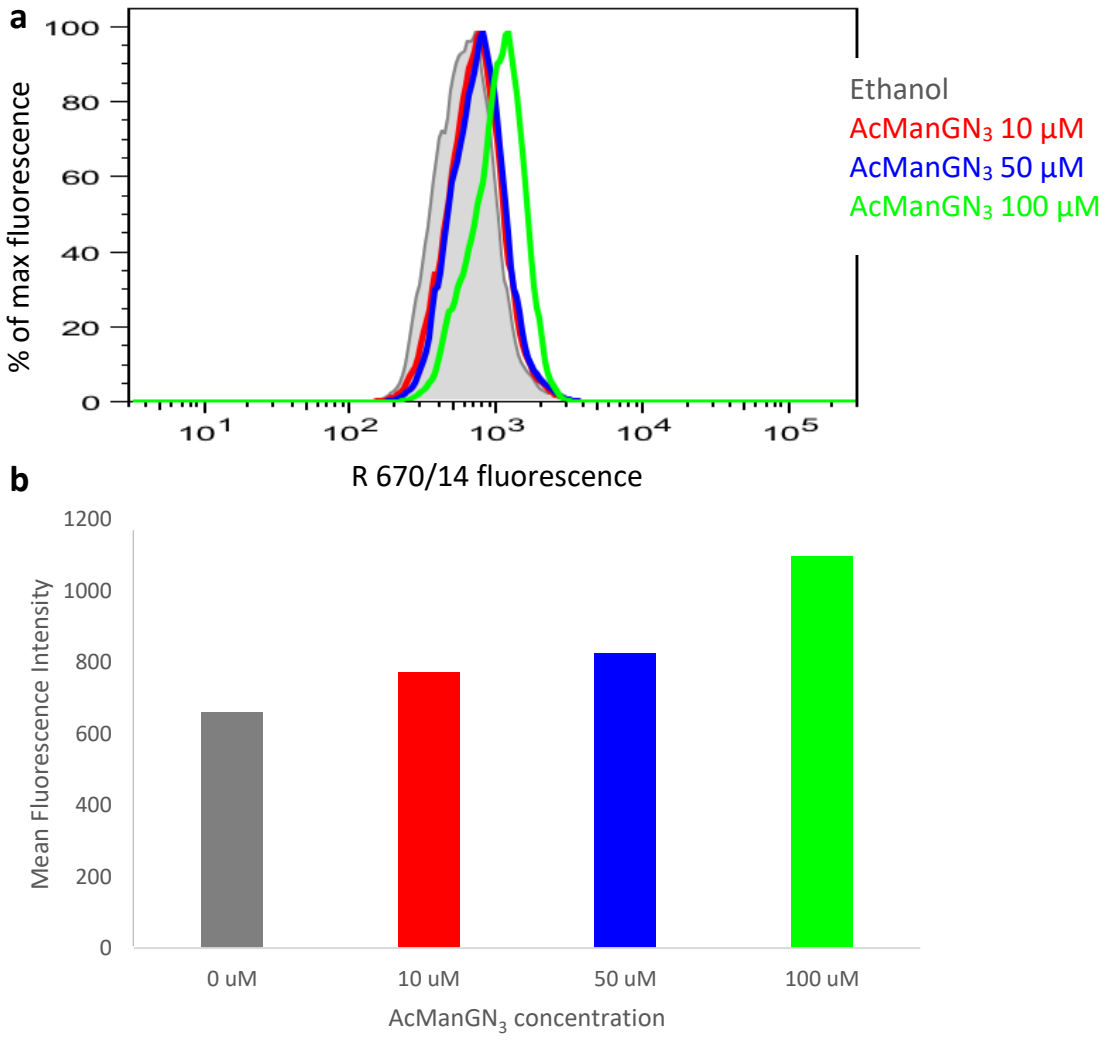


Figure 2.19: (a) Fluorescence shift due to incorporation of $AcManGN_3$ at different concentrations. (b) Averaged mean fluorescence intensity (MFI).

Cy5-DBCO labelling of $AcManGGN_3$ **26** also shows the largest increase in fluorescence at 100 µM (Figure 2.20). Therefore, both sugars would be optimally added to 100 µM for sortase labelling.

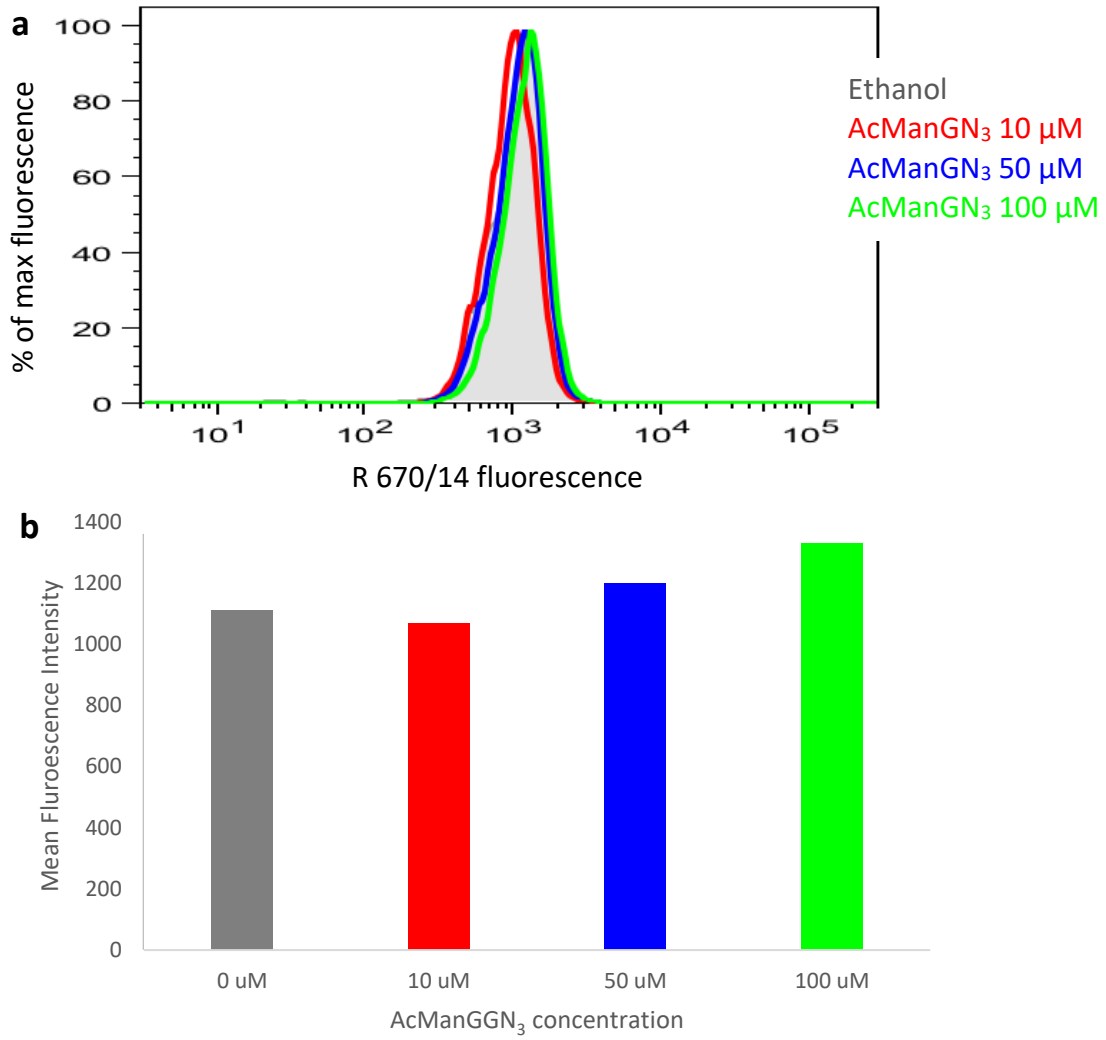


Figure 2.20: (a) Fluorescence shift due to incorporation of AcManGGN$_3$ at different concentrations. (b) Averaged mean fluorescence intensity (MFI).

It was also tested whether the labelling reaction could be performed at a higher temperature. Incubating at 4°C stops internalization of the surface glycans that are being labelled, however this will also slow down the click reaction, and especially slow down the sortase. It was initially planned to incubate the sortase at 4°C to prevent internalization of glycans and increase the sortase concentration to compensate for decreased activity, but it would be better to use sortase to label at its optimal temperature of 37°C [136]. Previously, click reactions have been performed at various temperatures. The Bertozzi group uses room temperature [68], with the secondary label (if used) added on ice to stop glycan internalization [61]. Boons *et. al.* found the click reaction between strained-alkyne probe and azides on CHO cells to be equally efficient

at 4°C or room temperature [174]. Yoon *et. al.* and Baskin *et. al.* have used click reactions at 37°C [171] [79]. Therefore, click labelling was tested at RT and 37°C, particularly to see if the temperature change had any effect on the cell viability.

Cell viability ranged from 27-48% at 4°C, 25-35% at 22°C, and 22-37% at 37°C, so the temperature increase only slightly decreases cell viability. The labelling at 4°C resulted in an increase in fluorescence that was not statistically significant (Figure 2.21a),. The labelling at 22°C appears slightly better (Figure 2.21b). The labelling gives the best signal-to-noise ratio at 37°C, as both sugars cause about a 2-fold increase in fluorescence (Figure 2.21c). Labelling at a higher temperature has still not given the 10-fold difference between sugars and control that have been previously reported [73]. However, as this labelling protocol used FITC-avidin, which has 3-6 fluorophores per avidin, so the signal would be amplified. Therefore, the cell labelling appeared to work and it also confirmed the cells can survive a labelling protocol at 37°C, which will be optimal for sortase activity (Figure 2.22).

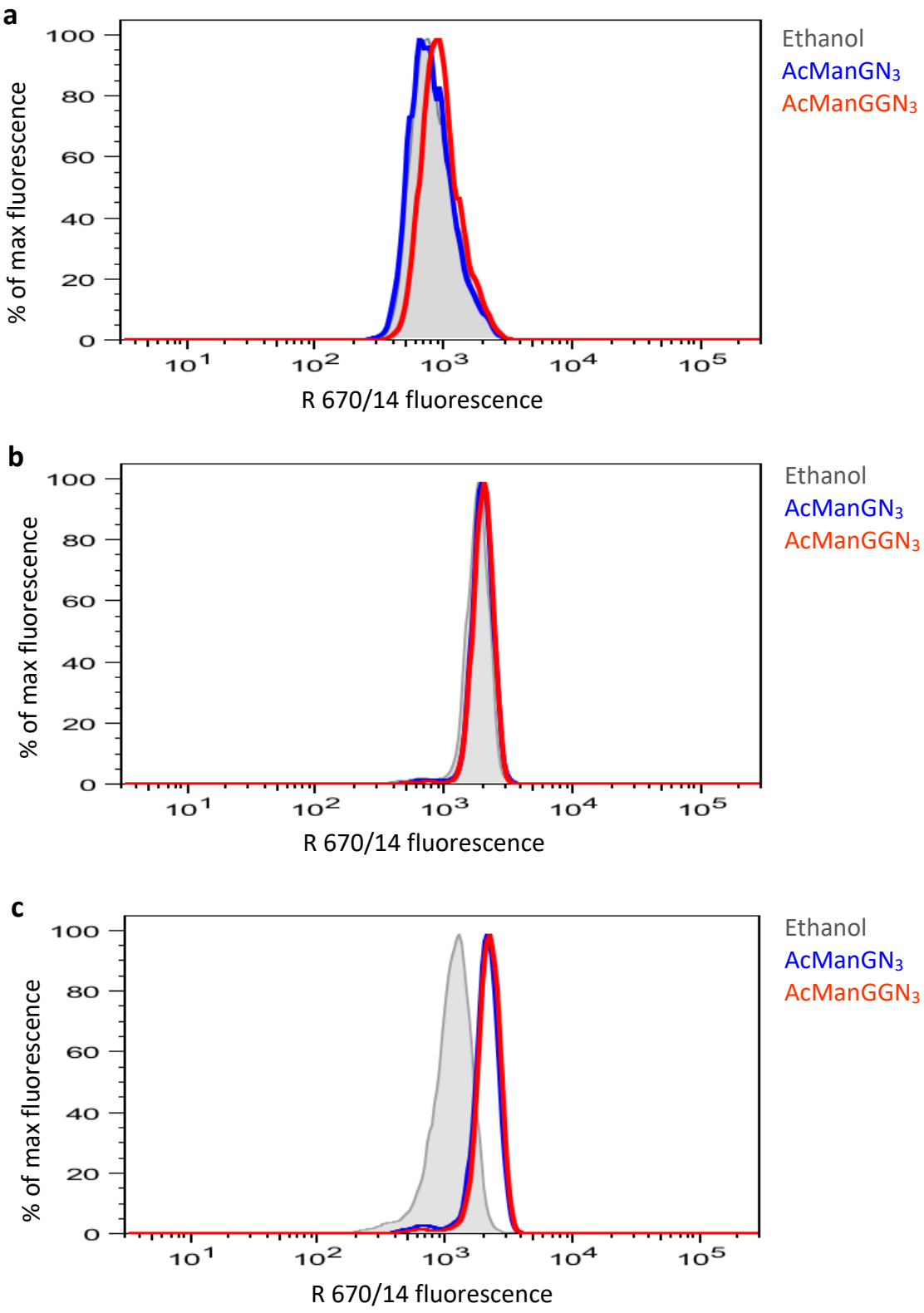


Figure 2.21: Fluorescence shifts due to incorporation of azidopeptide sugars and labelling by click reaction performed at (a) 4°C, (b) 22°C, and (c) 37°C.

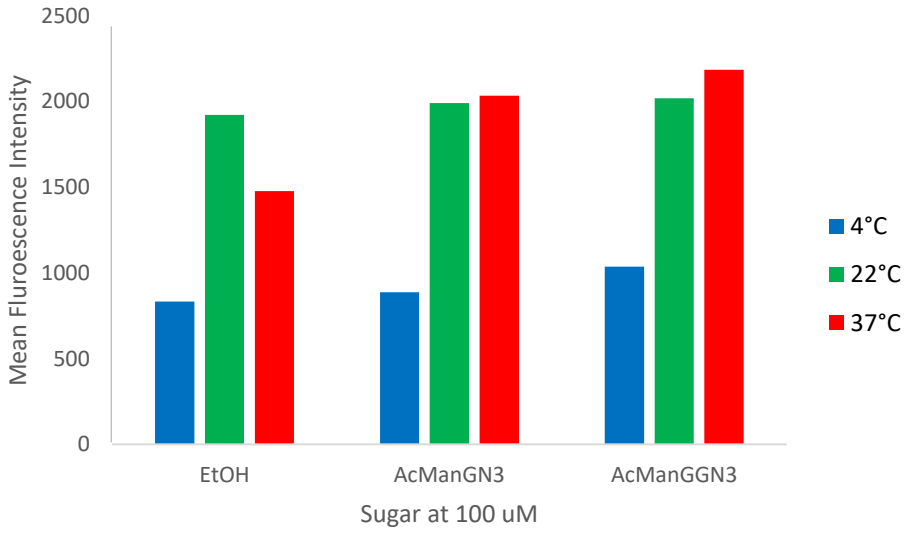


Figure 2.22: Averaged mean fluorescence intensity (MFI) of HEK cells after click labelling at various temperatures.

Incorporation of glucosamine derivatives

$AcGlcGN_3$ **6** and $AcGlcAN_3$ **28** were tested for their incorporation into intracellular glycoproteins. At the first labelling attempt, cells labelled with $AcGlcGN_3$ **6** grew to about 60% confluence, and cells labelled with $AcGlcAN_3$ **28** to about 50% confluence, with the ethanol (control) to about 80% confluence in 72 hours. Cells were removed from the tissue culture dish by trypsin, which would also remove glycans from the cell surface and lysed by the addition of 100 µL lysis buffer for 30 min on ice. This was done to detect only intracellular glycans. The proteins were initially labelled using the Click-iT TAMRA Protein Analysis Detection Kit (ThermoFisher) and run on SDS-PAGE. This protocol failed to label the proteins. A second protocol used copper-catalysed labelling with a TAMRA-alkyne with THPTA ligand [113] to accelerate the reaction. This successfully labelled proteins, and $AcGlcGN_3$ **6** was shown to incorporate into glycoproteins, but incorporation of $AcGlcAN_3$ **28** appears to be very low as the fluorescence is slightly above the background (Figure 2.23).

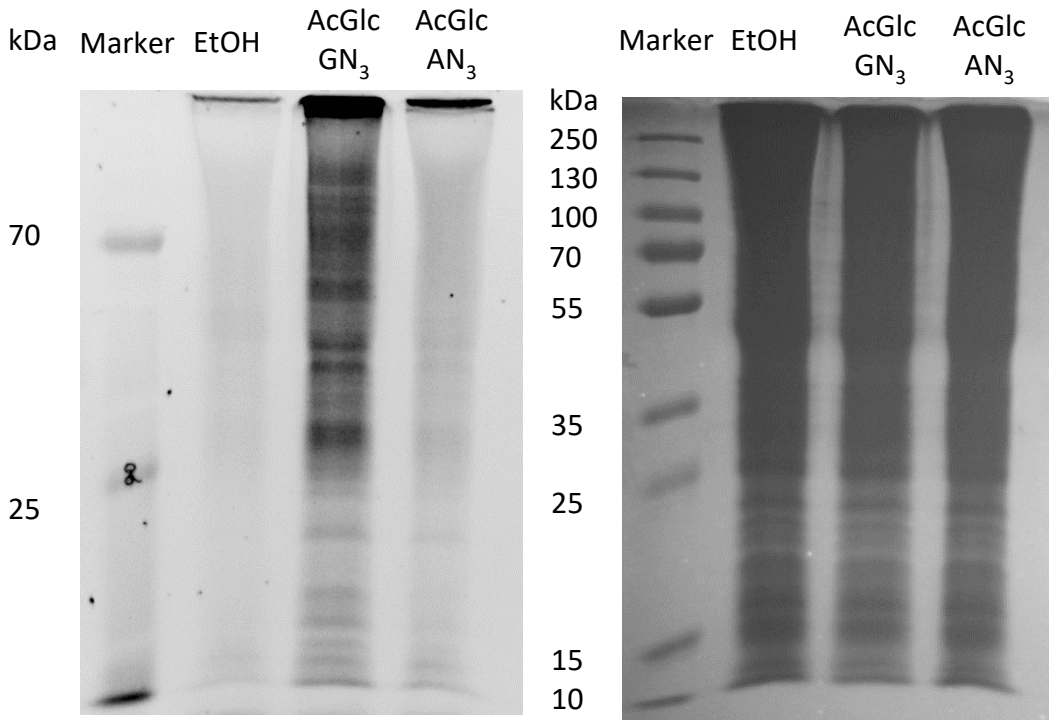


Figure 2.23: Incorporation of GlcNAc derivatives into intracellular *O*-linked glycoproteins. The fluorescent labelling with TAMRA on the left gel (colours inverted for clarity) shows AcGlcGN$_3$ has incorporated well, but AcGlcAN$_3$ is only slightly higher than the background fluorescence. The Coomassie-stained gel on the right shows equal amounts of protein having been loaded in each lane.

CHO cells

Labelling was tested in Chinese Hamster Ovary (CHO) cells on the basis that incorporation of unnatural sugars varies greatly between cell lines, with HeLa, HL-60, Jurkat and CHO varieties all incorporating AcManGN$_3$ **4** and subsequently giving at least a 10-fold increase in fluorescence after labelling with phosphine-FLAG followed by FITC-anti-FLAG antibody [61]. When azide sugars were fed to CHO K1 cells subsequent attempts at labelling failed, as no fluorescence above background could be detected. The copper-free reaction between AcManGN$_3$ **4** and Cy5-DBCO was confirmed to work *in vitro* (Figure 2.24). This suggested that the CHO cells were not incorporating the sugars, although it was not clear why. It is possible that different strains of CHO cells have been used in the literature, and not the CHO K1 cells tested.

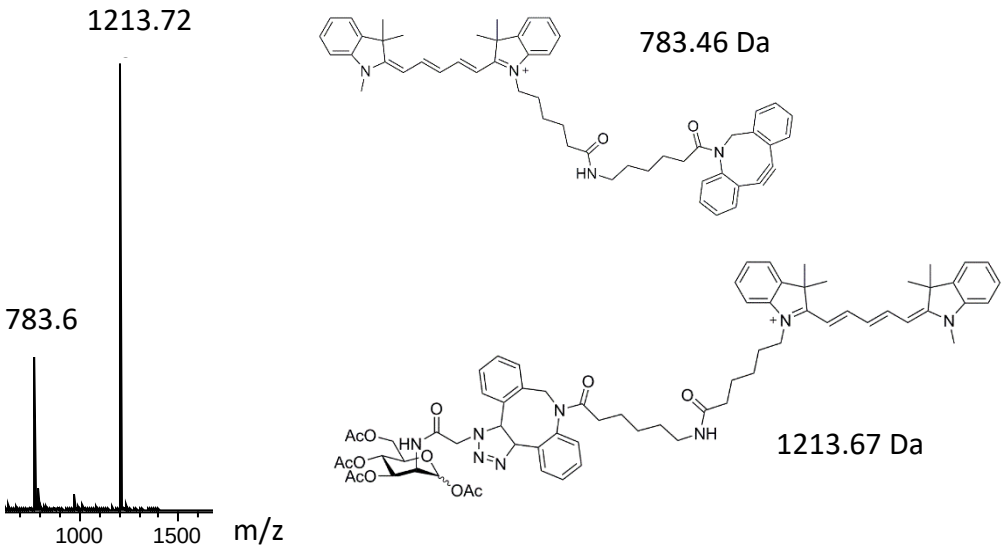


Figure 2.24: Reaction between AcManGN$_3$ and Cy5-DBCO *in vitro*. After 1 h the majority of the strained-alkyne dye has reacted with the azide group on the sugar.

Conclusion

Here we showed that synthesised sugar analogues combined with "on-cell" click chemistry allows for specific labelling of mammalian cell surface glycans, although the efficiency of the process remains low and further optimisation is required.

The novel sugars AcManGGN$_3$ **26**, AcManThz **27** and AcGlcAN$_3$ **28** were tested for their effects on viability in HEK cells, having not been tested before in mammalian cells. The sugars are all well tolerated up to 100 µM. AcManThz **27** is tolerated up to at least 300 µM, which is higher than most other sugars [124]. The azide sugars were then tested for incorporation and labelling by SPAAC with a DBCO-Cy5 dye, followed by FACS analysis of cells. AcManGN$_3$ **4** and AcManGGN$_3$ **26** were successfully incorporated onto cell surface glycoproteins, although at low levels compared with most sugars in the literature [171] [79]. AcGlcGN$_3$ **6** incorporation into intracellular glycoproteins was confirmed by gel analysis and THPTA-catalysed reaction with an alkyne dye, but AcGlcAN$_3$ **28** was not suggesting a possible issue of sensitivity that remains to be resolved. The next step was to use a reducing agent on the incorporated azide sugars to convert to amine sugars. This was to allow sortase to recognise the G and GG motifs on the cell glycoproteins, and ligate a probe bearing a fluorophore and the LPETG motif in order to fluorescently label sugars in live cells.

Chapter 3 : Sugar incorporation, sortase-catalysed labelling and OPAL labelling

Use of sortase on incorporated sugars

The previous chapter showed that sugars carrying peptide motifs that might be recognized by sortase can be detected by incorporating the azide sugar into cells and using a strained-alkyne probe to fluorescently label the cells. The next step of the project was to use sortase to label the cells. To label sugars on the cell surface with sortase, the sugars must be incorporated and reduced to the amine form (Figure 3.1).

Figure 3.1: Reduction of azide-bearing sugar to amine that can be recognised by sortase.

Previously, 2-(Diphenylphosphino) benzoic acid (2DPBA) has been used as a water-soluble reducing agent to reduce unnatural azide-bearing amino acids incorporated into cell surface proteins [144]. The conversion of the azide to amine transformed the unnatural amino acid into a GGK unnatural side chain, with the amino group of the exposed glycine accepted by sortase. 20 µM of sortase successfully catalysed labelling with a ubiquitin carrying the LPLTG sortase motif. It was therefore hoped that the AcManGN$_3$ **4** and AcManGGN$_3$ **26** sugars shown to incorporate in the previous chapter would be able to ligate fluorescent LPXTG probe in the presence of sortase after reduction to amine (Figure 3.2).

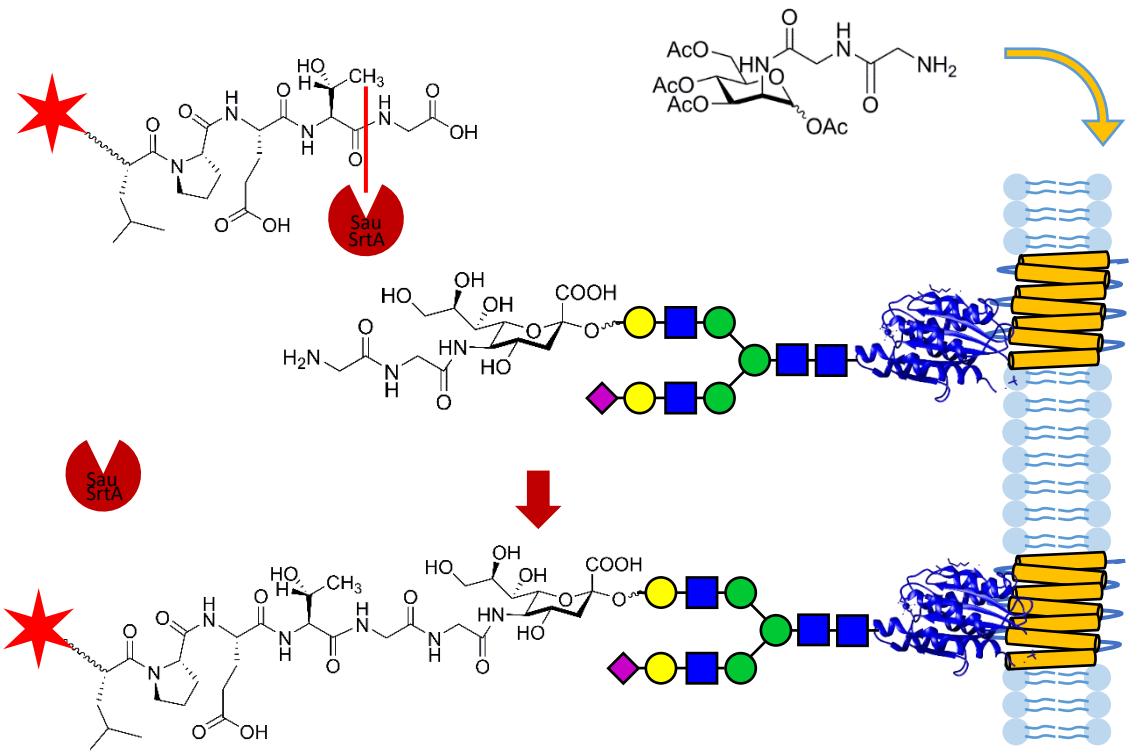


Figure 3.2: Sortase labelling of cells. MOE incorporation of AcManGGNH$_2$ will allow sortase-catalysed ligation between LPETG probe and glycine linked to sialic acid, labelling the cell without the use of "click" chemistry. This will allow reversible labelling of the glycan.

Sortase has previously been engineered by many groups to create fast variants and alter the specificity of the enzyme [153]. The fast mutant SrtA 5M has 5 mutations that overall give a 140-fold improvement in activity [150]. SrtA 7M is derived from SrtA 5M and has 2 further mutations that allow activity in the absence of Ca^{2+} ions [156]. Both of these accept the LPXTG motif (Figure 1.15).

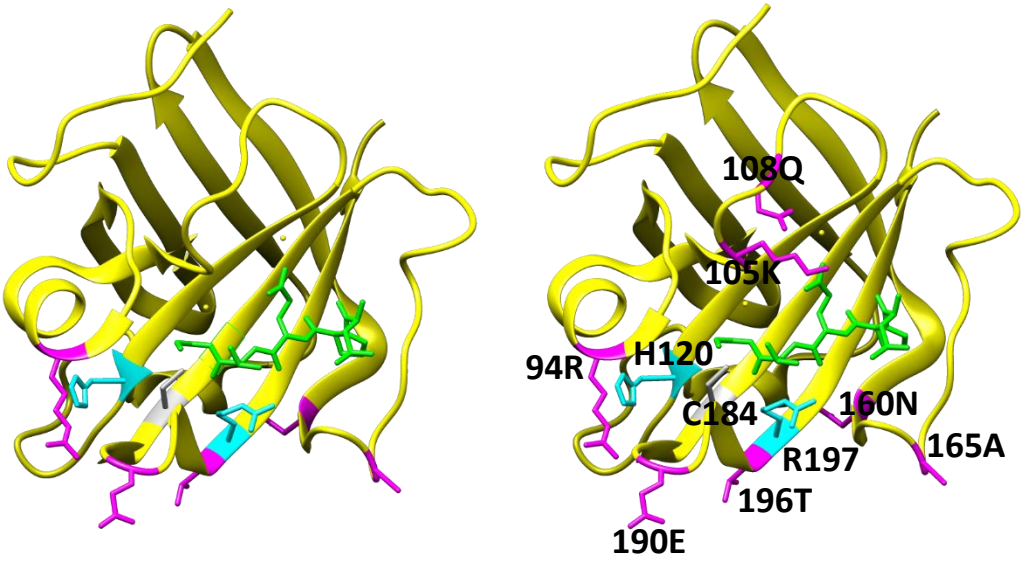


Figure 3.3: Stereo view comparing the structure of SrtA 5M (left) and SrtA 7M (right) and showing the mutations at P94R, E105K and E108Q (both in SrtA 7M only), D160N, D165A, K190E, K196T (magenta). The LPXTG peptide (green) is shown interacting with the conserved H120 and R197 residues (cyan) and the catalytic C184 (grey).

Spy SrtA (Figure 3.4), a variant from *Streptococcus pyogenes*, accepts an LPXTA motif [162].

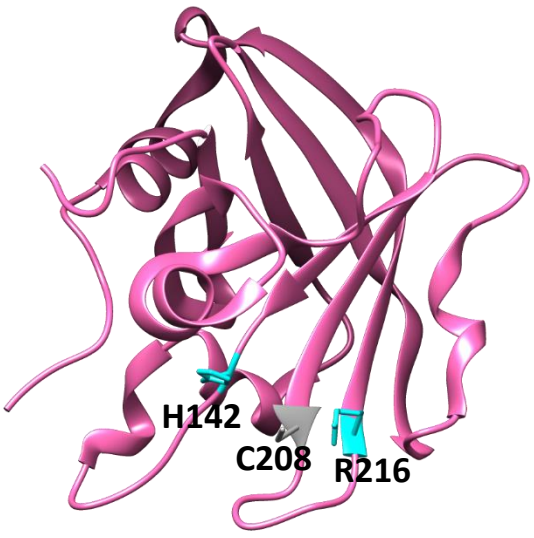


Figure 3.4: The structure of Spy SrtA, with the conserved H142 and R216 residues in cyan and the catalytic C208 in grey.

The use of multiple sortases could therefore potentially allow orthogonal labelling, as two different sugars incorporated into cells carrying different peptide motifs would in theory allow labelling by two sets of sortases and fluorescent probes in the same experiment.

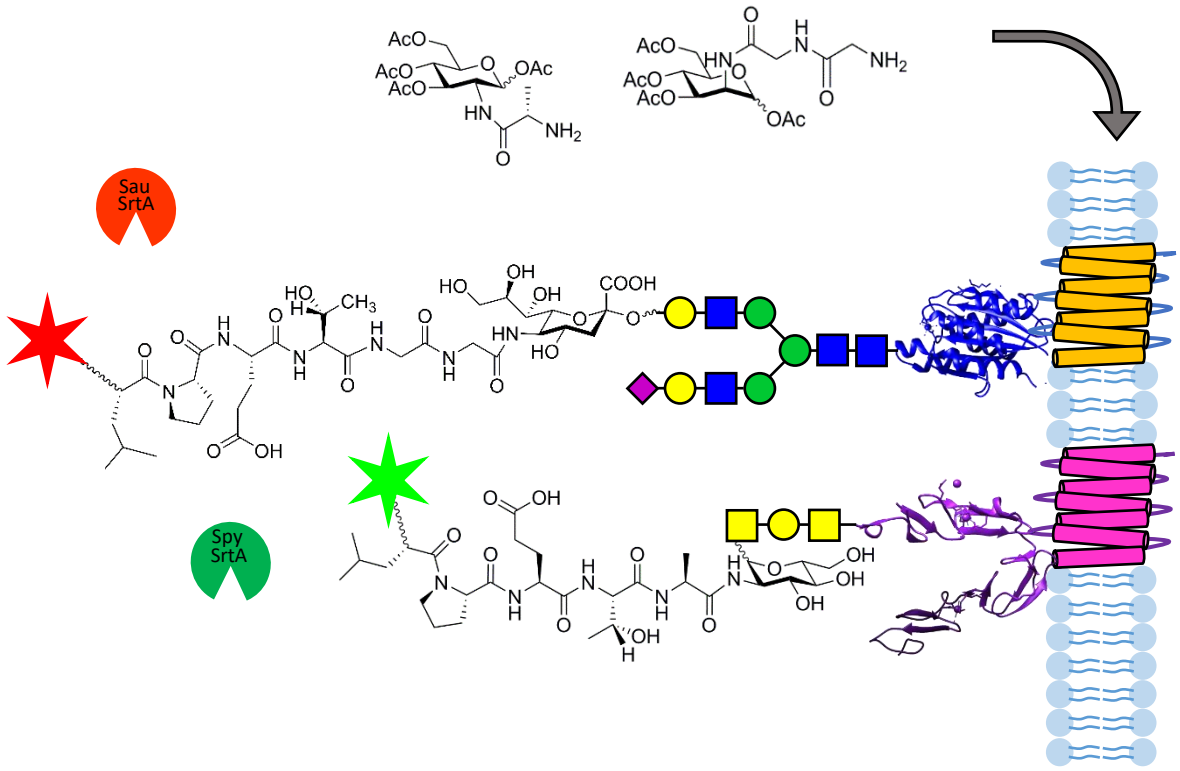


Figure 3.5: Incorporation of $AcGlcANH_2$ and $AcManGGNH_2$ would allow two probes to label two different sugars, catalysed by the sortase variants Spy SrtA and SrtA 5M.

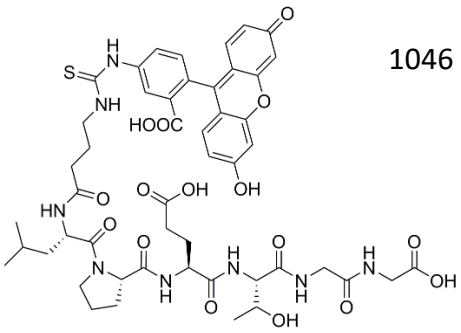
The three sortases were obtained from Addgene as bacterial stabs, which were inoculated in 10 mL LB overnight, and DNA extracted using the QIAprep miniprep protocol. The plasmids were verified by sequencing by GATC Biotech, and transformed into electrocompetent *E. coli* BL21 (DE3) for protein expression. The sortases were expressed [160], purified on a nickel column, and dialysed in HEPES ligation buffer (Table 1).

Sortase	Concentration (µM)	Volume (mL)
Spy SrtA	450	2.5
5M Sau SrtA	800	3
7M Sau SrtA	475	3.4

Table 1: Yields from the production of 3 sortases.

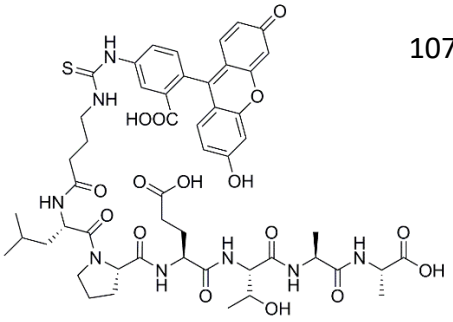
Testing 3 Sortases

First, the reaction of tripeptides to fluorescent probes carrying sortase motifs (Figure 3.6) was tested *in vitro* [145]. The probe containing LPETGG **30** was tested with SrtA 5M and SrtA 7M, the probe containing LPETAA **31** was tested with Spy SrtA. Reactions were set up in 100 µL HEPES ligation buffer with 60 µM sortase, 600 µM fluorescent probe and 600 µM tripeptide, incubated at 37°C for 1 h and samples run on LC-MS. A second set of reactions increased tripeptide concentration to 6 mM.



1046 Da

30



1074 Da

31

Figure 3.6: Fluorescein-Gaba-LPETGG and Fluorescein-Gaba-LPETAA probes for sortase-catalysed ligation to glycine-bearing and alanine-bearing sugars respectively.

The reaction catalysed by SrtA 5M successfully ligated tripeptide and fluorescent probe at 0.1 mol eq. (Figure 3.7b). If probe and tripeptide are present in equal amounts, the reaction does not go to completion as sortase also catalyses the hydrolysis of the threonine-glycine bond. The ligation reaction can be forced to completion by adding 10 times the concentration of tripeptide as fluorescent probe (Figure 3.7c).

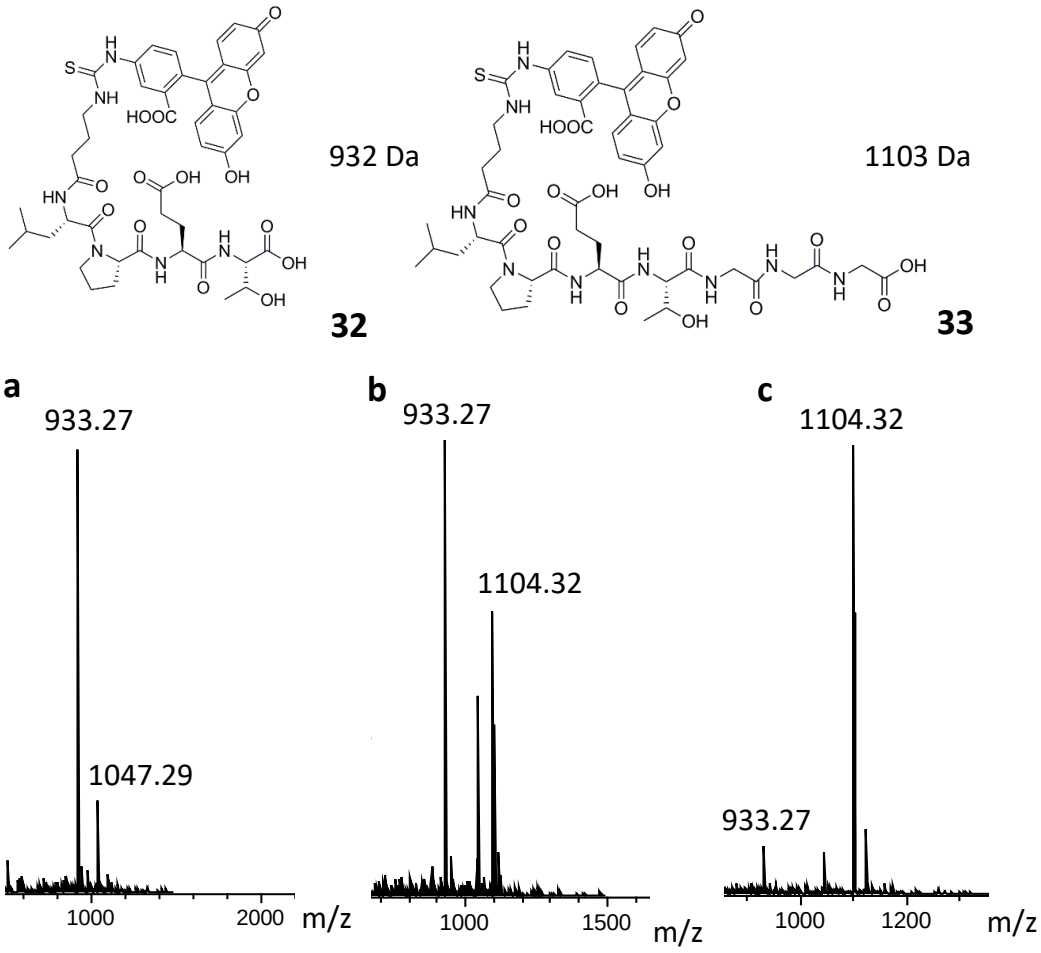
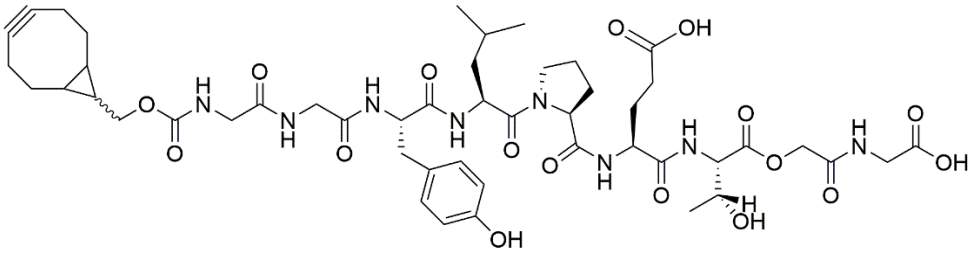


Figure 3.7: Ligation of tripeptide and fluorescent probe by SrtA 5M. (a) With only fluorescent probe present, sortase hydrolyses the threonine-glycine bond, hydrolysed product being seen at 933 Da. (b) With fluorescent probe and tripeptide present, the ligated product can be seen at 1104 Da. (c) Adding 10 times the concentration of tripeptide forces the reaction to completion, as very little hydrolysed product or intact probe can be seen.

The reaction catalysed by Spy SrtA was similarly tested. However, the reaction proved to be very slow, as the probe could be detected but no hydrolysed or ligated products. The turnover of Spy SrtA is just 0.014 s^{-1} [175], whereas SrtA 5M is 5.4 s^{-1} [150]. Increasing the concentration of tripeptide by 10 times and doubling the concentration of sortase present still did not give ligated product.

The sortase-catalysed ligation between the glycine residue at the *N*-terminus of galactose-binding protein (gMBP) and alkyne depsipeptide **25** (kindly synthesised by Gintare Bucaite) was tested *in vitro* as further proof that sortase would be able to ligate fluorescent probes onto incorporated peptide sugars on the surface of cells. This reaction used 6 µM SrtA 5M, 60 µM gMBP and 600 alkyne depsipeptide. After 6 h modified gMBP could be seen (Figure 3.8).



25

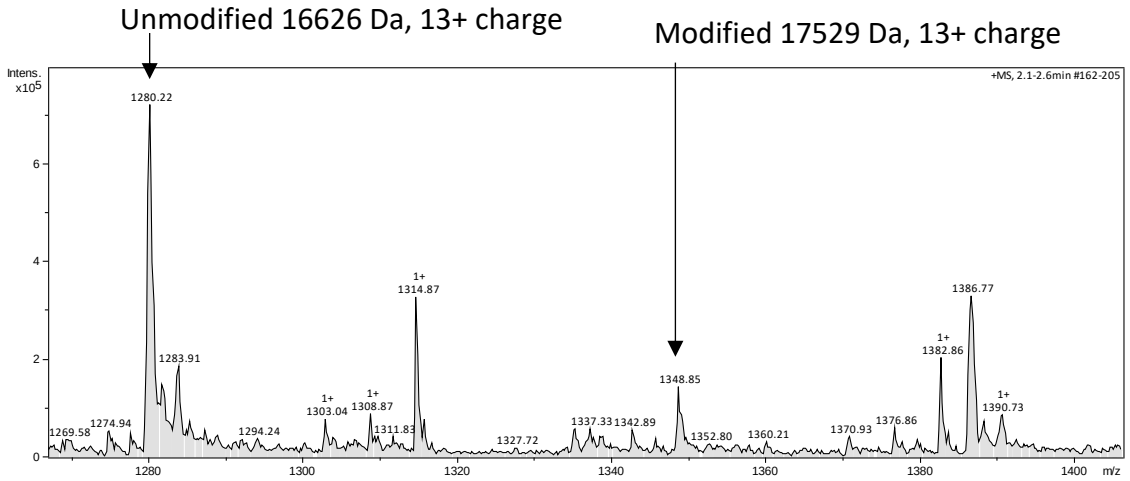


Figure 3.8: LC-MS of 5M SrtA-mediated ligation between gMBP and alkyne depsipeptide showing a small amount of modified protein after 6 hrs.

Sortase testing on HEK cells

As in the previous chapter, HEK cells were incubated in media containing either 100 µM AcManGN$_3$ **4** or 100 µM AcManGGN$_3$ **26** for 72 h. To reduce the sugars, cells were resuspended in PBS with 500 µM 2DPBA for 20 min on ice and then resuspended in sortase ligation buffer (50 mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$ at pH 7.5). Previously, some click reactions were performed on ice to slow internalization of the surface glycans, so the sortase concentration was increased to 160 µM to account for decreased

activity at the lower temperature, and fluorescein-Gaba-LPETGG probe **30** added to 1600 µM. After 30 min incubation with sortase on ice, the cells were washed and analysed by flow cytometry. There was no difference in fluorescence between the control cells and the cells with sugars incorporated (Figure 3.9).

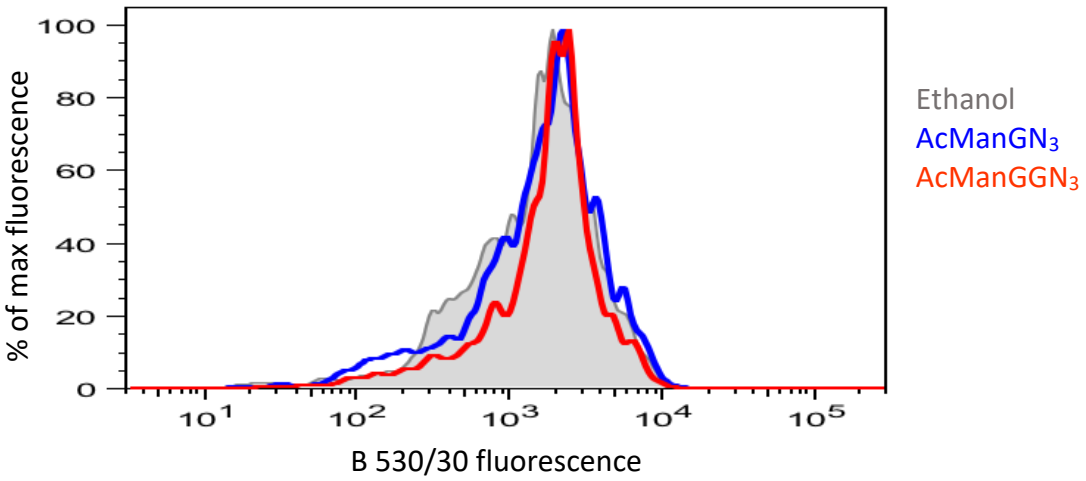


Figure 3.9: Fluorescence shift showing no difference between HEK cells with ManNAc derivatives incorporated and negative control cells.

In this initial experiment, samples appeared to have very low viability, between 7-9%, with a lot of debris visible on the flow cytometer. Previously, after the click reaction protocol HEK cells showed viability around 30-40%. It was thought that the sortase labelling protocol would not work well if it killed most of the cells. It is possible the 2DPBA affected the cells, or possibly one of the ligation buffer components. Calcium ions are used to transfect DNA into chemically competent cells, and the HEPES ligation buffer contains 5 mM $CaCl_2$. Therefore, the calcium present could be disrupting the HEK cell membranes and making the cells more fragile.

The click reaction had been shown in the previous chapter to work well at 37°C without killing the cells or internalising the surface glycans. A higher temperature would allow the sortase reaction to work faster and the sortase labelling protocol was therefore repeated on HEK cells at 37°C. The HEK cells had been incubated in AcManGGN$_3$ **26** for 72 h. However, the viable cell population normally visible (Figure 3.10b and c) appeared to have decreased significantly (Figure 3.11a).

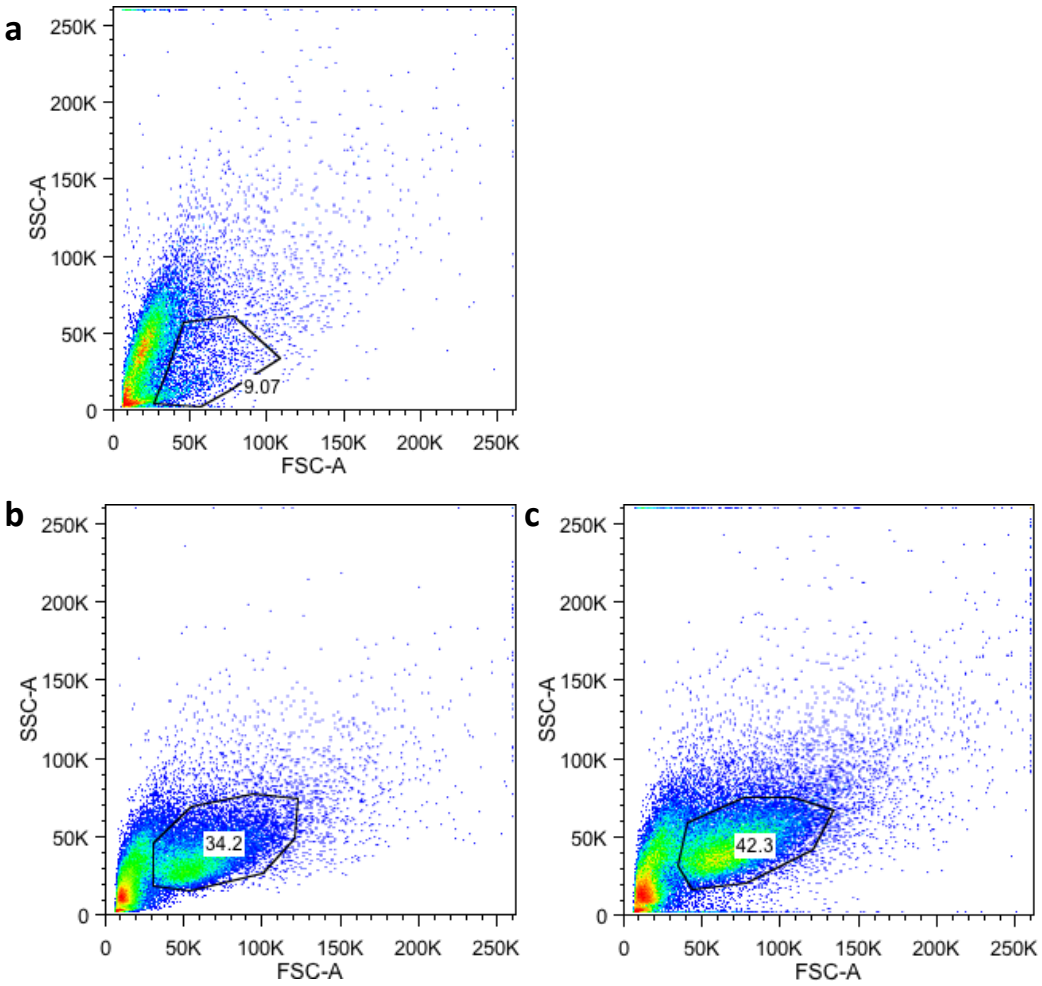


Figure 3.10: (a) HEK cells after the sortase labelling protocol, showing an unusually low viability. (b) and (c) HEK cells after the click reaction protocol on $AcManGN_3$ and $AcManGGN_3$ respectively, showing a larger percentage of viable cells.

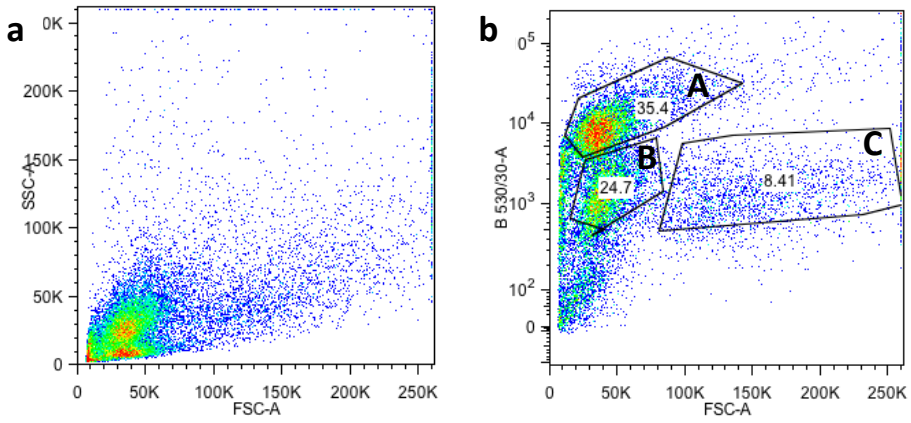


Figure 3.11: Cell populations after labelling with sortase at 37°C. (a) The live cell population is no longer clear. (b) Sorting according to fluorescence and size shows 3 populations that may be live.

Plotting cell fluorescence against size suggested 3 main cell populations, plus debris in the lower left (Figure 3.11b). It was not clear which of these populations was the live cell population, as cells move to the right and up as they increase in size and granularity, before dying and decreasing to the small size of debris.

Comparing each gated population in turn showed that in subset A, cells incubated with sortase at 37°C had the same fluorescence intensity as the controls without sugars or 2DPBA reduction. The population incubated with sortase at 4°C shows a big decrease, so the low temperature, which was to slow sortase activity, is also slowing non-specific binding (Figure 3.12a). Subset B also showed no statistically significant differences between samples (Figure 3.12b). Subset C had no statistically significant differences, although as before the cells incubated in sortase 37°C have a slightly higher fluorescence intensity (Figure 3.12c).

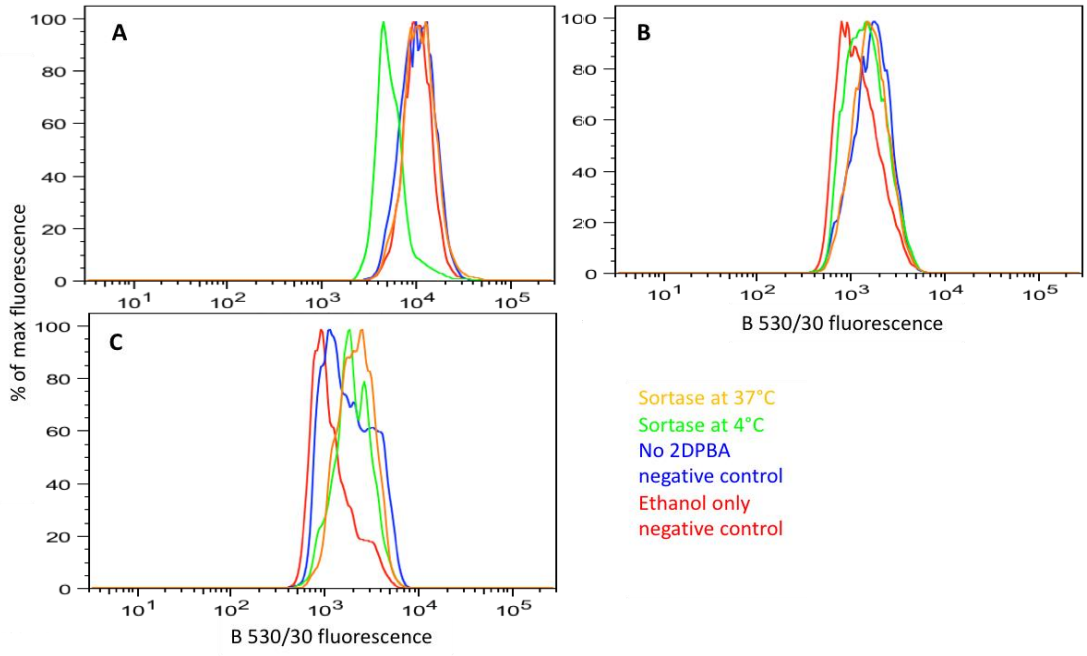


Figure 3.12: Fluorescence shifts due to incorporation of azidopeptide sugars and labelling by sortase at 37°C. Subset A is the largest gated population, B the second largest gate, and C the third largest.

The results from the 3 batches were normalised to MFI of the lowest ethanol control and pooled to compare the fluorescence intensity of subset C from each batch (Figure 3.13). Sortase labelling appears more fluorescent than the ethanol control, however the no 2DPBA control has still slightly higher fluorescence intensity than the sortase. There is also an increase in labelling without sortase.

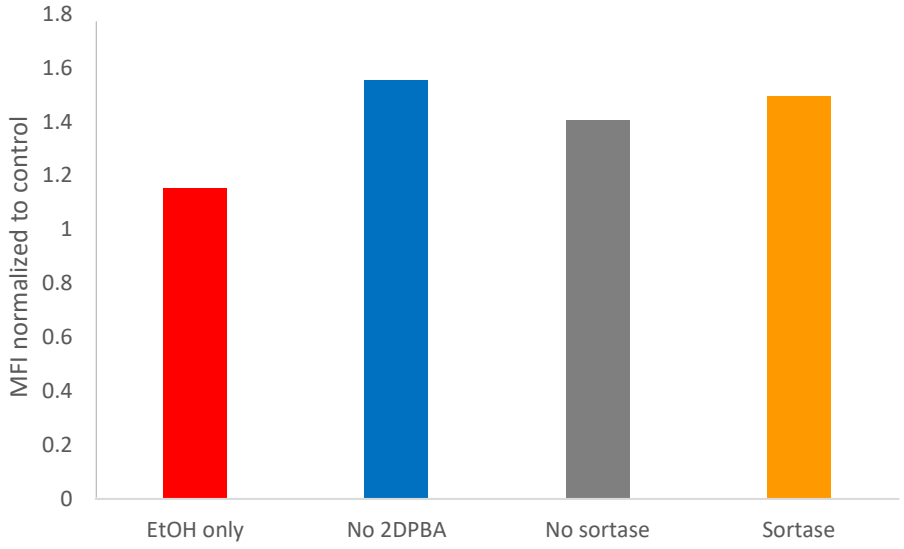


Figure 3.13: Averaged MFI of HEK cells after ligation of fluorescein-Gaba-LPETGG probe by sortase.

There are a few possible causes for this result where the sortase or the azide reduction is not working as intended. It is unlikely the sortase is recognising azide sugars without requiring reduction to the amine, as the NH_2 group is the attacking nucleophile that ligates to the thioacyl intermediate [143]. Perhaps the azide is being reduced by another mechanism in the cell, although the azide-alkyne click reaction has been used by many groups in the literature without this problem being reported [79, 174, 176]. Alternatively, the 2DPBA is not completely reducing the azide, giving a relatively small population of aminopeptide sugars to be ligated by sortase. It was also concerning that subset C was no more than 10% of the cells detected, and this suggested very low viability.

Further testing of sortase labelling on HEK cells at 4°C (no sortase activity) or 37°C (sortase active) showed a shift in the size of cells that may be related to the low viability. The population of large blebbed cells (C) decreases at 37°C, while the increase in the smallest cells (B) may indicate an increase in dead cells (Figure 3.14).

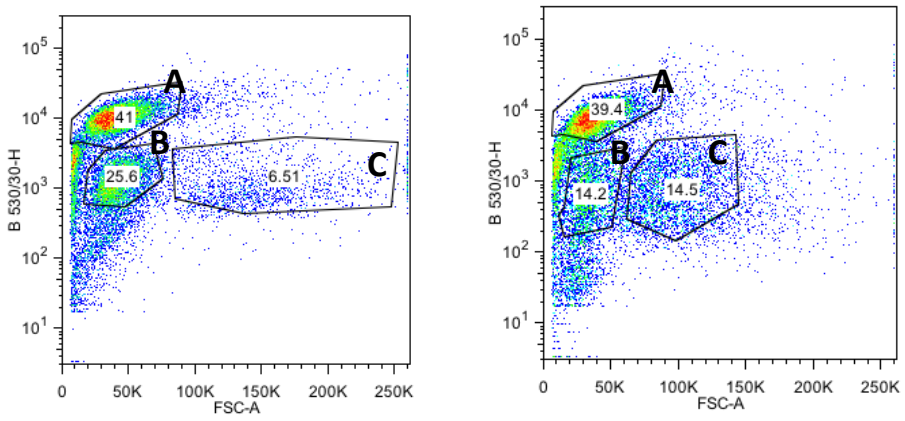


Figure 3.14: Cell populations after labelling with sortase at 37°C (left) or 4°C (right).

The results show a statistically significant increase (P = 0.043) in fluorescence when the cells are labelled with sortase at 37°C after 2DPBA reduction. However, there is also a significant increase (P = 0.018) in labelling in the absence of 2DPBA reducing agent.

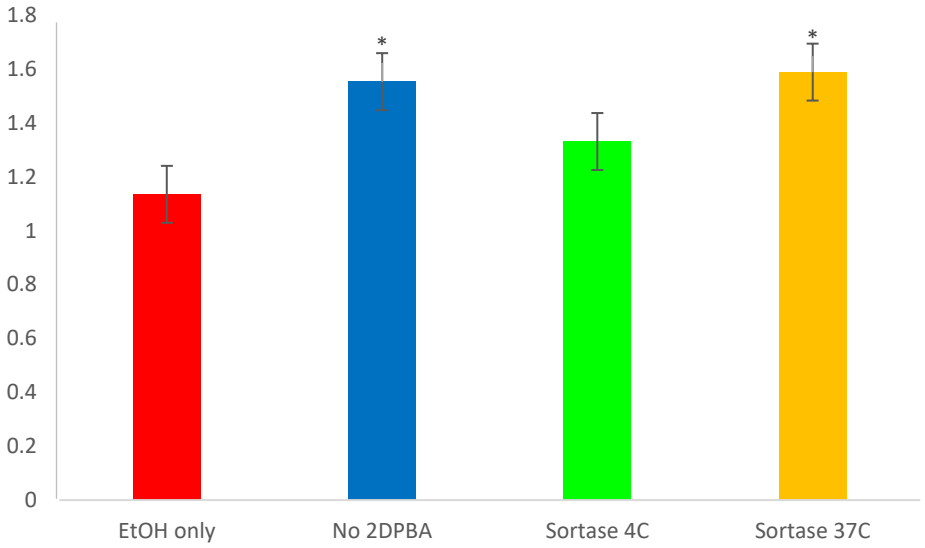


Figure 3.15: Averaged MFI of HEK cells after ligation of fluorescein-Gaba-LPETGG probe by sortase. Sortase and the control with no 2DPBA reducing agent both have a significantly increased fluorescence intensity. Error bars represent mean ± S.E.M. of biological triplicates (n=3). * = P < 0.05, compared with no sugar treatment.

The sortase at 4°C (no activity) control is consistently lower than these samples. This is not a significant difference, and may be due to non-specific binding of the probe, which is expected to be more rapid at 37°C.

To check if the no reducing agent control result could be replicated with a different reducing agent, HEK cells were incubated in TCEP for 24 h as with 2DPBA, or for 15 min after detaching from the plate. This resulted in similar populations to the 2DPBA results, and again the 3 largest population subsets were analysed (Figure 3.16).

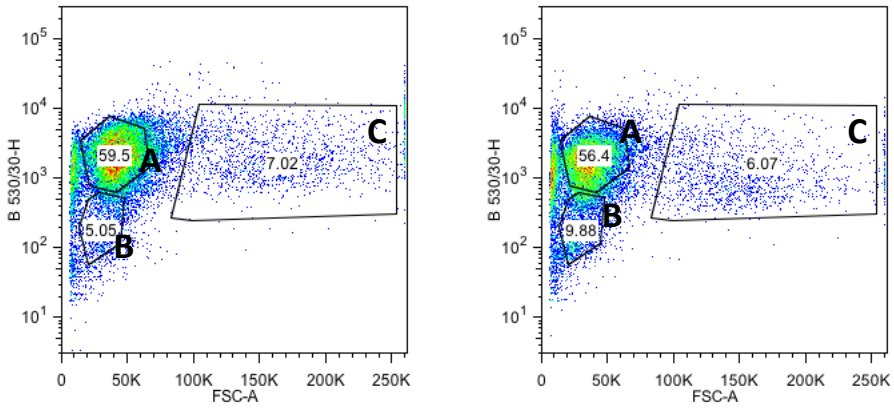


Figure 3.16: Cell populations following reduction with TCEP for 24 h (left) or 15 min after cell harvest (right) and labelling with sortase at 37°C.

Subsets A and B resulted in no differences between conditions tested. Subset C showed the ethanol control to have the lowest fluorescence intensity, but the other conditions overlapped and showed no significant difference between them (Figure 3.17). There is non-specific labelling, as it is not dependent on the presence of reducing agent or sortase.

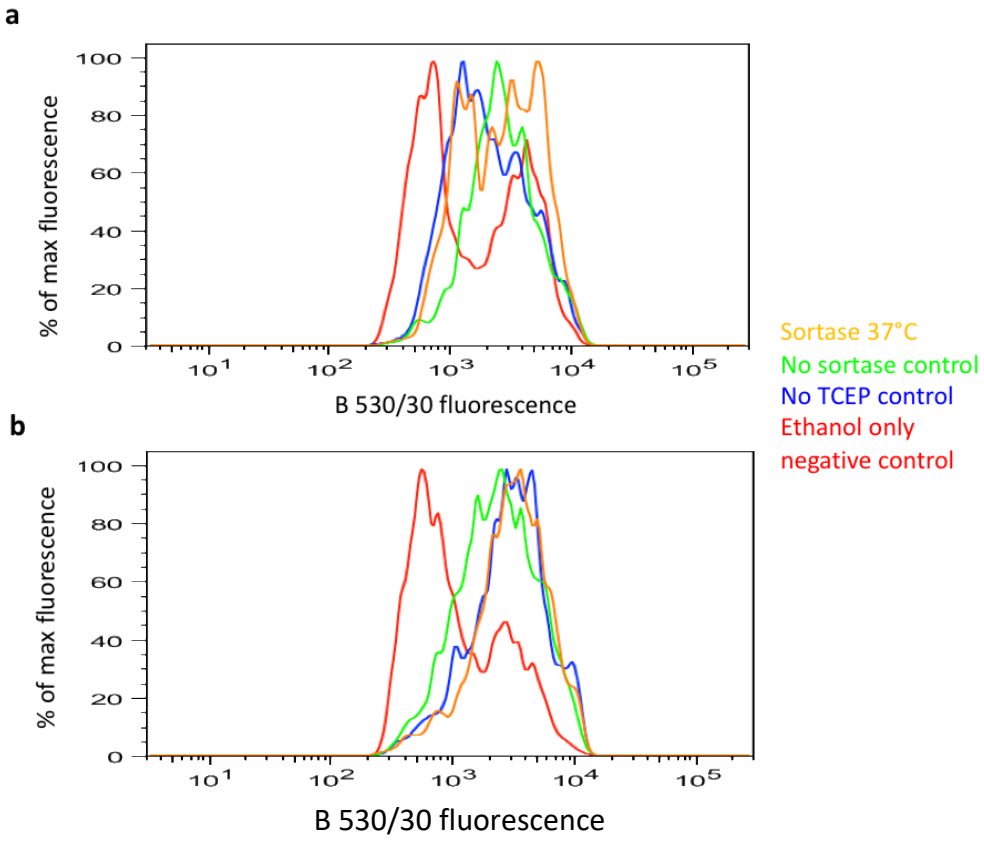


Figure 3.17: Fluorescence shift of two separate batches of HEK cells, subset C, following reduction with TCEP for 24 h (a) or 15 min after cell harvest (b) and labelling with sortase at 37°C.

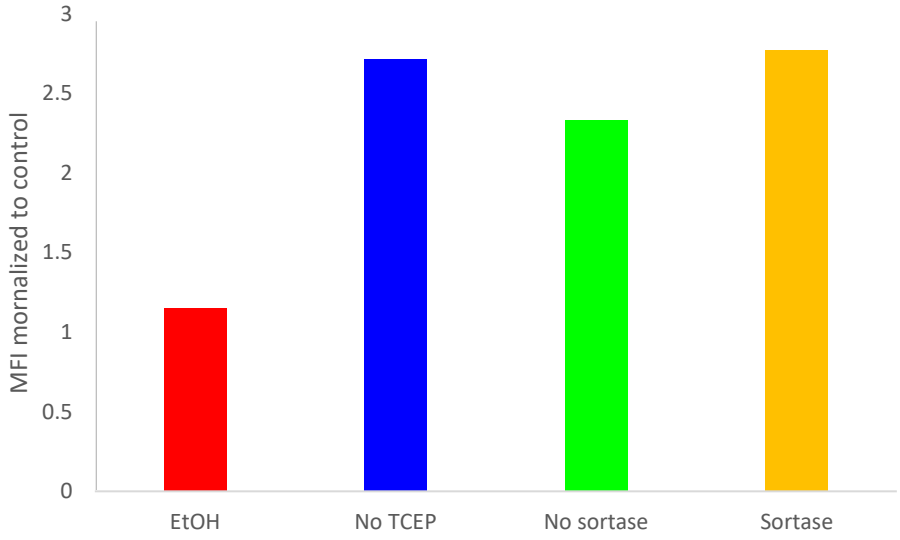


Figure 3.18: Averaged MFI of HEK cells after reduction by TCEP and ligation of fluorescein-Gaba-LPETGG probe by sortase. The controls without TCEP or sortase and the reaction with all components all have an increased fluorescence intensity compared with the control without AcManGGN$_3$.

As it was still unclear if subset C was the live cell population, sortase labelling was repeated with 4',6-diamidino-2-phenylindole (DAPI) stain to differentiate live and dead cells added immediately before performing FACS on each sample. The data was analysed by plotting DAPI stain fluorescence intensity against sortase probe fluorescence intensity. DAPI binds more strongly to dead cells, as it passes through the membrane more easily [177]. Without DAPI stain, all cells appear in the lower portion of the scatter plot (Figure 3.19a). With DAPI stain added, dead cells are positive for DAPI fluorescence and the dead cell population moves. The lower left is the live population gate and the upper right is the dead population gate (Figure 3.19b,c).

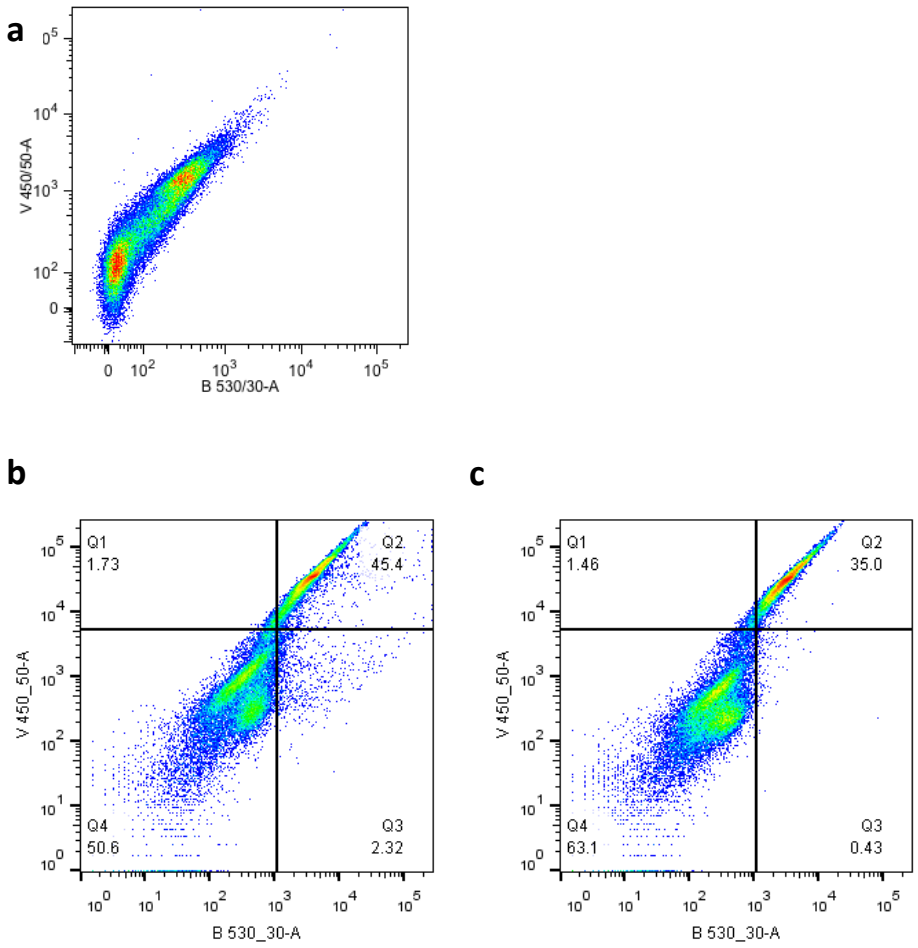


Figure 3.19: HEK cell populations plotted according to DAPI stain (y axis) and sortase probe (x axis). (a) Cells incubated at 4°C. Without DAPI stain all cells are in the lower half. (b) Cells incubated at 4°C. With DAPI stain, the live cells are in the lower left quadrant. (c) Cells incubated at 37°C.

The lower quadrants were gated for each sample and representative samples shown below (Figure 3.20). The three large peaks to the left for HEK cells only at 4°C (pink) and 37°C (purple) and for sortase added but no probe (grey), all show very low fluorescence intensity. Next, the addition of probe but no sortase at 4°C (yellow peak) shows an increase due to non-specific binding by the probe. The rate of this binding increases with temperature, as shown by the peak for probe but no sortase at 37°C (green). Probe + sortase at 4°C (blue) has a higher fluorescence intensity than probe, no sortase at 4°C, suggesting a low level of activity by the sortase at 4°C. Probe + sortase at 37°C (orange) is more fluorescent still, but does not exceed the non-specific binding by probe with no

sortase at 37°C. The addition of probe + sortase to EtOH control cells at 4°C (cyan) is equal to probe + sortase to the sugar-incorporated cells at 4°C, so this is non-specific labelling. But the addition of probe + sortase to EtOH control cells at 37°C (red) has fluorescence intensity equal to or possibly exceeding probe + sortase labelling at 37°C. Therefore, the sortase labelling hasn't worked as intended. It appears the sortase is active but is non-specifically ligating probe to the cell surface at 37°C.

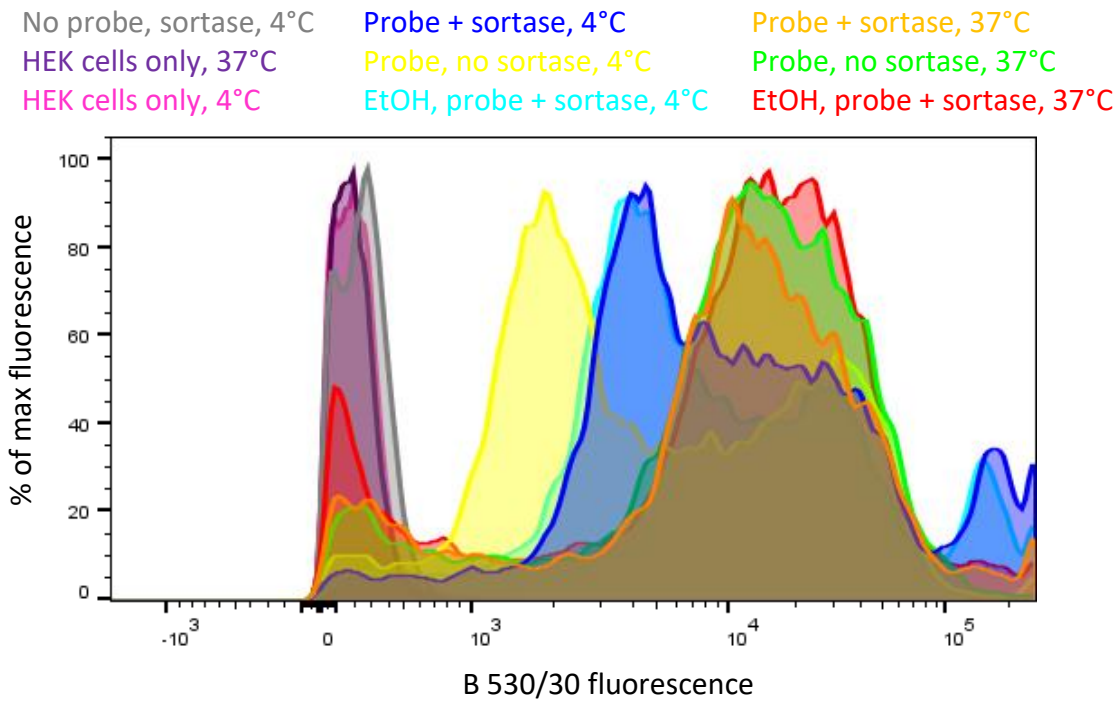


Figure 3.20: Fluorescence shift of confirmed live population of HEK cells, following reduction with 2DPBA for 24 h and incubation with sortase and probe.

The result at 4°C where sortase activity appeared to be labelling cells was further investigated. Cells were tested again with EtOH (no sugar) and no sortase controls at 4°C and 37°C (Figure 3.21). The no sortase controls both show non-specific labelling by the probe (blue and cyan peaks). When sortase was added at 4°C, this increased fluorescence whether the cells had been treated with EtOH (red) or $AcManGGN_3$ **26**. When sortase was added at 37°C the increase in fluorescence appears larger. But again, this applies for EtOH (orange) or sugar (pink). So the sortase is labelling the cells, and ligates faster at the higher temperature, and appeared to be non-specifically labelling exposed glycines both times. However, an unexpected result was found at 4°C. A

comparison of sugar + no sortase against EtOH + sortase showed a decrease without sortase in line with previous results. But the sugar + sortase result at 4°C appeared to be higher than the EtOH + sortase result and this was not true of the comparison at 37°C (Figure 3.22).

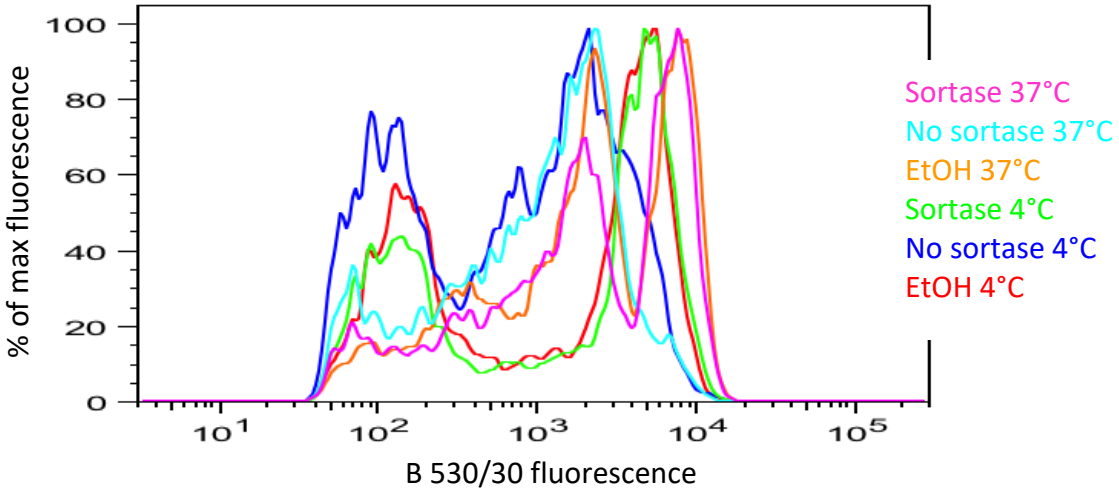


Figure 3.21: Fluorescence shift of HEK cells due to labelling by sortase at various temperatures.

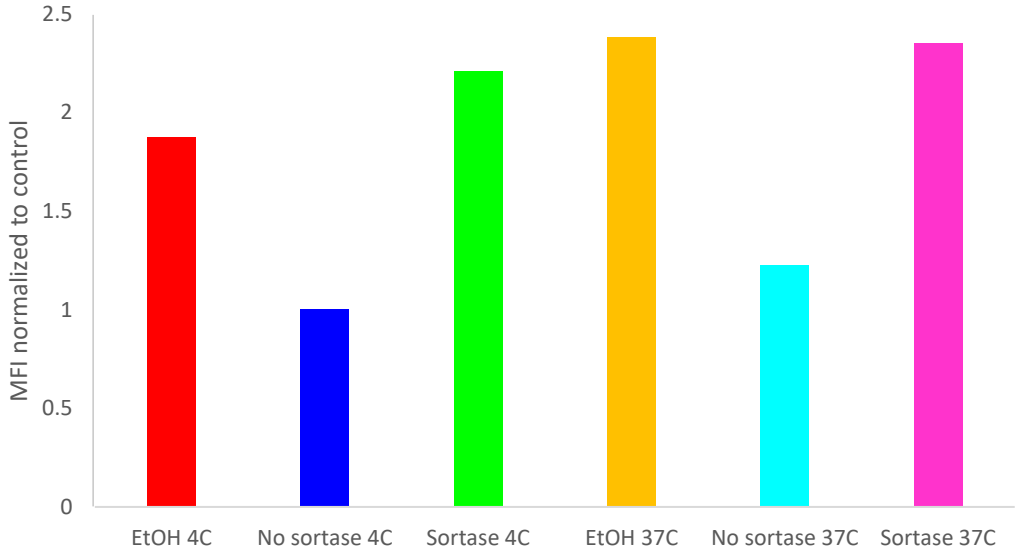


Figure 3.22: Averaged MFI of HEK cells after ligation of fluorescein-Gaba-LPETGG probe by sortase at 4°C or 37°C. Both controls without sortase show a decrease in labelling compared with the ethanol control. At 4°C, the sugar-fed cells also show an increase over the ethanol control.

It is possible that at the higher temperature the sortase rapidly labels available sugars and non-specifically labels glycines on the cell surface, so the background labelling at 37°C is higher and the increased labelling due to sugar is not detectable.

To further test if the 2DPBA reduction was working as expected, the sugars were reduced in 500 µM 2DPBA for 30 min before addition to the media so that aminopeptide sugars would be incorporated by the cells and the reduction step no longer had to be performed on live cells. The azidopeptide sugars were fed to cells first, on the basis that the previous experiments with click-labelling showed the azidopeptide sugars are incorporated into cells. It is not certain that the aminopeptide sugars can be incorporated. The cells were harvested and labelled with sortase as before, with 3 separate experiments, each with 3 replicates (Figure 3.23).

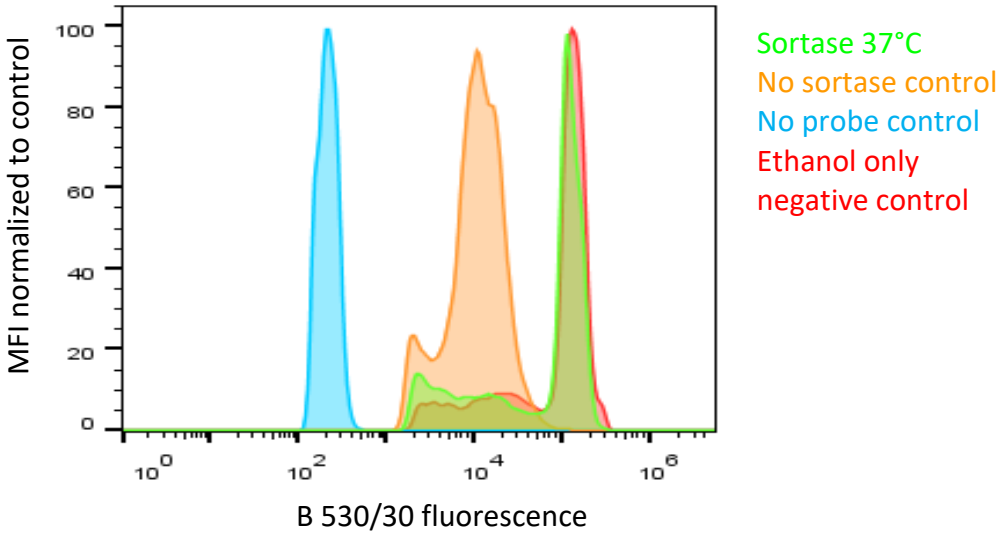


Figure 3.23: Fluorescence shift due to incorporation of aminopeptide sugars and labelling by sortase at 37°C.

The final batch was stained with DAPI as previously to facilitate identification of the live cell population. This showed a population in the lower right quadrant that significantly increases compared with no sortase when sortase is added (Figure 3.24), however the increase in fluorescence also occurs in the ethanol control sample. Therefore, the live cells are again being non-specifically labelled by sortase. Both the cells incubated with aminopeptide sugar and EtOH control cells show a statistically significant increase over the control without sortase (Figure 3.25).

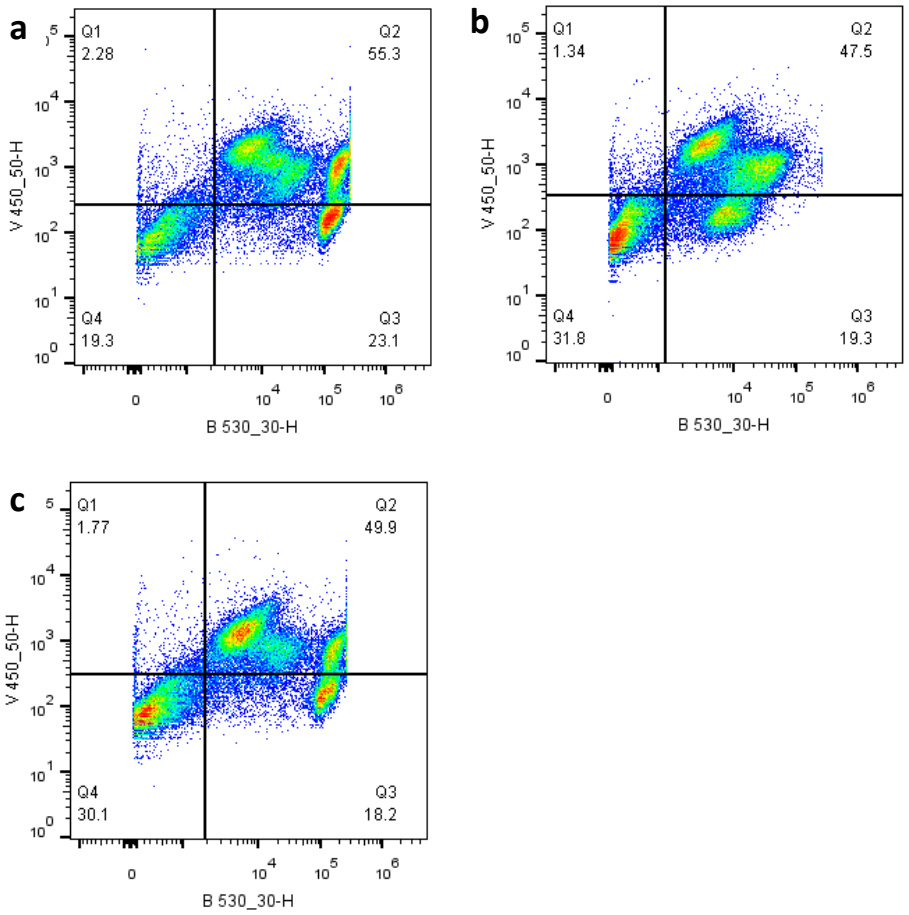


Figure 3.24: HEK cell populations plotted according to DAPI stain (y axis) and sortase probe (x axis). The live cells are in the lower two quadrants. (a) EtOH negative control, (b) no sortase control, (c) sortase incubation at 37°C.

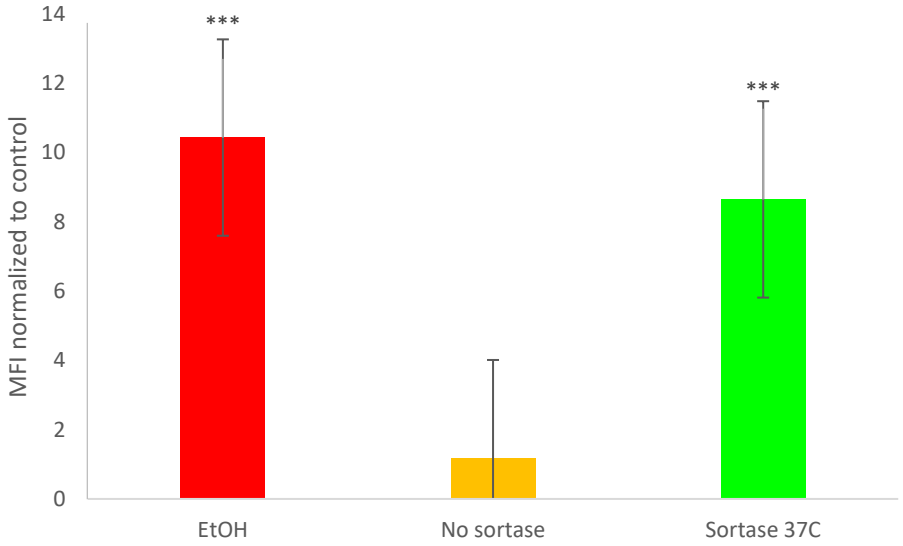


Figure 3.25: Increase in fluorescence intensity due to labelling by sortase. Error bars represent mean ± S.E.M. of biological replicates (n=3). *** = P < 0.001, compared with the no sortase control.

It is not certain the aminopeptide sugars were being incorporated, as unlike the azidopeptide sugars they cannot be tested by first reacting with a fluorescent click probe. However the sugars ManLev **1** [71] and Ac_4ManBeoc **10** [69] have been successfully incorporated into mammalian cells and are larger than the AcManG derivative **29**, and of comparative size to the AcManGG derivative **34** (Figure 3.26).

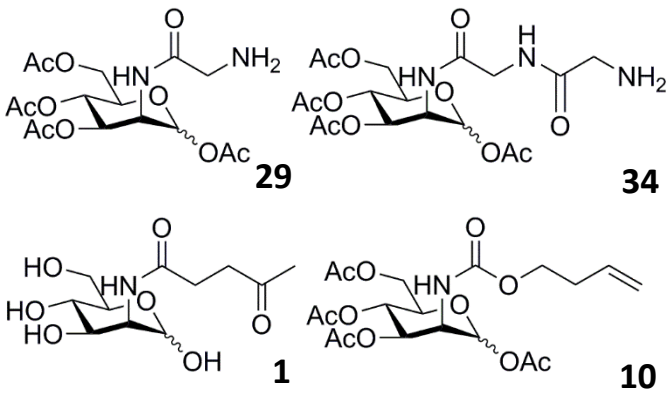


Figure 3.26: Size comparison of the mannosamine-glycine derivatives and the previously incorporated ManLev (lower left) and Ac_4ManBeoc.

As an attempt to decrease the background labelling, the sortase protocol was modified to include a blocking step. Cells that had been fed with azide sugars were incubated with sortase and non-fluorescent Gaba-LPETGG probe for 1 h, washed, reduced with 2DPBA for 20 min, washed, and then incubated with fluorescein-Gaba-LPETGG probe **30** for 1 h. It was hoped that the initial sortase incubation would label glycine residues on the cell surface with non-fluorescent probe. With these sites blocked, after reduction of azide to amine the only sites available for ligation during the second sortase incubation would be the amine sugars, and this would hopefully improve the specificity of the labelling. The samples without blocking appeared to have the highest fluorescence (Figure 3.27a), suggesting the blocking decreased the number of sites available for sortase ligation. However, the labelling was still not improved over the background fluorescence (Figure 3.27b).

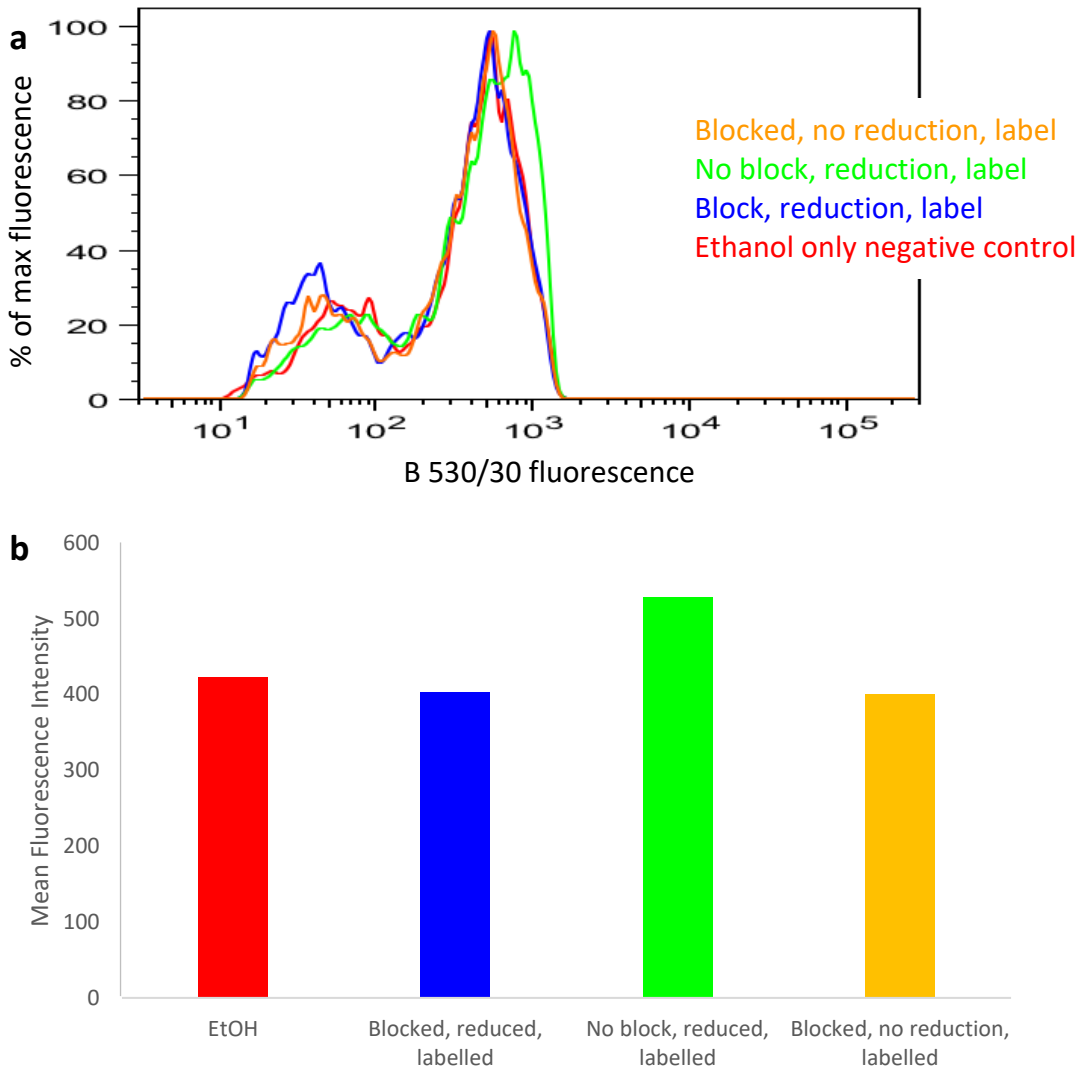
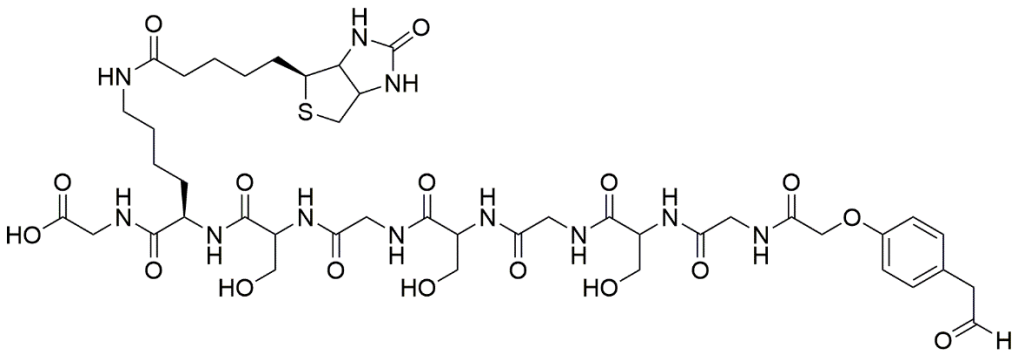


Figure 3.27: The addition of a blocking step before sortase labelling of HEK cells does not decrease background labelling.

Despite several attempts to optimise the labelling of sugars using sortase, the results were not as successful as hoped and it was decided to investigate an alternative sugar, AcManThz **27**, which had been shown in Chapter 2 to be well tolerated by cells up to 300 µM.

Organocatalyst-mediated Protein Aldol Ligation

Previously, the Organocatalyst-mediated Protein Aldol Ligation (OPAL) method has been used to label proteins in cell lysate and has not been tested on live cells [167]. Using allylpalladium (II) chloride dimer **35** as a catalyst, the ring opens and gives an aldehyde that can be fluorescently labelled by an aminooxy probe (synthesised with the kind help of Amanda Noble) (Figure 3.28). The reaction is sped up by addition of (*S*)-(-)-5-(2-Pyrrolidinyl)-1*H*-tetrazole (Pro-tetrazole), a proline tetrazole catalyst **36** (Figure 3.29). The OPAL method has previously been reported to label the incorporated sialic acid analogue 9-deoxy-9-*N*-carboxy-benzaldehyde NeuAc with an aminooxy biotin probe **37** that could be detected by a streptavidin-fluorescein conjugate [168] (Figure 3.29).



37

Figure 3.28: The OPAL probe, consisting of an aldehyde, peptide linker and biotin.

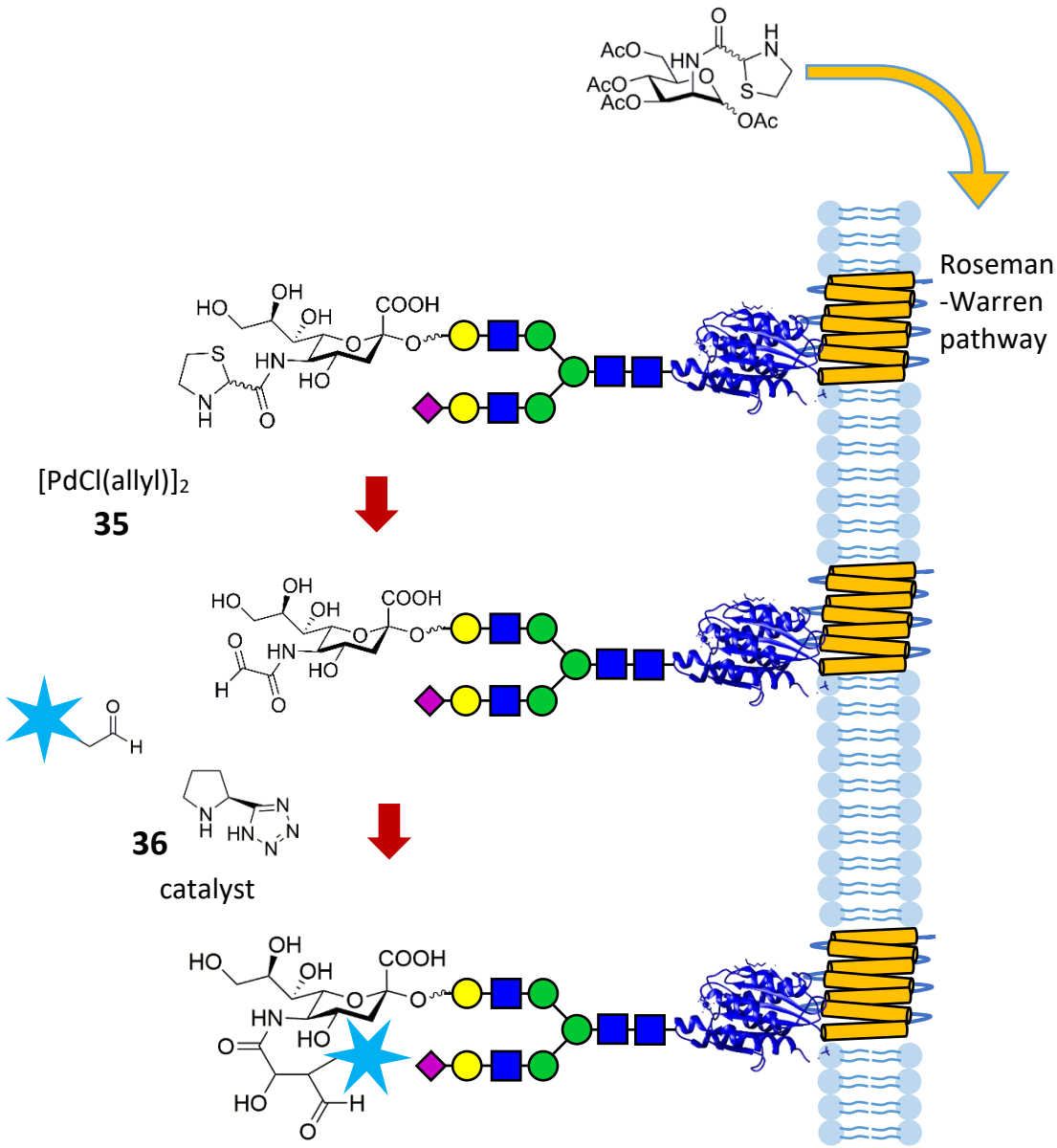


Figure 3.29: Incorporation of mannosamine-thiazolidine gives sialic acid residues with a caged aldehyde. The aldehyde is exposed by a palladium catalyst and can be conjugated to aminooxybiotin probe using Pro-tetrazole catalyst.

OPAL testing on HEK cells

The AcManThz **27** sugar was therefore tested in HEK cells for incorporation and labelling as the previous sugars were. After 72 h incorporation, the cells were harvested, incubated in 300 µM Pd [(allyl)Cl_2] **35** for 30 min, then incubated with Pro-tetrazole **36** OPAL probe **37** for 60 min, and finally incubated with streptavidin-647 conjugate for 60 min. Again, DAPI staining was used to establish the live cell population. Initially, the

labelling appeared successful, as the highest fluorescent peak was the reaction with all components and the peak without sugar was far lower (Figure 3.30).

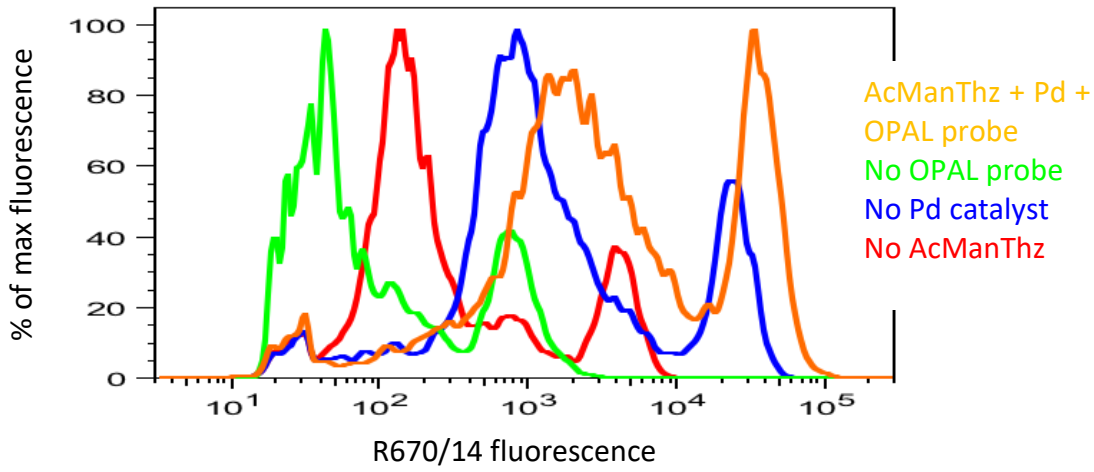


Figure 3.30: Fluorescence shift of live cells due to labelling of AcManThz-fed cells by OPAL method.

As the cells with no OPAL probe show a large decrease in fluorescence compared with the ethanol control cells, which have no sugar but OPAL probe and Pd catalyst, the SA-647 was unable to bind to the cells without OPAL probe (Figure 3.31). There was also a large increase in the sample given all components of the reaction. However, the no Pd control still shows an increase in fluorescence, which remained when the experiment was repeated a second time. Therefore, the cells were being labelled even when the Pd catalyst was not present to decage the aldehyde.

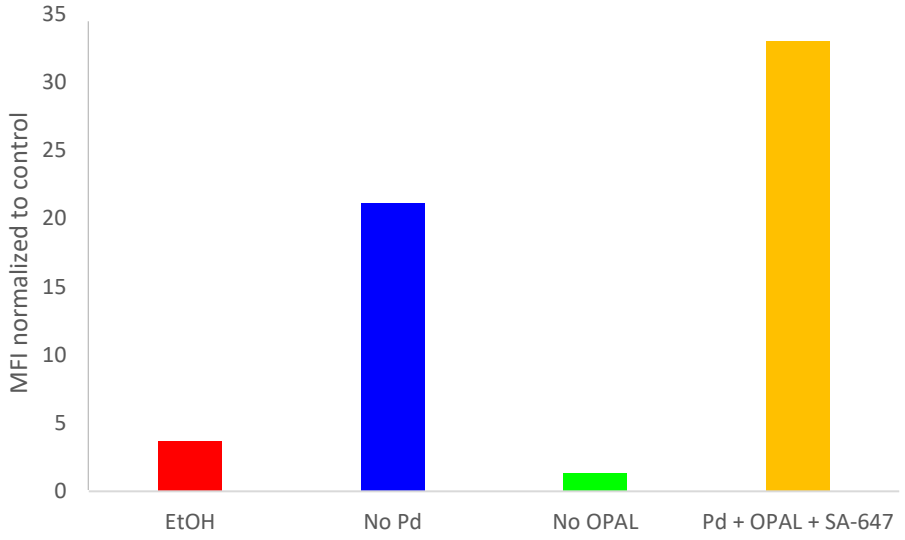


Figure 3.31: Labelling of HEK cells using the OPAL method. Although the cells with all reaction components shows an increase compared with the ethanol control, the increased fluorescence of the no Pd control suggested non-specific labelling.

As this was possibly due to non-specific labelling or side reactions, it was decided to attempt a blocking step by adding phenylacetaldehyde and Pro-tetrazole **36** catalyst. This would hopefully react with exposed aldehydes and then after Pd decaging the only aldehydes present on the cell surface would be from the decaged sugar, and would therefore allow specific labelling by the OPAL probe **37**. The live cell population was again selected using DAPI negative cells. Again, the controls lacking OPAL probe or SA-647 had only background cell autofluorescence (Figure 3.32a). The control without AcManThz **27** sugar was also very low. However, the control without Pd catalyst **35** was still higher than the reaction with all components. The no blocking control was higher than any other, suggesting that the blocking did slightly decrease background fluorescence, but not enough to solve the labelling taking place without Pd catalyst (Figure 3.32b).

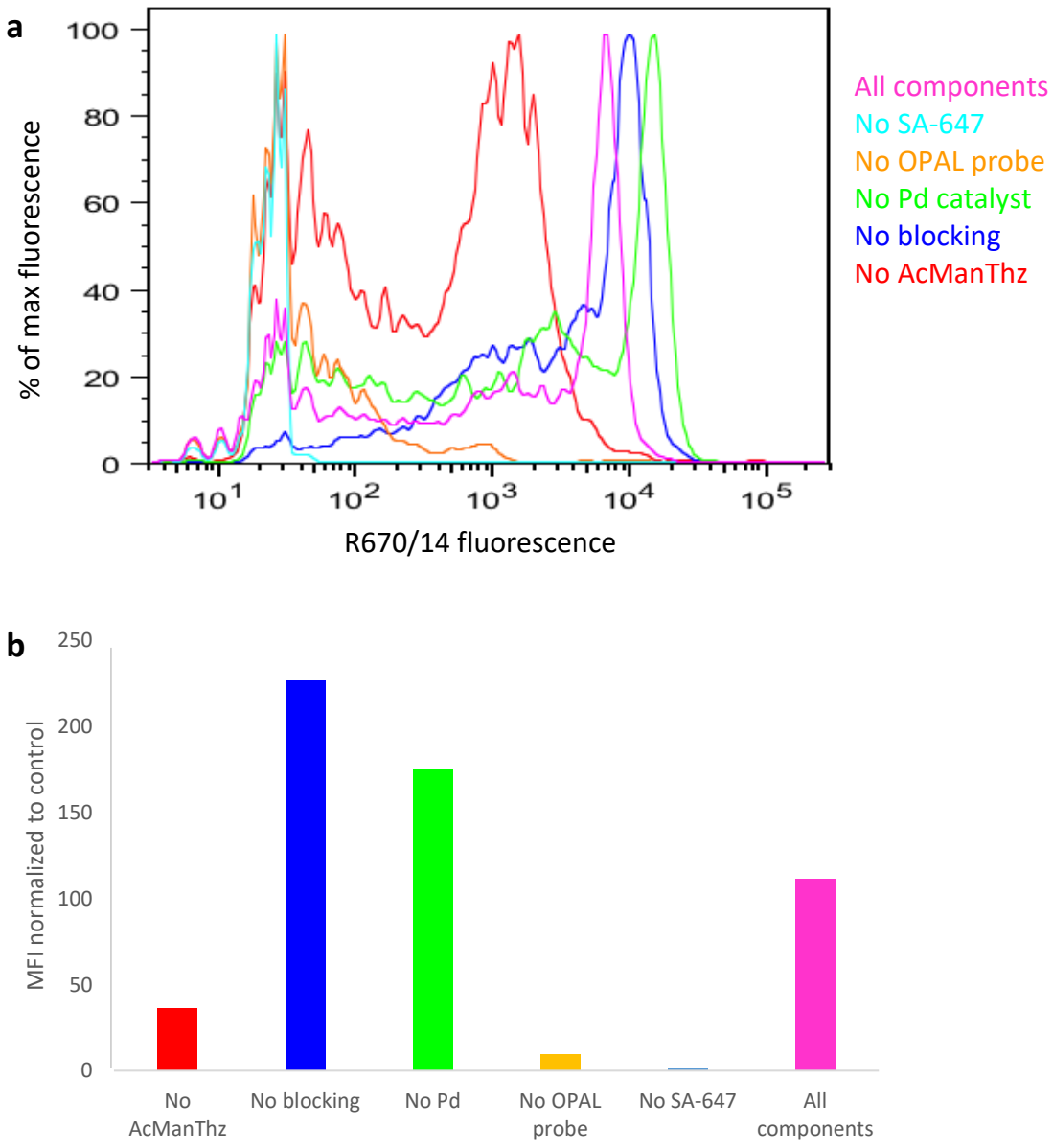


Figure 3.32: Labelling of live AcManThz cells using the OPAL method shows increased fluorescence without blocking by phenylacetaldehyde compared to the control without AcManThz.

At this point, it was hypothesised that the cells were not incorporating enough of the AcManThz **27** sugar to have a significant amount of labelling. The feeding protocol was modified by replacing the media with starvation media (DMEM supplemented with 1% (v/v) Fetal Bovine Serum, 1% (v/v) L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin) 24 h before seeding cells into well plates pre-coated with sugar. The decreased amount of growth factor (cells were normally grown in DMEM supplemented

with 10% (v/v) Fetal Bovine Serum) caused all the cells to synchronise cell division starting when the normal medium was added, and it was hoped that this would improve uptake of AcManThz **27**. As before, DAPI stain was used to identify live cells. The FACS graph suggested an improvement, as the sample with all components had the highest fluorescence and the controls without AcManThz **27** and without Pd catalyst **35** are lower (Figure 3.33a). This suggested that starving the cells prior to feeding sugars improved labelling (Figure 3.33b). and it was decided to add the starvation step for all future experiments.

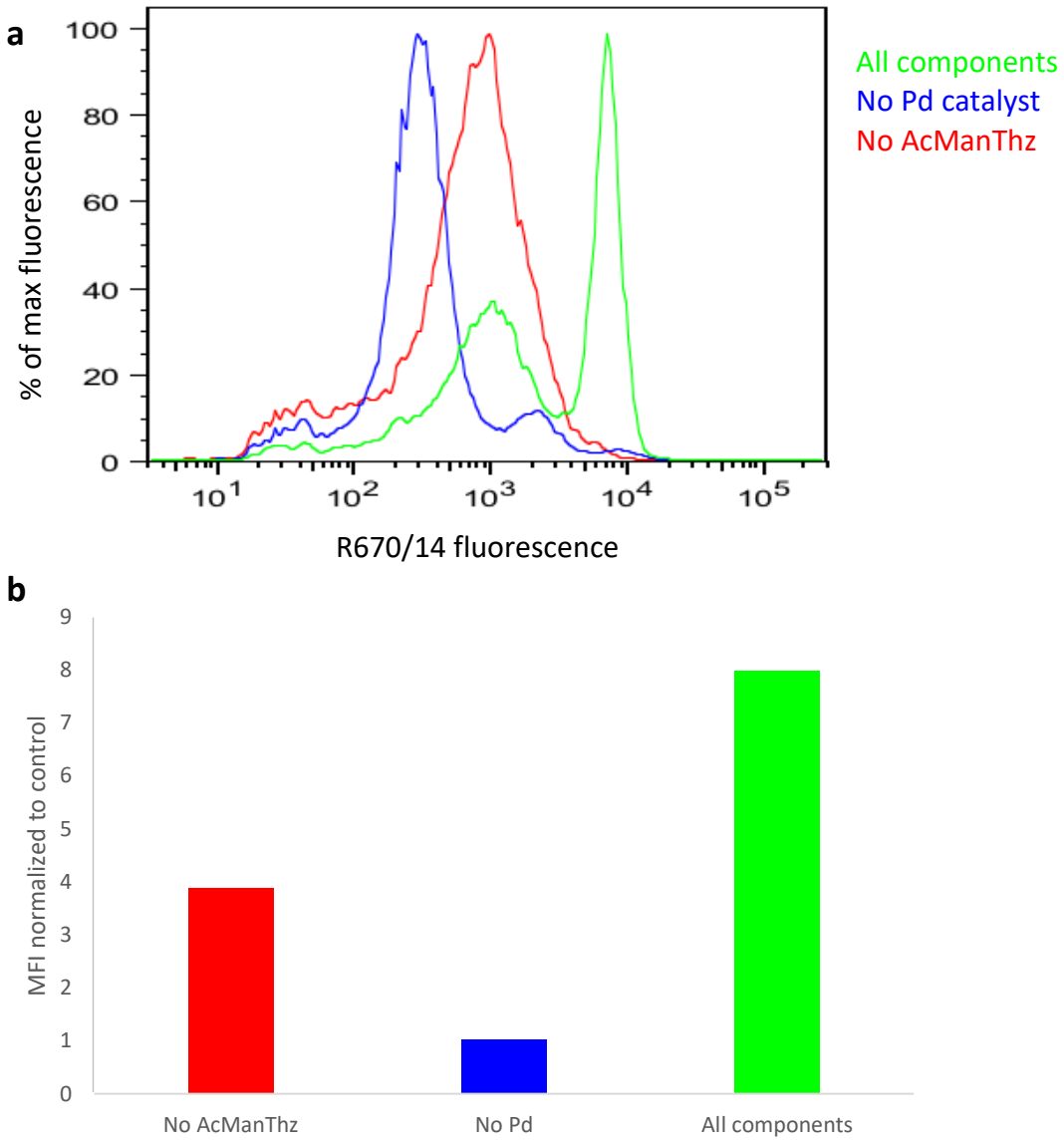


Figure 3.33: Labelling of starved cells fed AcManThz using the OPAL method. The cells with all labelling components have the highest fluorescence (a), and this appears to be an improvement over non-starved cells (b).

To further improve the labelling, we decided to see if the reaction could be performed with a lower concentration of Pd, as higher concentrations may be toxic to the cells. $PdCl_2$ has been shown to have toxic effects on the mitochondria at 200 μM, although there is little information on how organic compounds of Pd compare [178]. Concentrations of Pd [(allyl)Cl_2] **35** were tested from 0-300 μM on starved cells with and without AcManThz **27** feeding. The cells showed a certain level of background fluorescence (Figure 3.34a) at 0 μM that did not differ significantly whether AcManThz **27** was present or absent. At 100 μM (Figure 3.34b) the presence of AcManThz **27** caused increased fluorescence. The peak of background fluorescence appeared to split into 3 peaks – presumably unlabelled, partially labelled and completely labelled. At 300 μM (Figure 3.34c) all samples shifted right to the completely labelled peak, suggesting that this high concentration of palladium catalyst killed the cells and caused a large amount of background labelling on the damaged cell surfaces.

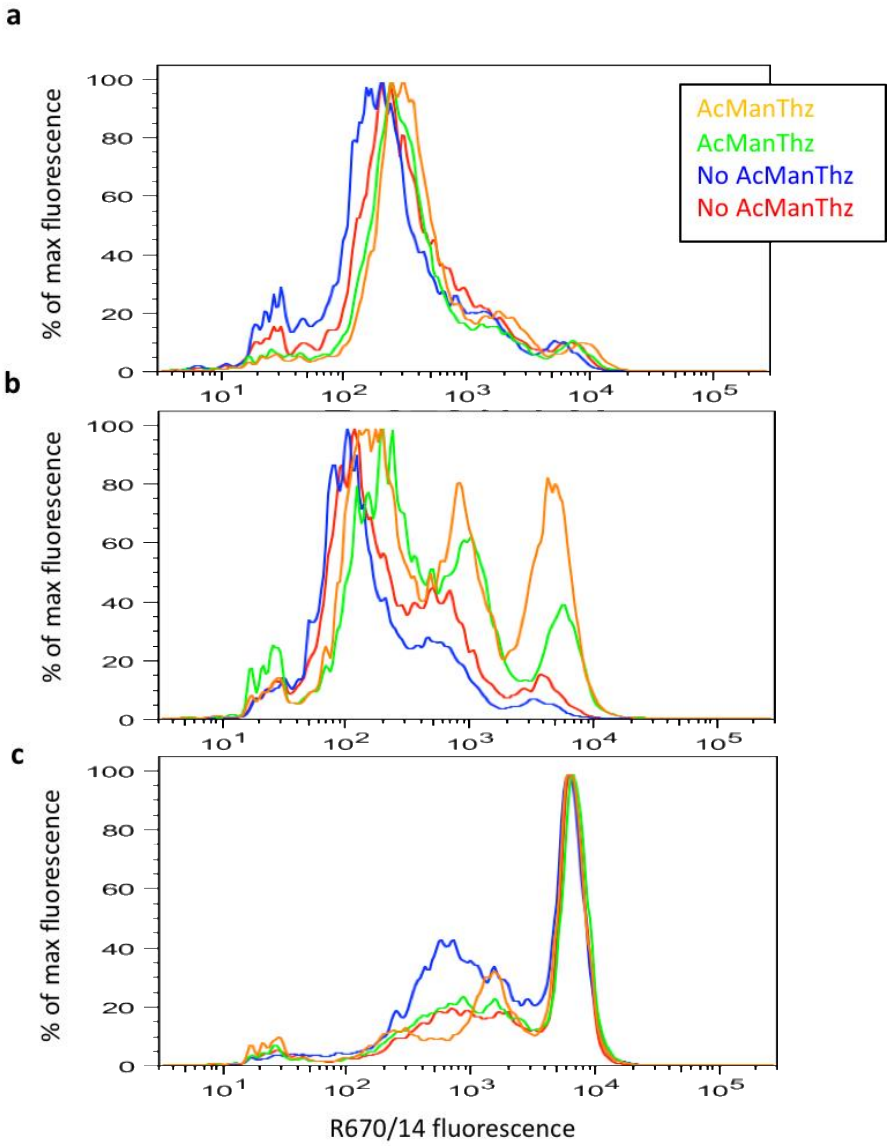


Figure 3.34: Fluorescence shift of duplicate samples due to labelling of AcManThz-fed cells by OPAL method with [Pd] of 0 µM (a), 100 µM (b) and 300 µM (c).

The optimal Pd [(allyl)Cl_2] **35** concentration for labelling live HEK cells therefore appeared to be 100 µM (Figure 3.35). The fluorescence was normalized to control cells that had no treatments – no sugar, Pd or OPAL probe added.

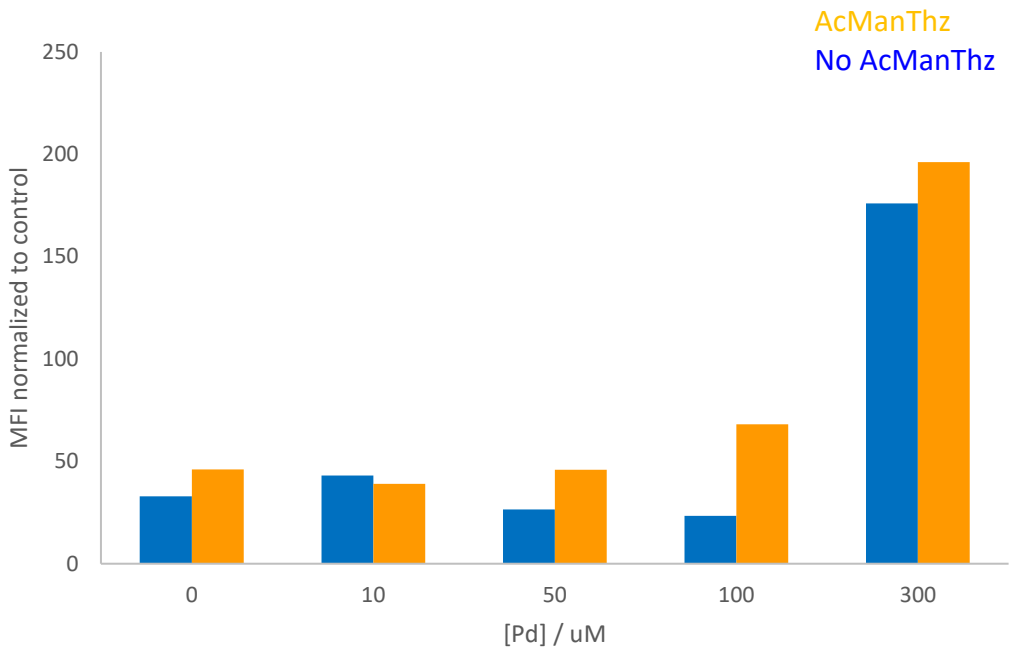


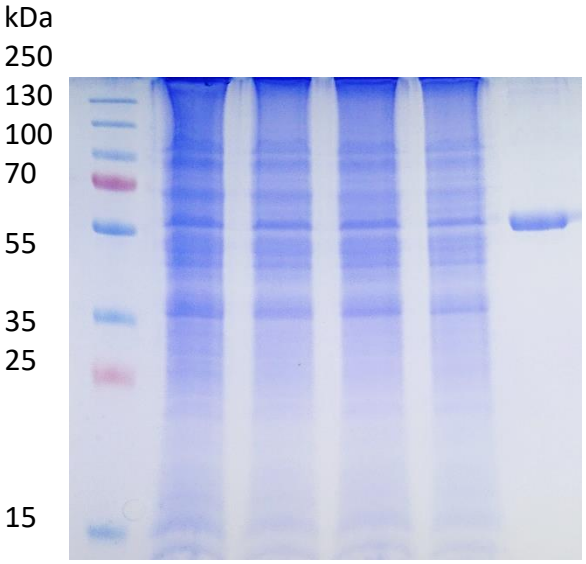
Figure 3.35: Labelling of AcManThz cells using the OPAL method at different concentrations of Pd [(allyl)Cl_2] suggested the optimal concentration to be 100 µM as this yielded the largest difference between cells fed AcManThz and control cells.

Future work would look to optimise the OPAL labelling protocol further, perhaps by varying the concentration of OPAL probe **37** and incubation time.

OPAL testing by western blot

As well as testing on live cells, AcManThz **27** was tested for incorporation into intracellular glycoproteins. The starved cells were fed AcManThz **27** to a concentration of 300 µM for 72 h as before, then the cells were lysed and centrifuged and the supernatant proteins underwent the labelling protocol.

For the first attempt, proteins were decaged with 50 µM Pd allyl chloride **35** at RT for 30 min and desalted with a PD G25 MiniTrap column. The OPAL probe **37** was added to a final concentration of 1 mM, and Pro-tetrazole **36** catalyst to 25 mM, and the reaction left at RT for 60 min before 20 µL samples were run on SDS-PAGE. This was analysed by western blot (Figure 3.36). At first no lanes aside from the biotinylated positive control showed an increase in labelling, after 20 min all lanes showed background labelling.



–	+	+	+	–	AcManThz
+	–	+	+	–	$[PdCl(allyl)]_2$
+	+	+	+	–	OPAL probe
+	+	–	+	–	Pro-tetrazole catalyst
–	–	–	–	+	Biotinylated BSA positive control

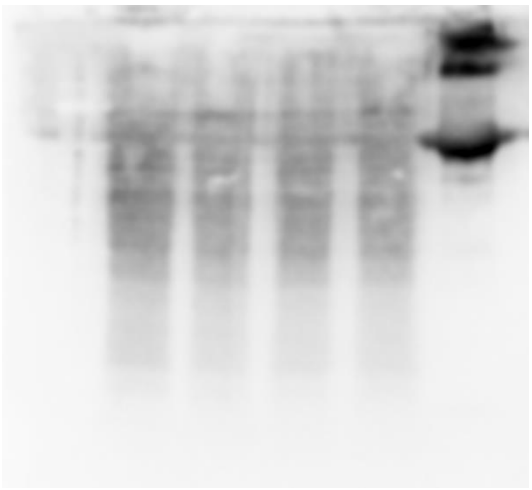


Figure 3.36: Coomassie and western blot of labelling of AcManThz in glycoproteins. The labelling has not worked properly as all lanes show background labelling.

The second attempt used 50 µM Pd allyl chloride **35** at RT for 60 min for the decaging reaction, and all other conditions were as before. This improved labelling, as the bands appeared more rapidly (5 min) and labelling was absent in the controls without OPAL

probe **37** or Pro-tetrazole **36** catalyst, indicating that the OPAL reaction was working as expected (Figure 3.37).

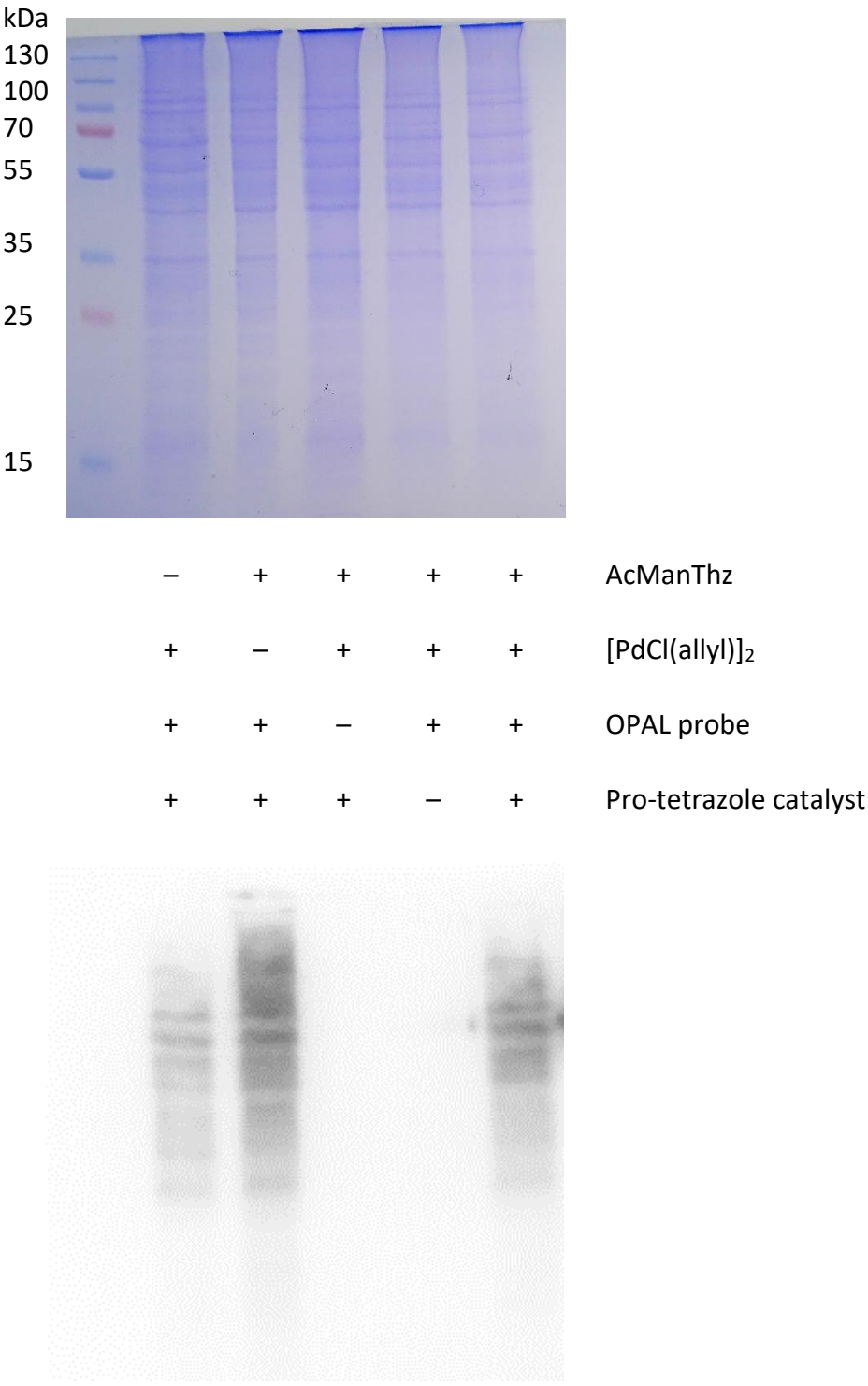


Figure 3.37: Coomassie and western blot of labelling of AcManThz in glycoproteins. The absence of labelling in the controls without OPAL probe or Pro-tetrazole was an improvement.

Another attempt added a blocking step. First, the harvested proteins were reacted with 1 mM phenylacetaldehyde and 25 mM Pro-tetrazole **36** catalyst for 60 min at RT, then the buffer was removed and 300 µM Pd allyl chloride **35** was added at RT for 60 min RT and the protocol continued as before. Again, there was significant background labelling across all lanes (Figure 3.38).

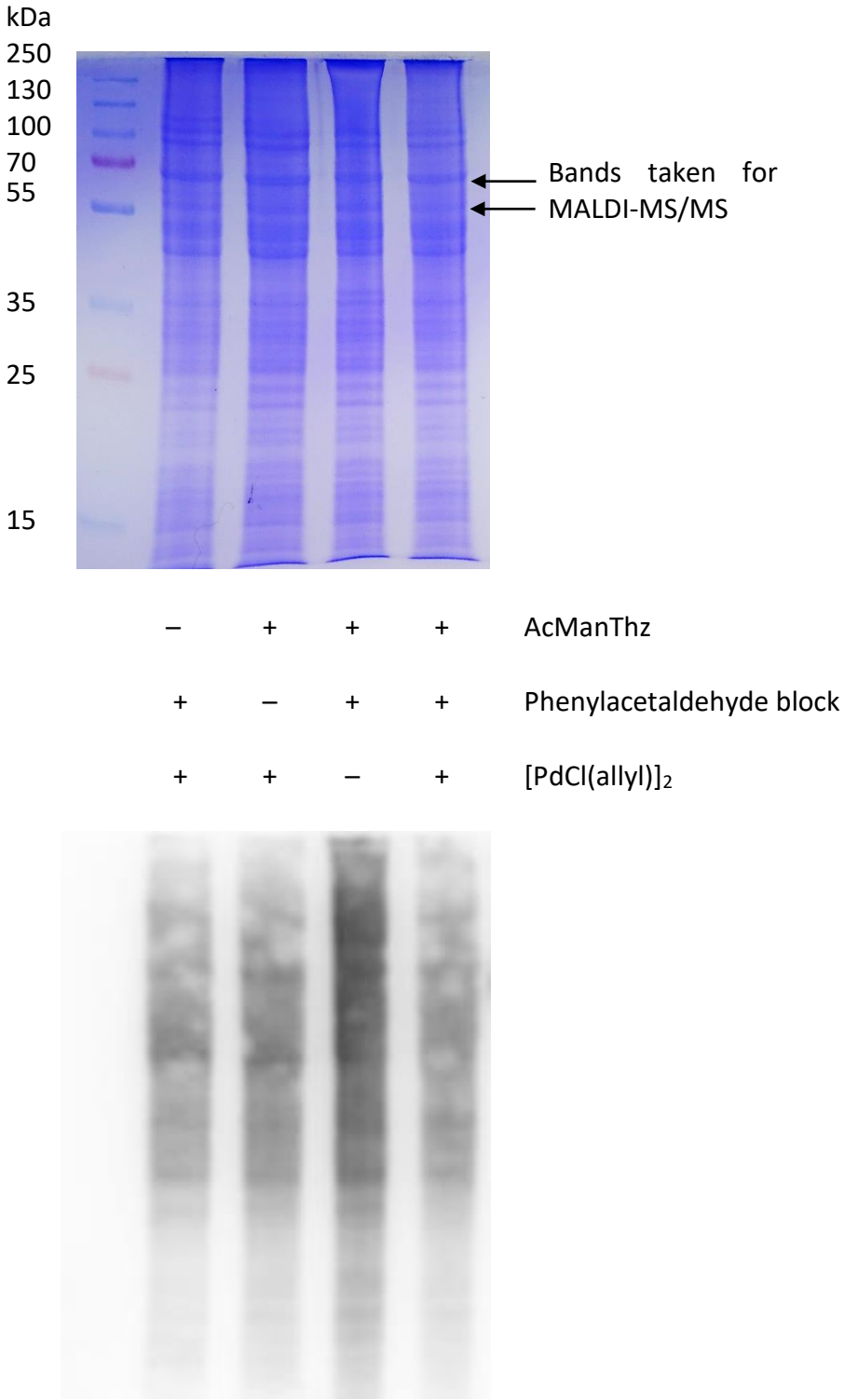


Figure 3.38: Coomassie and western blot of labelling of AcManThz in glycoproteins with a blocking step added. Again, there is significant background labelling across all lanes.

As a further investigation, the two marked bands from the EtOH control lane were removed for protein identification by MALDI-MS/MS (kindly performed by Chris Taylor). The matches were Heat shock 70 kDa protein 1A and 60 kDa mitochondrial heat shock protein. These are commonly found in cells under any stress, so it is likely that the OPAL probe **37** is non-specifically labelling many available proteins.

The labelling of AcManThz **27** glycoproteins by OPAL probe **37** and western blot was not optimised, but it is possible that with further variations in Pd concentration and reaction duration the specific labelling could be improved. The decaging of thiazolidine has also been shown to be catalysed by copper [179], which is present in some metalloenzymes and therefore the cells may be decaging some of the AcManThz without requiring the addition of Pd catalyst. This would also explain the difficulty in optimising the Pd concentration for labelling of live cells.

Conclusion

Building on our initial click chemistry approach, we tested new methods with the aim of improving live cell labelling efficiency. The first one exploiting bacterial sortase A ligation for labelling was unsuccessful. SrtA 5M was made and successfully ligated peptides and gMBP in vitro, but while AcManGN$_3$ **4** and AcManGGN$_3$ **26** were incorporated into HEK cells, subsequent "on cell" mediated ligation of amine sugars to fluorescein-Gaba-LPETGG **30** probe did not work. Modifications to the protocol including testing the reducing agents 2DBPA or TCEP, varying the temperature and introduction of a blocking step did not improve the labelling by sortase. The second labelling method, OPAL, was tested on AcManThz **27**. The OPAL probe **37** was successfully ligated to the surface of HEK cells with the optimal Pd concentration at the decaging step. Evaluating sugar incorporation via the OPAL method on total cell glycoproteins by western blot was unsuccessful, despite varying Pd concentration and addition of a blocking step, and the method requires further optimisation. It would be possible to test the incorporation of AcManThz **27** into cells using radiolabelled sugar, as has previously been done with GlcNAc derivatives [70].

Chapter 4 : Production of sortase variants

Sortase classification

There are currently 6 classes of sortase [180]. The sortases in class A are the most well-characterized and carry out a “housekeeping” role where the enzyme ligates surface proteins containing an LPXTG motif to the cell surface [181]. Class B sortases recognize a different motif, NPQTN, and specifically ligate iron transporter proteins [137]. Class C sortases recognise the LPXTG motif, but the nucleophile is the ε-amine group from a lysine residue on a pilin protein, rather than the oligoglycine used by class A sortase [138]. Class D sortases recognise the similar LPNTA motif, and ligate proteins involved in spore formation in bacteria [139]. Sortases in class E have been more recently studied, recognising the LAXTG motif and linked to the formation of hyphae in *Streptomyces coelicolor* [164, 182]. Finally, class F sortases have also only been investigated more recently but are thought to be another housekeeping sortase class that accepts LPXTG [141].

The previous chapter showed unsuccessful attempts to label cells using the SrtA 5M mutant. These 5 mutations are P94R, D160N, D165A, K190E, and K196T [150]. SrtA 5M, has a 20-fold increase in catalytic conversion of –GGG substrate compared with the wild-type enzyme. It is possible that a mannosamine sugar carrying a GGG peptide would have been more reactive towards sortase if incorporated into the cell surface, but attempts to synthesis the mannosamine-GGG sugar were unsuccessful. It would also remove the advantage of being able to add GGG peptide in excess to remove LPXTG tag, which would be the more favourable reaction for SrtA 5M and therefore create a reversible labelling system.

Although the OPAL system was shown to label the surface of cells, it was hoped the investigation of sortase variants would yield novel enzymes that would also be usable in labelling peptide sugars. One method for obtaining sortase variants would be directed evolution, where enzymes are screened for activity against a desired substrate, and the fastest enzymes have mutations randomly introduced and are again screened for activity against the substrate [160]. Previously, a library of sortase enzymes displayed on the surface of yeast cells were selected for their ability to ligate two substrates, yielding a sortase with a 140-fold increase in activity after 8 rounds of mutation and screening [150]. However, another method of obtaining sortase variants with different

sequences is to harvest sortase from different species of Gram-positive bacteria, as it appears in many species [183].

Selection of Sortase Variants

The industrial partner Prozomix has an extremely large database of collected metagenomic data samples accessed using the Prozomigo software. This metagenomic library was created from samples such as soil, sewage and seawater taken from the environment and the DNA sequenced directly. This yields many new DNA sequences, as the vast majority of bacteria in the environment have never been cultured in the laboratory and have not been investigated for enzymes that may function as useful biocatalysts [184]. A trawl through the database therefore allows the identification of putative enzymes of a particular class of interest. These enzymes are selected for functional diversity and cloned and produced on site (Figure 4.1).

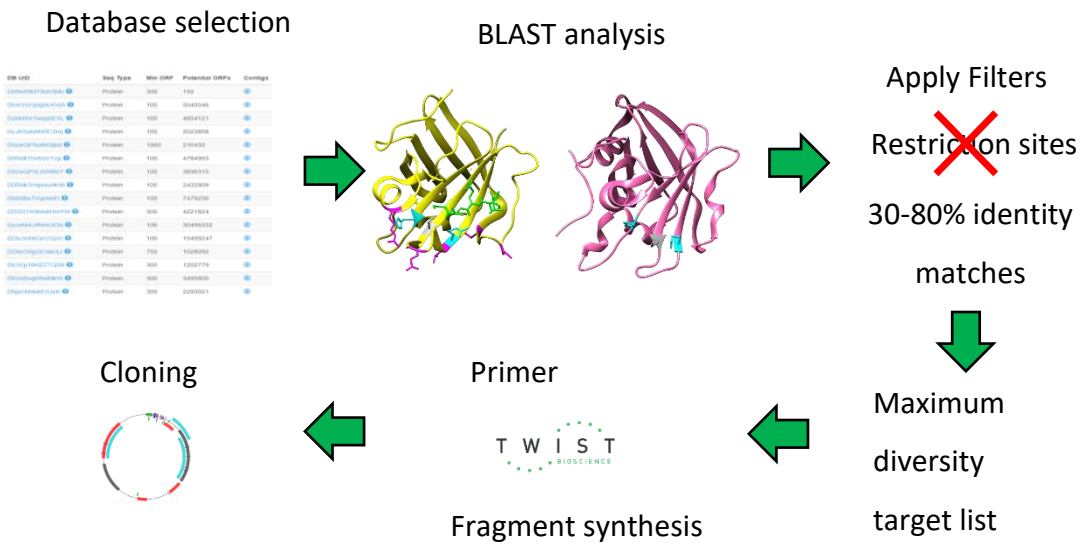


Figure 4.1: An overview of the Prozomix process of creating novel enzymes, from initial database search to cloning and protein production.

The initial searches checked all metagenomes for any sequence in any reading frame that ran for at least 150 base pairs from start codon to stop codon. Three searches were performed to check for matches against the two class A sortases from *Staphylococcus aureus* [135] and *Streptococcus pyogenes* [185], and for matches against the class B

sortase from *S. aureus* [186]. The sequences shown below have the TLXTC sortase active site motif highlighted in red.

SauSrtA

MKKWTNRLMTIAGVVLILVAAYLFAKPHIDNYLHDKDKDEKIEQYDKNVKEQASKDKKQQAKPQI
PKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNL
KAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVGVLDEQKGKDKQLTLITCDDYNEKTGVWEKR
KIFVATEVK

SpySrtA

MVKKQKRRKIKSMSWARKLLIAVLLILGLALLFNKPIRNTLIARNSNKYQVTKVSKKQIKKNKEAKST
FDFQAVEPVSTESVLQAQMAAQQLPVIGGIAIPELGINLPIFKGLGNTELIYGAGTMKEEQVMGGE
NNYSLASHHIFGITGSSQMLFSPLERAQNGMSIYLTDKEKIYEYIIKDVFTVAPERVDVIDDTAGLKE
VTLVTCTDIEATERIIVKGELKTEYDFDKAPADVLKAFNHSYNQVST

SauSrtB

MRMKRFLTIVQILLVVIIIIFGYKIVQTYIEDKQERANYEKLQQKFQMLMSKHQEHVRPQFESLEKIN
KDIVGWIKLSGTSLNYPVLQGKTNHDYLNLDFEREHRRKGSIFMDFRNELKNLNHNTILYGHHVG
DNTMFDVLEDYLKQSFYEKHKIIEFDNKYGKYQLQVFSAYKTTTKDNYIRTDFENDQDYQQFLDET
KRKSVINSDVNVTVKDRIMTLSTCEDAYSETTKRIVVVAKIIKVS

The initial BLAST analysis gives many sequences, so the Prozomigo software can make further adjustments to narrow down the list to a number that is practical to clone. Filters were applied to remove sequences that had restriction sites in the middle that would make cloning difficult. Sequences were also removed if they were below 30% identity or above 80%. The sequences that are below 30% identity are less likely to be in the desired enzyme class, so if cloned may not yield a sortase at all. In contrast, sequences above 80% are very likely to be sortase but are removed to maximise the diversity of the enzymes in the final list. Otherwise, the result could be an extremely long list of mainly 99% identity enzymes, which are very unlikely to give novel results.

On this basis, the initial 64 results from the SauSrtA trawl were filtered and decreased to 55 results. The PrimerGo software, which uses a brute-force approach, was used to create primer sequences for cloning and successfully created primers for 40 of these results. The initial 85 results from SpySrtA were similarly filtered down to 79 results, and after primers were created 58 results remained. SauSrtB gave a much larger 207 results,

which were filtered to 184 results and then after primers were created had 145 results left. The putative sortases found were plotted on a ClustalX tree [187].

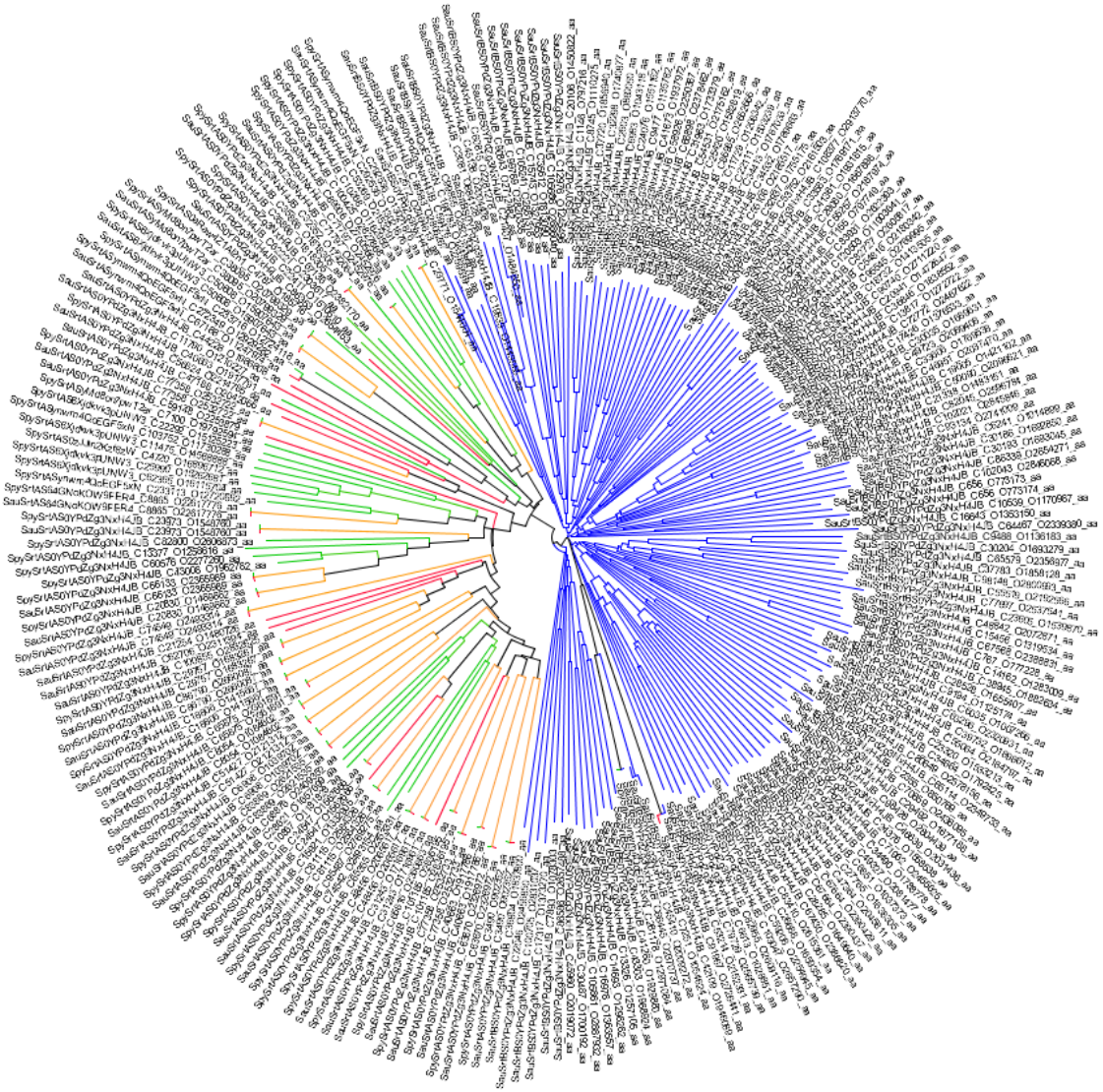


Figure 4.2: ClustalX 2.1 complete alignment of all putative sortase sequences after filters and primer selection. Matches to SauSrtA are in red, SpySrtA in green, and SauSrtB in blue. The orange branches lead to the same sequence matched twice, to SauSrtA and SpySrtA.

The branches represent the projected evolutionary relationship between the sequences. The sequences of putative class A sortases (red and green) and putative class B sortases (blue) clearly form two separate branches. The sequences that match SauSrtA and SpySrtA are not in two clearly defined clusters, but are mixed together. Furthermore, 28 sequences were picked up by both the SauSrtA search and the SpySrtA

search. These can be seen on the orange branches that end in a pair of one red branch and one green branch, pointing to sequences with identical serial numbers. These 28 “double hits” were thought to have the highest chance of being sortase enzymes and were therefore chosen as the most suitable candidates for cloning and expression.

Cloning of Sortase Variants

Next, the sequences were prepared for cloning by checking the sequences for protein domains that can cause problems when cloned into the usual *E. coli* strain at Prozomix. In particular, signal peptides found at the *N*-terminus of proteins are normally removed. Transmembrane helices are also usually removed, as they reduce protein solubility and it would be difficult to purify an insoluble protein. The SignalP and TMHMM servers were used to identify these domains (Figure 4.3), and the EMBL SMART server was used to check if each protein had a sortase domain.

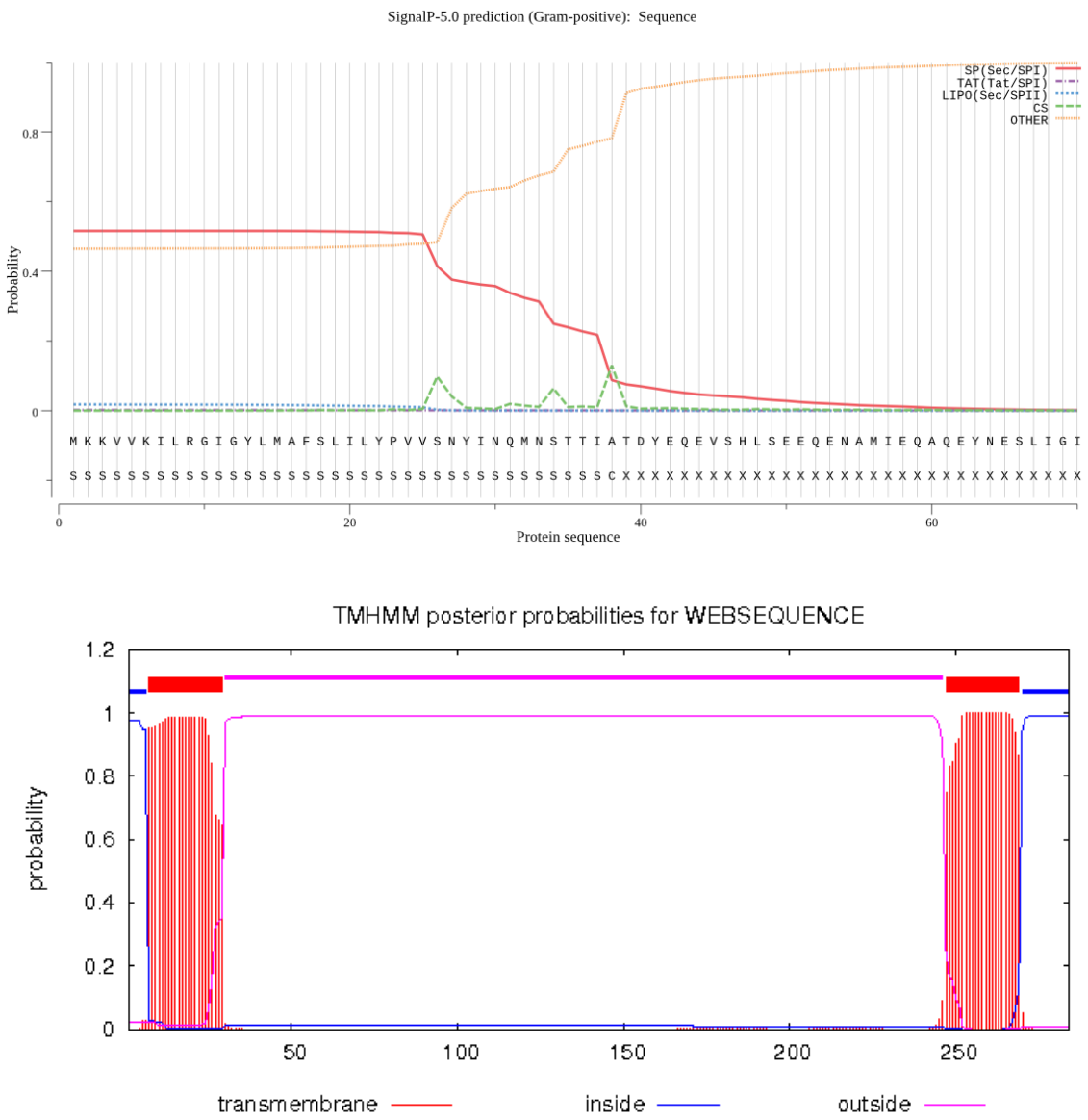


Figure 4.3: Prediction of a signal peptide by SignalP (upper) and prediction of two transmembrane helices by TMHMM (lower), both in Srt006.

Below are the 28 sequences with signal peptides predicted by Signal P highlighted in red, TM helix domains are highlighted or underlined in blue, and the sortase domain present in each sequence highlighted in orange. The proteins are mainly between 16-26 kDa, although there are two outliers of 49 kDa and 58 kDa.

Srt001

MKQKIINIVLGLFFLAGLGITLYPMFSDWYISRHQGQVVDNYDKKAAQMSQKEINEALEKAQEYN
EELLGNVVLTDPFDPAAVEKQNEDYDNLLNIGGDGVMGSVEIPSINVYLPIYHGTDSKSLEKGAGH
LENSSLPIGGKGTHAIISAHTGLPSAKMFDDLTEVKEGDIFYIHVLNRTLAYEVNQIKVVLPENVSDLL

IDKNKDYVTLVTCTPYGINSHRLLVRGERIPYKEAKKKDNSDKKGSNMWKAYALAAVLAVVLIILLW
YYYHIRRKRRRKKARKKRVRRNEK*

Srt002

MTVSLYLKVRKHISCFQIQGRSRKMKNKIINIFLVLILLCGIGVLSYPFISNILQDRKQDQILTEYNEEM
EKLSDAEIEDAKEKAKQYNDSLSNTVVISDPFDPDAADDMSADYISALNLEKNGIMAYIEIPRIDVYE
PVYHGTSEEVLAKGVGHLEGTSLPIGGESTHTVLSGHTGLPEAEIFTKLESVKEGDIFLIHVLNETLAY
KVDQIKVVEPSETDDLKIVPGYDYATLVTCTPYGINSHRLLVRGIRTEYTQDVSDEAAGQAEEGPDG
GGWKAMYEKAIIEGILFAVILVVLINVIVRKFGKKNKSEVAIYDEKEKKEKKPCKKAGKRRRRKSKRT
ER*

Srt003

MCTIDIPKIGVYLPVQHGTGAETLERAVGHVVGTSLPVGGSGTHAVLSAHSGMASAKLFSDIDQL
AEGDMFYIHVLGEVLAYKVDAIHTVLPTDTSLLQIEDGKDQVTLVTCTPFGVNTHRLLVRGHRVPY
TPEQEATAAAEKPVASSWTQHYLTGLGIGLGVVAVIGGGYFLIRRKRHA*

Srt004

MLKKIGKFLLVLVVLGISMGLILYPFVSNYLFENRADGIIDTVEKTADDADEEKYKEEIEAAQKYNAEL
ATGHVVLKDPFVEEKLDEDAKEYNSLLNMADGGVMGFIKIPCIDVSLPIYHGTSAEILELGAGHLQG
TSLPIGGESTHSVITGHTGLSSAKLFTDLTELEEGDMFFLHVMGEKLAYKVGQISVILPEEMDKLTIE
NGKDYCTLVTCTPYGVNTHRLLVRGERTEYTEDKEEEAGTEVKKTQSKWMEEYTKSIIISSSVLAVM
LIVLLIWRHFAGRRRRKRRKKKQTRNNQKRRKKKSSANTNGKRESPDRQKSEARAGKKKRKPREE
PD*

Srt005

MLEKAREYNEKLASSHVVLTDPFSEETNGGISEKEYYQLLNLDNTGVMCSLEIPSINVDLPVYHGTS
NSVLEKGVGHLEGTSLPVGGKDTHAVFTGHTGLNKAKLFTDLTELQKGDQFYIRVLDKILAYEVCQI
DVVLPEDTSKLSVVDGQDLVTLVTCTPYGQNTHRLLVRGKRTKYSPKAYEKERAKKKAGRSQWM
RTYQKAILLGIGITSIVLFIISFIRKKLQKR*

Srt006

MKKVVKILRGIGYLMAFSLILYPVVSNYINQMNSTTIATDYEQEVSHLSEEQENAMIEQAQEYNESL
IGIGTIADPFSESNENQTEDDEYNKLLKIDDTGMMGYIDIPKLDVVLPVYHGTSEKVLQSGVGHLKN
TSLPVGGESCHAVLSGHRGLANAKIFTDLNKMEVGDVFYIKVLHHTFAYQVDQILTVLPSDIDSLQI

EKGKDYVTLVTCTPYAVNTHRLLVRGTRIPYEEAQKIDEEVGLHHTIPFNLILLIGGIVALIIIWVSIKSH
KRRKEKKHEQNK*

Srt007

MLLYNSRENGIKMRKKNVQKKKFYIGDIFRIIVLLIALSVLLYPTVSNYLYEKNGARVISSYDENAVRL
SESEKQAMLEAARQYNRELLGNIELLDPFSPLKKEVDARYQSLLNTNEAGMMGYIRIPKIDVELPIY
HGTEERILQSGVGHFEGTSLPVGGESSHTVLTGHRGLPSKLLFTDLDQMKEGDIFYLKILGETFAYKI
DQILTVLPENTKALTIEPGKDYATLVTCTPYAVNTHRLLVRGIRIPYEEAVRQVPDEKITPTLPFQVKV
LLAAIGVLVLIFIVYQLWKKTRKRKEMRRKRK*

Srt008

MKKKVSKFFIVIIFLAGLSLLLYPFVANQWNNHRQKQLISSYEDNLTQLTEAGDIDYAKELKKAQAY
NDALVPSILPDSFAVADAREEEDSAYMNCLNLTGDGMMGIVEIPKIAIKLPIYHGTSDEVLQQAAG
HLEGSSLPIGGESTHAVISAHRGLPSASLFTDLDQMKIGDHFLIHVLDNTLCYEVDQILVVEPEDTDA
LAVEEGEDLVTLLTCTPYGVNTQRLLVRGHRVDYVADEVAAEQTPLSGISLHTNYLLWVIVGILITG
VFIPFLFIREKRFRKKAAREKETEE*

Srt009

MILPDSFAIADAKEEKDRFYESCLNITGDGIMGMVEIPKIDVELPIYHYTTEEVLKNAAGHLEGSSLP
VGGKSTHSVISAHRGLPSAILFTDLDKMKKGDHFMIHVLDDILCYEVDRISIVKPEDTSDLNVEKGK
DLMTLLTCTPYGVNTERLLVRGHRVPYKEAYSSEKDSSISLETNYILWVVIGLLLTAALIGGMYVRER
KKKHEA*

Srt010

MREKRKSRKRHNKRGQILTPVIFLTSLLLFAAGVGIFIYPALSNYLAQREQKEVIEEYAQTVEQSDKD
KMARQWELAEEYNETLLGDPVHDPFIPGTGYALPDNYESVLNVNKDGVMGYLKIPKIKVDLPIYH
GTSEEVLEKGAGHVDVTALPIGGVNRHPVISAHRGLPSAELFTRLDELEKGDRFFLHILDKTLAYKV
DQVRVIKPEELEQLQTYHDKDYVTLLTCTPYGVNTHRLLVRGERIPYEVAEENGGEAFDDEKNQG
MPQWVKEYVVMIIAGILLLVFAGRMFAARK*

Srt011

MKRKLSILLIVLIFVAGIGIMSYPLVSSAINNYVSRSMVKEYTNRVAQMPSEKTQKMFEEATKYNNS
LSNNMIITDPFDEKAFQKIGANYEKTLNIDDNGLIGYIDVPKINVYLPIYHGTSEEILSKGAGHLQNTS
LPVGGASTHSVISAHSGFPGETFFDYLTDMKVGDEFYVHILDRTLKYEVDQIEVVLPSEINSLRIVDG

EDLVTLLTCTPYGINTHRLLVRGKRVDYDDTKYIETGESLAKFDNGYIFFLGYKIPYAVAIGIIVGFVAL
VVFVVVFLLKRKKKKSNDAKRLRDDSADENQNDSGGDSDD*

Srt012

MFMKKKLPIIAIILIFIIGLGIMSYPLVSSMINNSKFKDGMNDYTETVKKLDKKDNTKLFKSAKKYNHS
LTQTSIITDPFDEEAYKAIGAHYNDVLNVDGKGLIGYVVVPRIDVNLPIYHGSSKKVLEKGAGHLQN
TSMPIGGKSTHAVISAHTGFPDQTFFDNLTDLVKGDIFYIKVLDKTLAYKVDQIKVVLPEDTNDLRII
PNEDHVTLLTCTPYGINTHRLLVRGVRTKYVPEEVAKNKLVKQAPFEKCFYFLGYKIPYWVAGAVIG
GFVLLVVIIVLIAVRKNKKKSSLKHIDDNSEKKVESDEK*

Srt013

MDTTTKLKVGFAALMILSGIGVASYPVISNAVAQRHASEAIQNYDDTVKSMDTEKLDAAKEAARR
YNQQLDSAVERDAAGEAEDTGTSYVDLVDVGESLGYITIPKIDVNLPIYEGTSDDVLQRGVGHLEG
SSYPLGGESTHSVLTGHRGLPEAVLFTDLDKLEEGDCFYLHILDETLAYQVDQIKVVEPEHTEDLEIT
AGQDYCTLVTCTPYAINTHRMLVRGTRIPYNGEEETNTPQAVQYQALRTGTVVKRVVDAWPWIA
VTGSIVIGGEALLLLLLLHRCKKREEDD*

Srt014

MTKDDIATAKEAADEYNSSGNTGNFYNAVVKDGIVGYINIPAIDVNLPIYDNTKDETLELGAGLLEK
SSLPTGKNGTHSVITGHSGLTMSKMFSDLTELTTEDCFYIYFLDEKIKYQVYSIETVKPSEVTNHLGY
NRSEGYCTLMTCTPIGVNTHRLLVHGKRVEESENESVNAKSTSQSNSSSTKTTSNIIKKKRL*

Srt015

MENDRKNQTTRTKKRKKMLTGVLFVTGLLLCIYPAFASVIDRKHQQNVIQTYQGNIEANSEEKLQE
MLGEAERYNEMLWQTNGVLIGDIEQGILEEESYQSQLNLSGTGVMGTLSIPKINVDLPIYHGTEEEI
LANGVGHLQESSLPVGGENTHCILTGHRGLPNAKLFTRLDEMEQGDLFFLTVCGEKLAYEVTKIEIV
HPEDVEGLRIQAEKDLVSLITCTPYGLNTKRLVVTGERIPYTEKQEQEIVPGSMSFRELVFTALPFLYL
VVGIGSMIRKKRGKRSSNETQKD*

Srt016

MKKQKKIWNKIIFIIGLLLVSYPLVSSMIERQHQKDTVATYQGSIEEEDKSRIQDAIAKASEYNNMLF
QTQGASIGDLQNGILSEENYENLLNLSGTGVMGSIEIPKINVDIPIYHGTSEEVLASGVGHFQDSSLP
VGGNNTRCILTGHRGLPNSKLFTRLDELEKEDLFFISTCGETLAYRITEIEVVEPEEAELLEILPEKDLCT
LITCTPYGINTQRLVITGERVPYEKAEYDSIERKLPSFREAFFMVLPFIFMMAGLISFFRRKRHGKKNE
Q*

Srt017

MISTYESEVKQTDKSKVKEQIQSAQKYNDMLFQTRGASVGNIDTEILSDENYESILNLTGKGIMGSI
EIPKIGVDLPIYHGTSDDVLSNGVGHLQNSSFPVGGENTRTVLTGHRGLPNAKLFTRLDELKKDDLF
YIHVGNKTLAYQIYKIEIVKKEEAPDVLGIEEGKDLATLLTCTPYGVNTHRLILTGKRVPYSKKKKEAIE
PEMMSWRELLFTALPFLIVFMLIVRFIFKTRKERKLKS*

Srt018

MALVVGPKQNVAPMKTSATSNDQLQQIVTWVSGQQDRDFGLDMVHFSQTGTSRKITVYTTKFA
SGGRFDTTGVSSPFSWSKDSRHTERWELWIIVSDCDEWYGHGEFNGDEDPDASYIVPEGQNETL
FGFVAVSTSTGDITSGNLLDAVQFRQYYRMDFQTPPSGSGTYSTDGETQVAFDSAASKSDYALVG
SDVTVTAKPVDASHHFLGAYVGDTFVDADQWTANADGSYSLTRTLTANLSVRLMFSANTVVYNV
NGGDPYDSDNPETGHEIYIKRGSSYTNKTAATKRNDDGWRFTDWKYDDDHILGVNHKITYGSKT
NDLSISQNDQTVVSGIPAEKGITLTAQWTYRQAFATTPEEDAEYTSFLNEDPNGVMAYINIPRLKST
LPIYHYTIPESLQKGVGHLRGSALPVGGKNTHAVLSAHRGLPTSRLFTDLDKVRKGDHFYITVLNRTL
AYEVDRISVVKPDQTKELSVQPGKDLVTLVTCTPYGVNTQRLLVRGHRVPYNAAQAKHEAADSAV
SSFTVYGFVATYGTLAIALAGIAVLRRNAAARTPHHAADWPHSLTVSVR*

Srt019

MALLHRRSTSKRASQKMWENNAAMPDDASANDNVSSARPEPAAFVPTPFDEVIDISDLVAERRR
LQRNLWAMRIITIVLLIAAIAVACFPLALQFESDRNLAATTATTAKEVAGWPYPQAEDKLTAARAY
NKKLAESGQPILGEAVDPFAAAQGGSQASGEDSASKKDKEYQSLLNTGNGVMGTIKVPKQSINLP
FYHGTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEALMFTRLDEVKEGDFFYIEVMGETL
GYKVDRISVILPDDTSKLKIVPGEDRVTLMTCTPYGVNTHRLLISGHRVAIPMPAPEPNDVLDARNI
ALGVGLGILAAGLFIIWLARRHKAARIIRHGAFWPWMRRKGESANDTLA*

Srt020

MAFLHRRPKSKHARSRANAPADASAAAIADASVDTAAVDTTTNDSAAADEAAASNTFTQDPFNE
VVDVSDLAAERRRLQRNLLAMRIITILLLIAAIGAACFPLALQFQSSQELAATAATSAKQVAGWPYP
QAEDKLKVARAYNKKLADSGQPVLGEAVDPFSAAQGGSRASGEDSASKKDKEYQSLLDAGNGV
MGTIKVPKQSINLPIYHGTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEAMMFTRLDEVK
KGDFFYIEVMGETLGYKVDNISVILPDDTSKLRIAPGEDRATLMTCTPYGVNTHRLLISGHRVSIPVP
APEPRDVHDGRTIGIAVGVGILLAGLFIIRLLRRHKPARIMHHAAFWPWHRRDGESANDTLRVLAE
PVKKTVGELPPSV*

Srt021

MVAKARAYNRRLAATPQVIGELSGEDGAVKGDFGFKSDREYQSLLDFGDGIMATIEIPSIGVDLPV
RHGADAYALDNGLGHLHGTSLPVGGTSTHSVITGHTGVADKALFTRLTELRKGDVFYVKVAAQTL
AYKVTRIRTVDPDDLRSVRVEPGRDLVTLVTCTPIFLNTYRLLVTGERASMPGDAPYPEDAPKTSRR
DMRPYLVSGVALLAGLPAAVVASRRKTRPYGRHGRKD*

Srt022

MTEDEHEAKMSPRHARTTPRRPATTRRSGAPFLQRLLIVLLALAGSSTLLYPSAGNWFSDRAHASV
TTGYTEVVASMPAGERAAELDRARAYNARLGQGSLTDPYSSSFSGQVDPAERADYLDQLSLTGSG
TIAGIRVPSVGIALPVYHGTSEQTLTLGVGHLEGSALPVGGTGTNSVLTGHSGVPQARLFTDLHGV
EAGDLVYLDVMGETFAYEIDRTDVVLPTETDLLQPVKGEDLVTLVTCTPVGVNSHRLVVQAHRVE
VVDEVLGAATVAGRDGAGFPWWVVALAVVLALAGVFLFRPAAPGRERRSRRRPAPRHARGRHD
GDEREGRRTSHGMSTTLRAGRLEARA*

Srt023

MTVPRSQRKGDAMKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSKGIDSDIT
SEPVKTSDKSDSDKNDTADVTGQPANESDNKSETADDDKEKTDTSSSKNNTVTNTPKPTSHELAS
EGDDNAEAEQQTAGNTDSTENTSEDSETALSENTDSDSDASDDEESATRPYQTLYGEMEKYNKD
LTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIRLPLMLGASTKNLEKGAAVLSETS
MPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVYITNPWETLVYECTSTQVIYPDDVQSILIQ
PGKDMVTLFTCHPYVLGGGPYRYLVFCERVDTQKRKEADGILNPDATTPAPAEEADNVVVTEPEE
VTSSDDSTSESTDEDLSVENPAQNPGNNDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDEN
LENDPEVKKAEERGKRLLALEQTLRYMLPVVLIALSAIIILFRANSKPKRRKRKRKNTVNRKTKPRPKK
KEK*

Srt024

MKQRQVIGLLLLSAGILFMLWPIWSTYQKQATTTDLKQQWKKSLQTVDAKETTKPLATKGDGLLT
IPSLNFEQVILEGASTDVLDRSIGHIKETGAPGKGNYALAGHRSFTKGLHFNKLPQLKTGADVIVTTK
DHRYTYRMESSKLVKPTDLSVLDQDVKQPMITLITCDPPETATNRLIKQGVLIKTETLD*

Srt025

MKRRQILGLLLLTAGILLIIWPLYTSHQKQAAAEELKQRWTQDLKAVDAKETKRPVATSGAGLLTIP
SLDFEQVILEGASTDILDQSIGHIKQTGSPGKGNYALAGHRSFTKGLHFNRLPELKKGAHVIVTTKSH
RYTYKMMTSQLVKPTDVSVLNQDVKQATITLITCDPPETATNRLIKQGTLIKTESL*

Srt026

MKIRKSRLVIAQQRKVDKRKKVGTIVAIVALILIGIVIYCIANFDHIQGQAANYVATNHLSRQRKNEQ
KKKPSYNMKAVQPVSPQSLANAYQHRRDYRAVGQIAIRDHNVLLNIYRGVGNVELNLGAGTMN
QNQKMGEGNYALAGHNMDDGRSFFSPLYTAKVRGNLPNGTTILLTDYKKVYYYKITSSRFISVYNL
RLAWNNKEFKKKPVISLFTCDWTGQGRLFIRGKYTGSQDYKGASKYVRSSFNF*

Srt027

MDAQELPVVGGIAIPEVGINLPIFKGLGNTELTYGAGTMKEDQVMGGENNYSLASHHVFGIAGAS
DMLFSPLDKAKEGMKIYLTDKNKVYTYVISEVKVVQPTEVAVVDDTPGKSEVTLVTCTDAEATQRT
IVKGELKSQVDFDKASSDIIEAFNKSYNQFQN*

Srt028

MEYEKHDTEFVDASKINQPDLAEVASASLDKKQVIGRISIPSISLELPVLKASTEKNLLSGAATVREN
QVMGQGNYALAGHNMSKKGVLFSDVSTLKKNDKIYLYDNANEYEYTVTGVSEVTPDKWEVVED
HGKDEITLITCVSVKDNSKRFVITGDLVGTKAKK*

Previously, the *N*-terminal hydrophobic signal peptide of sortase was shown to have no amino acid conservation [149]. In the case of SrtA5M the deletion of residues 2-59 makes a soluble sortase clone without affecting catalytic activity. Furthermore, there is no overlap in any of the sequences between any transmembrane helices and the predicted sortase domain. Therefore, it was decided that signal peptides and transmembrane helices could all be deleted. Similarly, if transmembrane helices were removed, the small section just before the first helix and the small positively-charged tail after the second helix were also removed. Srt006 is shown as an example of a finished sequence, see appendix for all finished sequences.

Srt006

MTDYEQEVSHLSEEQENAMIEQAQEYNESLIGIGTIADPFSESNENQTEDDEYNKLLKIDDTGMM
GYIDIPKLDVVLPVYHGTSEKVLQSGVGHLKNTSLPVGGESCHAVLSGHRGLANAKIFTDLNKMEV
GDVFYIKVLHHTFAYQVDQILTVLPSDIDSLQIEKGKDYVTLVTCTPYAVNTHRLLVRGTRIPYEEAQ
KIDEEVGLHH

The 28 finished sequences were ordered from Twist Bioscience and received as synthetic linear DNA. The DNA sequences were reverse engineered from the desired protein sequences to make use of optimised codons for *E. coli*. Each sequence also had a His tag to aid purification inserted at the start of the protein, and the sequences were

bookended by restriction sites for NdeI and XhoI to allow straightforward cloning (Figure 4.4).

Figure 4.4: Restriction sites for NdeI and XhoI.

These were cloned into pET28a plasmid (Figure 4.5) and transformed into *E. coli* using the proprietary GRASP protocol. Briefly, restriction digestion with NdeI and XhoI was performed on inserts and pET28a plasmid, and an NZY Gelpure kit used to clean up the DNA. After overnight ligation with DNA ligase, the DNA was transformed into Top10 *E. coli* heat-shock competent cells and plated onto kanamycin-resistant plates.

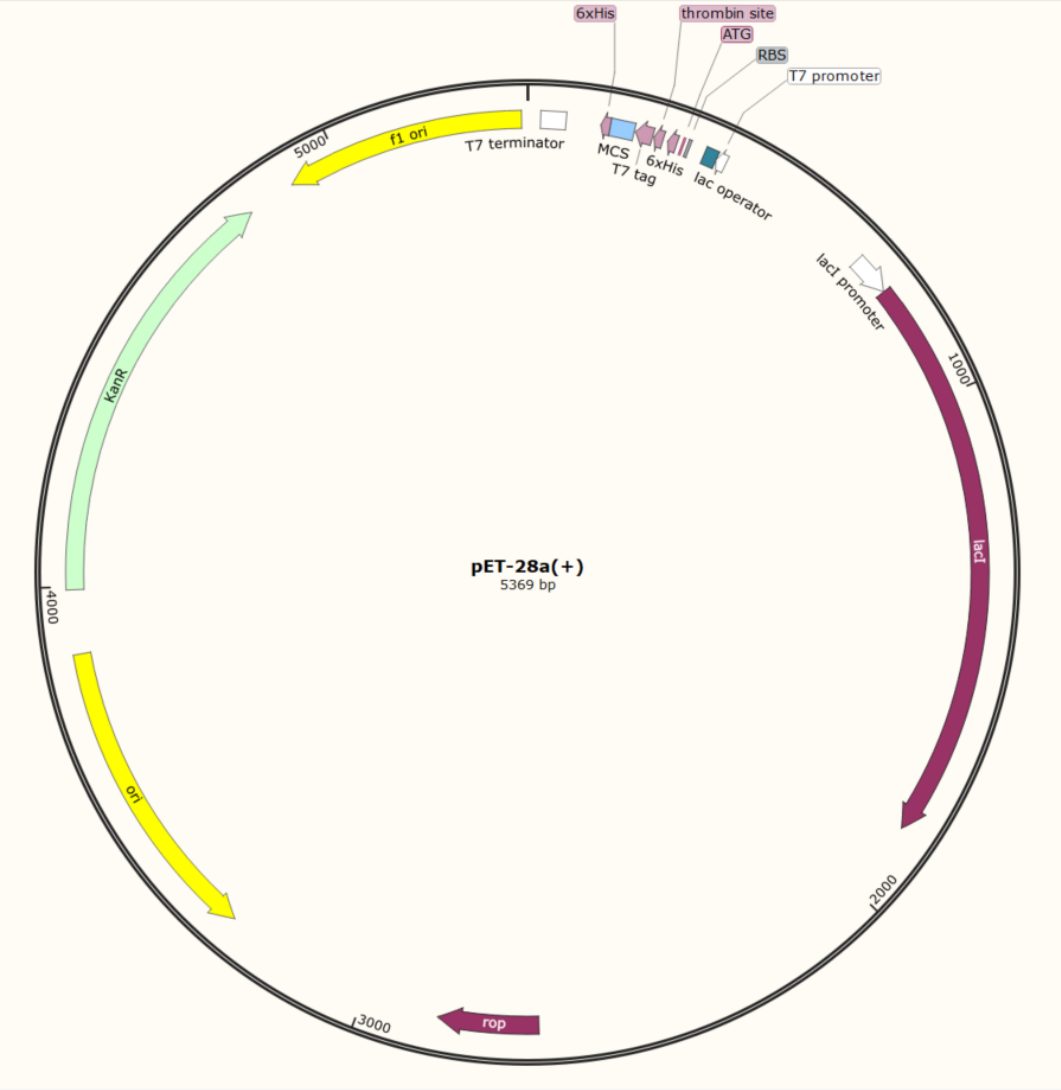


Figure 4.5: A simplified diagram of pET-28a plasmid. The kanamycin resistance gene is in green, the lac operon in purple. In addition, there are many unmarked restriction sites in addition to the tight cluster at the multiple cloning site in blue.

48 colonies were picked using a sterile tip and sent to LGC for sequencing. Purified plasmids were sent back with DNA sequences. These were checked against the 28 desired sequences to find clones that were a 100% match. Srt002, Srt014, Srt018 and Srt028 were not found in any of the sequences, so were lost at the cloning step. In addition, Srt013 frameshifted at the K28 codon from AAA to AAAA and Srt020 underwent an apparent deletion at the A2 codon that also resulted in a frameshift, so both of these sequences had no usable clones. The remaining 22 sequences were verified and cloned into BL21 (DE3) *E. coli* heat-shock competent cells for protein expression.

Expression of Sortase Variants

The *E. coli* was grown in 20 mL LB_{Kan} overnight starter cultures and inoculated into 750 mL flasks of LB_{Kan}, then grown at approx. 25°C for 24 h. Protein pellets were harvested by centrifugation, sonicated and the cell lysate cleared by centrifugation. Lysate was purified on a Ni-column and the Bradford assay used to check fractions for protein. Srt021 is shown for example (Figure 4.6), see experimental for all sortase purification gels.

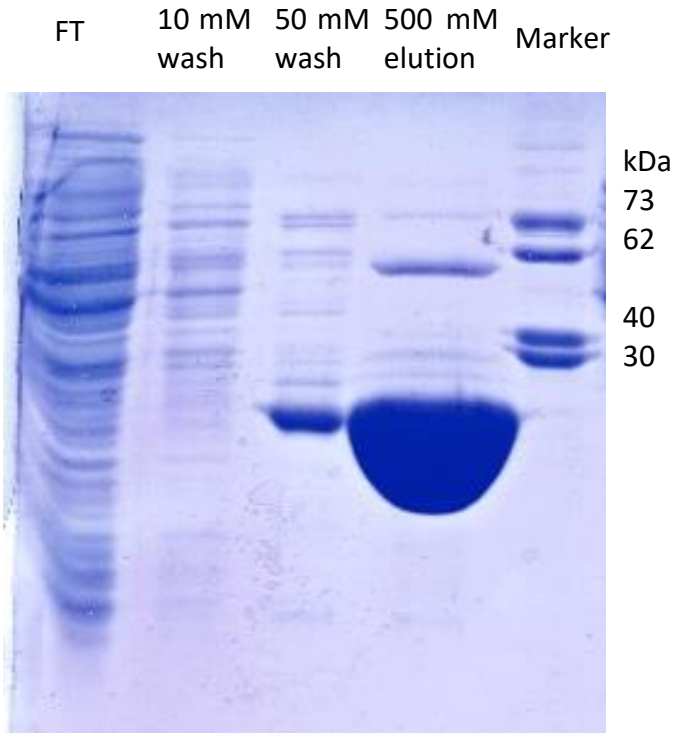


Figure 4.6: Purification of Srt021. The lanes show flowthrough, 10 mM imidazole and 50 mM imidazole wash steps, and a very large protein band about 20 kDa in the 500 mM imidazole elution fractions.

The purified proteins had 50% (w/v) ammonium sulphate added and were stored by refrigeration. All 22 proteins were successfully produced (Table 2).

	Class	Size (Da)	Amount (mg)	Protein conc. (mg / mL)	Protein conc. (µM)	Notes
SrtA001	C	27434	123	15.4	561	
SrtA003	C	18194	154	22	1209	
SrtA004	C	28176	155	15.5	550	
SrtA005	C	24369	157	19.7	808	
SrtA006	C	25555	97	12.2	477	
SrtA007	C	27853	26	5.3	190	Flask grew poorly and was harvested after 2 days
SrtA008	C	27107	98	8.9	328	
SrtA009	C	21850	146	11.2	513	
SrtA010	C	28415	134	13.4	472	
SrtA011	C	28413	34	3.4	120	Gel showed it was largely in the insoluble fraction
SrtA012	C	28448	242	16.2	569	
SrtA015	C	26723	94	9.4	352	
SrtA016	C	26653	31	6.2	233	
SrtA017	C	25883	233	23.3	900	
SrtA019	C	27526	119	10.8	392	
SrtA021	C	24033	258	25.8	1074	
SrtA022	C	25438	181	9.5	373	
SrtA023	C	50998	17	2.8	55	The large size may have led to poor expression
SrtA024	D	20025	167	23.8	1189	
SrtA025	D	19661	137	22.9	1165	
SrtA026	A	25815	10	1.5	58	
SrtA027	A	19791	5	1.5	76	Low concentration - missed the main peak as it eluted

Table 2: Novel sortases and the amounts produced at Prozomix.

Most of the proteins produced are classified as class C sortases. It is likely that the filters applied when choosing sequences from the database eliminated class A sortases when removing sequences above 80% identity. So far, few class C sortases have been engineered to work *in vitro*. The class C sortase of *Corynebacterium diphtheriae* has previously been altered to function *in vitro* by creating a double mutant [163]. The crystal structure of the sortase showed a rigid "lid" partially blocking the active site. This lid contained a conserved DPW sequence, and mutating this sequence to GPG allowed the enzyme to polymerise pilin proteins *in vitro*. Of the 18 class C sortase sequences, 11 have a conserved DPF or DPY sequence, which would suggest low *in vitro* activity. 7 of

the class C sortases, highlighted in Table 2 in green, lack the DPW lid and therefore may have a more exposed active site that is better suited to *in vitro* activity.

Sortase bioinformatics

The 22 sortase sequences were all checked against BLAST for the closest match. The genome of the organism with the matching protein was searched for possible sortase substrates and other possible sortases in order to investigate the likelihood that the sortase made at Prozomix was a genuine sortase with substrates available to ligate to the cell wall. Searches were performed with the wildcard LP?TG string to find the normal class A / class C substrates and LP?TA was also searched for in case of a sortase variant that accepts the motif used by SpySrtA. Another search, TL?TC, checked the genome for any other sortases, any proteins tagged as "sortase" in the genome were also noted.

Srt001

Srt001 is a 100% match to a predicted class C sortase from *Anaerobutyricum hallii* (previously called *Eubacterium hallii*) [188]. *E. hallii* is a Gram-positive intestinal microbe that metabolises lactate, butyrate and glycerol. The whole genome shotgun sequence for strain JCM 31263 has an exact match to Srt001, named "class C sortase". The genome was searched for LPXTG motifs, yielding 17 possible sortase substrates, and the search for LPXTA yielded 14 possible substrates. Of these, the product named "cell surface protein", which has an LPATG motif and is located immediately before Srt001, is a likely substrate for this sortase. The TLXTC search yielded 9 potential sortase enzymes.

```
 E. hallii 1   MKQKIINIVLGLFFLAGLGITLYPMFSDWYISRHQGQVVDNYDKKAAQMSQKEINEALEK  60
               MKQKIINIVLGLFFLAGLGITLYPMFSDWYISRHQGQVVDNYDKKAAQMSQKEINEALEK
      Srt001 1 MKQKIINIVLGLFFLAGLGITLYPMFSDWYISRHQGQVVDNYDKKAAQMSQKEINEALEK  60

 E. hallii 61  AQEYNEELLGNVVLTDPFDPAAVEKQNEDYDNLLNIGGDGVMGSVEIPSINVYLPIYHGT  120
               AQEYNEELLGNVVLTDPFDPAAVEKQNEDYDNLLNIGGDGVMGSVEIPSINVYLPIYHGT
    Srt001 61  AQEYNEELLGNVVLTDPFDPAAVEKQNEDYDNLLNIGGDGVMGSVEIPSINVYLPIYHGT  120

E. hallii 121  DSKSLEKGAGHLENSSLPIGGKGTHAIISAHTGLPSAKMFDDLTEVKEGDIFYIHVLNRT  180
               DSKSLEKGAGHLENSSLPIGGKGTHAIISAHTGLPSAKMFDDLTEVKEGDIFYIHVLNRT
   Srt001 121  DSKSLEKGAGHLENSSLPIGGKGTHAIISAHTGLPSAKMFDDLTEVKEGDIFYIHVLNRT  180
```

```
E. hallii 181 LAYEVNQIKVVLPENVSDLLIDKNKDYVTLVTCTPYGINSHRLLVRGERIPYKEAKKKDN  240
              LAYEVNQIKVVLPENVSDLLIDKNKDYVTLVTCTPYGINSHRLLVRGERIPYKEAKKKDN
   Srt001 181 LAYEVNQIKVVLPENVSDLLIDKNKDYVTLVTCTPYGINSHRLLVRGERIPYKEAKKKDN  240

E. hallii 241 SDKKGSNMWKAYALAAVLAVVLIILLWYYYHIRRKRRRKKARKKRVRRNEK  291
              SDKKGSNMWKAYALAAVLAVVLIILLWYYYHIRRKRRRKKARKKRVRRNEK
   Srt001 241 SDKKGSNMWKAYALAAVLAVVLIILLWYYYHIRRKRRRKKARKKRVRRNEK  291
```

Srt003

The closest match is a class C sortase from *Oscillibacter* sp. that has 91% identity. There are a few substitutions, mainly towards the ends of the protein. The *Oscillibacter* genus contains Gram-negative anaerobic gut bacteria [189]. As the matching organism has been classified to genus but not to species, the *Oscillibacter* sp. isolate CIM:MAG 164 sequence was chosen for the genome search. The genome has 21 proteins with LPXTG motifs. 4 of these marked “hypothetical protein” directly follow 3 separate class C sortases and are almost certainly virulence factors. 24 proteins have LPXTA motifs, one of these is flagged as a virulence factor. The genome also has 4 class C sortases and 1 class B sortase flagged. All of the class C sortases have a TL(I/L/V)TC motif. The class B contains a TLSTC motif. A final search of the genome for TLXTC motifs gave one other hit, a hypothetical predicted protein with a TLSTC motif and therefore another possible sortase.

```
 Oscilli. 30 MAYVDISKINVYLPVQHGTDADTLERAVGHVVGTSLPVGGSSTHAVLSAHSGMASSKLFS  89
             M  +DI KI VYLPVQHGT A+TLERAVGHVVGTSLPVGGS THAVLSAHSGMAS+KLFS
    Srt003 1 MCTIDIPKIGVYLPVQHGTGAETLERAVGHVVGTSLPVGGSGTHAVLSAHSGMASAKLFS  60

 Oscilli. 90 DIDQLTEGDVFYIHVLGEVLAYKVDAIHTVLPTDTSLLQIEDGKDQVTLVTCTPFGVNTH  149
             DIDQL EGD+FYIHVLGEVLAYKVDAIHTVLPTDTSLLQIEDGKDQVTLVTCTPFGVNTH
   Srt003 61 DIDQLAEGDMFYIHVLGEVLAYKVDAIHTVLPTDTSLLQIEDGKDQVTLVTCTPFGVNTH  120

Oscilli. 150 RLLVRGHRVPYVPEQEATAAAEKPVASSWTRHYLTGLGVGLGVVAVAGGGYFLMRRKRHA  209
             RLLVRGHRVPY PEQEATAAAEKPVASSWT+HYLTGLG+GLGVVAV GGGYFL+RRKRHA
  Srt003 121 RLLVRGHRVPYTPEQEATAAAEKPVASSWTQHYLTGLGIGLGVVAVIGGGYFLIRRKRHA  180
```

Srt004

The closest match is a 99% identity class C sortase that is again from an unclassified organism, one that belongs to the *Ruminococcus* genus, *Ruminococcus* species are anaerobic, Gram-positive gut microbes [190]. The genome for *Ruminococcus* sp. AM40-10AC was searched and found to contain 11 proteins with the LPXTG motif, one of them flagged as a cell wall anchor. 8 proteins with LPXTA motifs were also found. The sortase search gave 4 sortases flagged as class B, one flagged as class C, and one unclassified sortase. 14 proteins in total contain a TLXTC motif.

```
  Ruminoco. 1 MLKKIGKFLLVLVVLGISMGLILYPFVSNYLFENRADGIIDTVEKTADDADEEKYKEEIE  60
              MLKKIGKFLLVLVVLGISMGLILYPFVSNYLFENRADGIIDTVEKTADDADEE+YKEEIE
     Srt004 1 MLKKIGKFLLVLVVLGISMGLILYPFVSNYLFENRADGIIDTVEKTADDADEEQYKEEIE  60

 Ruminoco. 61 AAQKYNAELATGHVVLKDPFVEEKLDEDAKEYNSLLNMADGGVMGFIKIPCIDVSLPIYH  120
              AAQKYNAELA GHVVLKDPFVEEKLDEDAKEYNSLLNMADGGVMGFIKIPCIDVSLPIYH
    Srt004 61 AAQKYNAELAAGHVVLKDPFVEEKLDEDAKEYNSLLNMADGGVMGFIKIPCIDVSLPIYH  120

Ruminoco. 121 GTSAEILELGAGHLQGTSLPIGGESTHSVITGHTGLSSAKLFTDLTELEEGDMFFLHVMG  180
              GTSAEILELGAGHLQGTSLPIGGESTHSVITGHTGLSSAKLFTDLTELEEGDMFFLHVMG
   Srt004 121 GTSAEILELGAGHLQGTSLPIGGESTHSVITGHTGLSSAKLFTDLTELEEGDMFFLHVMG  180

Ruminoco. 181 EKLAYKVGQISVILPEEMDKLTIENGKDYCTLVTCTPYGVNTHRLLVRGERTEYTEDKEE  240
              EKLAYKV QISVILPEEMDKLTIENGKDYCTLVTCTPYGVNTHRLLVRGERTEYTEDKEE
   Srt004 181 EKLAYKVDQISVILPEEMDKLTIENGKDYCTLVTCTPYGVNTHRLLVRGERTEYTEDKEE  240

Ruminoco. 241 EAGTEVKKTQSKWMEEYTKSIIISSSVLAVMLIVLLIWRHFAGRRRRKRRKKKQTRNNQK  300
              EAGTEVKKTQSKWMEEYTKSIIISSSVLAVMLIVLLIWRHFAGRRRRKRRKKKQTRNNQK
   Srt004 241 EAGTEVKKTQSKWMEEYTKSIIISSSVLAVMLIVLLIWRHFAGRRRRKRRKKKQTRNNQK  300

Ruminoco. 301 RRKKKSSANTNGKRESPDRQKSEARAGKKKRKPREEPD  338
              RRKKKSSANTNGKRESPDRQKSEARAGKKKRK REEPD
   Srt004 301 RRKKKSSANTNGKRESPDRQKSEARAGKKKRKTREEPD  338
```

Srt005

The closest match is a class C sortase from *A. hallii*, 99% identity. This is the same organism that matched Srt001.

```
   A. hall. 1 MLEKAREYNEKLASSHVVLTDPFSEETNGGISEKEYYQLLNLDNTGVMCSLEIPSINVDL  60
              MLEKAREYNEKLASSHVVLTDPFSEETNGGISEKEYYQLLNLDNTGVMCSLEIPSINVDL
    Srt005 54 MLEKAREYNEKLASSHVVLTDPFSEETNGGISEKEYYQLLNLDNTGVMCSLEIPSINVDL 113

  A. hall. 61 PVYHGTSNSVLEKGVGHLEGTSLPVGGKDTHAVFTGHTGLNKAKLFTDLTELQKGDQFYI 120
              PVYHGTSNSVLEKGVGHLEGTSLPVGGKDTHAVFTGHTGLNKAKLFTDLTELQKGDQFYI
   Srt005 114 PVYHGTSNSVLEKGVGHLEGTSLPVGGKDTHAVFTGHTGLNKAKLFTDLTELQKGDQFYI 173

 A. hall. 121 RVLDKILAYEVCQIDVVLPEDTSKLSVVDGQDLVTLVTCTPYGQNTHRLLVRGKRTKYSP 180
              RVLDKILAYEVC+IDVVLPEDTSKLSVVDGQDLVTLVTCTPYGQNTHRLLVRGKRTKYSP
   Srt005 174 RVLDKILAYEVCRIDVVLPEDTSKLSVVDGQDLVTLVTCTPYGQNTHRLLVRGKRTKYSP 233

A. hallii 181 KAYEKERAKKKAGRSQWMRTYQKAILLGIGITSIVLFIISFIRKKLQKR  229
              KAYEKERAKKKAGRSQWMRTYQKAILLGIGITSIVLFIISFIRKKLQKR
   Srt005 234 KAYEKERAKKKAGRSQWMRTYQKAILLGIGITSIVLFIISFIRKKLQKR  282
```

Srt006

The closest match is a 98% identity class C sortase from an unclassified member of the *Blautia* genus. The *Blautia* genus contains Gram-positive gut bacteria [191]. The genome searched was from the strain Blautia sp. TM10-2. It contains 24 proteins with the LPXTG motif, two of which are flagged as isopeptide-forming pilin proteins that are almost certainly virulence factors. 10 proteins contain the LPXTA motif. There are 7 sortases in total, the others are all stated to be class B. 7 proteins in total have the TLXTC sortase motif, but curiously, some of the proteins flagged as sortase have a mutated SLSTC motif and therefore may not function as sortase.

```
  Blautia 1 MKKVVKILRGIGYLMAFSLILYPVVSNYINQMNSTTIATDYEQEVSHLSEEQENAMIEQA  60
            MKKVVKILRGIGYLMAFSLILYPVVSNYINQMNSTTIATDYEQEVSHLSE+QENAMI+QA
   Srt006 1 MKKVVKILRGIGYLMAFSLILYPVVSNYINQMNSTTIATDYEQEVSHLSEDQENAMIKQA  60

 Blautia 61 QEYNESLIGIGTIADPFSESNENQTEDDEYNKLLKIDDTGMMGYIDIPKLDVVLPVYHGT  120
            QEYNESLIGIG+IADPFSESNENQTEDDEYNKLLKIDDTGMMGYIDIPKLDVVLPVYHGT
  Srt006 61 QEYNESLIGIGSIADPFSESNENQTEDDEYNKLLKIDDTGMMGYIDIPKLDVVLPVYHGT  120
```

```
Blautia 121 SEKVLQSGVGHLKNTSLPVGGESCHAVLSGHRGLANAKIFTDLNKMEVGDVFYIKVLHHT  180
            SEKVLQSGVGHLKNTSLPVGGESCHAVLSGHRGLANAKIFTDLNKMEVGDVFYIKVLHHT
 Srt006 121 SEKVLQSGVGHLKNTSLPVGGESCHAVLSGHRGLANAKIFTDLNKMEVGDVFYIKVLHHT  180

Blautia 181 FAYQVDQILTVLPSDIDSLQIEKGKDYVTLVTCTPYAVNTHRLLVRGTRIPYEEAQKIDE  240
            FAYQVDQILTVLPSD D+LQIEKGKDYVTLVTCTPYAVNTHRLLVRGTRIPYEEAQKIDE
 Srt006 181 FAYQVDQILTVLPSDTDALQIEKGKDYVTLVTCTPYAVNTHRLLVRGTRIPYEEAQKIDE  240

Blautia 241 EVGLHHTIPFNLILLIGGIVALIIIWVSIKSHKRRKEKKHEQN  283
            EVGLHHTIPFNLILLIGGIVALIIIWVSIKSHKRRKEKKHEQN
 Srt006 241 EVGLHHTIPFNLILLIGGIVALIIIWVSIKSHKRRKEKKHEQN  283
```

Srt007

The closest match is a class C sortase from an unclassified *Coprococcus* genus bacterium. The sequence matches with 99% identity starting from Srt007 M13, but the first 12 residues aren't present in the match. *Coprococcus* are also anaerobic gut microbes [192]. The genome searched was for the strain AM25-15LB. 2 proteins contain LPXTG, one flagged as a cell wall anchor domain-containing protein and as an isopeptide-forming fimbrial protein, making them likely sortase substrates. There is also a single LPXTA-containing protein. The search for TLXTC yielded one other hit, a class B sortase.

```
 Coproco. 13 MRKKNVQKKKFYIGDIFRIIVLLIALSVLLYPTVSNYLYEKNGARVISSYDENAVRLSES  72
             MRKKNVQKKKFYIGDIFRIIVL IALSVLLYPTVSNYLYEKNGARVISSYDENAVRLSES
    Srt007 1 MRKKNVQKKKFYIGDIFRIIVLFIALSVLLYPTVSNYLYEKNGARVISSYDENAVRLSES  60

 Coproco. 73 EKQAMLEAARQYNRELLGNIELLDPFSPLKKEVDARYQSLLNTNEAGMMGYIRIPKIDVE  132
             EKQAMLEAARQYNRELLGNIELLDPFSPLKKEVDARYQSLLNTNEAGMMGYIRIPKIDVE
   Srt007 61 EKQAMLEAARQYNRELLGNIELLDPFSPLKKEVDARYQSLLNTNEAGMMGYIRIPKIDVE  120

Coproco. 133 LPIYHGTEERILQSGVGHFEGTSLPVGGESSHTVLTGHRGLPSKLLFTDLDQMKEGDIFY  192
             LPIYHGTEERILQSGVGHFEGTSLPVGGESSHTVLTGHRGLPSKLLFTDLDQMKEGDIFY
      Srt007 LPIYHGTEERILQSGVGHFEGTSLPVGGESSHTVLTGHRGLPSKLLFTDLDQMKEGDIFY  180

Coproco. 193 LKILGETFAYKIDQILTVLPENTKALTIEPGKDYATLVTCTPYAVNTHRLLVRGIRIPYE  252
              LKILGETFAYKIDQILTVLPENTKALTIEPGKDYATLVTCTPYAVNTHRLLVRGIRIPYE
   Srt007 181 LKILGETFAYKIDQILTVLPENTKALTIEPGKDYATLVTCTPYAVNTHRLLVRGIRIPYE  240
```

```
Coproco. 253 EAVRQVPDEKITPTLPFQVKVLLAAIGVLVLIFIVYQLWKKTRKRKEMRRKRK  305
             EAVRQVPDEKITPTLPFQVKVLLAAIGVLVLIFIVYQLWKKTRKRKEMRRKRK
  Srt007 241 EAVRQVPDEKITPTLPFQVKVLLAAIGVLVLIFIVYQLWKKTRKRKEMRRKRK  293
```

Srt008

This matches a class C sortase from *Eubacterium ramulus* with 99% identity. *E. ramulus* is a Gram-positive anaerobic human intestinal bacterium [193]. The genome of strain ATC 29099 was searched. 13 LPXTG motif proteins are present, including one flagged as "LPXTG cell wall anchor domain-containing protein", and 2 proteins marked as isopeptide fimbrial proteins that are likely substrates for class C sortase. There are also 18 LPXTA motif proteins. Searching for TLXTC yields 3 other class C sortases and 2 class B sortases.

```
  E. ramul. 1 MKKKVSKFFIVIIFLAGLSLLLYPFVANQWNNHRQKQLISSYEDNLTQLTEAGDIDYAKE  60
              MKKKVSKFFIVIIFLAGLSLLLYPFVANQWNNHRQKQLISSYEDNLTQLTEAGDIDYAKE
     Srt008 1 MKKKVSKFFIVIIFLAGLSLLLYPFVANQWNNHRQKQLISSYEDNLTQLTEAGDIDYAKE  60

 E. ramul. 61 LKKAQAYNDALVPSILPDSFAVADAREEEDSAYMNCLNLTGDGMMGIVEIPKIAIKLPIY  120
              LKKAQAYNDALVPSILPDSFAVADAREEEDS YMNCLNLTGDGMMGIVEIPKIAIKLPIY
    Srt008 61 LKKAQAYNDALVPSILPDSFAVADAREEEDSTYMNCLNLTGDGMMGIVEIPKIAIKLPIY  120

E. ramul. 121 HGTSDEVLQQAAGHLEGSSLPIGGESTHAVISAHRGLPSASLFTDLDQMKIGDHFLIHVL  180
              HGTSDEVLQQAAGHLEGSSLPIGGESTHAVISAHRGLPSASLFTDLDQMKIGDHFLIHVL
   Srt008 121 HGTSDEVLQQAAGHLEGSSLPIGGESTHAVISAHRGLPSASLFTDLDQMKIGDHFLIHVL  180

E. ramul. 181 DNTLCYEVDQILVVEPEDTDALAVEEGEDLVTLLTCTPYGVNTQRLLVRGHRVDYVADEV  240
              DNTLCYEVDQILVVEPEDTDALAVEEGEDLVTLLTCTPYGVNTQRLLVRGHRVDYVADEV
   Srt008 181 DNTLCYEVDQILVVEPEDTDALAVEEGEDLVTLLTCTPYGVNTQRLLVRGHRVDYVADEV  240

E. ramul. 241 AAEQTPLSGISLHTNYLLWVIVGILITGVFIPFLFIREKRFRKKAAREKETEE  293
              AAEQTPLSGISLHTNYLLWVIVGILITGVFI  LFIREKRFRKKAAREKETEE
   Srt008 241 AAEQTPLSGISLHTNYLLWVIVGILITGVFILILFIREKRFRKKAAREKETEE  29
```

Srt009

This is another 100% match, to a class C sortase from *Clostridium scindens*. *C. scindens* is a Gram-positive anaerobe, found in the human colon [194]. The searched genome was

taken from strain ATCC 35704. This yielded 10 LPXTG-containing proteins, with one flagged as a cell wall anchor protein, one as a drug export protein, and one as a peptidoglycan-binding protein, all cell surface proteins. 14 proteins contain the LPXTA motif. The search for TLXTC motifs showed one other class C sortase, and 2 class B sortases.

```
  C. scind. 1 ILPDSFAIADAKEEKDRFYESCLNITGDGIMGMVEIPKIDVELPIYHYTTEEVLKNAAG  60
              MILPDSFAIADAKEEKDRFYESCLNITGDGIMGMVEIPKIDVELPIYHYTTEEVLKNAAG
    Srt009 74 MILPDSFAIADAKEEKDRFYESCLNITGDGIMGMVEIPKIDVELPIYHYTTEEVLKNAAG  133

 C. scind. 61 HLEGSSLPVGGKSTHSVISAHRGLPSAILFTDLDKMKKGDHFMIHVLDDILCYEVDRISI  120
              HLEGSSLPVGGKSTHSVISAHRGLPSAILFTDLDKMKKGDHFMIHVLDDILCYEVDRISI
   Srt009 134 HLEGSSLPVGGKSTHSVISAHRGLPSAILFTDLDKMKKGDHFMIHVLDDILCYEVDRISI  193

C. scind. 121 VKPEDTSDLNVEKGKDLMTLLTCTPYGVNTERLLVRGHRVPYKEAYSSEKDSSISLETNY  180
              VKPEDTSDLNVEKGKDLMTLLTCTPYGVNTERLLVRGHRVPYKEAYSSEKDSSISLETNY
   Srt009 194 VKPEDTSDLNVEKGKDLMTLLTCTPYGVNTERLLVRGHRVPYKEAYSSEKDSSISLETNY  253

C. scind. 181 ILWVVIGLLLTAALIGGMYVRERKKKHEA  209
              ILWVVIGLLLTAALIGGMYVRERKKKHEA
   Srt009 254 ILWVVIGLLLTAALIGGMYVRERKKKHEA  282
```

Srt010

This matches with 99% identity a class C sortase from *Ruminococcus torques*. *R. torques* is a Gram-positive gut microbe [192]. The genome of strain L2-14 was searched. There are 14 proteins in total with an LPXTG motif, including 3 pilin proteins each named “Cna protein B-type domain” that are likely substrates for class C sortase. There are also 8 proteins with an LPXTA motif. The search for sortase gives 2 other sortases, both class B.

```
  R. torques 1 MREKRKSRKRHNKRGQILTPVIFLTSLLLFAAGVGIFIYPALSNYLAQREQKEVIEEYAQ  60
               MREKRKSRKRHNKRGQILTPVIFLTSLLLFAAGVGIFIYPALSNYLAQREQKEVIEEYAQ
     Srt010 1 MREKRKSRKRHNKRGQILTPVIFLTSLLLFAAGVGIFIYPALSNYLAQREQKEVIEEYAQ  60

 R. torqu. 61 TVEQSDKDKMARQWELAEEYNETLLGDPVHDPFIPGTGYALPDNYESVLNVNKDGVMGYL  120
              TVEQ DKDKMARQWELAEEYNETLLGDPVHDPFIPGTGYALPDNYESVLNVNKDGVMGYL
   Srt010 61 TVEQIDKDKMARQWELAEEYNETLLGDPVHDPFIPGTGYALPDNYESVLNVNKDGVMGYL  120
```

```
R. torqu. 121 KIPKIKVDLPIYHGTSEEVLEKGAGHVDVTALPIGGVNRHPVISAHRGLPSAELFTRLDE 180
              KIPKIKVDLPIYHGTSEEVLEKGAGHVDVTALPIGGVNRHPVISAHRGLPSAELFTRLDE
   Srt010 121 KIPKIKVDLPIYHGTSEEVLEKGAGHVDVTALPIGGVNRHPVISAHRGLPSAELFTRLDE 180

R. torqu. 181 LEKGDRFFLHILDKTLAYKVDQVRVIKPEELEQLQTYHDKDYVTLLTCTPYGVNTHRLLV 240
              LEKGDRFFLHILDKTLAYKVDQVRVIKPEELEQLQTYHDKDYVTLLTCTPYGVNTHRLLV
   Srt010 181 LEKGDRFFLHILDKTLAYKVDQVRVIKPEELEQLQTYHDKDYVTLLTCTPYGVNTHRLLV 240

R. torqu. 241 RGERIPYEVAEENGGEAFDDEKNQGMPQWVKEYVVMIIAGILLLVFAGRMFAARK 295
              RGERIPYEVAEENGGEAFDDEKNQGMPQWVKEYVVMIIAGILLLVFAGRMFAARK
   Srt010 241 RGERIPYEVAEENGGEAFDDEKNQGMPQWVKEYVVMIIAGILLLVFAGRMFAARK 295
```

Srt011

The closest match is a class C sortase from *Ruminococcus bromii*, 99% identity. *R. bromii* is another Gram-positive gut microbe [190]. The genome of strain AM31-32 was searched. There are 3 proteins in total with an LPXTG motif. There are also 3 proteins with an LPXTA motif. Searching for the TLXTC motif gave 1 other class C sortase with a TMFTC motif, 1 class D, and 1 class B.

```
 R. bromii 1 MKRKLSILLIVLIFVAGIGIMSYPLVSSAINNYVSRSMVKEYTNRVAQMPSEKTQKMFEE 60
             MKRKLSILLIVLIFVAGIGIMSYPLVSSAINNYVSRS VKEYT+RVAQMPSEKTQKMFEE
   Srt011 38 MKRKLSILLIVLIFVAGIGIMSYPLVSSAINNYVSRSRVKEYTDRVAQMPSEKTQKMFEE 97

R. bromii 61 ATKYNNSLSNNMIITDPFDEKAFQKIGANYEKTLNIDDNGLIGYIDVPKINVYLPIYHGT 120
             ATKYNNSLSNNMIITDPFDEKAFQKIGANYEKTLNIDDNGLIGYIDVPKINVYLPIYHGT
   Srt011 98 ATKYNNSLSNNMIITDPFDEKAFQKIGANYEKTLNIDDNGLIGYIDVPKINVYLPIYHGT 157

R. bromii 121 SEEILSKGAGHLQNTSLPVGGASTHSVISAHSGFPGETFFDYLTDMKVGDEFYVHILDRT 180
              SEEILSKGAGHLQNTSLPVGGASTHSVISAHSG+PGETFFDYLTDMKVGDEFYVHILDRT
   Srt011 158 SEEILSKGAGHLQNTSLPVGGASTHSVISAHSGYPGETFFDYLTDMKVGDEFYVHILDRT 217

R. bromii 181 LKYEVDQIEVVLPSEINSLRIVDGEDLVTLLTCTPYGINTHRLLVRGKRVDYDDTKYIET 240
              LKYEVDQIEVVLPSEINSLRIVDGEDLVTLLTCTPYGINTHRLLVRGKRVDYDDTKYIET
   Srt011 218 LKYEVDQIEVVLPSEINSLRIVDGEDLVTLLTCTPYGINTHRLLVRGKRVDYDDTKYIET 277
```

```
R. bromii 241 GESLAKFDNGYIFFLGYKIPYAVAIGIIVGFVALVVFVVVFLLKRKKKKSNDAKRLRDDS  300
              GESLAKFDNGYIFFLGYKIPYAVAIGIIVGFVALVVFVVVFLLKRKKKKSN AKRLRDDS
   Srt011 278 GESLAKFDNGYIFFLGYKIPYAVAIGIIVGFVALVVFVVVFLLKRKKKKSNGAKRLRDDS  337

R. bromii  01 ADENQNDSGGDSDD  314
              ADENQNDSGGD DD
   Srt011 338 ADENQNDSGGDCDD  351
```

Srt012

This gave another match from a *Ruminococcus* genus bacterium, with no species classification. The match is 98% identity. Without an identified species, the searches through the previous 2 genomes for *R. torques* and *R. bromii* will be considered sufficient.

```
 Ruminoco. 1 MFMKKKLPIIAIILIFIIGLGIMSYPLVSSMINNSKFKDGMNDYTETVKKLDKKDNTKLF  60
             MFMKKKLPIIAIILIFIIGLGIMSYPLVSSMINNSKFKDGMNDYTETVK LDKKDNTKLF
    Srt012 1 MFMKKKLPIIAIILIFIIGLGIMSYPLVSSMINNSKFKDGMNDYTETVKTLDKKDNTKLF  60

 Ruminoco. 61 KSAKKYNHSLTQTSIITDPFDEEAYKAIGAHYNDVLNVDGKGLIGYVVVPRIDVNLPIYH  120
              KSAKKYNHSLTQTSIITDPFDEEAYKAIGAHYNDVLNVDGKGLIGYVVVPRIDVNLPIYH
   Srt012 61 KSAKKYNHSLTQTSIITDPFDEEAYKAIGAHYNDVLNVDGKGLIGYVVVPRIDVNLPIYH  120

Ruminoco. 121 GSSKKVLEKGAGHLQNTSMPIGGKSTHAVISAHTGFPDQTFFDNLTDLVKGDIFYIKVLD  180
              GSSKKVLEKGAGHLQNTSMPIGGKSTHAVISAHTGFPDQTFFDNLTDLVKGDIFYIKVLD
   Srt012 121 GSSKKVLEKGAGHLQNTSMPIGGKSTHAVISAHTGFPDQTFFDNLTDLVKGDIFYIKVLD  180

Ruminoco. 181 KTLAYKVDQIKVVLPEDTNDLRIIPNEDHVTLLTCTPYGINTHRLLVRGVRTKYVPEEVA  240
              KTLAYKVDQIKVVLPEDTNDLRIIPNEDHVTLLTCTPYGINTHRLLVRGVRTKYVPEEVA
   Srt012 181 KTLAYKVDQIKVVLPEDTNDLRIIPNEDHVTLLTCTPYGINTHRLLVRGVRTKYVPEEVA  240

Ruminoco. 241 KNKLVKQAPFEKCFYFLGYKIPYWVAGAVIGGFVLLVVIIVLIAVRKNKKKSSLKHIDDN  300
              KNKLVKQAPFEKCFYFLGYKIPYWVAGAVIGGFVLLVVIIVLIAVRKNKKKSSLKHIDDN
   Srt012 241 KNKLVKQAPFEKCFYFLGYKIPYWVAGAVIGGFVLLVVIIVLIAVRKNKKKSSLKHIDDN  300

Ruminoco. 301 SEKKVESDEK  310
              SEKKVESDEK
   Srt012 301 SEKKVESDEK  310
```

Srt015

This is a 100% match to a class C sortase from a bacterium in the *Lachnospiraceae* genus. This genus is also composed of anaerobic gut bacteria [195]. The genome of the strain MGYG-HGUT-02492 was searched. 9 LPXTG motif proteins are present, one flagged as "isopeptide-forming domain-containing fimbrial protein" and another as "LPXTG cell wall anchor domain-containing protein", both of which are likely sortase substrates. There are also 12 LPXTA motif proteins. Searching for TLXTC yields 2 sortases, 1 each from class C and class B.

```
 Lachnosp.   1 MENDRKNQTTRTKKRKKMLTGVLFVTGLLLCIYPAFASVIDRKHQQNVIQTYQGNIEANS  60
               MENDRKNQTTRTKKRKKMLTGVLFVTGLLLCIYPAFASVIDRKHQQNVIQTYQGNIEANS
      Srt015   1 MENDRKNQTTRTKKRKKMLTGVLFVTGLLLCIYPAFASVIDRKHQQNVIQTYQGNIEANS  60

Lachnosp.  61 EEKLQEMLGEAERYNEMLWQTNGVLIGDIEQGILEEESYQSQLNLSGTGVMGTLSIPKIN  120
               EEKLQEMLGEAERYNEMLWQTNGVLIGDIEQGILEEESYQSQLNLSGTGVMGTLSIPKIN
     Srt015  61 EEKLQEMLGEAERYNEMLWQTNGVLIGDIEQGILEEESYQSQLNLSGTGVMGTLSIPKIN  120

Lachnos.  121 VDLPIYHGTEEEILANGVGHLQESSLPVGGENTHCILTGHRGLPNAKLFTRLDEMEQGDL  180
               VDLPIYHGTEEEILANGVGHLQESSLPVGGENTHCILTGHRGLPNAKLFTRLDEMEQGDL
    Srt015 121 VDLPIYHGTEEEILANGVGHLQESSLPVGGENTHCILTGHRGLPNAKLFTRLDEMEQGDL  180

Lachnos.  181 FFLTVCGEKLAYEVTKIEIVHPEDVEGLRIQAEKDLVSLITCTPYGLNTKRLVVTGERIP  240
               FFLTVCGEKLAYEVTKIEIVHPEDVEGLRIQAEKDLVSLITCTPYGLNTKRLVVTGERIP
    Srt015 181 FFLTVCGEKLAYEVTKIEIVHPEDVEGLRIQAEKDLVSLITCTPYGLNTKRLVVTGERIP  240

Lachnos.  241 YTEKQEQEIVPGSMSFRELVFTALPFLYLVVGIGSMIRKKRGKRSSNETQKD  292
               YTEKQEQEIVPGSMSFRELVFTALPFLYLVVGIGSMIRKKRGKRSSNETQKD
    Srt015 241 YTEKQEQEIVPGSMSFRELVFTALPFLYLVVGIGSMIRKKRGKRSSNETQKD  292
```

Srt016

This matches a class C sortase from *Blautia hansenii* with 99% identity. This is another Gram-positive gut bacterium [191]. The genome of the strain DSM 20583 was searched. The genome has 14 proteins with LPXTG motifs. Several are flagged as proteins that are likely sortase substrates, including "peptidoglycan-binding protein", "isopeptide-forming domain-containing fimbrial protein", "cell wall anchor protein", "SpaH/EbpB family LPXTG-anchored major pilin" and "Cna B-type domain-containing protein". 8

proteins have LPXTA motifs. The genome also has 3 class B sortases and one other class C sortase.

```
  B. hanse. 1 MKKQKKIWNKIIFIIGLLLVSYPLVSSMIERQHQKDTVATYQGSIEEEDKSRIQDAIAKA  60
              MKKQKKIWNKIIFIIGLLLVSYPLVSSMIERQHQKD VATYQGSIEEEDKSRIQDAIAKA
     Srt016 1 MKKQKKIWNKIIFIIGLLLVSYPLVSSMIERQHQKDAVATYQGSIEEEDKSRIQDAIAKA  60

  B. hanse. 61 SEYNNMLFQTQGASIGDLQNGILSEENYENLLNLSGTGVMGSIEIPKINVDIPIYHGTSE  120
              SEYNNMLFQTQGASIGDLQNGILSEENYENLLNLSGTGVMGSIEIPKINVDIPIYHGTSE
    Srt016 61 SEYNNMLFQTQGASIGDLQNGILSEENYENLLNLSGTGVMGSIEIPKINVDIPIYHGTSE  120

B. hanse. 121 EVLASGVGHFQDSSLPVGGNNTRCILTGHRGLPNSKLFTRLDELEKEDLFFISTCGETLA  180
              EVLASGVGHFQDSSLPVGGNNTRCILTGHRGLPNSKLFTRLDELEKEDLFFISTCGETLA
   Srt016 121 EVLASGVGHFQDSSLPVGGNNTRCILTGHRGLPNSKLFTRLDELEKEDLFFISTCGETLA  180

B. hanse. 181 YRITEIEVVEPEEAELLEILPEKDLCTLITCTPYGINTQRLVITGERVPYEKAEYDSIER  240
              YRITEIEVVEPEEAELLEILPEKDLCTLITCTPYGINTQRLVITGERVPYEKAEYDSIER
   Srt016 181 YRITEIEVVEPEEAELLEILPEKDLCTLITCTPYGINTQRLVITGERVPYEKAEYDSIER  240

B. hanse. 241 KLPSFREAFFMVLPFIFMMAGLISFFRRKRHGKKNEQ  277
              KLPSFREAFFMVLPFIFMMAGLISFFRRKRHGKKNEQ
   Srt016 241 KLPSFREAFFMVLPFIFMMAGLISFFRRKRHGKKNEQ  277
```

Srt017

This matches a class C sortase belonging to an unclassified bacterium from the *Eubacterium* genus with 98% identity. This was already covered at Srt001.

```
  Eubacter. 1 MISTYESEVKQTDKSKVKEQIQSAQKYNDMLFQTRGASVGNIDTEILSDENYESILNLTG  60
              MISTYESEVKQTDKSKVKEQIQSAQKYNDMLFQTRGASVGNIDTEILSD+NYESILNLTG
    Srt017 23 MISTYESEVKQTDKSKVKEQIQSAQKYNDMLFQTRGASVGNIDTEILSDKNYESILNLTG  82

 Eubacter. 61 KGIMGSIEIPKIGVDLPIYHGTSDDVLSNGVGHLQNSSFPVGGENTRTVLTGHRGLPNAK  120
              KGIMGSIEIPKIGVDLPIYHGTSDDVLSNGVGHLQNSSFPVGGENTRTVLTGHRGLPNAK
    Srt017 83 KGIMGSIEIPKIGVDLPIYHGTSDDVLSNGVGHLQNSSFPVGGENTRTVLTGHRGLPNAK  142

Eubacter. 121 LFTRLDELKKDDLFYIHVGNKTLAYQIYKIEIVKKEEAPDVLGIEEGKDLATLLTCTPYG  180
              LFTRLDELKKDDLFYIHVGNKTLAYQIYKIEIVKKEEAPDVLGIEEGKDLATLLTCTPYG
   Srt017 143 LFTRLDELKKDDLFYIHVGNKTLAYQIYKIEIVKKEEAPDVLGIEEGKDLATLLTCTPYG  202
```

```
Eubacter. 181 VNTHRLILTGKRVPYSKKKKEAIEPEMMSWRELLFTALPFLIVFMLIVRFIFKTRKERKL  240
              +NTHRLILTGKRVPYS+KKKEAIEPEMMSWRELLFTALPFLIVFMLIVRFIF  RKERKL
   Srt017 203 INTHRLILTGKRVPYSEKKKEAIEPEMMSWRELLFTALPFLIVFMLIVRFIFNKRKERKL  262

Eubacter. 241 KS  242
              KS
   Srt017 263 KS  264
```

Srt019

A 100% match to a class C sortase from *Bifidobacterium longum*. *B. longum* is another anaerobic gut bacterium [196]. The genome for strain 157F was searched and contains 13 LPXTG proteins, including 5 marked “putative cell surface protein”, and 32 LPXTA proteins. 4 TLXTC motifs appear, 3 in unclassified sortases and one in a putative oxidoreductase.

```
 B. longum 1 MALLHRRSTSKRASQKMWENNAAMPDDASANDNVSSARPEPAAFVPTPFDEVIDISDLVA  60
           MALLHRRSTSKRASQKMWENNAAMPDDASANDNVSSARPEPAAFVPTPFDEVIDISDLVA
    Srt019 1 MALLHRRSTSKRASQKMWENNAAMPDDASANDNVSSARPEPAAFVPTPFDEVIDISDLVA  60

 B. longum 61 ERRRLQRNLWAMRIITIVLLIAAIAVACFPLALQFESDRNLAATTATTAKEVAGWPYPQA  120
              ERRRLQRNLWAMRIITIVLLIAAIAVACFPLALQFESDRNLAATTATTAKEVAGWPYPQA
    Srt019 61 ERRRLQRNLWAMRIITIVLLIAAIAVACFPLALQFESDRNLAATTATTAKEVAGWPYPQA  120

B. longum 121 EDKLTAARAYNKKLAESGQPILGEAVDPFAAAQGGSQASGEDSASKKDKEYQSLLNTGNG  180
              EDKLTAARAYNKKLAESGQPILGEAVDPFAAAQGGSQASGEDSASKKDKEYQSLLNTGNG
   Srt019 121 EDKLTAARAYNKKLAESGQPILGEAVDPFAAAQGGSQASGEDSASKKDKEYQSLLNTGNG  180

B. longum 181 VMGTIKVPKQSINLPFYHGTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEALMF  240
              VMGTIKVPKQSINLPFYHGTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEALMF
   Srt019 181 VMGTIKVPKQSINLPFYHGTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEALMF  240

B. longum 241 TRLDEVKEGDFFYIEVMGETLGYKVDRISVILPDDTSKLKIVPGEDRVTLMTCTPYGVNT  300
              TRLDEVKEGDFFYIEVMGETLGYKVDRISVILPDDTSKLKIVPGEDRVTLMTCTPYGVNT
   Srt019 241 TRLDEVKEGDFFYIEVMGETLGYKVDRISVILPDDTSKLKIVPGEDRVTLMTCTPYGVNT  300

B. longum 301 HRLLISGHRVAIPMPAPEPNDVLDARNIALGVGLGILAAGLFIIWLARRHKAARIIRHGA  360
              HRLLISGHRVAIPMPAPEPNDVLDARNIALGVGLGILAAGLFIIWLARRHKAARIIRHGA
   Srt019 301 HRLLISGHRVAIPMPAPEPNDVLDARNIALGVGLGILAAGLFIIWLARRHKAARIIRHGA  360
```

```
B. longum 361 FWPWMRRKGESANDTLA  377
              FWPWMRRKGESANDTLA
   Srt019 361 FWPWMRRKGESANDTLA  377
```

Srt021

The first hit is a sortase from *Bifidobacterium adolescentis* with 100% identity. The Blast protein has 84 extra residues at the start. *B. adolescentis* is another Gram-positive intestinal bacterium [197]. The genome searched was 487B. There are 15 proteins with LPXTG motifs, including two flagged as "isopeptide-forming domain-containing fimbrial protein" that are likely pilus proteins. Another likely sortase C candidate is a protein marked "SpaH/EbpB family LPXTG-anchored major pilin", and another is "LPXTG cell wall anchor domain-containing protein". Searching for LPXTA motifs gives 15 proteins. Searching for TLXTC gives 7 sortases, 5 stated to be class C, 1 class E and 1 unclassified.

```
B. adoles.   1 MVAKARAYNRRLAATPQVIGELSGEDGAVKGDFGFKSDREYQSLLDFGDGIMATIEIPSI  60
               MVAKARAYNRRLAATPQVIGELSGEDGAVKGDFGFKSDREYQSLLDFGDGIMATIEIPSI
    Srt021  88 MVAKARAYNRRLAATPQVIGELSGEDGAVKGDFGFKSDREYQSLLDFGDGIMATIEIPSI  147

B. adoles.  61 GVDLPVRHGADAYALDNGLGHLHGTSLPVGGTSTHSVITGHTGVADKALFTRLTELRKGD  120
               GVDLPVRHGADAYALDNGLGHLHGTSLPVGGTSTHSVITGHTGVADKALFTRLTELRKGD
   Srt021 148 GVDLPVRHGADAYALDNGLGHLHGTSLPVGGTSTHSVITGHTGVADKALFTRLTELRKGD  207

B. adoles.  21 VFYVKVAAQTLAYKVTRIRTVDPDDLRSVRVEPGRDLVTLVTCTPIFLNTYRLLVTGERA  180
               VFYVKVAAQTLAYKVTRIRTVDPDDLRSVRVEPGRDLVTLVTCTPIFLNTYRLLVTGERA
   Srt021 208 VFYVKVAAQTLAYKVTRIRTVDPDDLRSVRVEPGRDLVTLVTCTPIFLNTYRLLVTGERA  267

B. adoles.181 SMPGDAPYPEDAPKTSRRDMRPYLVSGVALLAGLPAAVVASRRKTRPYGRHGRKD  235
               SMPGDAPYPEDAPKTSRRDMRPYLVSGVALLAGLPAAVVASRRKTRPYGRHGRKD
   Srt021 268 SMPGDAPYPEDAPKTSRRDMRPYLVSGVALLAGLPAAVVASRRKTRPYGRHGRKD  322
```

Srt022

MTEDEHEAKMSPRHARTTPRRPATTRRSGAPFLQRLLIVLLALAGSSTLLYPSAGNWFSDRAHASV
TTGYTEVVASMPAGERAAELDRARAYNARLGQGSLTDPYSSSFSGQVDPAERADYLDQLSLTGSG
TIAGIRVPSVGIALPVYHGTSEQTLTLGVGHLEGSALPVGGTGTNSVLTGHSGVPQARLFTDLHGV
EAGDLVYLDVMGETFAYEIDRTDVVLPTETDLLQPVKGEDLVTLVTCTPVGVNSHRLVVQAHRVE

VVDEVLGAATVAGRDGAGFPWWVVALAVVLALAGVFLFRPAAPGRERRSRRRPAPRHARGRHD
GDEREGRRTSHGMSTTLRAGRLEARA*

Srt022 is shown above as it is very different to anything in the NCBI database. The closest hit is a class C sortase from *Cellulomonas hominis*, 53% identity, alignment shown below. There are a lot of substitutions, plus two gaps of 4 and 5 amino acids. Also, the first 11 amino acids are not in the NCBI sequence, which starts at the underlined proline. The other red sequence of 49 amino acids at the end is also not present in the NCBI sequence. This is very likely to be a novel enzyme, possibly from an unclassified species of *Cellulomonas*.

```
C. homin. 12  PRHARTTPRRPATTRRSGAPFLQRLLIVLLALAGSSTLLYPSAGNWFSDRAHASVTTGYT  71
              PR   + PR    TRR+    L+R   VL+AL G   LLYPSAGNWF+ RA A+ T  Y
      Srt022 PRAGSSGPRGDGVTRRA----LRRAPTVLVALVGVGLLLYPSAGNWFATRAQAASTNSYV  92

C. homin. 72  EVVASMPAGERAAELDRARAYNARLGQGSLTDPYSSSFSGQVDPAE--RADYLDQLSLTG  129
                V  +P  +R A L RAR YNAR+G  +LTDPY+      VD  E   A+Y DQL + G
      Srt022 AAVERIPQADREAVLRRAREYNARIGVQTLTDPYT-----DVDEVEATSAEYDDQLRVEG  147

C. homin. 130 SGTIAGIRVPSVGIALPVYHGTSEQTLTLGVGHLEGSALPVGGTGTNSVLTGHSGVPQAR  189
              +  +A +R+P++ + LPV+HGTSE+TLT GVGHL+GS+LPVGG  TN+VLTGHSG+P A
      Srt022 AEAMARVRIPAIDVRLPVFHGTSEETLTAGVGHLQGSSLPVGGPDTNAVLTGHSGLPNAI  207

C. homin. 190 LFTDLHGVEAGDLVYLDVMGETFAYEIDRTDVVLPTETDLLQPVKGEDLVTLVTCTPVGV  249
              LFTDLH V  GDLVY+DV+GET AY +D  + VLPTETDLL+ V GEDL+TLVTCTPVGV
      Srt022 LFTDLHQVVTGDLVYIDVLGETLAYRVDLIETVLPTETDLLRVVAGEDLLTLVTCTPVGV  267

C. homin. 250 NSHRLVVQAHRVEVVD-EVLGAATVAGRDGAGFPWWVVALAVVLALAGVFLFRP  302
              NSHRLV++A RVE+ + +   A T        F WW++  A  +    V L  P
      Srt022 NSHRLVIRAERVELPEGDEAAAETSTEAQAVPFAWWLLGAAGAVLTGAVVLVVP  321
```

Srt023

MTVPRSQRKGDAMKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSKGIDSDIT
SEPVKTSDKSDSDKNDTADVTGQPANESDNKSETADDDKEKTDTSSSKNNTVTNTPKPTSHELAS
EGDDNAEAEQQTAGNTDSTENTSEDSETALSENTDSDSDASDDEESATRPYQTLYGEMEKYNKD

LTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIRLPLMLGASTKNLEKGAAVLSETS
MPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVYITNPWETLVYECTSTQVIYPDDVQSILIQ
PGKDMVTLFTCHPYVLGGGPYRYLVFCERVDTQKRKEADGILNPDATTPAPAEEADNVVVTEPEE
VTSSDDSTSESTDEDLSVENPAQNPGNNDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDEN
LENDPEVKKAEERGKRLLALEQTLRYMLPVVLIALSAIIILFRANSKPKRRKRKRKNTVNRKTKPRPKK
KEK*

When checked against the BLAST database in 2019, Srt023 was also very different to any sequence in the database. The best hit was a class C sortase from *Blautia* sp. BCRC 81119, 48% identity. Again, 12 residues from the start are not present, there are several substitutions, and 3 very large gaps in the NCBI sequence in addition. This was thought to be another novel enzyme.

```
Blautia 13  MKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSKGID--SDITSEP  70
            MK  +RV+IGG L+A GCG FLYPNFREWNTQREVE +I  FD+TY +G++    +T E
Srt023 1    MKNTARVLIGGALVACGCGFFLYPNFREWNTQREVEQIIETFDKTYDRGLNESGSVTHEN  60

Blautia 71  -VKTSDKSDSDKNDTADVTGQPANESDNKSETADDDKEKTDTSSSKNNTVTNTPKPTSHE  129
              KT+ K+D+ K++ A VT                                T T TP
Srt023 61   DSKTTKKADNKKDEHAKVT------------------------------TATATP------  85

Blautia 130 LASEGDDNAEAEQQTAGNTDSTENTSEDSETALSENTDSDSDASDDEESATRPYQTLYGE  189
              +EG D +  E    G T+  +N S  ++TA                    RPYQ LY +
Srt023 86   --AEGQDKSTDETAQNGQTNGDKNKS--AQTA-------------------RPYQALYDK  122

Blautia 190 MEKYNKDLTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIRLPLMLGAST  249
            ME YNKDL  NGQ I DAWSYEQQP DL+ + ID DNP IGYIEIPDMKIRLPL+LGAST
Srt023 123  MESYNKDLIDNGQYIADAWSYEQQPFDLTGLGIDSDNPAIGYIEIPDMKIRLPLLLGAST  182

Blautia 250 KNLEKGAAVLSETSMPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVYITNPWETLV  309
            +NLEKGAAV+S TSMPIGG DTNCVIAGHRGWEGSAYFQ+IENMKKGS+VY+TNPWETLV
Srt023 183  ENLEKGAAVMSNTSMPIGGTDTNCVIAGHRGWEGSAYFQYIENMKKGSRVYVTNPWETLV  242

Blautia 310 YECTSTQVIYPDDVQSILIQPGKDMVTLFTCHPYVLGGGPYRYLVFCERVDTQKRKEADG  369
            YEC   +VI P+DV SI+IQ GKDMVTL +CHPYVLGGGPYRY+VFCERVDTQKRKE DG
Srt023 243  YECADIEVINPNDVDSIMIQKGKDMVTLLSCHPYVLGGGPYRYVVFCERVDTQKRKETDG  302
```

```
 Blautia 370 ILNPDATTPAPAEEADNVVVTEPEEVTSSDDSTSESTDEDLSVENPAQNPGNNDDLSVDK  429
             ++NP +++P+          +  EE+  +D +   +T    SV +  Q
 Srt023 303 VMNPKSSSPS----------STSEEIGDTDSAEKNATSASNSVSHTNQT-----------  341

Blautia 430 EEQSTQDEAGTGNTSTDISPTPAKETTLDENLENDPEVKKAEERGKRLLALEQTLRYMLP  489
                                        E+ P + + +E G  LLALEQTLRYMLP
 Srt023 342 --------------------------------ESSPAIVQNQE-GLDLLALEQTLRYMLP  368

Blautia 490 VVLIALSAIIILFR---ANSKPKRRKRKRKNTVNRKTKPRPKKKE  531
             +++I  SA+I+L R   +    K++ R+ K  +N+++K + K+++
 Srt023 369 IIVIMTSALIVLLRKPKSKHSKKQKSRRHKRNINKQSKNKNKRRK  413
```

A check of the database in 2021 revealed a new 99% match to a class C sortase from Ruminococcus sp. AM40-10AC, so the enzyme is no longer as novel as was previously thought. This organism was already covered at Srt004.

```
Blautia 1   MTVPRSQRKGDAMKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSK  60
            MTVPRSQRKGDAMKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSK
Srt023  7   MTVPRSQRKGDAMKKVSRVIIGGTLIALGCGCFLYPNFREWNTQREVESVIHKFDETYSK  66

Blautia 61  GIDSDITSEPVKTSDKSDSDKNDTADVTGQPANESDNKSETADDDKEKTDTSSSKNNTVT  120
            GIDSDITSEPVKTSDKSDSDKNDT DVTGQPANESDNKSETADDDKEKTDTSSSKNNTVT
Srt023  67  GIDSDITSEPVKTSDKSDSDKNDTTDVTGQPANESDNKSETADDDKEKTDTSSSKNNTVT  126

Blautia 121 NTPKPTSHELASEGDDNAEAEQQTAGNTDSTENTSEDSETALSENTDSDSDASDDEESAT  180
            NTPKPTSHELAS+GDD AEAEQQTAGNTDSTENTSEDSETALS NTDSDSDASDDEESAT
Srt023  127 NTPKPTSHELASKGDDKAEAEQQTAGNTDSTENTSEDSETALSGNTDSDSDASDDEESAT  186

Blautia 181 RPYQTLYGEMEKYNKDLTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIR  240
            RPYQTLYGEMEKYNKDLTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIR
Srt023  187 RPYQTLYGEMEKYNKDLTTNGQDIVDAWSYEQQPLDLSSVDIDEDNPVIGYIEIPDMKIR  246

Blautia 241 LPLMLGASTKNLEKGAAVLSETSMPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVY  300
            LPLMLGASTKNLEKGAAVLSETSMPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVY
Srt023  247 LPLMLGASTKNLEKGAAVLSETSMPIGGKDTNCVIAGHRGWEGSAYFQFIENMKKGSKVY  306

Blautia 301 ITNPWETLVYECTSTQVIYPDDVQSILIQPGKDMVTLFTCHPYVLGGGPYRYLVFCERVD  360
            ITNPWETLVYECTSTQVIYPDDVQSILIQPGKDMVTLFTCHPYVLGGGPYRYLVFCERVD
Srt023  307 ITNPWETLVYECTSTQVIYPDDVQSILIQPGKDMVTLFTCHPYVLGGGPYRYLVFCERVD  366
```

```
Blautia 361 TQKRKEADGILNPDATTPAPAEEADNVVVTEPEEVTSSDDSTSESTDEDLSVENPAQNPG  420
            TQKRKEADGILNPDATTPAPAEEADNVVVT+PEE  SS++STSE+TDEDLSVE+ AQ PG
Srt023  367 TQKRKEADGILNPDATTPAPAEEADNVVVTKPEEAESSNNSTSENTDEDLSVESTAQTPG  426

Blautia 421 NNDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDENLENDPEVKKAEERGKRLLAL  480
            +NDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDENLENDPEVKKAEERGKRLLAL
Srt023  427 SNDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDENLENDPEVKKAEERGKRLLAL  486

Blautia 481 EQTLRYMLPVVLIALSAIIILFRANSKPKRRKRKRKNTVNRKTKPRPKKKEK  532
            EQTLRYMLPVVLIALSAIIILFRANSKPKRRKRKRKNTVNRKTKPRPKKKEK
Srt023  487 EQTLRYMLPVVLIALSAIIILFRANSKPKRRKRKRKNTVNRKTKPRPKKKEK  538
```

Srt024

This is a class D sortase with a 99% identity match to *Exiguobacterium oxidotolerans*. *E. oxidotolerans* is a Gram-positive alkaliphile [198]. The genome for the strain JCM 12280 was searched. 13 LPXTG-containing proteins and 17 LPXTA-containing proteins are present. There are also 4 other sortases, 2 are flagged as class D, one of these has an unusual VISTC motif instead of TLXTC, and 2 class E or F sortases (the labelling for both is inconsistent) with the unusual motifs of TLISC motif and NLITC.

```
E. oxidoto.  1 MKQRQVIGLLLLSAGILFMLWPIWSTYQKQATTTDLKQQWKKSLQTVDAKETTKPLATKG  60
             MKQRQVIGLLLLSAGILFMLWPIWSTYQKQATTTDLKQQWKKSLQTVDAKETTKPLAT G
     Srt024  1 MKQRQVIGLLLLSAGILFMLWPIWSTYQKQATTTDLKQQWKKSLQTVDAKETTKPLATTG  60

E. oxidot. 61 DGLLTIPSLNFEQVILEGASTDVLDRSIGHIKETGAPGKGNYALAGHRSFTKGLHFNKLP  120
             DGLLTIPSL+FEQVILEGASTDVLDRSIGHIKETGAPGKGNYALAGHRSFTKGLHFNKLP
    Srt024 61 DGLLTIPSLDFEQVILEGASTDVLDRSIGHIKETGAPGKGNYALAGHRSFTKGLHFNKLP  120

E. oxido. 121 QLKTGADVIVTTKDHRYTYRMESSKLVKPTDLSVLDQDVKQPMITLITCDPPETATNRLI  180
             QLKTGADVIVTTKDHRYTYRMESSKLVKPTDLSVLDQDVKQPMITLITCDPPETATNRLI
   Srt024 121 QLKTGADVIVTTKDHRYTYRMESSKLVKPTDLSVLDQDVKQPMITLITCDPPETATNRLI  180

E. oxido. 181 KQGVLIKTETLD  192
             KQGVLIKTETLD
   Srt024 181 KQGVLIKTETLD  192
```

Srt025

This matches a class D sortase from *Exiguobacterium undae* with 98% identity. *E. undae* is a Gram-positive orange-pigmented species first isolated from pond water in Germany [199]. The genome for strain KCTC 3810 genome was searched. No LPXTG motifs, or LPNTA motifs were found in the genome, although 2 proteins with the LPXTA motif are present. The only sortase present is the 99% match.

```
  E. undae 1 MKRRQILGLLLLTAGILLIIWPLYTSHQKQAAAEELKQRWTQDLKAVDAKETKRPVATSG  60
             MKR++ILGLLLLTAGILLIIWPLYTSHQKQ AAEELKQRWTQDLKAVDAKETKRPVATSG
    Srt025 1 MKRQKILGLLLLTAGILLIIWPLYTSHQKQTAAEELKQRWTQDLKAVDAKETKRPVATSG  60

 E. undae 61 AGLLTIPSLDFEQVILEGASTDILDQSIGHIKQTGSPGKGNYALAGHRSFTKGLHFNRLP  120
             AGLLTIPSLDFEQVILEGASTDILDQSIGHIKQTGSPGKGNYALAGHRSFTKGLHFNRLP
   Srt025 61 AGLLTIPSLDFEQVILEGASTDILDQSIGHIKQTGSPGKGNYALAGHRSFTKGLHFNRLP  120

E. undae 121 ELKKGAHVIVTTKSHRYTYKMMTSQLVKPTDVSVLNQDVKQATITLITCDPPETATNRLI  180
             ELK+GAHVIVTTKSHRYTYKMMTSQLVKPTDVSVLNQDVKQATITLITCDPPETATNRLI
  Srt025 121 ELKQGAHVIVTTKSHRYTYKMMTSQLVKPTDVSVLNQDVKQATITLITCDPPETATNRLI  180

E. undae 181 KQGTLIKTESL  191
             KQGTLIKTESL
  Srt025 181 KQGTLIKTESL  191
```

Srt026

Srt026 is one of just two class A sortases made at Prozomix and matches a protein from *Lactobacillus crispatus* with 93% identity. *L. crispatus* is a Gram-positive anaerobe found in the gut and vagina [200]. The genome from strain MV-3A-US was searched. There are 4 proteins with the LPXTG motif and 3 with LPXTA. Srt026 has an unusual SLFTC sortase motif, which may impede its function as it is mutated. There is one other predicted class A sortase.

```
  L. crisp. 1 MKIRKSRLVIAQQRKVDKRKKVGTIVAIVALILIGIVIYCIANFDHIQGQAANYVATNHL  60
              MKI+KSRL I QQRK DK+KK+GTIVAIVALILIG+ IYCIANF+H+QG+AANYVATNHL
     Srt026 1 MKIKKSRLAINQQRKSDKKKKIGTIVAIVALILIGVAIYCIANFNHLQGKAANYVATNHL  60
```

```
 L. crisp. 61 SRQRKNEQKKKPSYNMKAVQPVSPQSLANAYQHRRDYRAVGQIAIRDHNVLLNIYRGVGN  120
              S QRKN+QKKKPS+NMKAVQPVSPQSLANAYQHRRDYRAVGQIAIRDHNVLLNIYRGVGN
    Srt026 61 SSQRKNKQKKKPSFNMKAVQPVSPQSLANAYQHRRDYRAVGQIAIRDHNVLLNIYRGVGN  120

L. crisp. 121 VELNLGAGTMNQNQKMGEGNYALAGHNMDDGRSFFSPLYTAKVRGNLPNGTTILLTDYKK  180
              VELNLGAGTMNQ QKMGEGNYALAGHNMDDGRSFFSPLYTAKVRGNLPNGTTILLTDYKK
   Srt026 121 VELNLGAGTMNQYQKMGEGNYALAGHNMDDGRSFFSPLYTAKVRGNLPNGTTILLTDYKK  180

L. crisp. 181 VYYYKITSSRFISVYNLRLAWNNKEFKKKPVISLFTCDWTGQGRLFIRGKYTGSQDYKGA  240
              VYYYKITSS+FISVYNLRLAWNNKEFKKKPVISLFTCDWTGQGRLFIRGKYTGSQDYKGA
   Srt026 181 VYYYKITSSQFISVYNLRLAWNNKEFKKKPVISLFTCDWTGQGRLFIRGKYTGSQDYKGA  240

L. crisp. 241 SKYVRSSFNF  250
              SKYVR SFNF
   Srt026 241 SKYVRGSFNF  250
```

Srt027

The other class A sortase matches with 99% identity to a sortase from *Streptococcus salivarius*. *S. salivarius* is a Gram-positive bacterium that lives in the oral cavity and upper respiratory tract [201]. The genome of the representative strain NCTC 8618 was searched. There are 23 predicted proteins in the genome that contain LPXTG motifs, including several marked as “cell wall anchor domain-containing protein” or “YSIRK-type signal peptide-containing protein”, which is also processed by sortase. The LPXTA motif appears in 7 proteins. There is one other sortase in the genome, a class B sortase.

```
  S. saliv. 1 MDAQELPVVGGIAIPEVGINLPIFKGLGNTELTYGAGTMKEDQVMGGENNYSLASHHVFG  60
              MDAQELPVVGGIAIPEVGINLPIFKGLGNTELTYGAGTMKEDQVMGGENNYSLASHHVFG
    Srt027 91 MDAQELPVVGGIAIPEVGINLPIFKGLGNTELTYGAGTMKEDQVMGGENNYSLASHHVFG  150

 S. saliv. 61 IAGASDMLFSPLDKAKEGMKIYLTDKNKVYTYVISEVKVVQPTEVAVVDDTPGKSEVTLV  120
              IAGASDMLFSPLDKAKEGMKIYLTDKNKVYTYVISEVKVVQPTEVAVVDDTPGKSEVTLV
   Srt027 151 IAGASDMLFSPLDKAKEGMKIYLTDKNKVYTYVISEVKVVQPTEVAVVDDTPGKSEVTLV  210

S. saliv. 121 TCTDAEATQRTIVKGELKSQVDFDKASSDIIEAFNKSYNQFQN  163
              TCTDAEATQRTIVKGELKSQVDFDKASSDIIEAFNKSYNQFQ+
   Srt027 211 TCTDAEATQRTIVKGELKSQVDFDKASSDIIEAFNKSYNQFQS  253
```

Overall, the sortases made at Prozomix primarily matched to known proteins from Gram-positive, anaerobic gut bacteria. The novel enzymes Srt022 and Srt023 were of particular interest. The next chapter deals with the testing of the Prozomix enzymes *in vitro*.

Conclusion

The search of the Prozomigo database yielded many potential sortases. These were narrowed down according to likelihood of being a sortase and diversity of sequence. 28 sequences that were homologous to both SpySrtA and SauSrtA were chosen for cloning. The subsequent cloning using the GRASP method yielded 22 sortase proteins. Many of the sortases were class C and matched sortases that have already been documented on protein databases, although most have not been studied in detail. One potential sortase was a novel enzyme that had no known match, and was therefore of particular interest.

Chapter 5 : Testing of sortase variants

In vitro Testing of Hydrolysis and Ligation

Chapter 3 demonstrated that the commercially available sortase SrtA 5M could rapidly ligate the peptides GGG and fluorescein-GABA-LPETGG **30** *in vitro*. Most of the 22 Prozomix sortase variants are class C, which are involved in pilus polymerisation [202]. The most well-studied class C sortase is from *Corynebacterium diphtheriae* [138]. The SpaABC pilus is assembled from a tip pilin, SpaC, and shaft pilins SpaA and SpaB (Figure 5.1), while the independently assembled SpaDEF pilus has the tip pilin SpaF and the shaft pilins SpaD and SpaE [203]. The *C. diphtheriae* sortase C is a pilin-specific sortase that cleaves the tip pilin at the LPXTG motif and forms an acyl sortase complex that undergoes nucleophilic attack from the lysine side chain at residue K190 of the shaft pilin SpaA [138]. SpaA also has an LPXTG motif towards its *C*-terminus, so can be further ligated to another shaft pilin and so on to polymerise a long chain. The ligation to the minor pilin SpaB at the residue K139 terminates polymerisation [204] and allows the complete pilin to be anchored to the cell wall by its motif LAFTG, which is recognised by the housekeeping sortase SrtF [205].

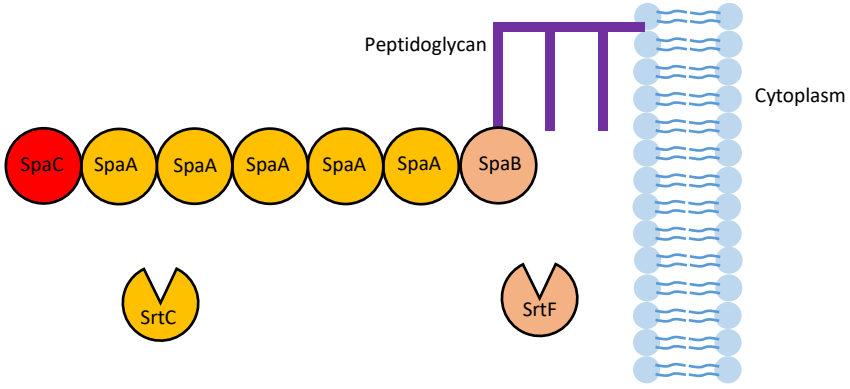


Figure 5.1: Assembly of the pilus in C. diphtheriae. SrtC ligates the tip pilin SpaC and the shaft pilin SpaA, and can further polymerise SpaA subunits to make a long chain. SrtF ligates the terminating SpaB subunit to the cell wall.

Previously, a class C sortase has been engineered to function *in vitro* [163]. The Ton-That group analysed the structure of the class C sortase from *C. diphtheriae* and found a conserved *N*-terminal "lid" over the active site that appeared to block access (Figure 5.2). The mutation of two residues D81G and W83G converted the DPW "lid" to GPG

and exposed more of the active site. The new sortase, $SrtA^{2M}$, successfully polymerized SpaA *in vitro*.

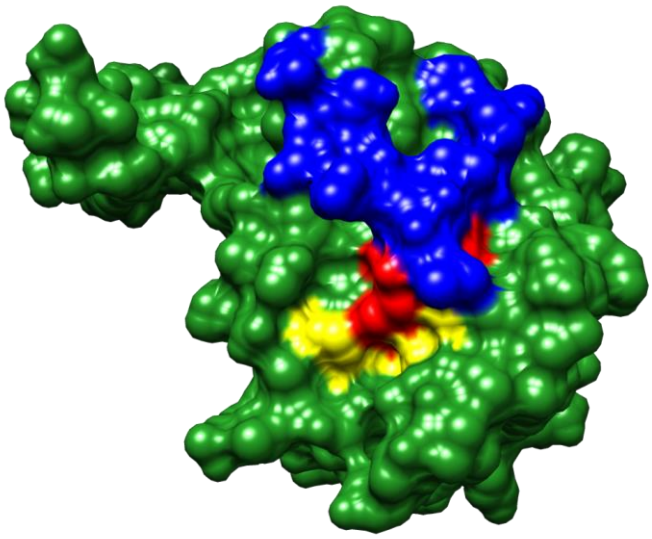


Figure 5.2: Sortase C engineered to function *in vitro* [163]. The "lid" (blue) is conserved and the DPW residues (red) partially block the active site (yellow). Mutation to GPG allows access to the active site resulting in a functional enzyme *in vitro*. PDB ID 5K9A.

The new sortases fall into two groups (Table 3). Eleven of the sortases have a conserved DPF or DPY sequence at this point, which is functionally very similar to the DPW sequence as the third residue in each case is a large hydrophobic amino acid. Five sortases lack this part of the sequence and therefore presumably lack the "lid" structure over the active site. Two other sortases already have a mutation to a smaller, more flexible lid, as in Srt015 DIE and Srt016 DLQ. It was therefore hypothesised that some of these sortases would work *in vitro* without further mutations being required.

Srt001	DPF
Srt003	None
Srt004	DPF
Srt005	DPF
Srt006	DPF
Srt007	DPF
Srt008	None
Srt009	DPW
Srt010	DPF
Srt011	DPF
Srt012	DPF
Srt015	DIE
Srt016	DLQ
Srt017	None
Srt019	DPF
Srt021	None
Srt022	DPY
Srt023	None

Table 3: The class C new sortases and the protein sequence corresponding to the DPW residues in *C. diphtheriae* sortase C.

The sortases were initially tested for hydrolysis and ligation activity against the Gaba-LPETGG and Gaba-LPETAA peptides tested on the previous sortases. As glycine is not the usual nucleophile for class C sortases during ligation with LPETGG, it was also necessary to make some peptides that mimicked pilin proteins. SpaA has a conserved sequence WXXXVXVYPK before the lysine that participates in bonding [138]. A peptide based on this sequence in SpaA, VYPKHG, was made using solid-phase peptide synthesis, with an acetyl group added to the *N*-terminus to prevent ligation by the amine group. Similar peptides were made, VYAKHG to see if the Pro was important to the reaction, and VYPRHG as a negative control without the lysine. The peptide FVAKNEG was also tested. The class D sortases, Srt024 and Srt025, were also tested for ligation between LPNTAG peptide and diaminopimelic acid, which is the natural substrate for class D sortase [206].

Hydrolysis of Gaba-LPETGG

The first test was to see if any of the new sortases showed activity when given the classic LPXTG motif of class A sortase. Hydrolysis reactions were run with 60 µM sortase and 600 µM Gaba-LPETGG **38** peptide (Figure 5.3), in 100 µL HEPES ligation buffer pH 7.5, incubated at 37°C. Samples were taken after 6 and 24 h. The sortase was precipitated out by addition of the same volume of ethanol and a 30 min incubation at -20°C, then

centrifuged at 10,000 *g* to remove the protein pellet. Supernatant was diluted 1/3 into LC-MS solvent and run on an Accucore C18 column.

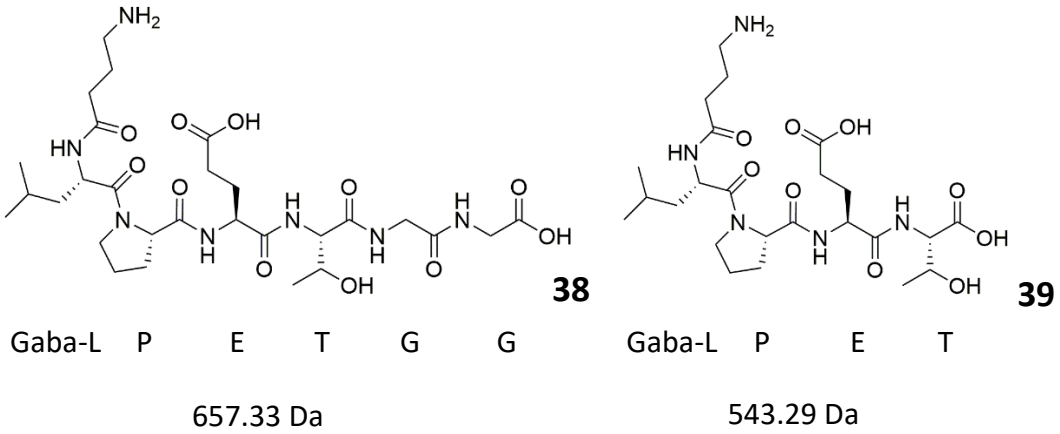


Figure 5.3: The Gaba-LPETGG peptide was tested for hydrolysis and ligation with GGG peptide in the presence of each new sortase. The expected hydrolysis product is Gaba-LPET.

No hydrolysis product **39** was detected for any of the 22 new sortases after 6 h. A positive control, SrtA 5M, showed hydrolysis product after 1 h. After 24 h, the negative control with only peptide showed the peptide $[M+H]^+$ peak at 658.31 Da with no hydrolysis product, while the positive control with SrtA 5M and peptide showed a large hydrolysis product $[M+H]^+$ peak at 544.27 Da (Figure 5.4). The enzymes Srt019, Srt021, Srt024, Srt025 and Srt027 showed a small peak of a possible hydrolysis product at 543.28 Da (Figure 5.5).

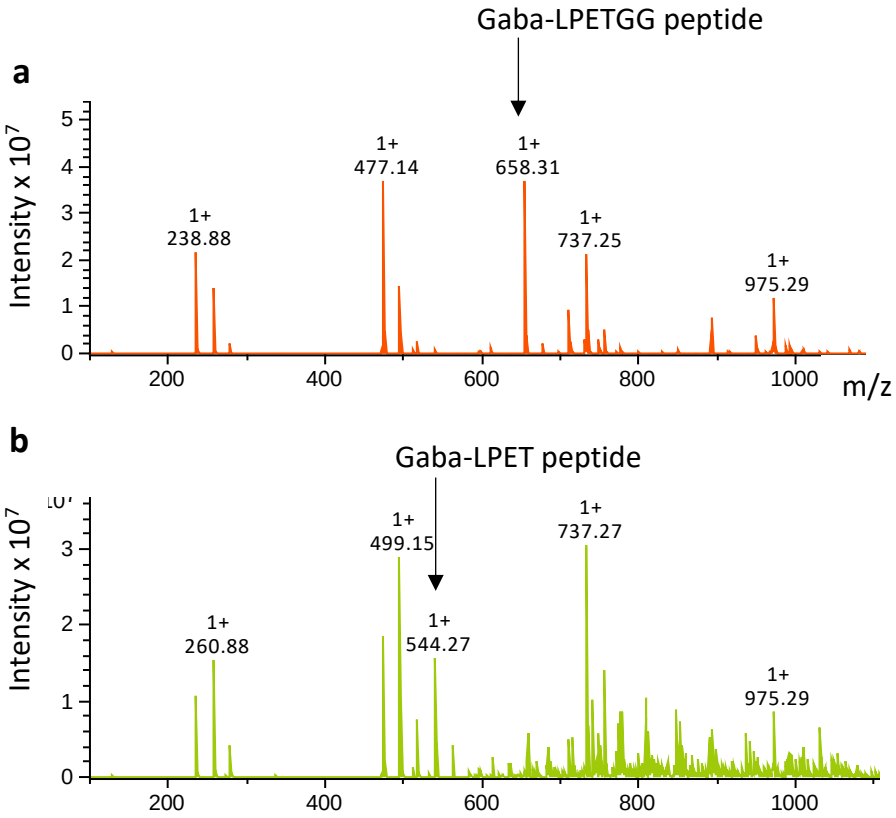
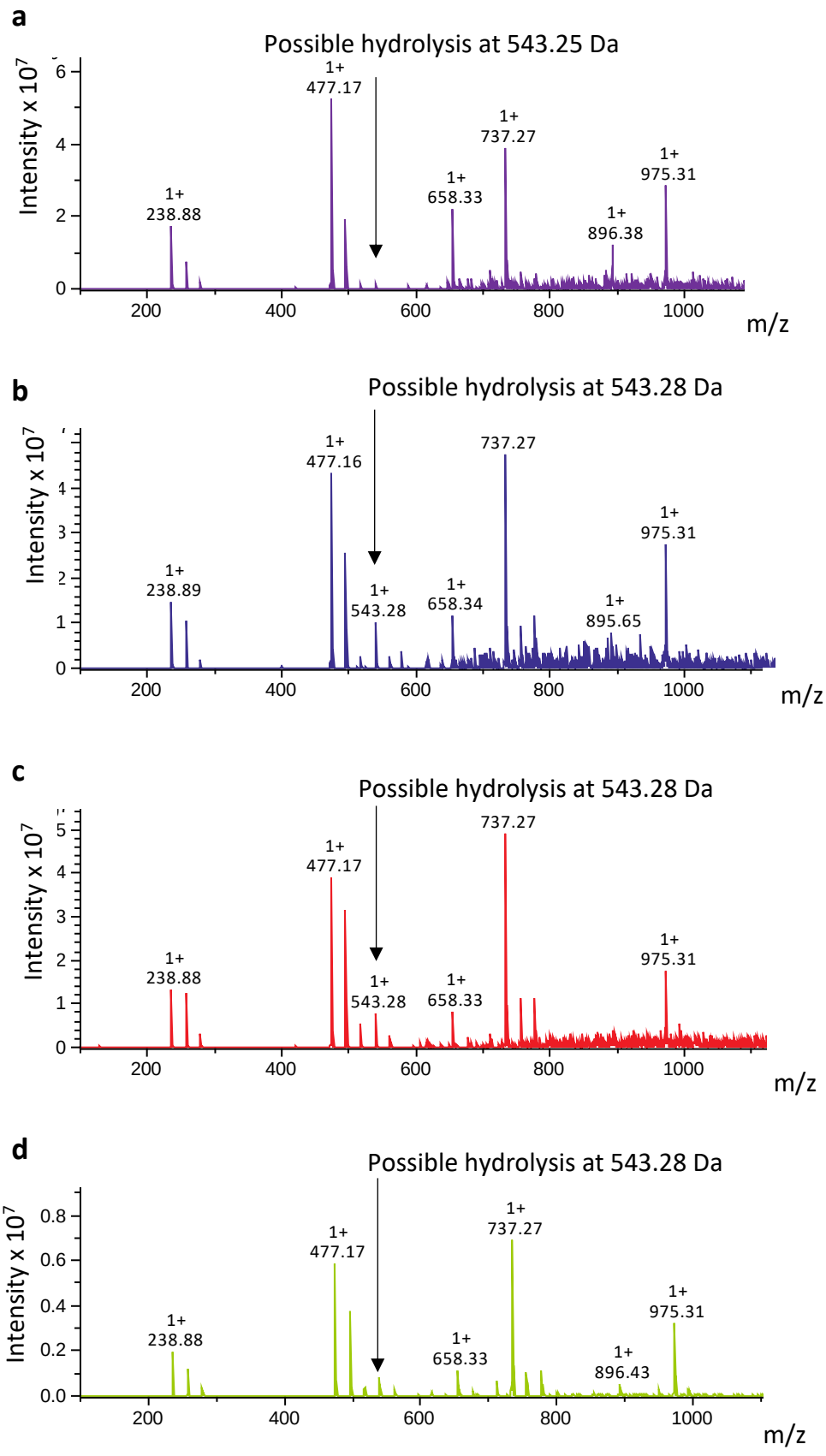


Figure 5.4: Negative control (a) and positive control (b) reactions with Gaba-LPETGG peptide and SrtA 5M. The negative control shows the peptide peak at 658.31 Da, in the SrtA 5M reaction this has been hydrolysed to 544.27 Da.


a
Possible hydrolysis at 543.25 Da
Intensity x 10^7
1+ 477.17
1+ 737.27
1+ 975.31
1+ 658.33
1+ 238.88
1+ 896.38
m/z
b
Possible hydrolysis at 543.28 Da
Intensity x 10^7
1+ 477.16
737.27
1+ 975.31
1+ 238.89
1+ 543.28
1+ 658.34
1+ 895.65
m/z
c
Possible hydrolysis at 543.28 Da
Intensity x 10^7
737.27
1+ 477.17
1+ 975.31
1+ 238.88
1+ 543.28
1+ 658.33
m/z
d
Possible hydrolysis at 543.28 Da
Intensity x 10^7
1+ 737.27
1+ 477.17
1+ 975.31
1+ 238.88
1+ 658.33
1+ 896.43
m/z

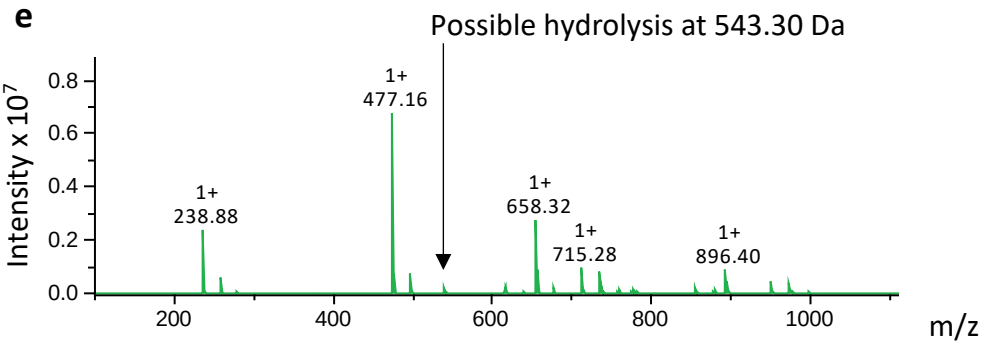


Figure 5.5: Sortases showing possible hydrolysis of Gaba-LPETGG peptide. Srt019 (a), Srt021 (b), Srt024 (c), Srt025 (d) and Srt027 (e) showed a possible hydrolysis product at 543.28 Da after 24 h.

The largest hydrolysis peaks were Srt021 and Srt024. The abundance of each peak was calculated and the ratio of 658.32 Da to 543.28 Da shows the percentage of peptide that has reacted (Table 4). Srt021, Srt024 and Srt025 appeared to be the fastest enzymes of the 22 novel sortases.

Sortase	% of Gaba-LPETGG peptide hydrolysed
Srt019	6.27
Srt021	46.8
Srt024	44.2
Srt025	42.1
Srt027	8.11

Table 4: Hydrolysis product Gaba-LPET yields after 24 h for Srt019, Srt021, Srt024, Srt025 and Srt027.

A second set of reactions was set up with 600 µM sortase and 600 µM peptide **38** – a 10-fold increase in sortase concentration, in order to see if the hydrolysis reaction could be pushed to completion. After 6 h Srt021, Srt025 and Srt027 showed a hydrolysis product at 543.28 Da (Figure 5.6).

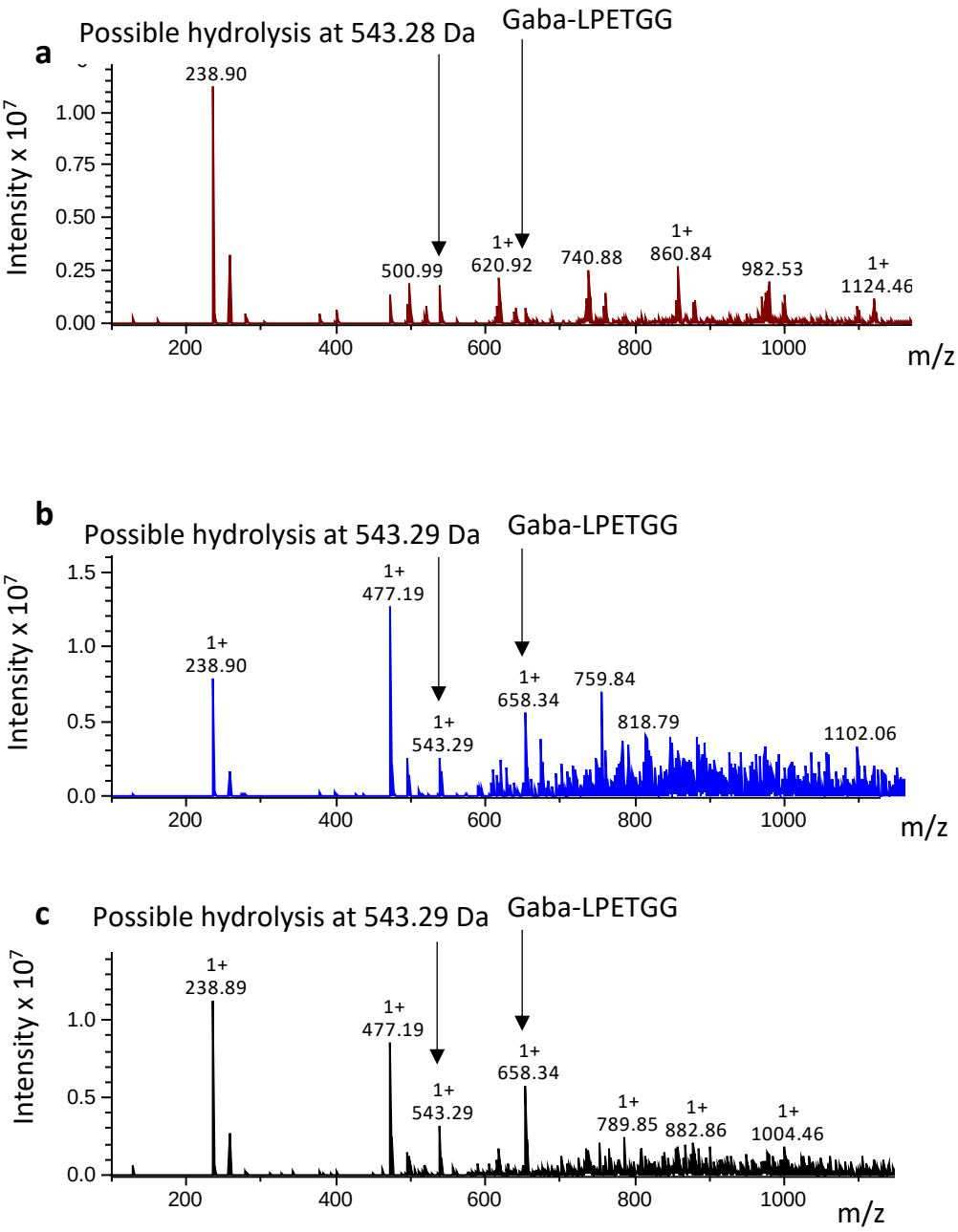


Figure 5.6: Sortases at increased concentration showing possible hydrolysis of Gaba-LPETGG peptide. Srt021 (a), Srt025 (b) and Srt027 (c) showed a possible hydrolysis product at 543.29 Da after 6 h when incubated with 600 µM sortase and 600 µM Gaba-LPETGG peptide.

After 24 h, most new sortases did not show any hydrolysis product **39**, although the SrtA 5M reaction again showed a hydrolysis product at the expected peak of 544.27 Da. The hydrolysis peak was not present in Srt001, Srt003, Srt004, Srt005, Srt006, Srt007, Srt008,

Srt009, Srt010, Srt011, Srt012, Srt015, Srt016, Srt017, Srt019, Srt022, Srt023, or Srt026. Srt008 is shown as an example (Figure 5.7).

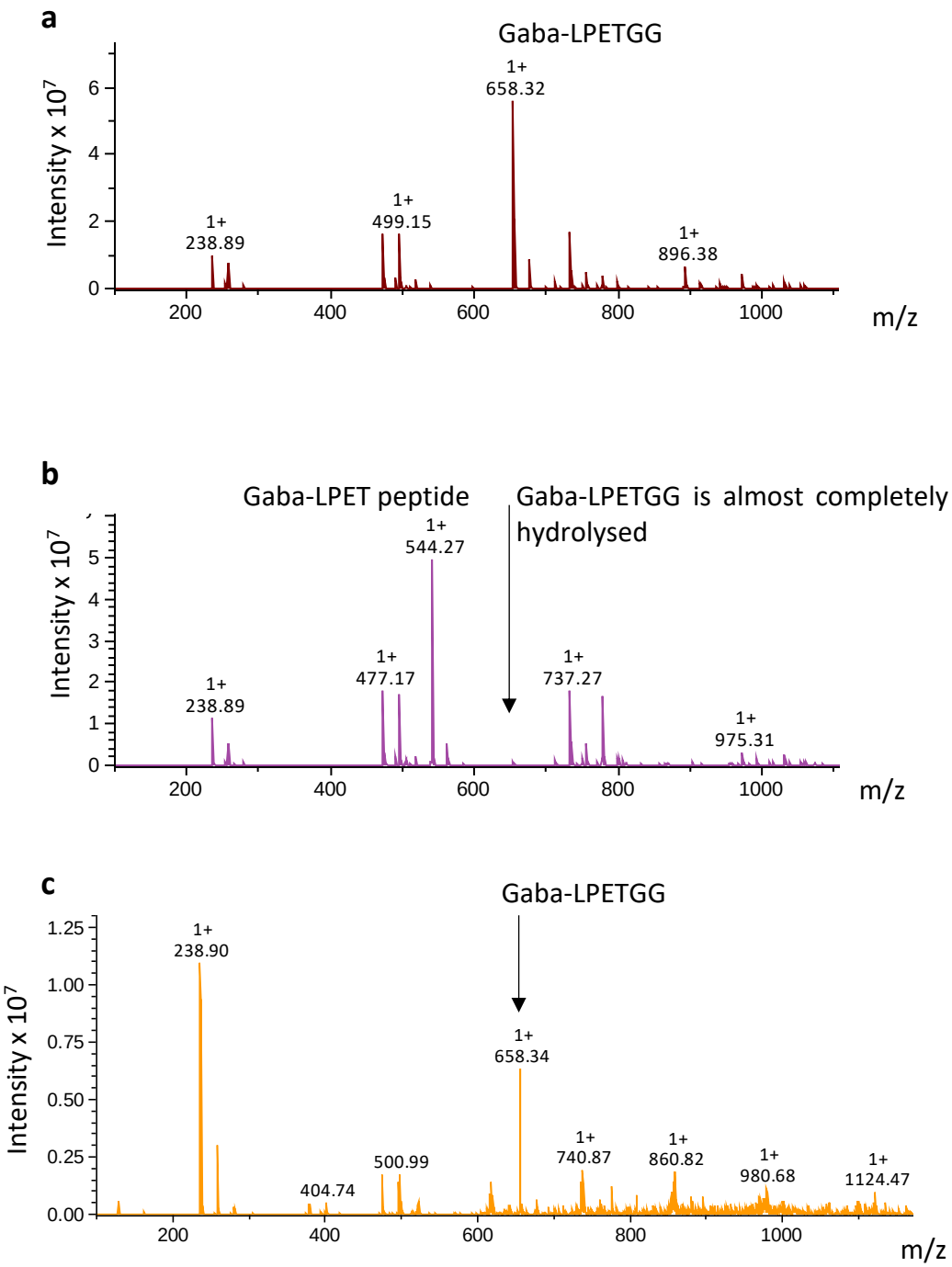


Figure 5.7: Negative control (a) and positive control (b) reactions with Gaba-LPETGG peptide and SrtA 5M. The negative control shows the peptide peak at 658.32 Da, in the SrtA 5M reaction this has been hydrolysed to 544.27 Da. (c) Srt008 has no hydrolysis product.

After 24 h more apparent hydrolysis product at 543.29 Da was present in Srt021, which had hydrolysed all Gaba-LPETGG **38** peptide, Srt025, which had partly hydrolysed the peptide, and Srt027, which had hydrolysed over half of the peptide. In addition, Srt024 also showed hydrolysis product (Figure 5.8).

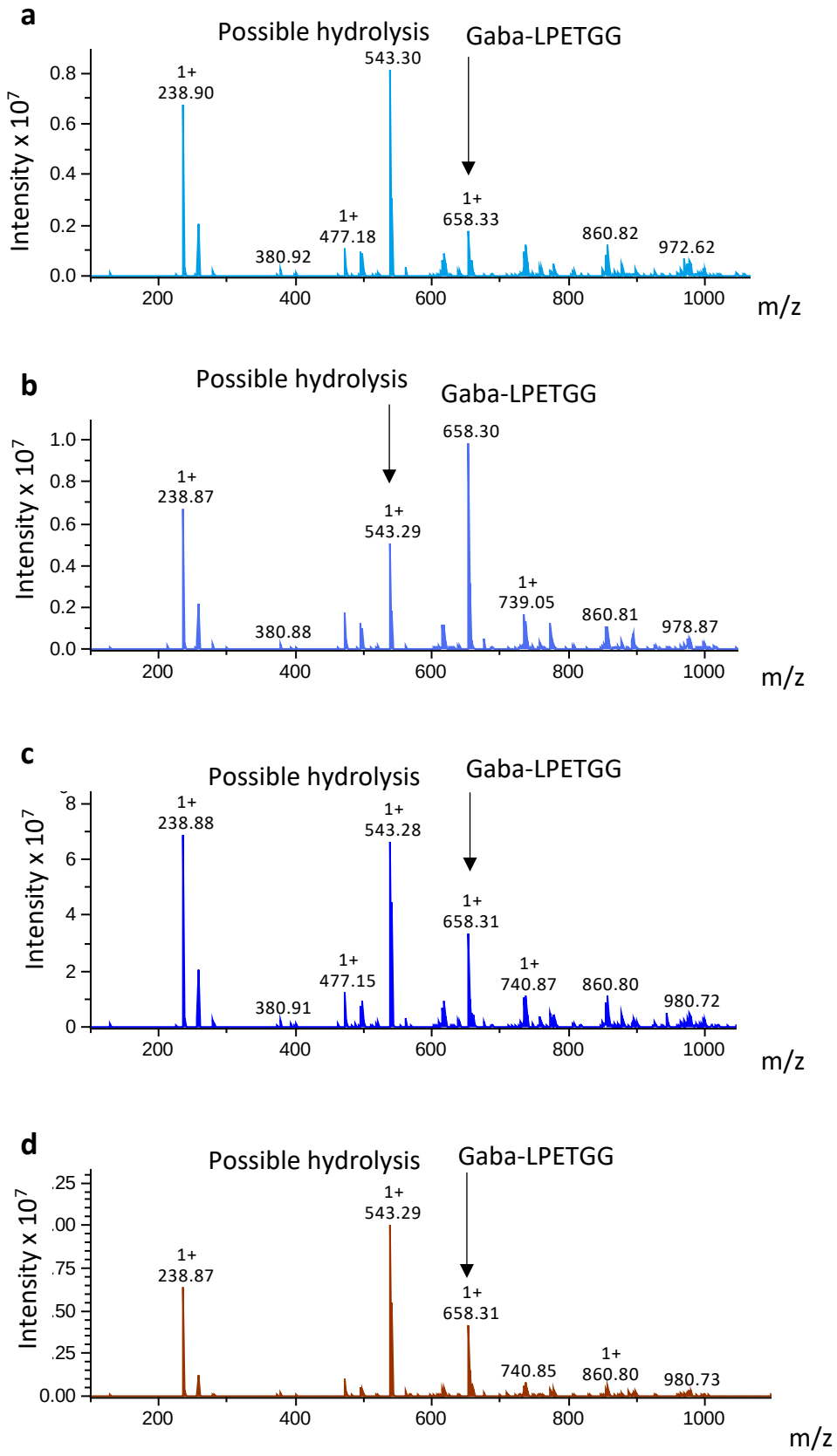


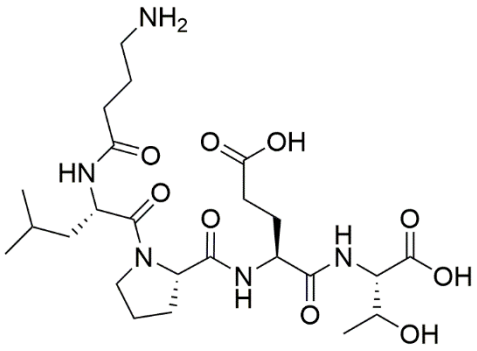
Figure 5.8: Sortase reactions showed increased hydrolysis product with a higher concentration of enzyme. Srt021 (a), Srt024 (b), Srt025 (c) and Srt027 (d).

The percentage of hydrolysis product has increased in all reactions at 24 h compared with 6 h (Table 5), which suggests it is due to enzyme activity.

Sortase	% hydrolysis after 6 h	% hydrolysis after 24 h
Srt021	69.8	91.6
Srt024	Negligible	39.4
Srt025	50.0	78.9
Srt027	35.8	74.3

Table 5: Hydrolysis product yields after 6 h and 24 h for Srt021, Srt024, Srt025 and Srt027.

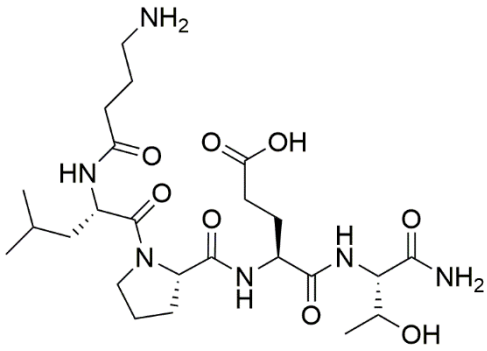
However, this hydrolysis product did not have the expected mass at 544.27 Da. If the product has a mass of 543.29 Da, it cannot be detected on the mass spectrometer at that mass without the addition of a charge – an H^+ for example, which then increases the mass by 1. It was hypothesised that this could be the result of ammonia acting as the nucleophile in the cleavage of intermediate from the thiol group of the enzyme. This would decrease the expected mass by 1 Da (Figure 5.9).



39

Gaba-L P E T

Hydrolysis product with H_2O 543.29 Da



40

Gaba-L P E T

Product with NH_3 542.31 Da

Figure 5.9: The expected hydrolysis product and the product that results from ammonia cleavage.

The enzymes were stored in 50% w/v ammonium sulphate solution after being made at Prozomix. They were reconstituted by centrifuging at 17,000 *g* for 5 min, removal of supernatant and resuspension in HEPES ligation buffer to the same volume. Therefore, there will still be traces of ammonium sulphate. This hypothesis was tested by adding

60 µM Srt017 (which is so far non-functional) to 60 µM SrtA 5M and 600 µM Gaba-LPETGG **38** peptide. This caused SrtA 5M to cleave with ammonia and yield the shifted hydrolysis peak **40** at 543.27 Da (Figure 5.10). Therefore, all new sortases were dialysed into three changes of HEPES buffer to remove the ammonium sulphate salt.

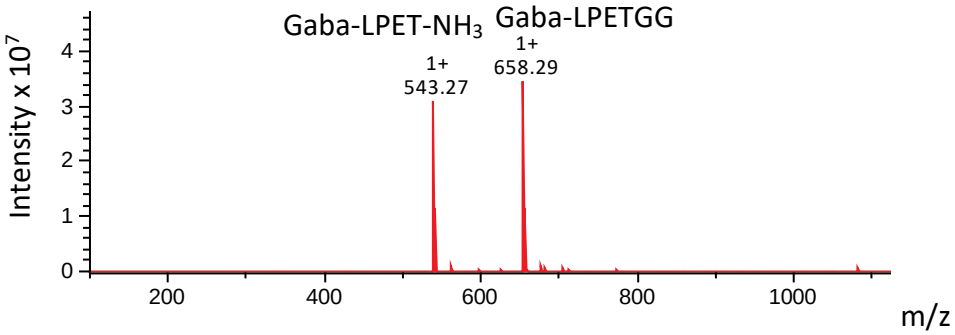


Figure 5.10: SrtA 5M cleaves Gaba-LPETGG peptide with NH_3 to yield the hydrolysis product at 543.27 Da if mixed with Srt017.

Ligation of Gaba-LPETGG with GGG

The next test was to use the now-cleaned sortases to ligate Gaba-LPETGG **38** with GGG **41** tripeptide. 60 µM sortase, 600 µM Gaba-LPETGG **38** peptide, and 6 mM tripeptide **41** were reacted in 100 µl HEPES ligation buffer pH 7.5, incubated at 37°C for 24 h. The tripeptide nucleophile was in excess to try to shift the reaction equilibrium to favour ligated product.

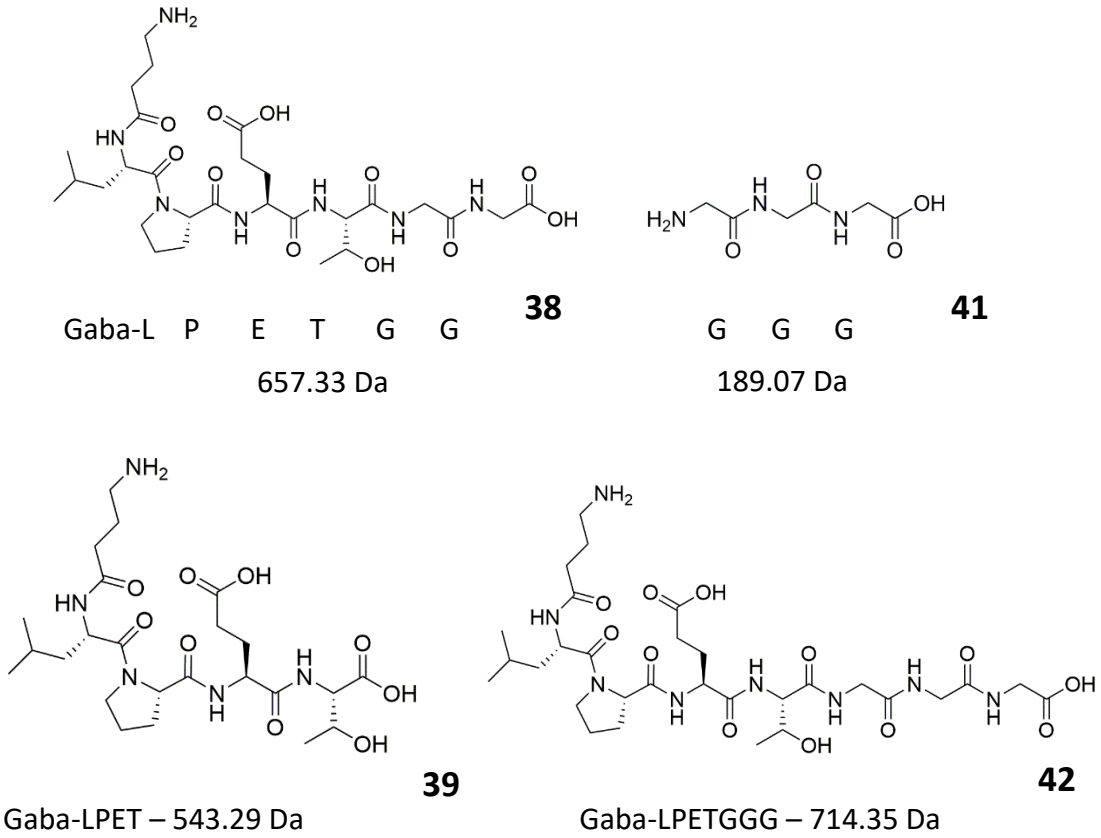
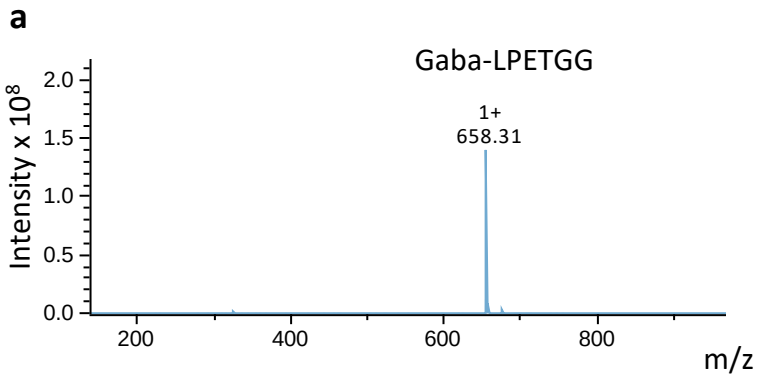


Figure 5.11: The initial Gaba-LPETGG and GGG peptides, with Gaba-LPET hydrolysis product and Gaba-LPETGGG ligation product.

The only new sortases to show ligation activity were Srt021 and Srt025, which both had a small peak **42** at 715.34 Da (Figure 5.12). Srt012 is shown as an example of a non-functional sortase.



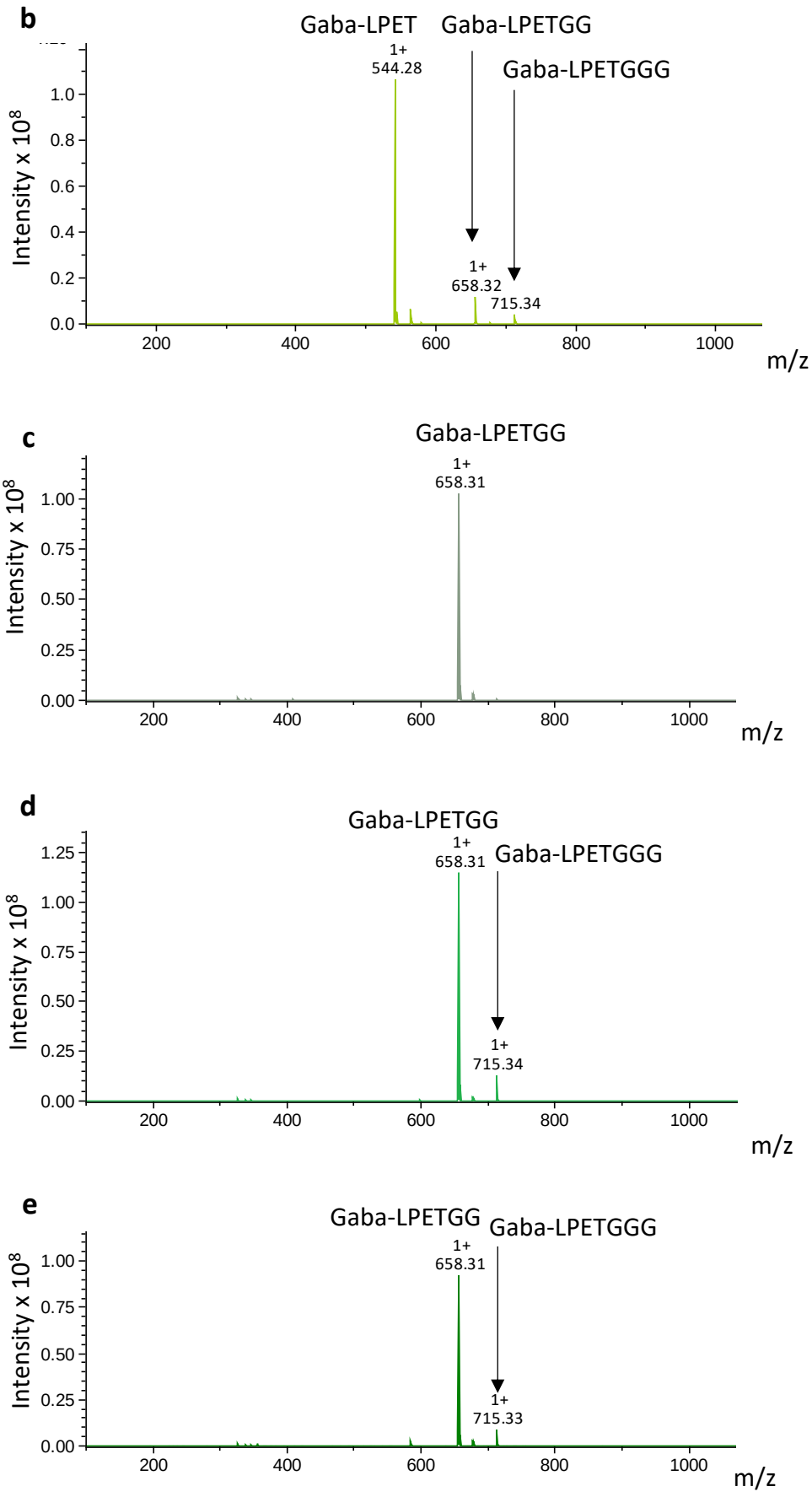


Figure 5.12: Sortase ligation of Gaba-LPETGG and GGG. Gaba-LPETGG and GGG negative control after 24 h (a), SrtA 5M showed hydrolysis product and ligated product

(b), Srt012 only has Gaba-LPETGG (c), Srt021 (d) and Srt025 (e) showed ligated product.

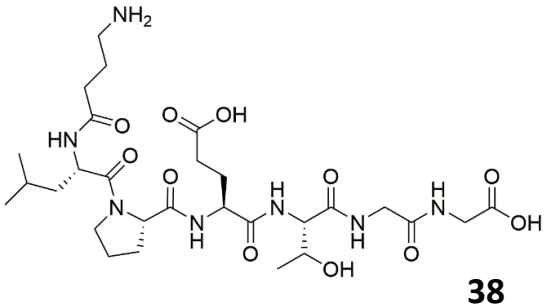
The lack of hydrolysis product (Table 6) is surprising as without GGG peptide Srt021 and Srt025 both cleaved with NH_3 and would be expected to cleave with H_2O in the absence of other nucleophiles.

Sortase	% peptide	% hydrolysis	% ligation
SrtA 5M	5.00	93.4	1.60
Srt021	89.7	0	10.3
Srt025	90.0	0	9.09

Table 6: Ligation product Gaba-LPETGGG yields after 24 h for Srt021 and Srt025.

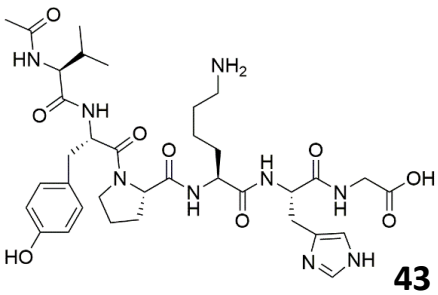
Ligation of Gaba-LPETGG with AcVYPKHG

The next test was to ligate Gaba-LPETGG **38** with the AcVYPKHG **43** sequence adapted from the SpaA pilin protein. As with the previous ligation, 60 µM sortase was combined with 600 µM Gaba-LPETGG **38** and 6 mM AcVYPKHG **43**, in HEPES ligation buffer incubated at 37°C for 24 h. Again, the nucleophile is in excess. The expected ligation product **44** was 1266.66 Da (Figure 5.13).



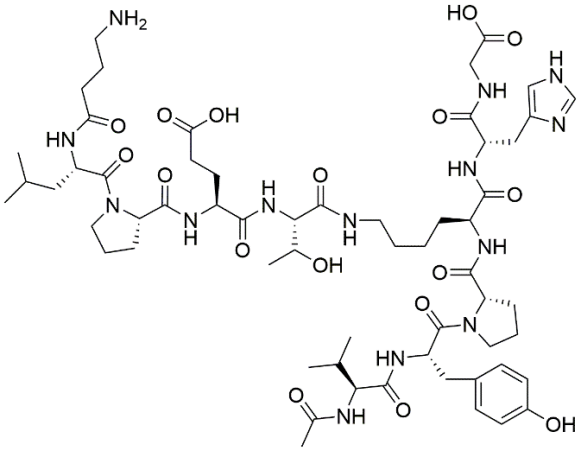
38

Gaba-LPETGG 657.33 Da



43

AcVYPKHG 741.38 Da



44

AcVYP(Gaba-LPETK)HG 1266.66 Da

Figure 5.13: The initial Gaba-LPETGG and AcVYPKHG peptides, with AcVYP(Gaba-LPETK)HG product.

Again, most sortases did not show any activity (Figure 5.14).

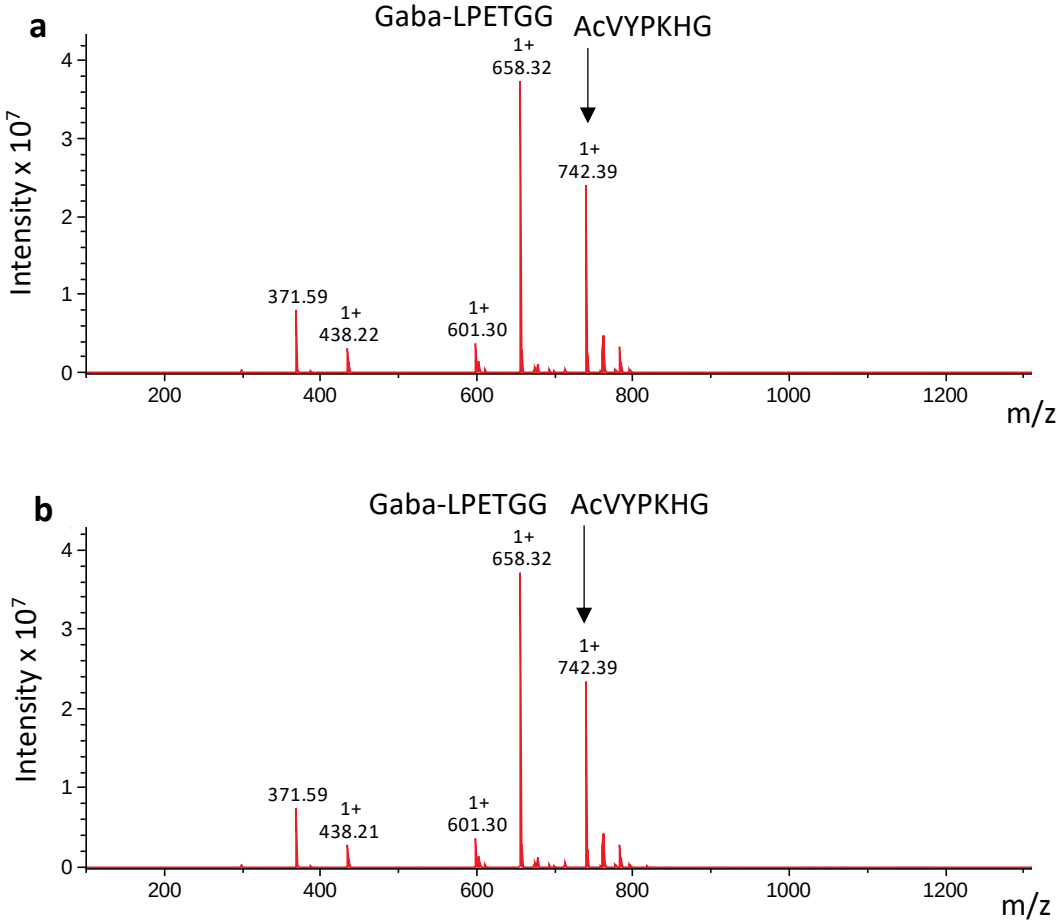
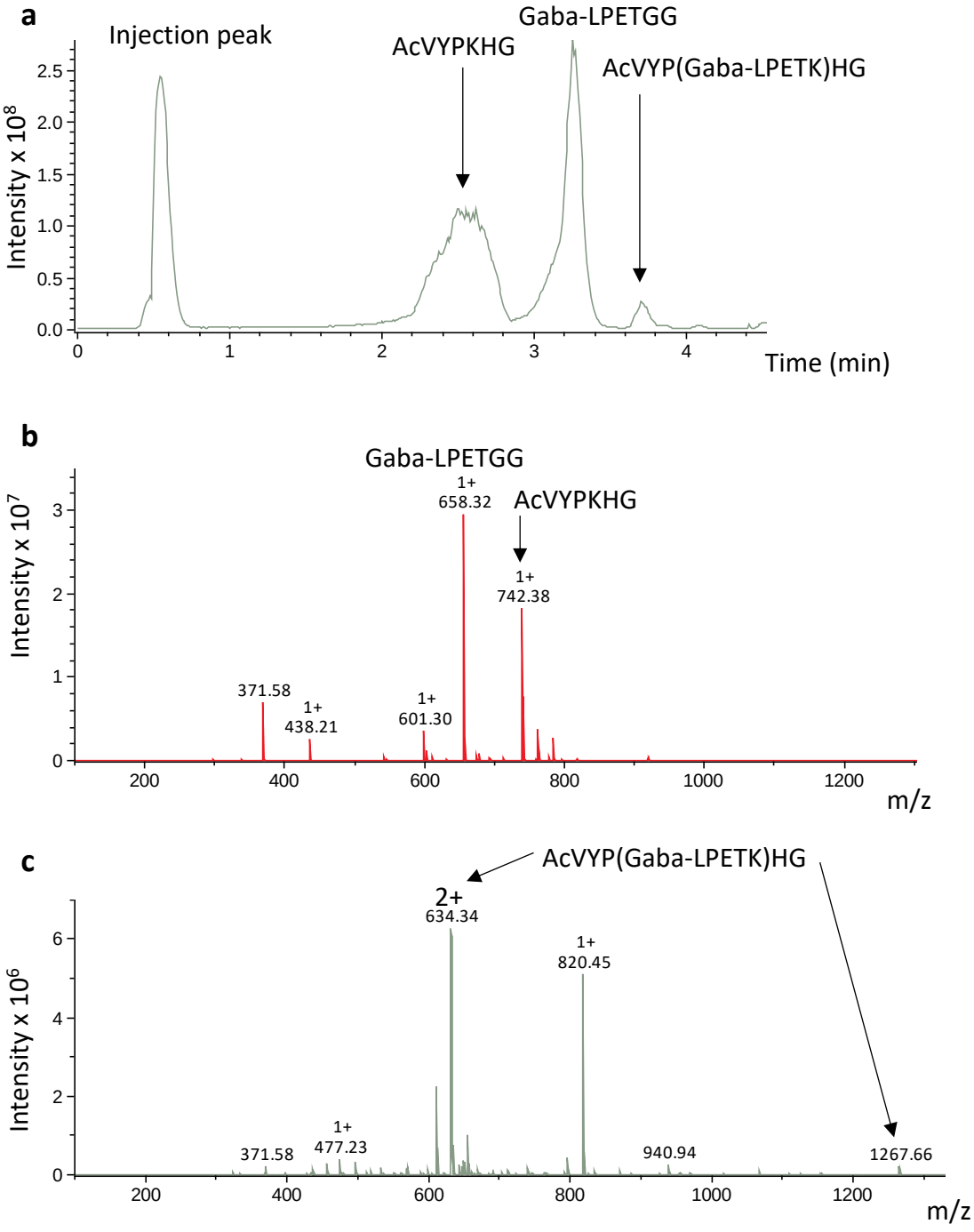


Figure 5.14: Tested ligation of Gaba-LPETGG and AcVYPKHG. Gaba-LPETGG and AcVYPKHG negative control after 24 h (a), Srt003 with no ligated product (b).

Srt021 and 025 showed ligation activity. The large majority of peptide has not ligated, and the ligation product cannot be seen on the overall peaks. However, the Accucore column separates out the compounds according to hydrophobicity and this allows compounds present in small quantities to be located. The MS chromatogram shows a large peak at the injection, then three further peaks as the peptides are eluted from the column (Figure 5.15a). The broad peak at 2.2 – 2.8 min is mainly composed of the AcVYPKHG **43** peptide. The taller peak at 3.0 – 3.4 min is mainly the Gaba-LPETGG **38** peptide. A small peak has appeared next to it at 3.6 – 3.8 min. This peak is shown to be

ligation product **44** by its compound spectra. Focusing solely on this chromatogram peak, there is a large peak on the compound spectrum at 634.50 Da, the +2 ligation peak, and a small +1 peak at 1267.66 (Figure 5.15c).



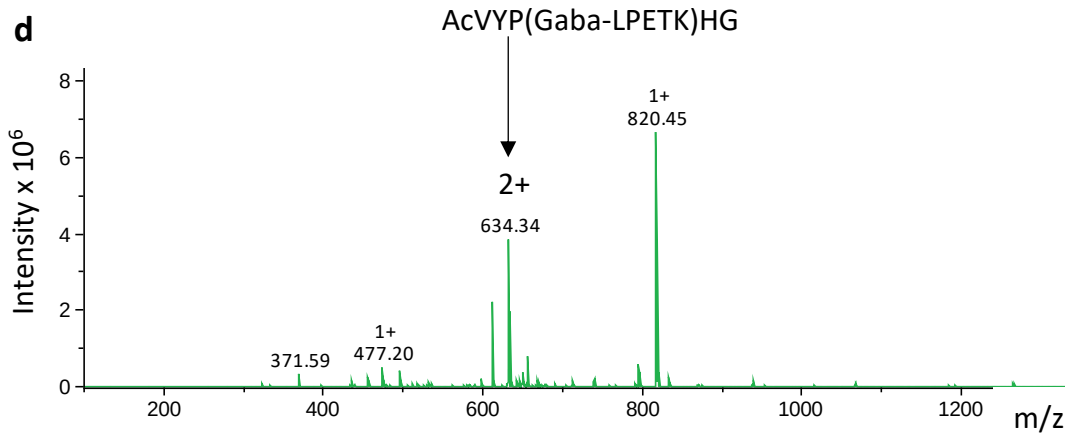


Figure 5.15: Successful ligation between Gaba-LPETGG and AcVYPKHG. Srt021 chromatogram trace (a) showing the presence of a new, ligation product peak at 3.6 – 3.8 min. Srt021 overall peaks (b) and ligation peak (c), showing a small amount of ligation product. Srt025 ligation peak (d) also shows a small amount of ligation.

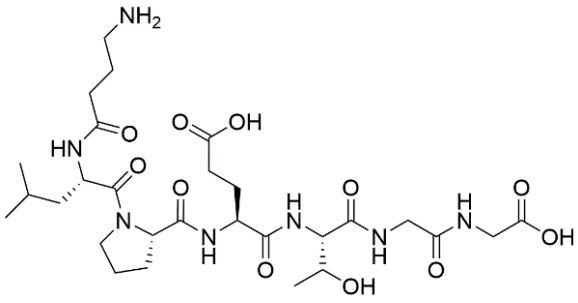
Srt021 and Srt025 therefore show *in vitro* ligation with the AcVYPKHG **43** peptide, although analysis of the peaks by percentage shows less than 1% ligation of Gaba-LPETGG **38**, which is the limiting peptide (Table 7).

Sortase	% peptide	% ligation
Srt021	98.7	0.45
Srt025	99.4	0.2

Table 7: Ligation product AcVYP(Gaba-LPETK)HG yields after 24 h for Srt021 and Srt025.

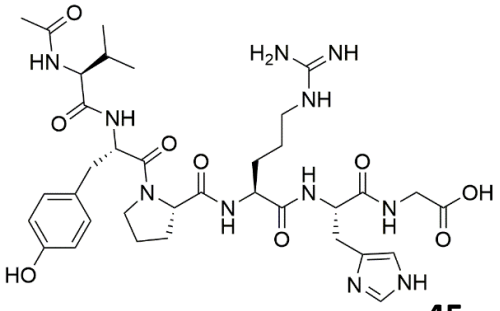
Ligation of Gaba-LPETGG with AcVYPRHG

A negative control ligation test was performed to check if the peptide was ligating as expected at the Lys residue. 60 µM sortase was incubated with 600 µM GaLPETGG **38** and 6 mM AcVYPRHG **45** peptide, where the Lys is replaced by Arg that should not undergo the reaction (Figure 5.16).



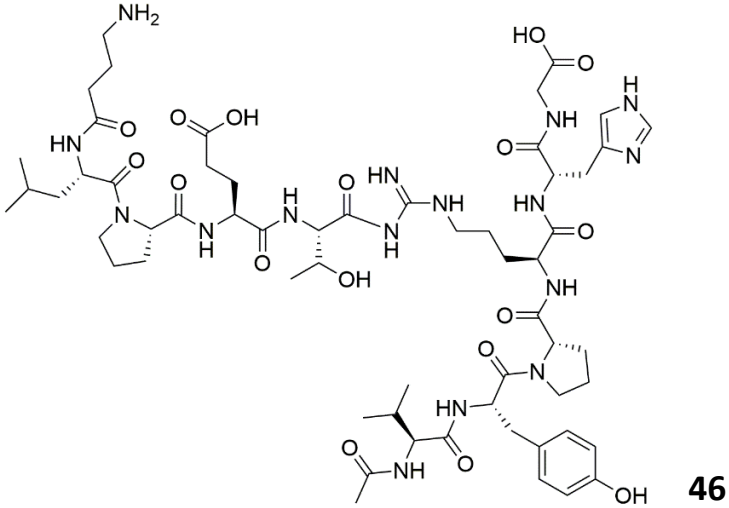
38

Gaba-LPETGG 657.33 Da.



45

AcVYPRHG 770.39 Da.



46

AcVYP(Gaba-LPETR)HG peak would appear at 1295.67 Da.

Figure 5.16: The initial Gaba-LPETGG and AcVYPRHG peptides, with the hypothetical AcVYP(Gaba-LPETR)HG product.

No sortases formed a ligation product **46** when the Lys residue was missing, as no ligation peaks could be seen at 1295 Da or 648 Da.

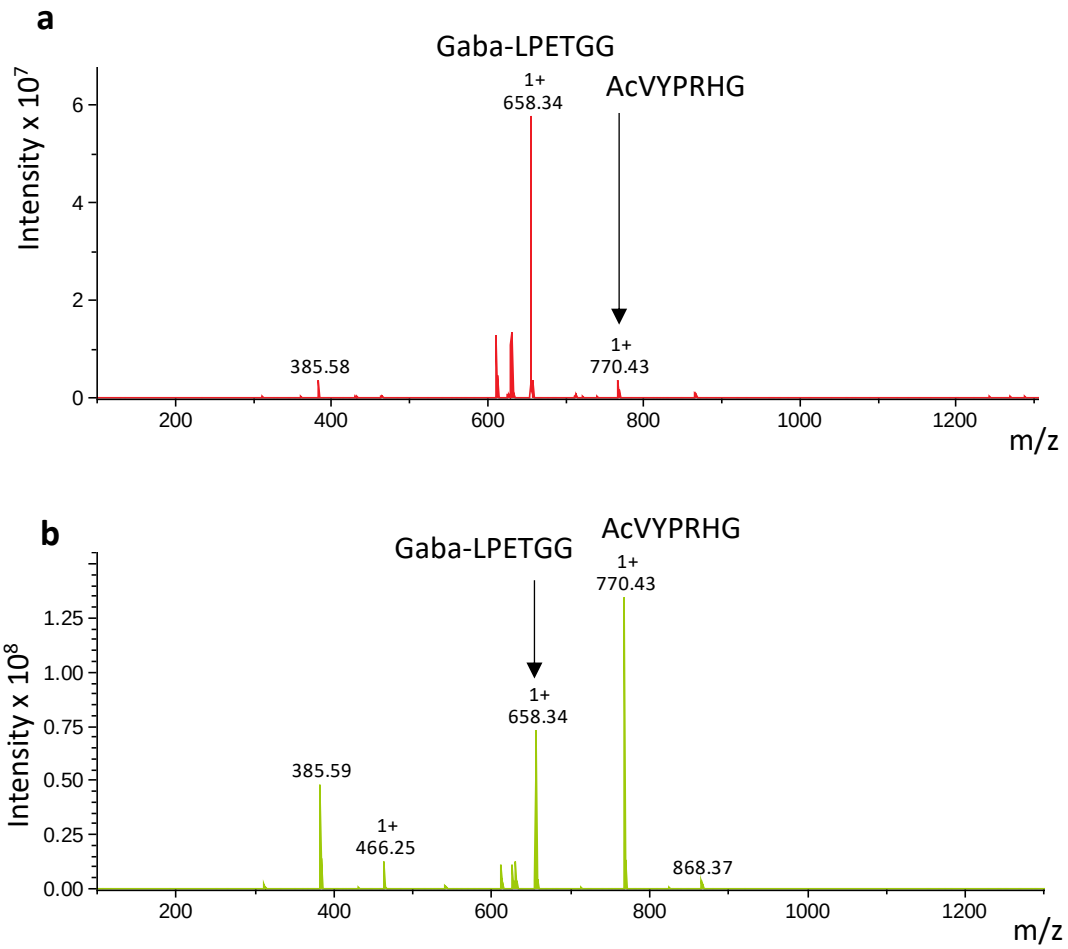
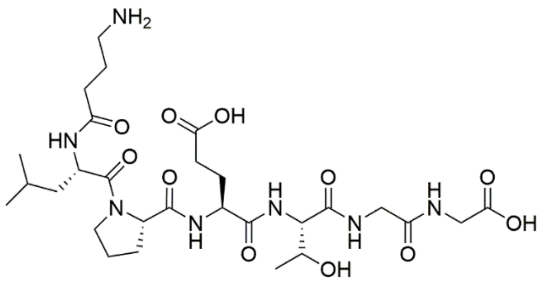


Figure 5.17: Ligation between Gaba-LPETGG and AcVYPRHG cannot occur. Gaba-LPETGG and AcVYPRHG negative control after 24 h (a), Srt021 with no ligated product (b).

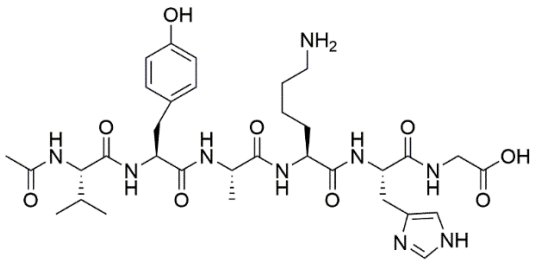
Ligation of Gaba-LPETGG with AcVYAKHG

A further ligation test investigated whether the sortases could ligate AcVYAKHG **47** peptide, where the replacement of Pro with Ala will give a differently shaped peptide backbone. As previously, 60 µM sortase was combined with 600 µM Gaba-LPETGG **38** and 6 mM AcVYAKHG **47**, in HEPES ligation buffer incubated at 37°C for 24 h. The expected ligation product **48** was 1240.65 Da.



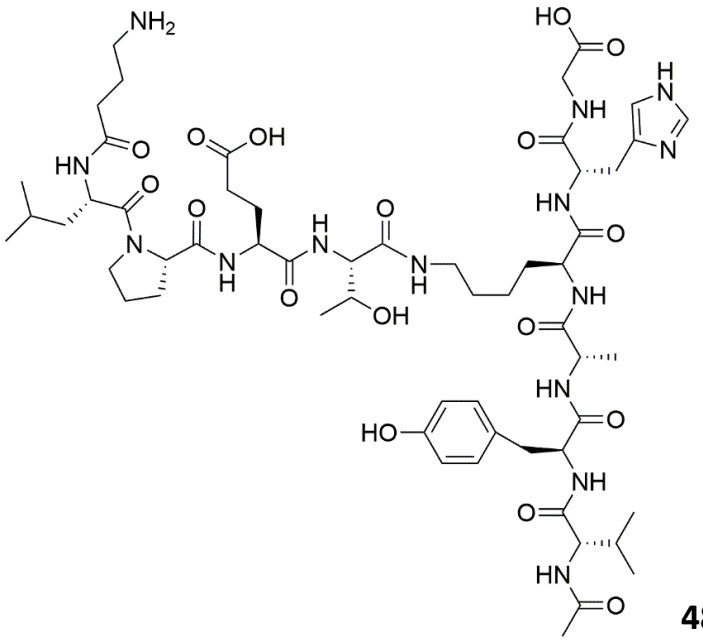
38

GaLPETGG – 657.33 Da



47

AcVYAKHG – 715.37 Da



48

AcVYAK(LPET)HG – 1240.65 Da

Figure 5.18: The initial Gaba-LPETGG and AcVYAKHG peptides, with the AcVYA(LPETK)HG ligated product.

Again, Srt021 and Srt025 show ligation activity. There is a +2 peak at 621.32 Da and a +1 peak at 1241.66 (Figure 5.19). As with the AcVYPKHG **43** peptide, the compound spectra shows only the small chromatogram peak and the overall ligation percentage is low, but it confirms sortase ligation activity.

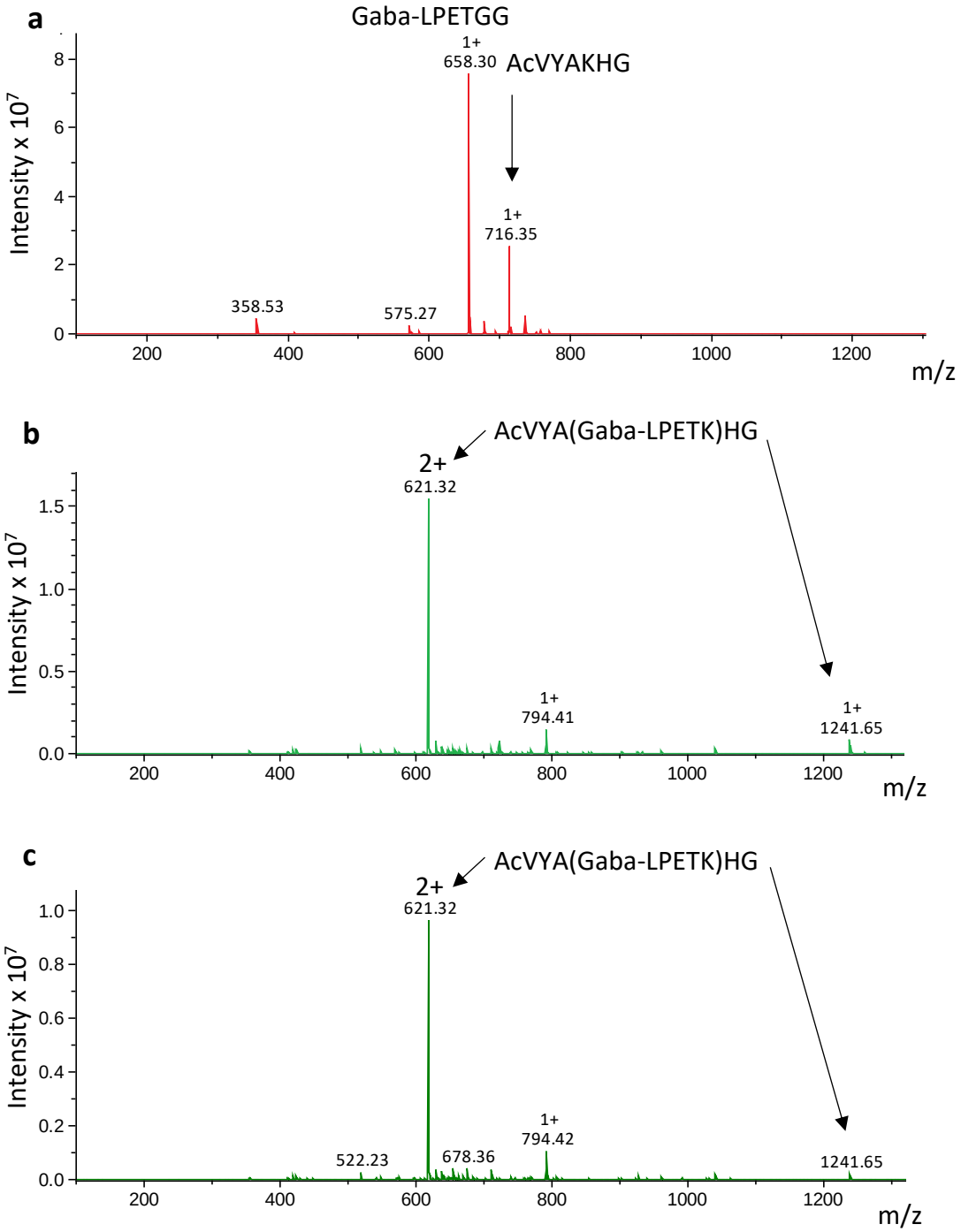


Figure 5.19: Successful ligation between Gaba-LPETGG and AcVYAKHG. Gaba-LPETGG and AcVYAKHG negative control after 24 h (a), Srt021 (b) and Srt025 (c) with AcVYA(Gaba-LPETK)HG ligated product.

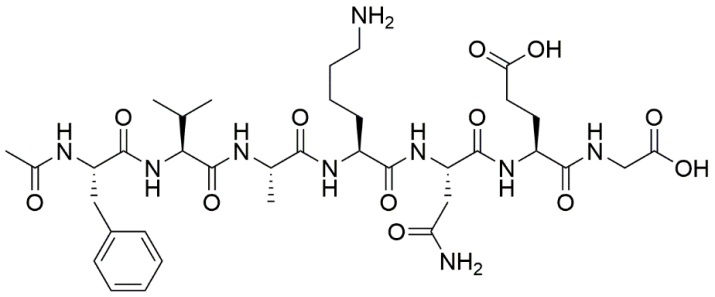
The overall ligation with AcVYAKHG **47** is still less than 1% for both Srt021 and Srt025 (Table 8).

Sortase	% peptide	% ligation
Srt021	99.5	0.47
Srt025	99.5	0.17

Table 8: Ligation product AcVYA(Gaba-LPETK)HG yields after 24 h for Srt021 and Srt025.

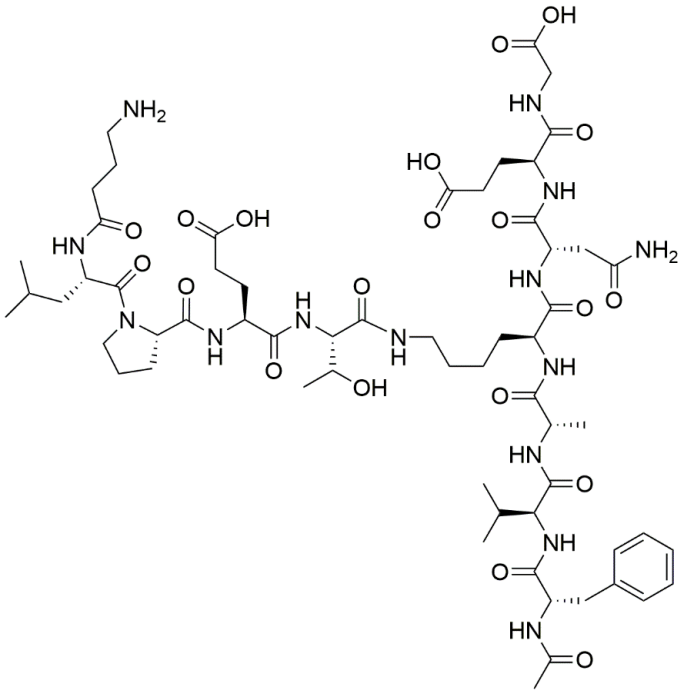
Ligation of Gaba-LPETGG with FVAKNEG

Another ligation test used 60 µM sortase + 600 µM GaLPETGG **38** + 6 mM FVAKNEG **49** peptide, in HEPES ligation buffer for 24 h at 37°C (Figure 5.20). No sortases showed any activity (Figure 5.21).



49

AcFVAKNEG – 805.40 Da



50

AcFVA(LPETK)NEG – 1330.68 Da

Figure 5.20: AcFVAKNEG peptide and the expected ligation product AcFVA(LPETK)NEG.

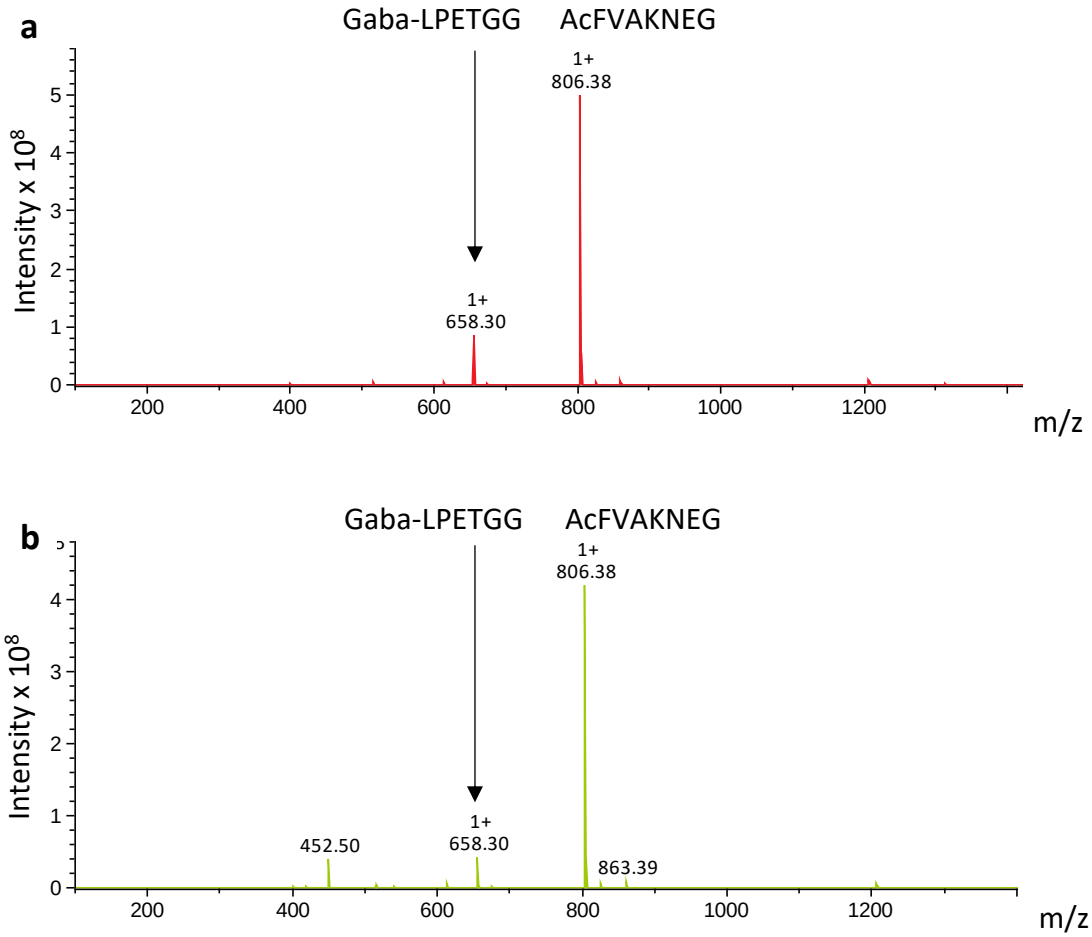


Figure 5.21: Ligation does not occur between Gaba-LPETGG and AcFVAKNEG. Gaba-LPETGG and AcFVAKNEG negative control after 24 h (a), Srt021 with no ligated product (b).

Hydrolysis of Gaba-LPETAA

As SpySrtA is a sortase that accepts an LPETA motif, the peptide Gaba-LPETAA **51** was tested for hydrolysis with the new sortases. 60 µM sortase with 600 µM Gaba-LPETAA **51** in HEPES ligation buffer was incubated at 37°C for 24 h. SpySrtA was included as a positive control and showed the same hydrolysis product, Gaba-LPET **39**, after 1 h (Figure 5.22). Most sortases (Srt001) for example did not show any hydrolysis product.

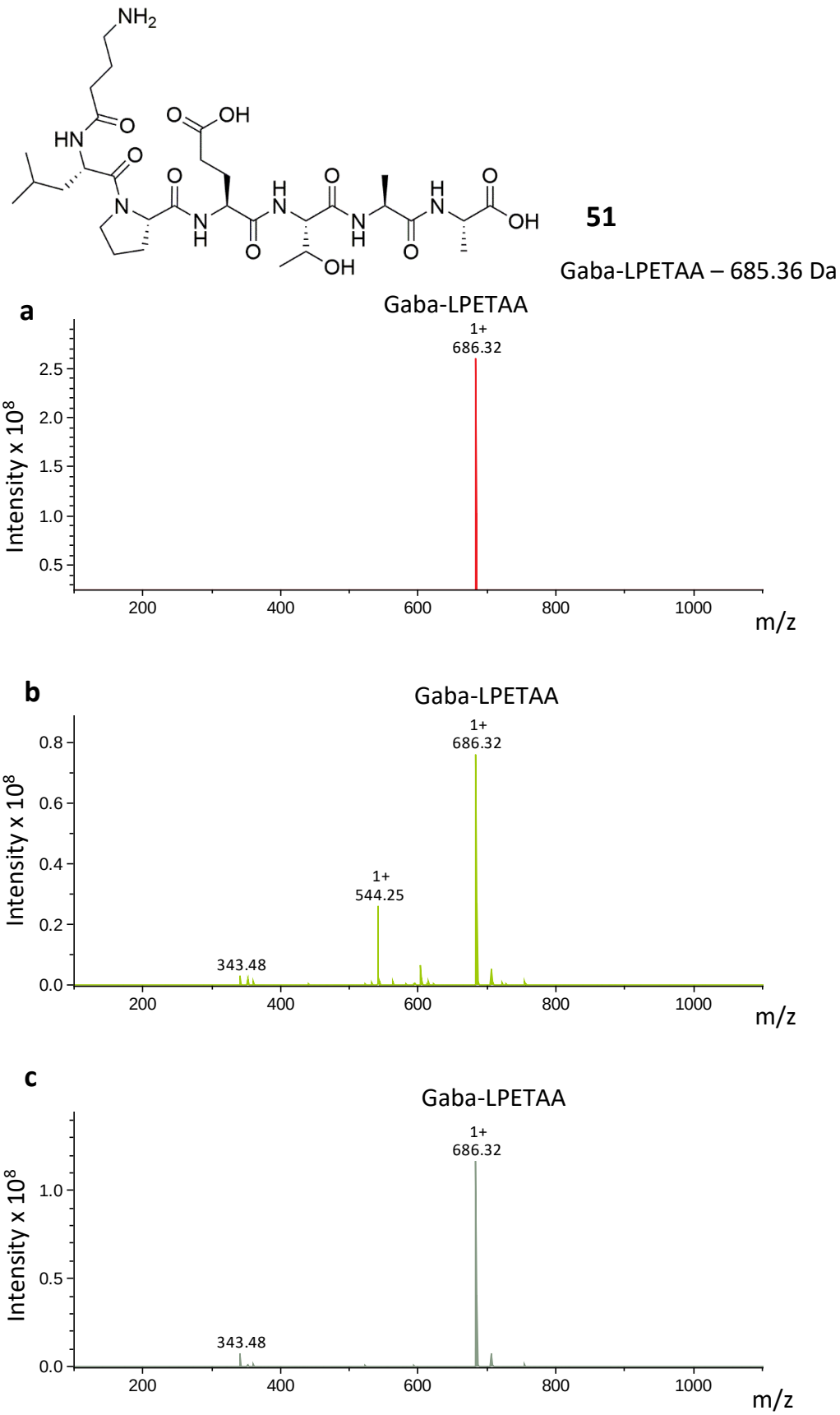


Figure 5.22: Testing hydrolysis of Gaba-LPETAA. Gaba-LPETAA negative control after 24 h (a), SpySrtA positive control with hydrolysis product at 544.25 Da (b), Srt001 showing no hydrolysis product (c).

Several enzymes show an unexpected pattern of peaks. There is a possible hydrolysis product at 548.27 Da, but this is 4 Da higher than the expected mass. This occurs for Srt004, Srt009, Srt010, Srt021, Srt022, Srt024. Notably, Srt025 does not show any reaction.

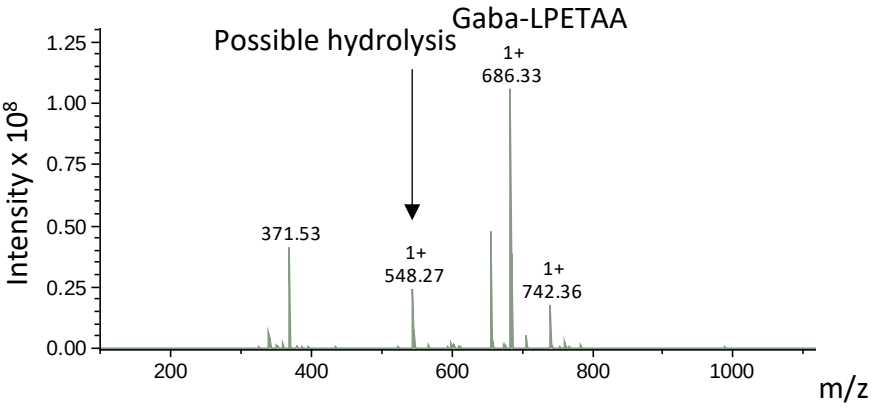
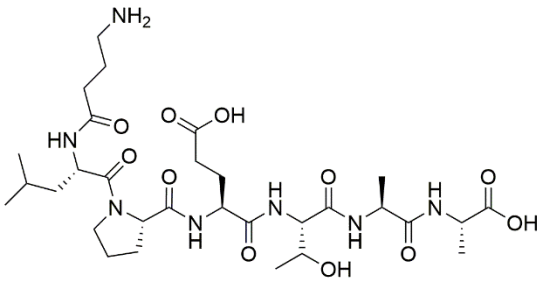


Figure 5.23: Srt021 with Gaba-LPETAA peptide showing a possible hydrolysis peak at 548.27 Da.

Ligation of Gaba-LPETAA with AAA

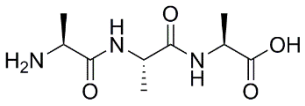
To further investigate the unexpected shift in the hydrolysis peak, ligation reactions were carried out with sortase at 60 µM, Gaba-LPETAA **51** 600 µM, AAA **52** 6 mM in HEPES ligation buffer, incubate at 37 °C for 24 h (Figure 5.24). Again, SpySrtA was included as a positive control (Figure 5.26).



Ga-L P E T A A

685.36 Da

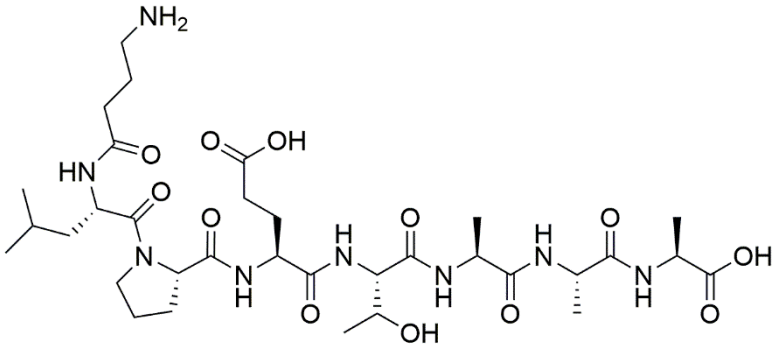
51



A A A

231.12 Da

52



53

Gaba-LPETAAA – 756.40 Da.

Figure 5.24: The initial Gaba-LPETAA and AAA peptides, with Gaba-LPETGGG ligation product.

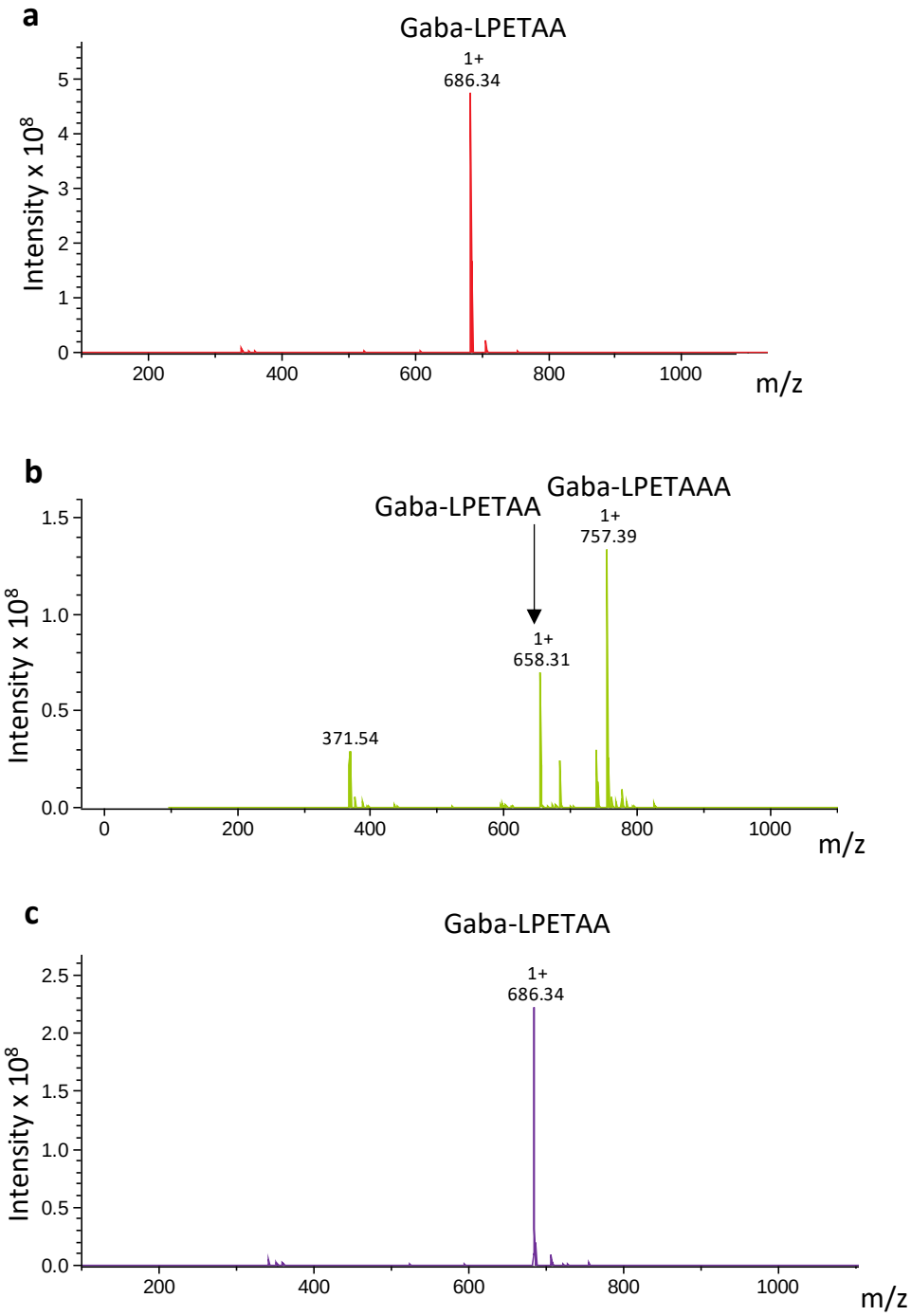


Figure 5.25: Testing ligation of Gaba-LPETAA and AAA. Gaba-LPETAA and AAA negative control after 24 h (a), SpySrtA positive control with ligation product at 757.39 Da (b), Srt006 showing no hydrolysis product (c).

Again, most sortases (Srt006 shown for example) only have unreacted Gaba-LPETAA **51** peptide. Several showed the same possible hydrolysis peak at 548.29 Da, although not all suggested activity in the previous hydrolysis test. These sortases were Srt001, Srt008, Srt009, Srt010 and Srt016 (Figure 5.26). Srt009 and Srt010 also showed this peak in the hydrolysis test.

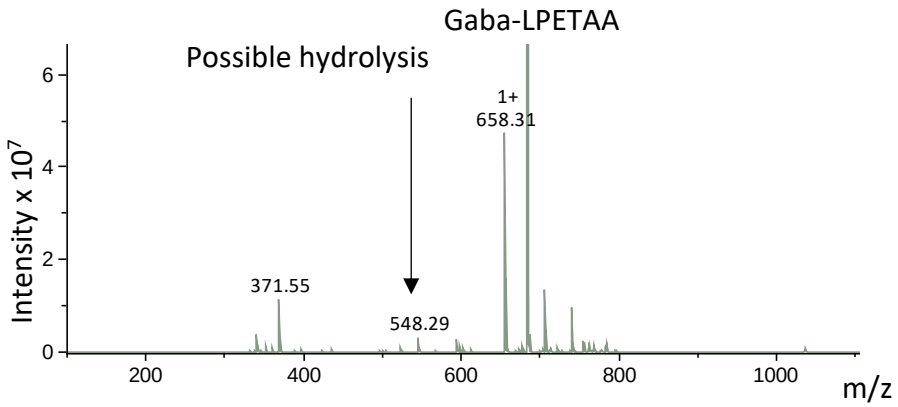


Figure 5.26: Srt001 with Gaba-LPETAA and AAA showing the same possible hydrolysis peak at 548.27 Da.

Ligation of Gaba-LPETAA with AcVYPKHG

As SpySrtA accepts the variant LPETA motif, the sortases were tested to see if the class C K peptide nucleophile could ligate to LPETAA. Ligation reactions were carried out with sortase at 60 µM, Gaba-LPETAA **51** 600 µM, AcVYPKHG **43** 6000 µM in HEPES ligation buffer, incubate at 37 °C for 24 h. The same AcVYP(Gaba-LPETK)HG **44** product mass as for ligation with Gaba-LPETGG **38** (1266.66 Da) would have been seen in the event of successful ligation, but no ligation peaks were seen in any of the reactions (Figure 5.27).

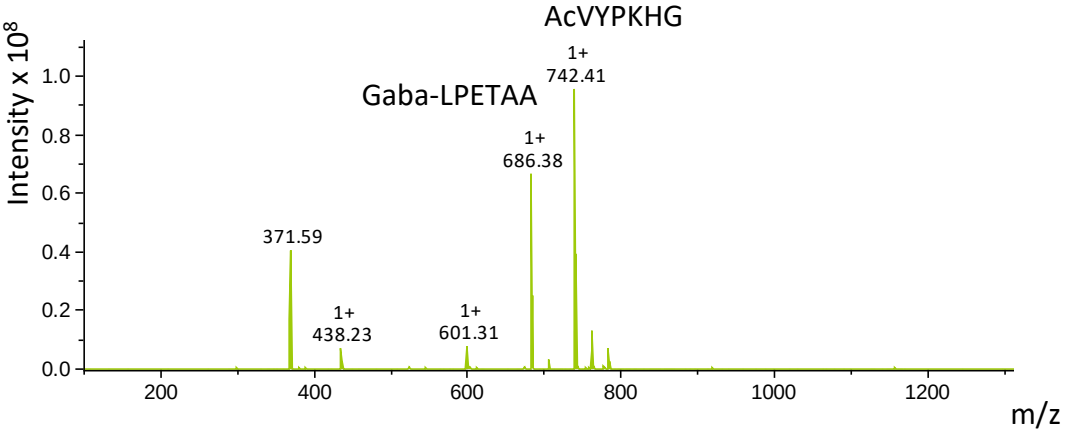
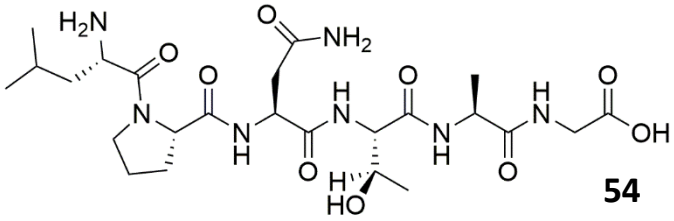


Figure 5.27: Srt025 with Gaba-LPETAA and AcVYPKHG peptides showing no ligation product.

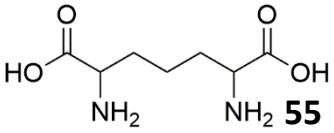
Class D sortases with LPNTAG

Class D sortase accepts the LPNTA motif and diaminopimelic acid (DAPA) as its natural substrates [207]. As Srt024 and Srt025 are class D sortases, these were tested with LPNTAG **54** peptide for hydrolysis and ligation with DAPA **55**. Srt021 was also tested, on the basis that it had been the most successful hit of the 22 new sortases.



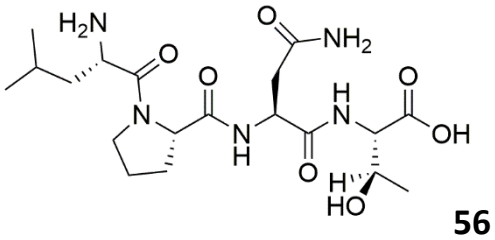
54

LPNTAG – 571.30 Da



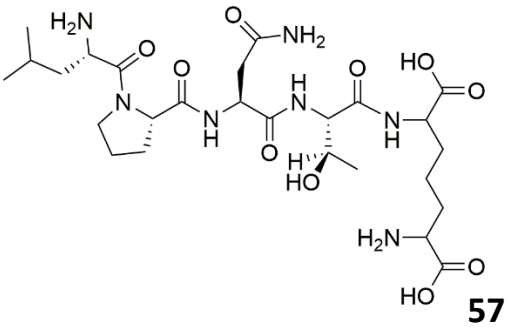
55

DAPA – 190.10 Da



56

LPNT – 443.24 Da



57

LPNT-(DAPA) – 615.32 Da

Figure 5.28: LPNTAG, DAPA, hydrolysed LPNT and ligated LPNT(DAPA).

None of the 3 sortases showed any activity with just LPNTAG peptide (Figure 5.29).

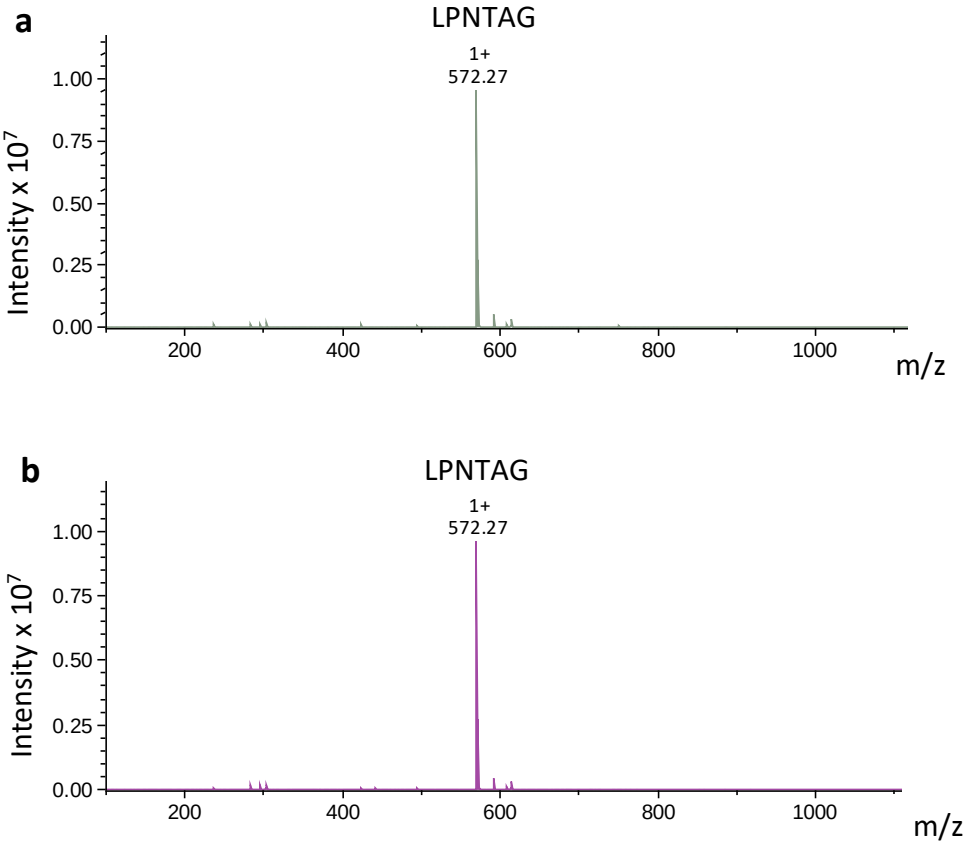


Figure 5.29: Testing of LPNTAG hydrolysis by class D sortase. LPNTAG negative control after 24 h (a), Srt025 showing no hydrolysis product (b).

All 3 sortases showed hydrolysis **56** and ligation product **57** when DAPA was added (Figure 5.30). The nucleophile may be needed to drive an otherwise unfavourable reaction.

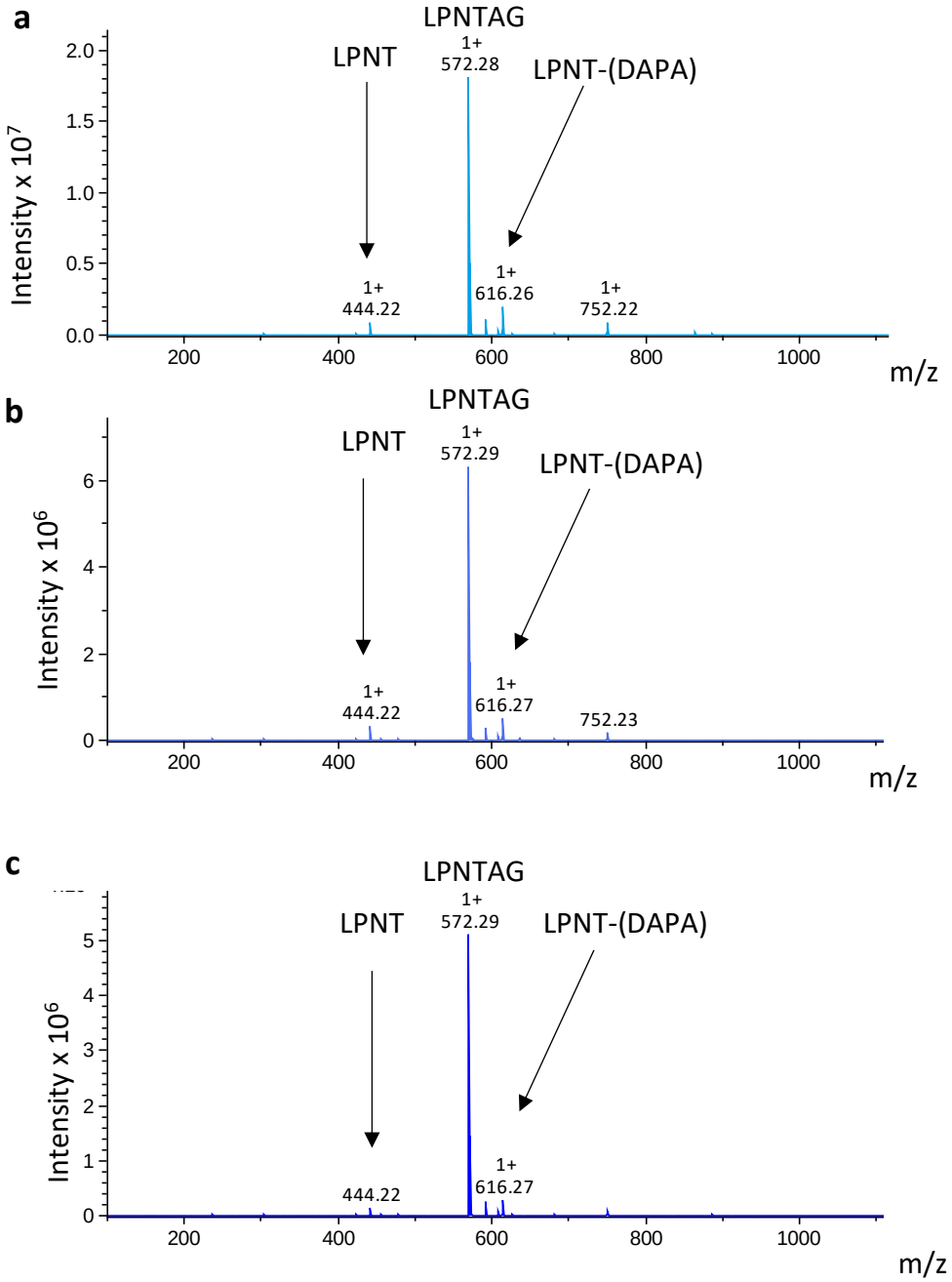


Figure 5.30: Successful ligation of LPNTAG and DAPA. LPNTAG and DAPA showing hydrolysis and ligation in the presence of Srt021 (a), Srt024 (b) and Srt025 (c).

The ligation amounts are very small (Table 9), but again the reaction works *in vitro*.

Sortase	% peptide	% hydrolysis	% ligation
Srt021	86.1	4.30	9.56
Srt024	88.6	4.38	6.99
Srt025	91.9	2.78	5.28

Table 9: Hydrolysis (LPNT) and ligation (LPNT-(DAPA)) yields after 24 h for Srt021, Srt024 and Srt025.

Sortase reactions summary

The 22 new sortases have shown several possible hits (Table 10). Srt019 has only shown activity in one initial Gaba-LPETGG **38** hydrolysis and no reactions since, so is not likely to be functional *in vitro*. Srt009 and Srt010 showed activity in Gaba-LPETAA **51** hydrolysis and ligation with AAA **52** and may be possible hits, although the hydrolysis product was unclear. Srt027 hydrolysed Gaba-LPETGG **38** but has not demonstrated any other reactions. However, as it is a class A sortase, it is unlikely the class C substrates would be suitable. The best hits found were Srt021 and Srt025.

Previously, the Ton-That group made a class C sortase mutant, SrtA 2M, that polymerises pilus proteins *in vitro* [163]. This sortase also took 24 h to show faint bands of reacted product, so the slow reaction rates of Srt021 and Srt025 are not unexpected. Srt021 lacks the DPW “lid” over the active site and therefore appears to mimic the more exposed active site of SrtA 2M. It is possible that obtaining the protein structure and adding further mutations will lead to more improvements in reaction rate. Srt025 is a class D sortase, a class that is poorly characterised and has not before been shown to work *in vitro*. Again, obtaining the protein structure would lead to the possibility of mutating residues close to the active site to create a faster variant.

	009	010	019	021	024	025	027
Gaba-LPETGG hydrolysis	-	-	+	+	+	+	+
Gaba-LPETGG hydrolysis 10x sortase	-	-	-	+++	++	++	++
Gaba-LPETGG + GGG ligation	-	-	-	+	-	+	-
Gaba-LPETGG + AcVYPKHG ligation	-	-	-	+	-	+	-
Gaba-LPETGG + AcVYPRHG ligation	-	-	-	-	-	-	-
Gaba-LPETGG + AcVYAKHG ligation	-	-	-	+	-	+	-
Gaba-LPETGG + FVAKNEG ligation	-	-	-	-	-	-	-
Gaba-LPETAA hydrolysis	+	+	-	+	+	-	-
Gaba-LPETAA + AAA ligation	+	+	-	+	-	-	-
Gaba-LPETAA + AcVYPKHG ligation	-	-	-	-	-	-	-
Fl-Gaba-LPNTAG hydrolysis				-	-	-	
Fl-Gaba-LPNTAG + DAPA ligation				+	+	+	

Table 10: A summary of possible hits from the 22 new sortases.

Optimal sortase conditions

As Srt021 and Srt025 accepted several peptide substrates, their tolerance of pH, temperature and buffer components was also tested. Reactions were set up to test hydrolysis and ligation of Gaba-LPETGG **38** and GGG **41** under varying conditions. The reactions were performed by Mohamed Abd El Bari. The abundance of each peak on the mass spectrum was calculated to give the percentage of reacted and unreacted peptide under each condition.

As pH increases, the sortase reaction equilibrium shifts. Hydrolysis was more rapid at lower pH (Figure 5.31a), and at pH 8 Srt021 showed no hydrolysis and Srt025 very little hydrolysis. Ligation was disfavoured at pH 6 but worked optimally at pH 7-8 (Figure 5.31b). In addition, Srt025 showed faster hydrolysis than Srt021 at all pH values tested, while Srt021 generally showed faster ligation. As they can accept the same peptides, this could help to optimise a reversible labelling system where Srt021 is used to ligate a probe and Srt025 is later added to favour probe removal.

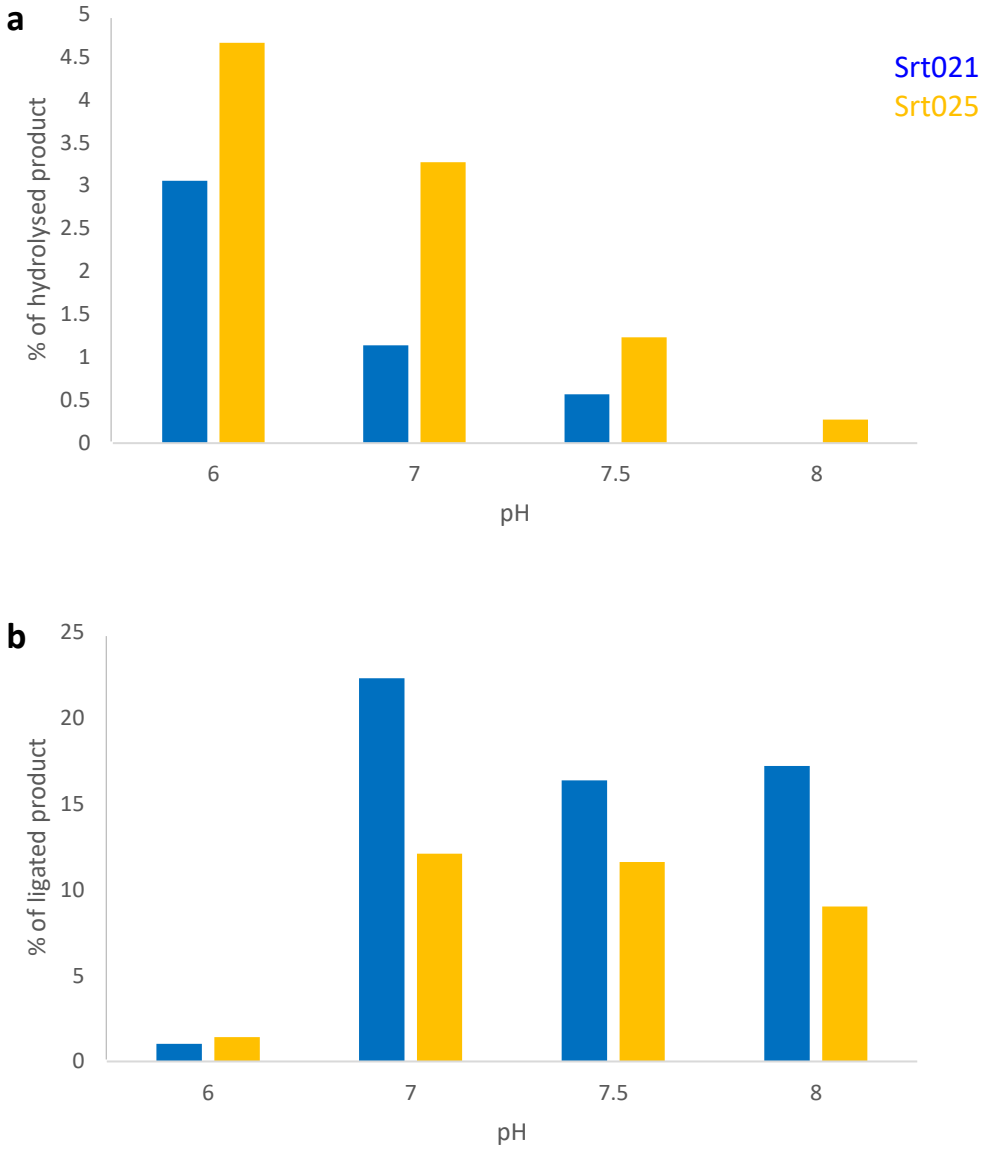


Figure 5.31: Srt021 and Srt025 hydrolysis (a) and ligation (b) at varying pH. Hydrolysis is favoured at acidic pH and ligation favoured at neutral or above pH. In addition, Srt021 performs faster ligation while Srt025 performs faster hydrolysis.

As temperature increases, Srt025 showed an increase in hydrolysis and ligation (Figure 5.32), suggesting an optimal temperature of 37°C as with most sortases [136]. Srt021 performed faster hydrolysis at higher temperatures but the ligation graph showed an unexpected result. The ligation at 25°C appeared to be much higher than the ligation at 37°C. It has been shown that mutation of the lid decreases the thermal stability of sortase [208], hence the lack of this lid in Srt021 may cause it to have an optimal temperature lower than most sortases.

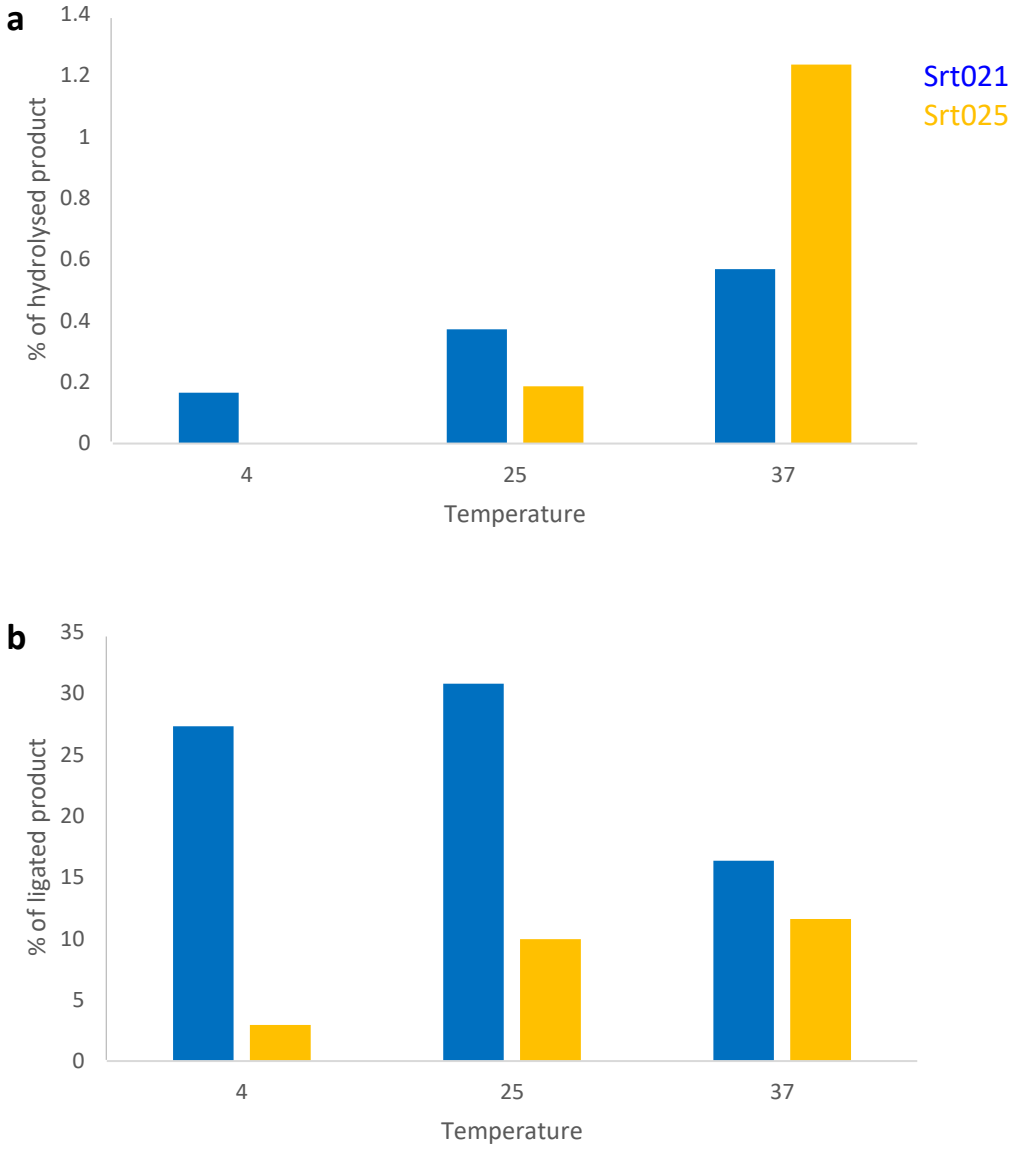


Figure 5.32: Srt021 and Srt025 hydrolysis (a) and ligation (b) at varying temperature. Srt021 is optimal at higher temperatures for hydrolysis but appears to have an optimal ligation temperature around room temperature. Srt025 has faster rates of hydrolysis and ligation with increasing temperature and the likely optimal is 37°C.

Testing of Srt021 and Srt025 in various buffers revealed that both sortases can function without calcium (Figure 5.33). SrtA requires calcium to bind Glu171 in the β6/β7 loop and acidic residues in the β3/β4 loop and form a stable closed conformation [124]. The results for the buffers without calcium suggested that Srt021 and Srt025 can form a stable closed conformation without calcium. In addition, both enzymes function in PBS, which precipitates calcium [209]. Srt021 performed hydrolysis most rapidly in Tris buffer

without calcium, and ligation in Tris buffer with or without calcium, and also in PBS. Srt025 performed hydrolysis most rapidly in Tris buffer with calcium, and ligation in HEPES without calcium. These results are preliminary, but suggest the two novel sortases will work under a wide range of conditions.

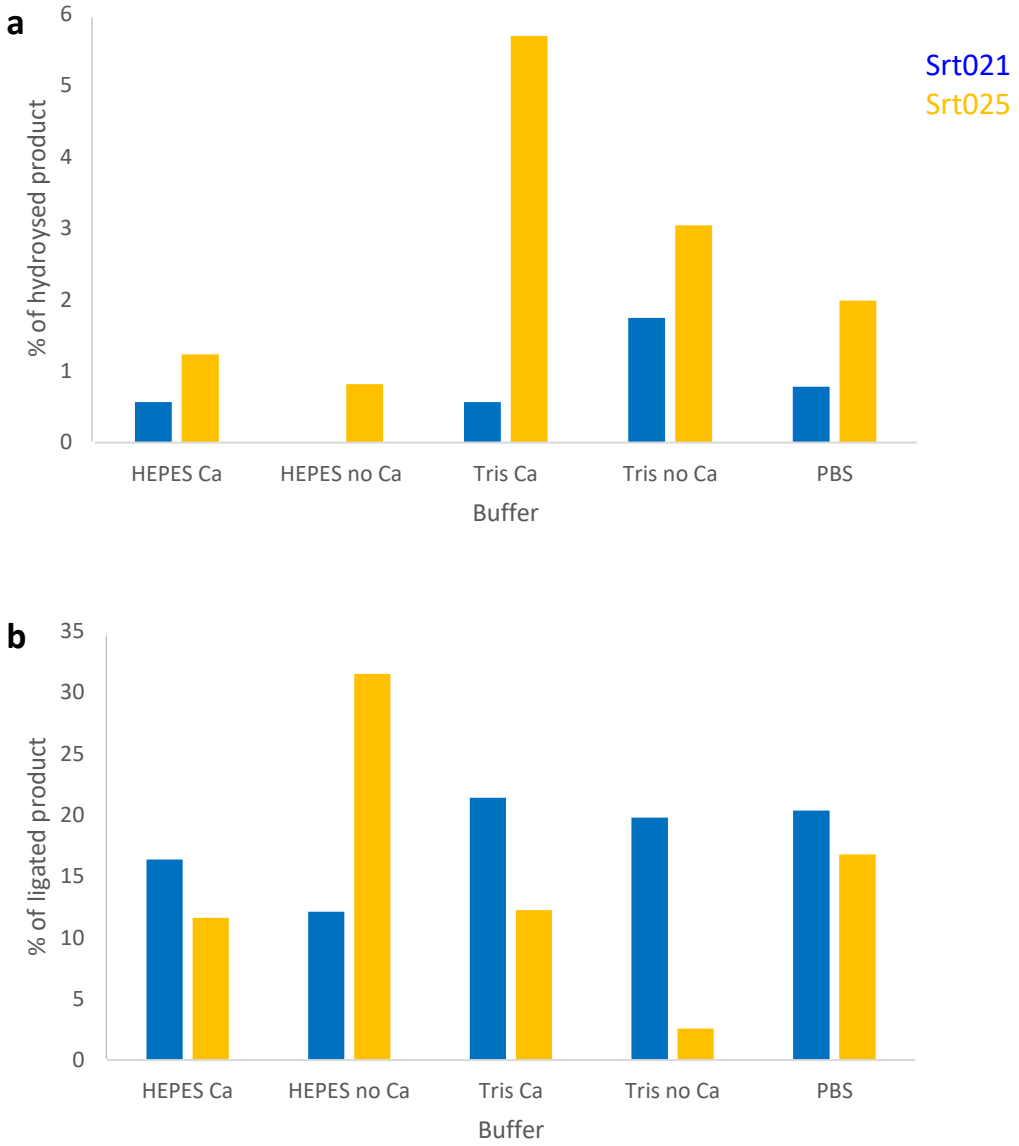


Figure 5.33: Srt021 and Srt025 hydrolysis (a) and ligation (b) in various buffers. Both enzymes show a tolerance for multiple buffers and do not require calcium to be active.

Structures of Srt021 and Srt025

A further investigation into the novel sortases was to attempt crystal growing and X-ray diffraction. A fresh batch of each of Srt021 and Srt025 were expressed in *E. coli*, purified by Ni column, and dialysed into HEPES ligation buffer. The proteins were then further purified by size-exclusion column and concentrated to 13.3 mg / mL (Srt021) and 10.3 mg / mL (Srt025). These proteins were set up for crystallisation trials in 96 well 2 drop plates (kindly performed by Wendy Robinson). The first crystallisation trial failed to produce crystals, so another batch of twice-purified protein was concentrated to at least 20 mg / mL and set up in another trial. The second attempt also failed to produce crystals.

As the experimental method for determining structures had not been successful, the proteins were modelled by computer. The Phyre server is a structure prediction algorithm that uses homology modelling to align and predict the structure of novel proteins based on the known crystal structure of a related protein [210]. Phyre2 identified the best model for Srt021 as Sortase A from *Corynebacterium diphtheriae*, modelling 92% of the sequence with 100% confidence. The best model identified for Srt025 was Sortase C-1 from *Streptococcus pneumoniae*, modelling 97% of the sequence with 100% confidence (Figure 5.34).

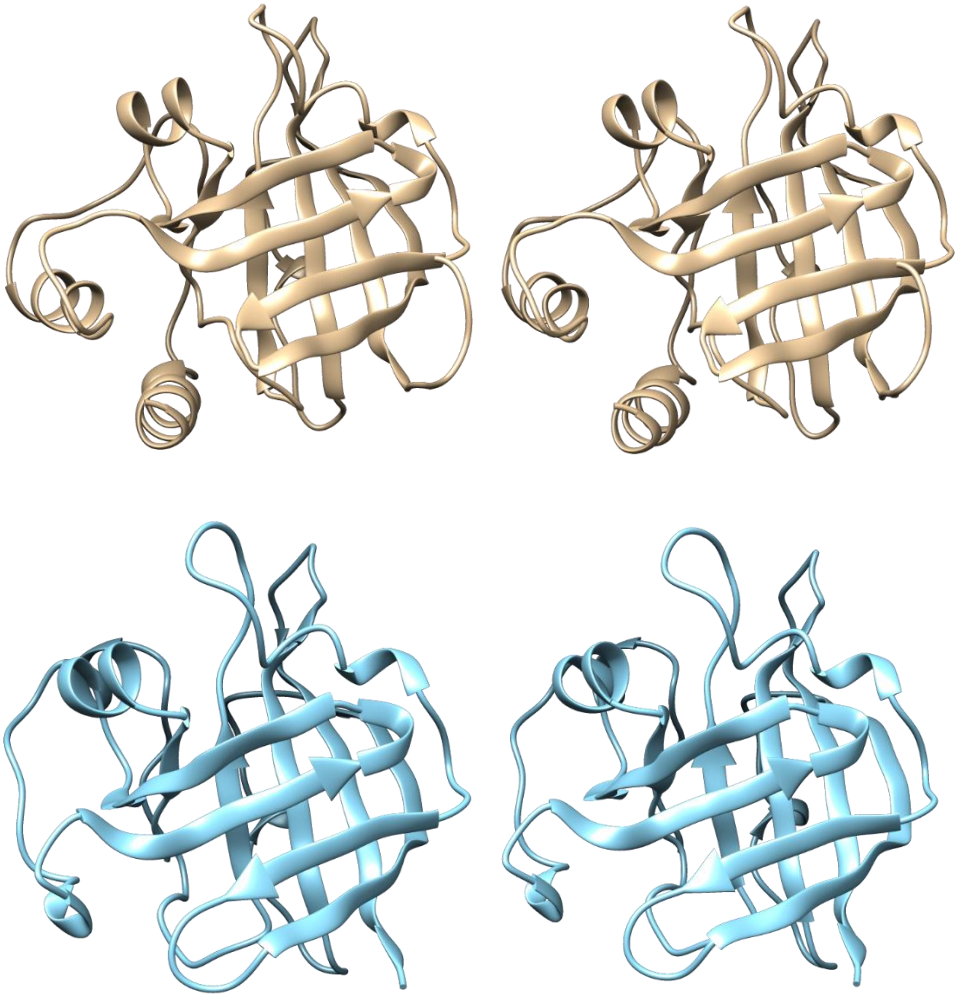


Figure 5.34: Stereo view of Srt021 (tan) and Srt025 (blue) structures modelled by Phyre 2.

Recently, Google's DeepMind has created an artificial intelligence for solving unknown protein structures without requiring a homologous structure, AlphaFold [211]. AlphaFold2 was used to model the structures of Srt021 and Srt025 (Figure 5.35).

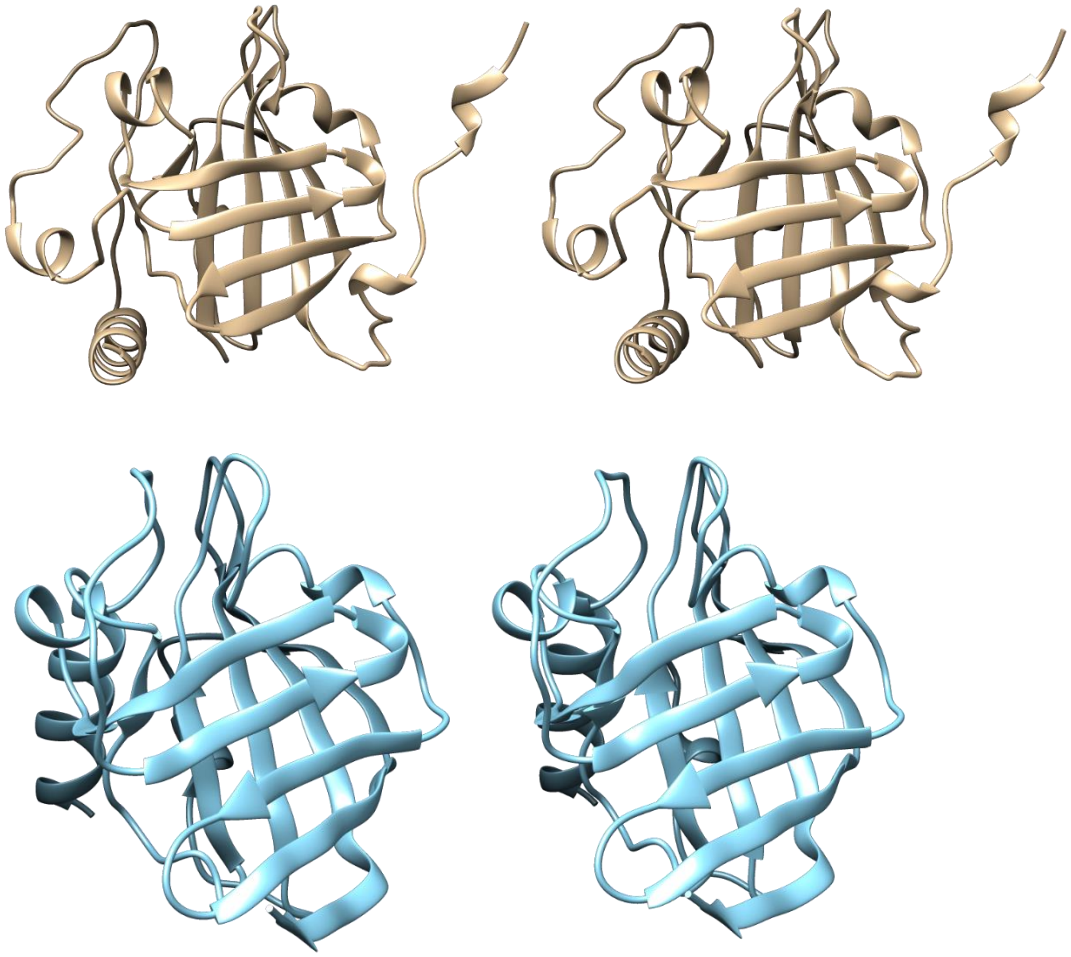


Figure 5.35: Stereo view of Srt021 (tan) and Srt025 (blue) structures modelled by AlphaFold2.

The two modelling methods are in good agreement. Srt021 has near-identical placement of the α-helices and β-sheets and the largest difference is in the placement of the long loop from approximately residues Ala14 to Gly34. Srt025 also has near-identical placement of the β-sheets and most of the α-helices, with a large difference in the placement of the first amino acids up to Gly28. AlphaFold predicts a long α-helix here while Phyre has a disordered loop. There are also slight differences in the placement of the other loops.

When overlaid, the structures of Srt021 and Srt025 share the eight-stranded β-barrel common to all sortases, and differ mainly in the placement of the first section of residues – an α-helix up to Ala14 in followed by a long disordered loop up to Gly50 in Srt021 (Figure 5.36). This corresponds to an α-helix up to Ala14 in followed by a long

disordered loop up to Gly28 in Srt025. The residues then match closely until diverging at Ala180 on Srt021 and Lys155 on Srt025, which is 4 residues from the end of Srt025. Srt021 is the larger protein and continues until Pro202 with a disordered loop.

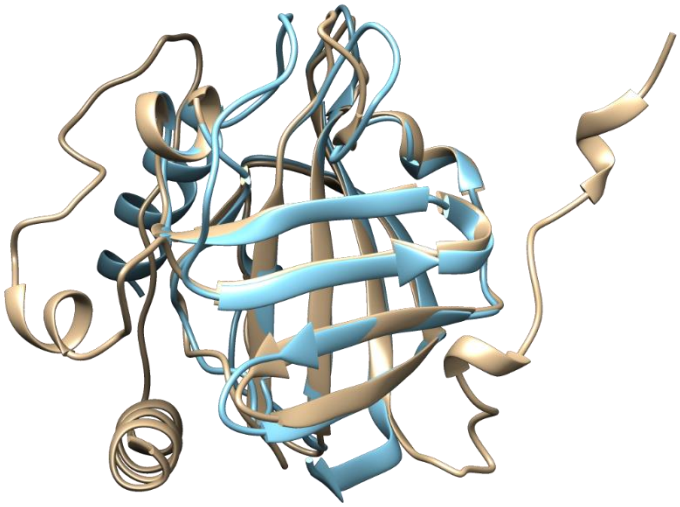


Figure 5.36: Overlay of Srt021 structure (tan) and Srt025 structure (blue) predicted by AlphaFold2 shows a similar β-barrel structure with some differences in the placement of the residues before the β-barrel fold.

When overlaid with the fast variant SrtA 5M, both AlphaFold sortase predictions show significant differences (Figure 5.37). The α-helix and long loop at the start of Srt021 appears to pass close to the active site. The end residues from Ser181 also diverge, but the β-barrel fold matches more closely. The predicted Srt025 structure differs even more from SrtA 5M, as the initial α-helix appears to be blocking where the active site of SrtA 5M would be. Again, the rest of the protein follows the β-barrel fold closely. It is possible that the α-helix flexes to give access to the active site, or that the Phyre2 structure that has a disordered loop in place of the α-helix is the more accurate.

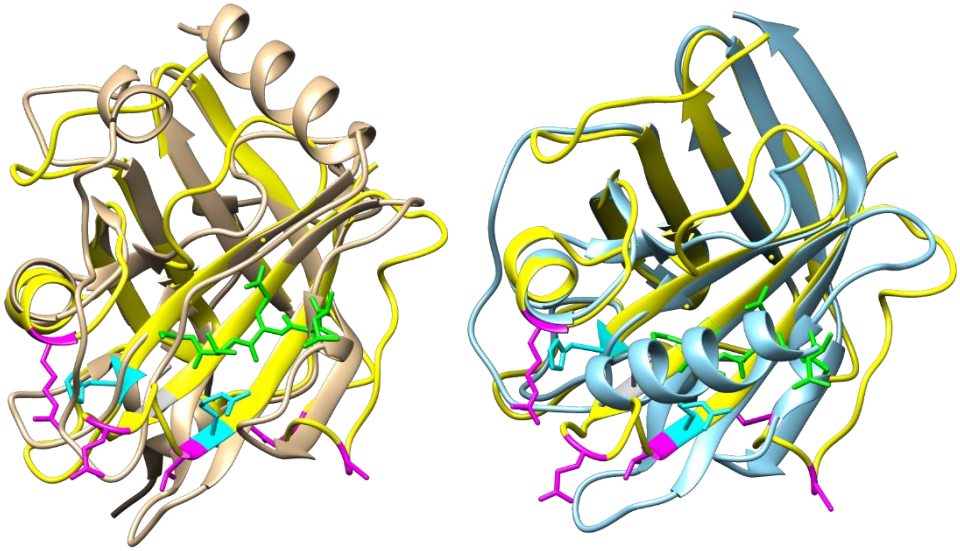


Figure 5.37: Overlay of Srt021 (tan) and Srt025 (blue) structures predicted by AlphaFold with SrtA 5M (yellow). Both structures differ in the residues before the β-barrel fold.

Finally, Srt021 and Srt025 can be overlaid with the class C SrtA 2M. The loop of Srt021 does not pass as closely to the active site as the DPW lid present in SrtA 2M, confirming that the lack of a lid region in Srt021 leaves the active site more exposed. The AlphaFold structure of Srt025 again suggests the first α-helix is in close proximity to the active site.

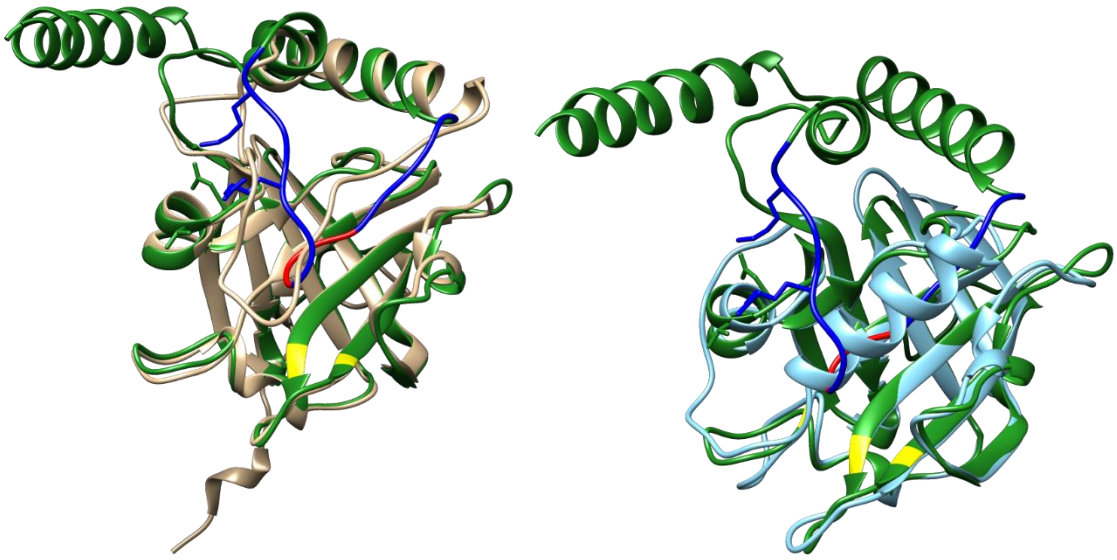


Figure 5.38: Overlay of Srt021 (tan) and Srt025 (blue) structures predicted by AlphaFold2 with SrtA 2M (green). Srt021 has a more exposed active site than SrtA 2M, while Srt025 may be less exposed due to an α-helix.

Conclusion

Two functional sortases have been found in the set of 22 sortases made at Prozomix. The sortase variants were tested for hydrolysis of peptides carrying the sortase motifs LPETG **38** / LPETA **51** / LPNTA **54**. They were also tested for ligation with nucleophiles including GGG **41** and AAA **52** tripeptide, AcVYPKHG **43** and similar variants taken from the pilus protein that is the natural nucleophile for class C sortase, and DAPA **55**, the natural nucleophile for class D sortase. Srt021 and Srt025 consistently showed activity, with Srt009, Srt010, Srt019, Srt024 and Srt027 also showing activity in at least one assay.

Previously, the *C. diphtheriae* class C sortase has been engineered to function in vitro by mutation of a DPW "lid" over the active site [163]. Srt021 naturally has no DPW lid and Srt025 is classified as class D, although its activity towards the same nucleophiles as Srt021 suggests it functions as a class C sortase. These two hits have been tested under varying conditions of pH, temperature and buffer. The results suggest an optimal pH around 7-7.5 for both enzymes. The optimal temperature of Srt021 appears to be closer to 25°C than 37°C, while Srt025 is optimal at 37°C. Both enzymes also tolerate a range of buffers and appear to be calcium-independent.

Chapter 6 : Conclusions and future work

Conclusions and future work

This project aimed to use sortase to specifically label sugars incorporated into glycoproteins and evolve sortases that would be specific to different sugar motifs. The novel sugars AcManGGN$_3$ **26**, AcManThz **27** and AcGlcAN$_3$ **28** are all tolerated by HEK cells up to 100 µM. AcManThz **27** is tolerated up to at least 300 µM, which is higher than most other sugars [124]. AcManGGN$_3$ **26** is incorporated into cell surface glycoproteins at levels comparable to AcManGN$_3$ **4**, which has been previously shown to be well incorporated [62], and can be labelled by SPAAC.

Unlike AcGlcGN$_3$ **6**, AcGlcAN$_3$ **28** does not appear to be incorporated into glycoproteins. There are several proteins that must tolerate a modified sugar for its successful incorporation into glycans. The sugars must be imported into the cell, by GLUT or SGLT transporter proteins. Several of these have been shown to tolerate multiple sugars, such as glucose and glucosamine [56]. It would be possible to test whether labelling of glycans by AcGlcGN$_3$ **6** was blocked by glucose transporter inhibitors such as cytochalasin B, which binds to GLUT1 [58]. This would help to work out how the sugars are being imported into the cell. Specific glucose transporters shown to be required for uptake of AcGlcGN$_3$ **6** could then be tested for binding affinity to AcGlcAN$_3$ **28**. This would help establish if the AcGlcAN$_3$ **28** is successfully diffusing into the cell. Following diffusion into the cell, sugars must enter the biosynthetic pathway to convert to nucleotide sugar donors. It is possible the Roseman-Warren pathway, which converts mannosamine derivates to sialic acid, does not need enzymes to be as stringent as sialic acid is the only 9-carbon sugar. In contrast, the hexosamine salvage pathway may require more stringent enzymes. It would be possible to test the incorporation of AcGlcAN$_3$ **28** into the glycans of homogenised cells [70]. This would specifically look at whether the biosynthetic enzymes accepted the unnatural sugar, as the membranes are disrupted. Finally, protein transporters must move the sugar donor into the Golgi apparatus, and the efficiency of sugar incorporation into glycans could also be tested with inhibitors of these transporters [212] [213]. Again, testing inhibitors confirmed to block AcGlcGN$_3$ **6** incorporation could help identify proteins that may be involved in uptake of AcGlcAN$_3$ **28**. Finally, radiolabelled AcGlcAN$_3$ **28** could be tested to check how much sugar is taken into the cytoplasm, incorporated into surface glycoproteins, or contained in proteins

secreted from the cell. It would also be possible to try butanoylated sugars to see if this helps incorporation [124].

The second aim of the project was to reduce sugars known to be incorporated with azide functional groups to amine to expose glycine motifs available for sortase ligation. Live-cell labelling with sortase was unsuccessful, despite several attempts at optimisation including different reducing agents, varying the temperature and introduction of a blocking step. As with $AcGlcAN_3$ **28**, it is not clear which step might present a bottleneck – initial sugar transport into the cell, biosynthetic pathways, and transport into Golgi apparatus. Again, testing of homogenised cell extract would verify if the biosynthetic pathways accepted the amine sugars, and transporter proteins suggested to be important in transporting the azide sugars by inhibitor studies could be tested for binding to the amine sugar. Future work could also use purified sialic acid synthase [214] to convert AcManG **29** and AcManGG **34** sugars to sialic acid bearing glycine motifs *in vitro*. This would verify if the enzyme can recognise the glycine motifs while linked to sialic acid.

The second labelling method, OPAL, was tested on AcManThz **27**. This method incorporates a thiazolidine that is decaged to aldehyde with allylpalladium (II) chloride dimer catalyst **35** [167]. This can be labelled by an aminooxybiotin probe **37** that is detected by fluorophore-streptavidin conjugate. After optimisation in particular of the Pd concentration, the OPAL probe appeared to be successfully ligated to the surface of HEK cells. Evaluating sugar incorporation via the OPAL method on total cell glycoproteins by western blot was unsuccessful, despite varying Pd concentration and addition of a blocking step. Future work would continue the optimisation of this method, possibly by investigating the concentrations of other components and reaction duration. Testing of the incorporation of radiolabelled AcManThz **27** would again shed light on how much sugar was being taken into the cell, and how much incorporated into surface glycoproteins. Alternatively, the protocol could be simplified by synthesis of an aminooxy probe with a fluorophore, which would remove a step and possibly improve the labelling.

In addition to testing new sugars for incorporation into glycans, the project aimed to develop new sortases for specific ligation between sugar residues and probes. 22 novel sortases were produced from a placement at Prozomix. These were drawn from a

database for their similarity to SpySrtA and SauSrtA, which both work *in vitro* [215] [142]. Many of the sortases were class C and matched sortases that have already been documented on protein databases, although most have not been studied in detail. Testing of the sortase variants revealed two functional sortases, Srt021 and Srt025. Sortases are grouped into classes based on their primary sequence [139], Srt021 is classified as class C, and Srt025 as class D. These accepted several nucleophiles for ligation with LPXTG **38** including GGG **41**, AcVYPKHG **43** and AcVYAKHG **47**. The latter two nucleophiles are based on pilin proteins, which are normally ligated by class C sortase [138, 216]. Srt021 also ligated LPXTA **51** with AAA **52** and Srt021, Srt024 and Srt025 showed ligation with the class D nucleophile [139] diaminopimelic acid **55**. Future work could test these sortases with depsipeptide to check if the ligation reaction can be forced to completion, and with AcManG and glycine-sialic acid sugars to see if the sortases could potentially be used to label incorporated sugars.

Srt021 and Srt025 were tested under varying conditions of pH, temperature, and buffer. The results suggested an optimal pH around 7-7.5 for both enzymes. SrtA works *in vitro* at optimal pH of 7.5 and temperature of 37°C [145]. The optimal temperature of Srt021 appeared to be closer to 25°C than 37°C, while Srt025 is optimal at 37°C. Both enzymes also tolerated a range of buffers and appeared to be calcium-independent, although other calcium-independent sortases already exist [156]. Future work could investigate the enzyme kinetics. It would be possible to ligate a fluorescent probe to a protein, then use SDS-PAGE to detect the increase in fluorescence in protein bands, as the ligation is a permanent bond. This would provide an estimate for the rate at which the enzyme converts to ligated product. It would also be possible to use isothermal titration calorimetry to analyse the binding energy between sortase and substrate. It would be desirable to mutate the sortases to bind substrate more strongly and react faster, and it may be possible to achieve this with CASTing [165]. This technique simultaneously mutates pairs of amino acids located close to the active site. The library of mutants produced may yield variants with improved binding or altered substrate preference.

Previously, the *C. diphtheriae* class C sortase has been engineered to function in vitro by mutation of a DPW “lid” over the active site [163]. Srt021 naturally has no DPW lid and Srt025 is classified as class D, although it shows activity towards the same class C nucleophiles as Srt021. Both sortases were modelled using Phyre2 and AlphaFold2 and

the structures suggested Srt021 had a more exposed active site, while Srt025 may be less exposed. Both attempts at growing crystals of Srt021 and Srt025 were unsuccessful. Future work would include another attempt at crystallization to obtain experimentally-determined structures and compare with the modelled structures. This would assist in determining site-directed mutations that could be tested for improvement to the reaction rate in order to develop faster sortase variants.

Ideally, Srt021 and Srt025 would be used as the basis to create two fast sortase variants that are orthogonal to each other like Srt 5M and Srt 2A [153]. The nucleophiles accepted could be glycine and lysine, as the current Srt021 and Srt025 accept both. This would allow multiple labelling of sugars in the same experiment. Future directed evolution of sortase variants would also aim to create variants that recognise the sugar the motif is attached to – first specific to glycine on sialic acid, and then further mutating so the variant is specific only to glycine on sialic acid that is α2,6-linked to GalNAc. This is sialyl-Tn antigen, found on many cancers [20]. Rapid identification of glycans associated with metastasis on cancer cells would improve diagnosis of aggressive cancers and enable earlier, more robust treatment.

Chapter 7 : Experimental

Peptide synthesis

Preparing Resin

Preloaded resin, either 0.54 mmol / g H-Gly-2-chlorotrityl or 0.64 mmol / g H-Ala-2-chlorotrityl (Sigma Aldrich) was weighed into an 8 mL plastic cartridge fitted with a filter and stopcock. 6 mL *N,N*-dimethylformamide (DMF) was added and left to swell for 30 min with rotation, then DMF was filtered off under vacuum.

Amino Acid Coupling

Reaction mixture containing Fmoc-amino acid (5 equiv) plus 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) (5 equiv) plus *N,N*-diisopropylethylamine (DIPEA) (10 equiv) dissolved in minimum DMF was added to the cartridge and left for 60 min with rotation. The mixture was drained, and resin washed 3 times with 6 mL DMF for 2 min with rotation.

Fmoc Deprotection

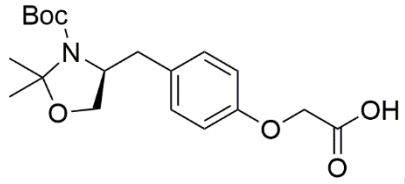
Resin was washed with 6 mL of 20% piperidine in DMF for 2 min with rotation and then piperidine filtered off, for a total of 5 washes. Resin was washed with 6 mL DMF for 2 min with rotation and then filtered, for a total of 5 washes. The peptide was elongated with further coupling and deprotection reactions as necessary.

N-terminal modifications

Some peptides had the *N*-terminal amine blocked by acetylation. Acetic anhydride (5 equiv) in minimal DMF was added to the cartridge and rotated for 5 min. DIPEA (10 equiv) was added to the cartridge and rotated for 30 min.

For the fluorescent peptides, after a final amino acid coupling with Boc-GABA, the resin was treated with a solution of fluorescein isothiocyanate (FITC) (6 equiv) and DIPEA (14 equiv) dissolved in a minimal amount of DMF. The reaction was spun at RT overnight in the dark. The resin was drained and washed with 6 mL DMF for 4 x 2 min spins, followed by peptide cleavage.

For the OPAL probe used in AcManThz labelling, added to the *N*-terminus was 5 equiv. (*S*)-2-(4-((3-(*tert*-butoxycarbonyl)-2,2-dimethyloxazolidin-4-yl)methyl)phenoxy) acetic acid **58** in the usual method for coupling amino acids. This compound was kindly synthesised by Amanda Noble [167]:



58

Cleavage and Isolation

Resin was isolated by washing 3 times with 6 mL DCM and 3 times with 6 mL methanol, then dried under high vacuum overnight. 6 mL of cleavage cocktail of 95% trifluoroacetic acid (TFA), 2.5% H_2O and 2.5% triisopropylsilane (TIS) was added and left for 60 min with rotation. This was drained into 40 mL ice-cold diethyl ether and centrifuged at 6000 g at 4°C for 10 min to pellet the peptide precipitate. Ether was carefully decanted and ether wash (40 mL), centrifuge and decant was performed on the pellet 3 more times. All non-fluorescent peptides were dissolved in H_2O with 10% glacial acetic acid, lyophilized and characterised by ESI-MS and LC-MS.

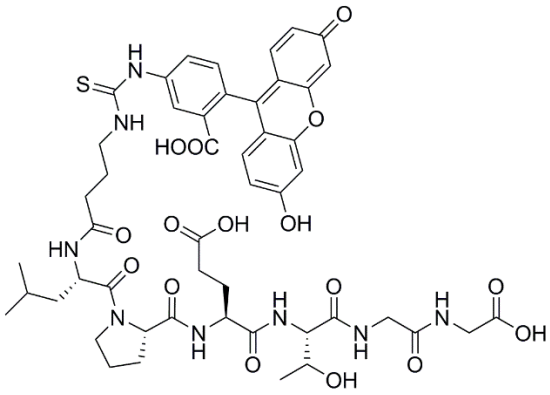
Fluorescent peptides were further purified on a Sephadex LH-20 size-exclusion column. The column was first equilibrated in 2 column volumes (CVs) (1000 mL) of methanol before the pellet was dissolved in minimal MeOH (5mL) and loaded. Fractions of 9 mL were collected at a flow rate of 70 mL per hour and the column run for 18 h. Fractions were monitored with UV for presence of fluorophore and the purity of FITC-peptide established by LC-MS. Pure fractions were dissolved in H_2O with 10% glacial acetic acid, lyophilized and characterised by ESI-MS and LC-MS.

The OPAL probe was further modified by oxidation with sodium periodate in 0.1 M PBS, 0.1 M NaCl, (pH 7.0). To 15 mg of cleaved peptide in 500 µL of buffer was added 210 µL of 112 mM $NaIO_4$. This was allowed to sit for 2 min on ice in the dark and quenched with the addition of 250 µL of 200 mM methionine. The reaction progress was checked on LC-MS and the oxidation repeated until the desired product (1038 Da peak) was seen. The solution was then loaded onto a 6 mL solid phase extraction cartridge (Supelclean LC-18) equilibrated with 2 mL acetonitrile followed by 2 mL water. The product was eluted over a gradient of acetonitrile and fractions containing product lyophilised to give the finished OPAL probe **37**.

LC-MS was performed using an UltiMate 3000 Rapid Separation HPLC and an HCT ultra Electron Transfer Dissociation II ion trap. The software used was Chromeleon Xpress, esquire and Compass HyStar. ESI-MS was performed using a Bruker micrOTOF

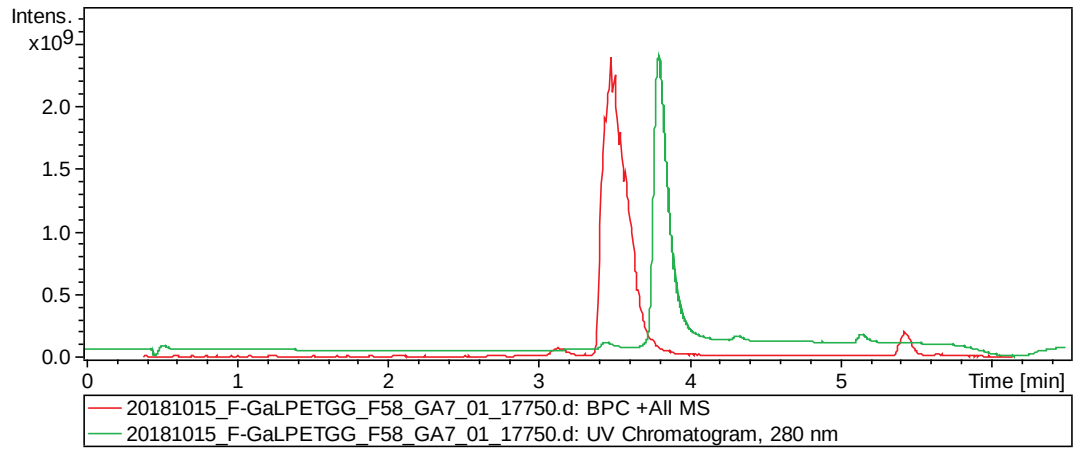
instrument. LC-MS and ESI-MS results were analysed using Compass DataAnalysis software.

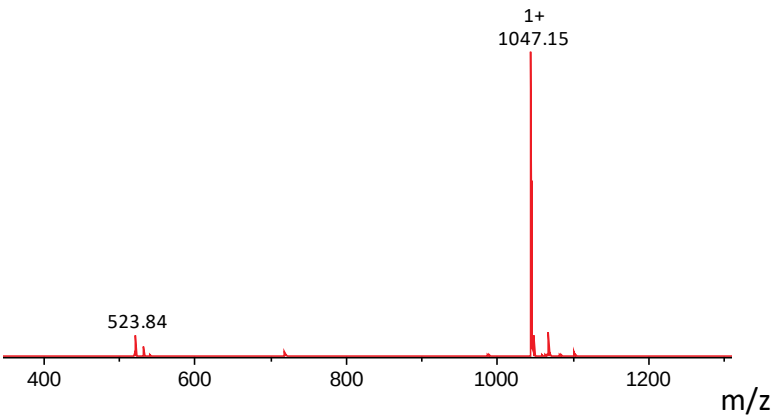
Fl-Gaba-LPETGG 30



200 mg of preloaded glycine resin loading 0.54. 60 mg, 0.057 mmol, 53% yield.

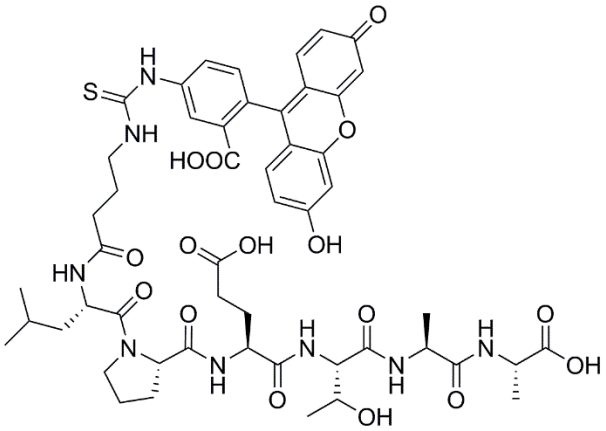
ESI-MS: Found $[M+Na]^+$ 1069.3579, $C_{49}H_{58}N_8NaO_{16}S$ requires 1069.3584.





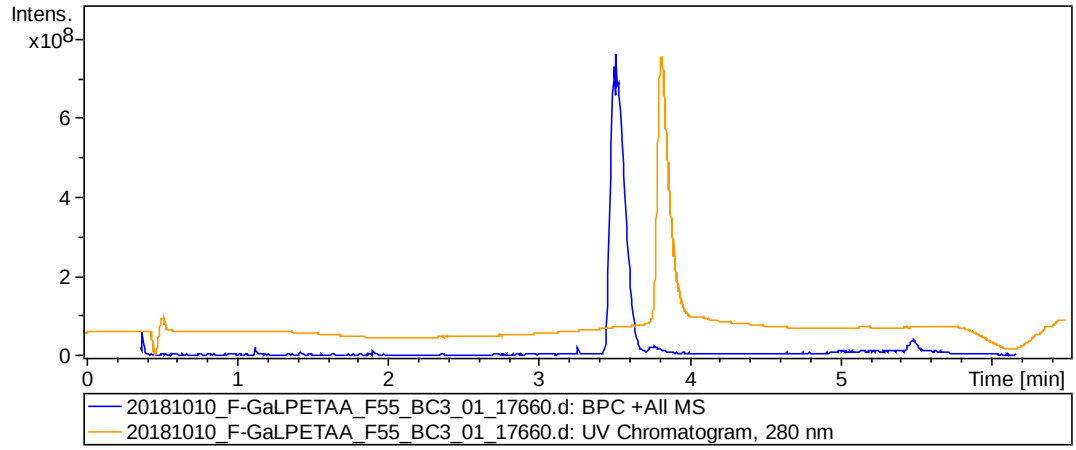
LC-MS: Found 1+ at 1047.15, 2+ at 523.84.

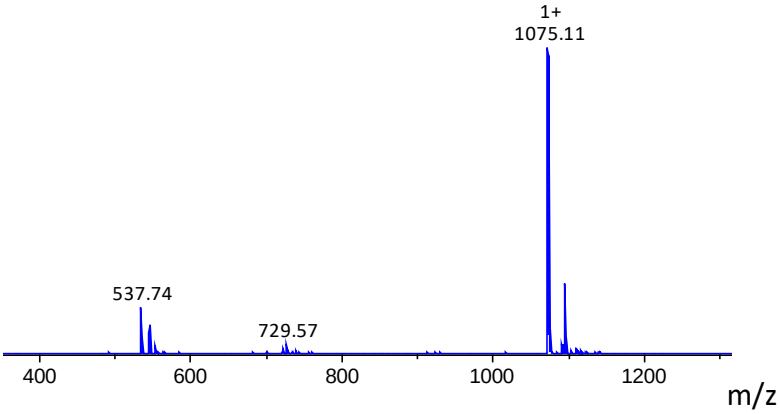
Fl-Gaba-LPETAA 31



150 mg of preloaded alanine resin loading 0.64. 43 mg, 0.040 mmol, 41% yield.

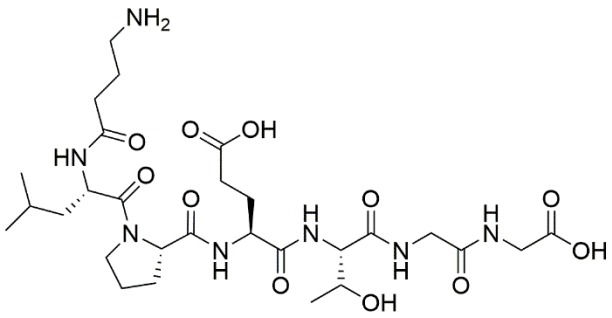
ESI-MS: Found $[M+Na]^+$ 1097.3912, $C_{51}H_{62}N_8NaO_{16}S$ requires 1097.3897.



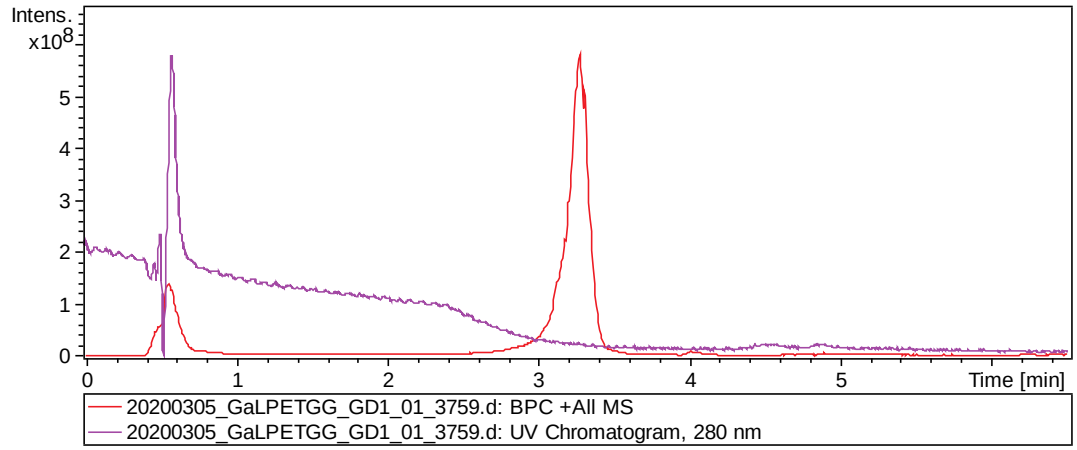


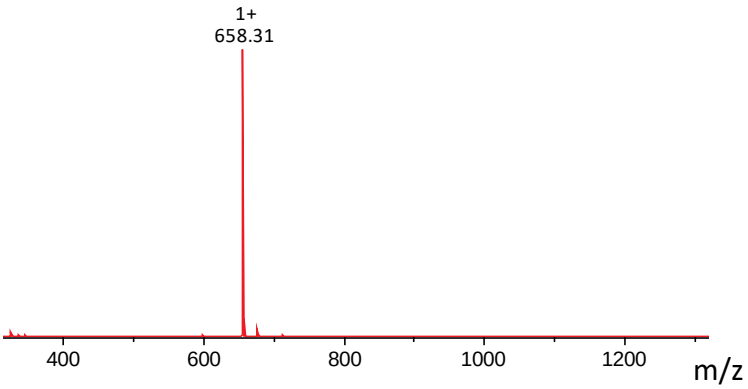
LC-MS: Found 1+ at 1075.11, 2+ at 537.74.

Gaba-LPETGG 38



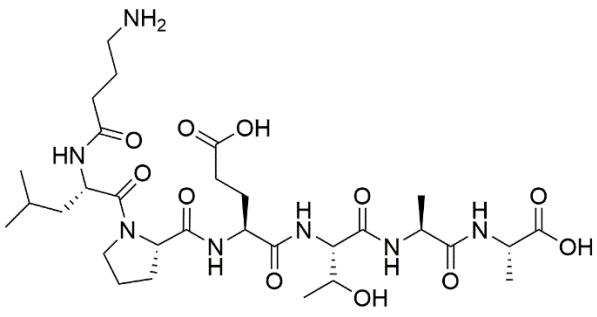
200 mg of preloaded glycine resin loading 0.54. 60 mg, 0.091 mmol, 85% yield.



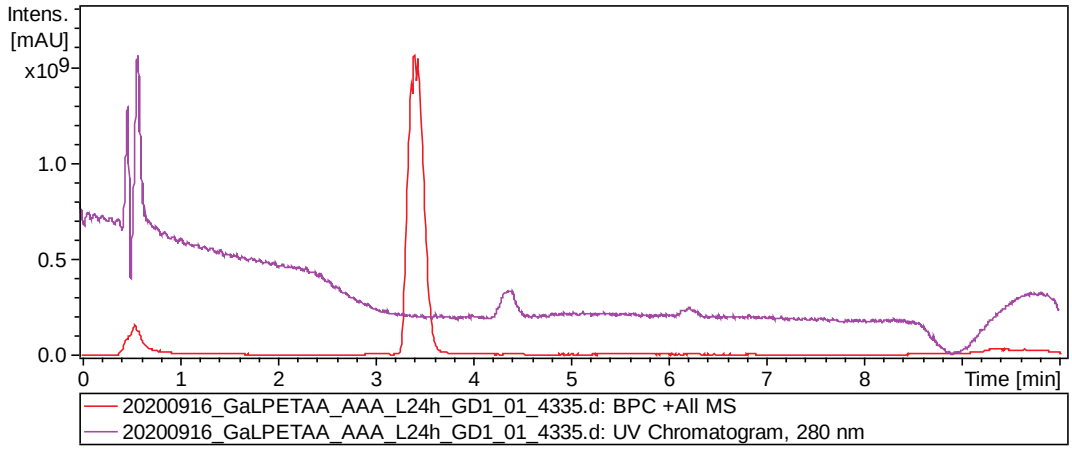


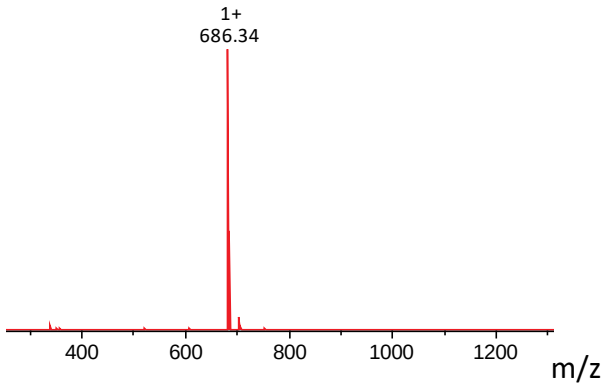
LC-MS: Found 1+ at 658.31.

Gaba-LPETAA 51



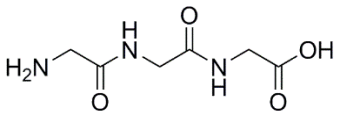
150 mg of preloaded alanine resin loading 0.64, 39 mg. 0.057 mmol, 59% yield.





LC-MS: Found 1+ at 686.34.

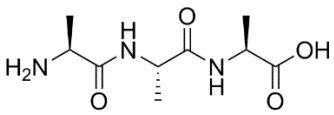
GGG 41



300 mg of preloaded glycine resin loading 0.52, 26 mg, 0.135 mmol, 87% yield.

ESI-MS: Found $[M+H]^+$ 190.0823, $C_6H_{12}N_3O_4$ requires 190.0822.

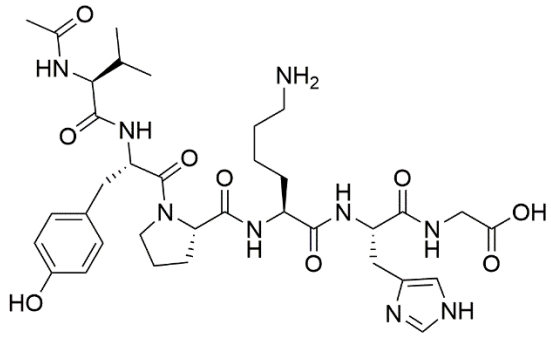
AAA 52



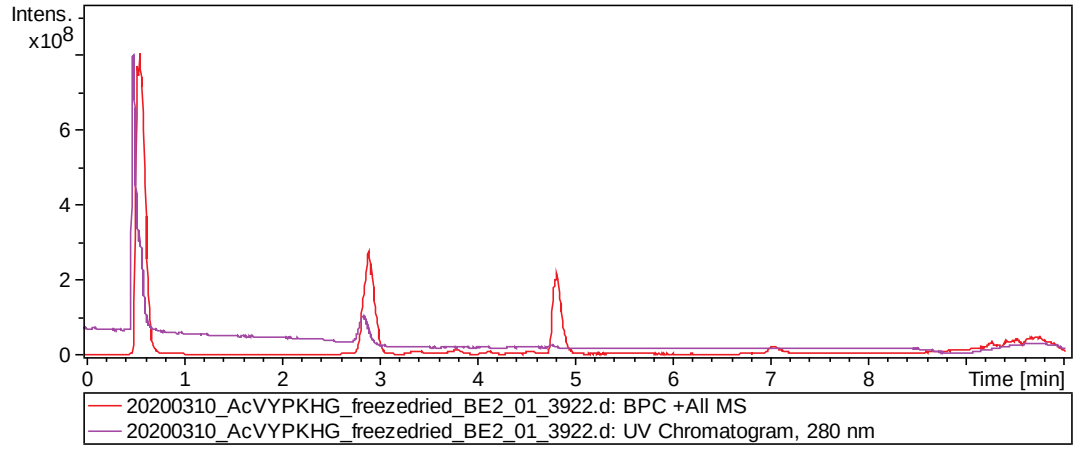
250 mg of preloaded alanine resin loading 0.64, 26 mg, 0.112 mmol, 70% yield.

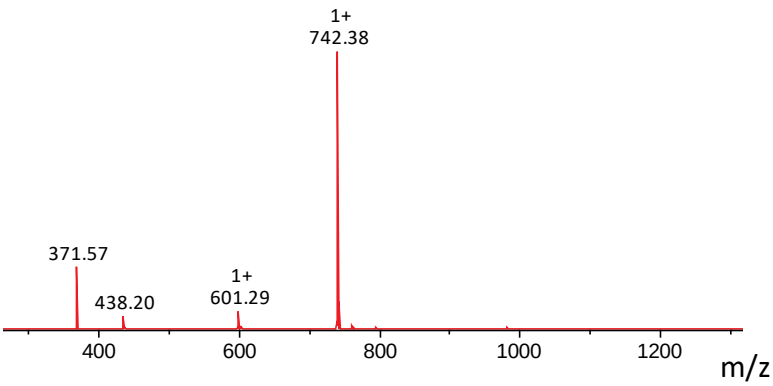
ESI-MS: Found $[M+H]^+$ 232.1287, $C_9H_{18}N_3O_4$ requires 232.1292.

AcVYPKHG 43



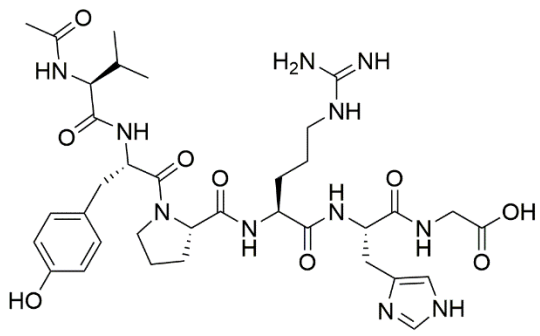
100 mg of preloaded glycine resin loading 0.79, 54 mg, 0.073 mmol, 92% yield.



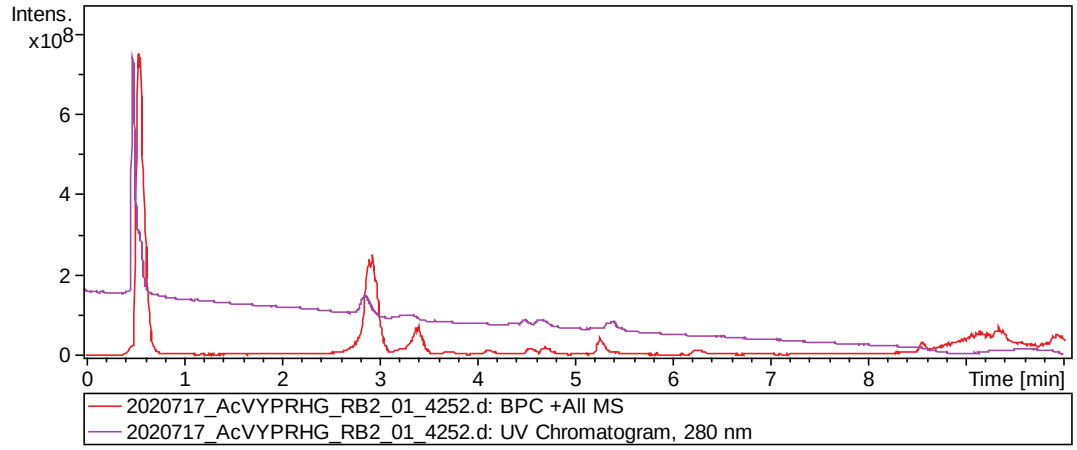


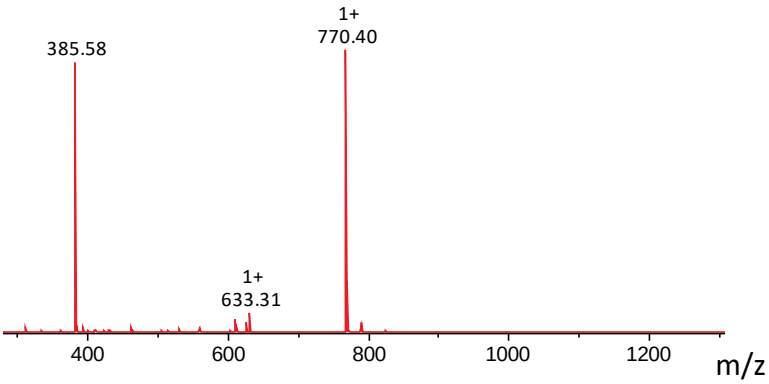
LC-MS: Found 1+ at 742.38, 2+ at 371.57.

AcVYPRHG 45



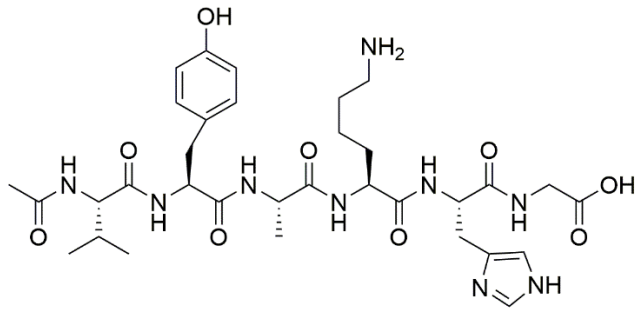
100 mg of preloaded glycine resin loading 0.79, 11 mg, 0.015 mmol, 19% yield.



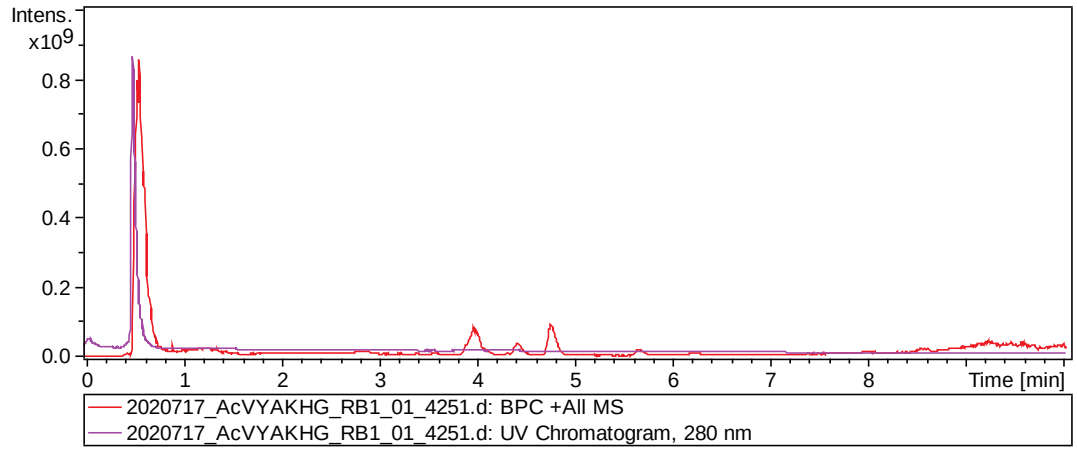


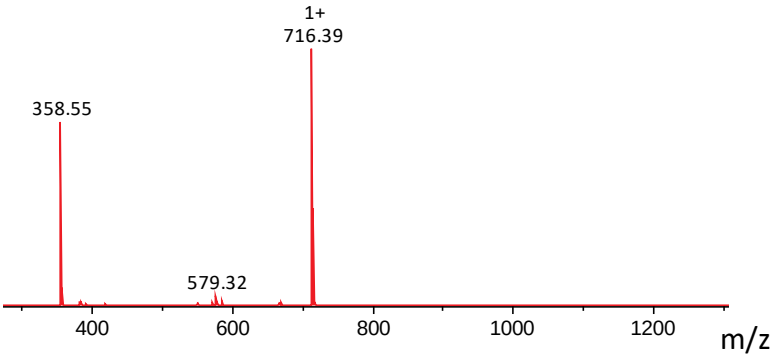
LC-MS: Found 1+ at 770.40, 2+ at 385.58.

AcVYAKHG 47



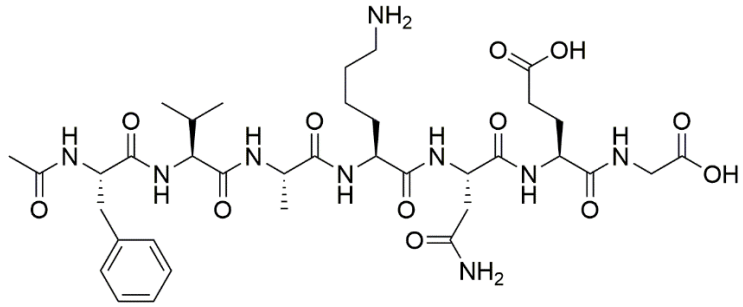
100 mg of preloaded glycine resin loading 0.79, 28 mg, 0.040 mmol, 51% yield.



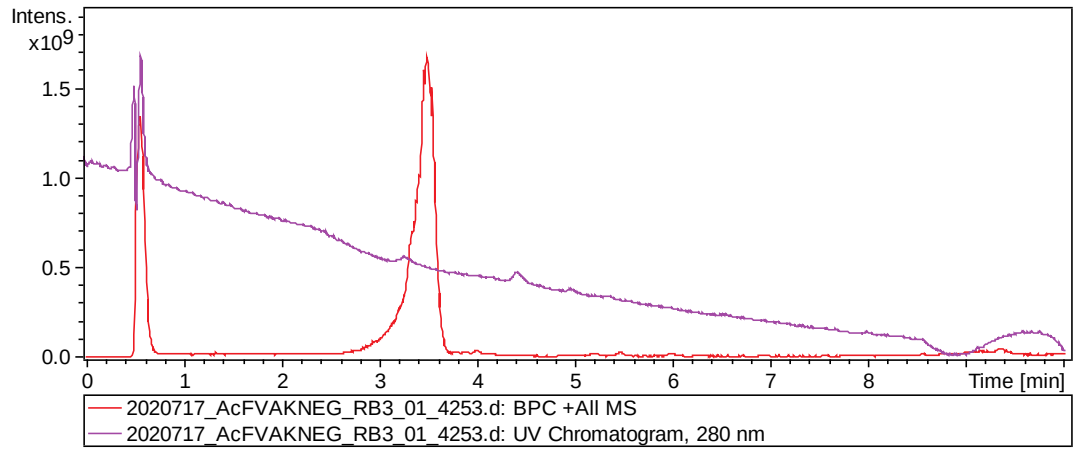


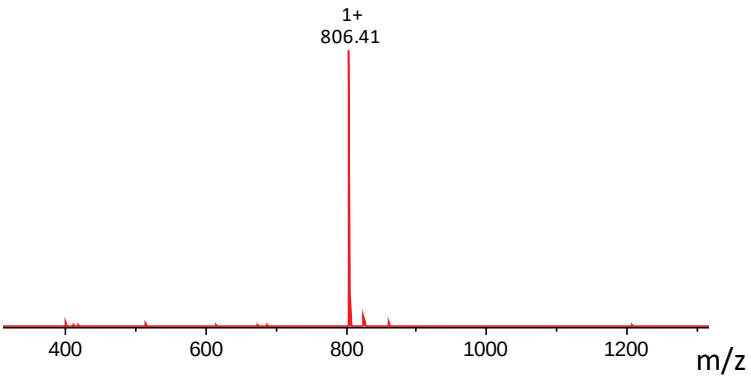
LC-MS: Found 1+ at 716.39, 2+ at 358.55.

AcFVAKNEG 49



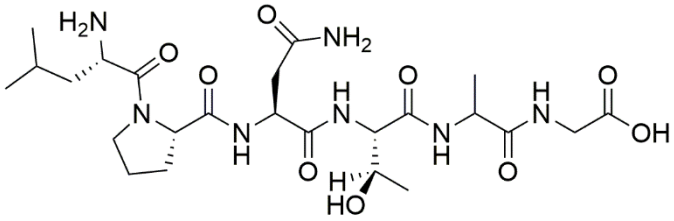
100 mg of preloaded glycine resin loading 0.79, 27 mg, 0.033 mmol, 42% yield.



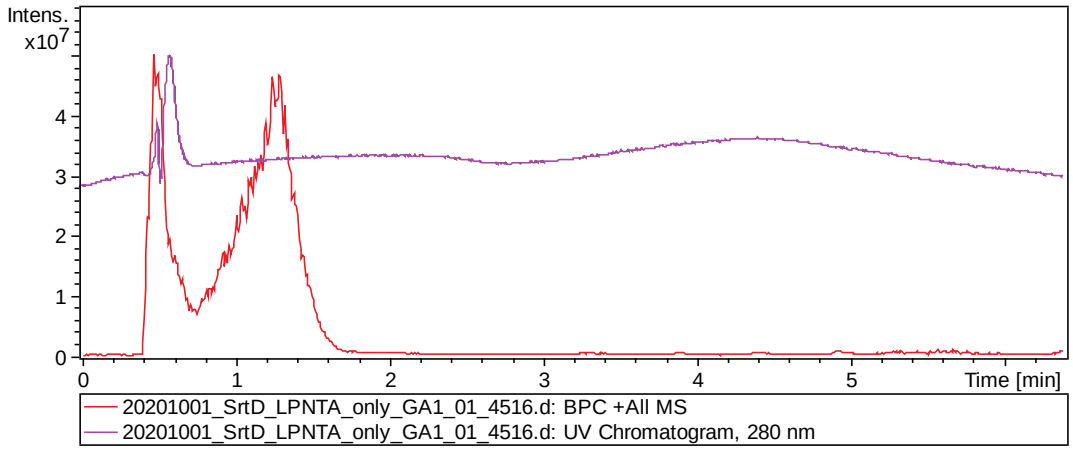


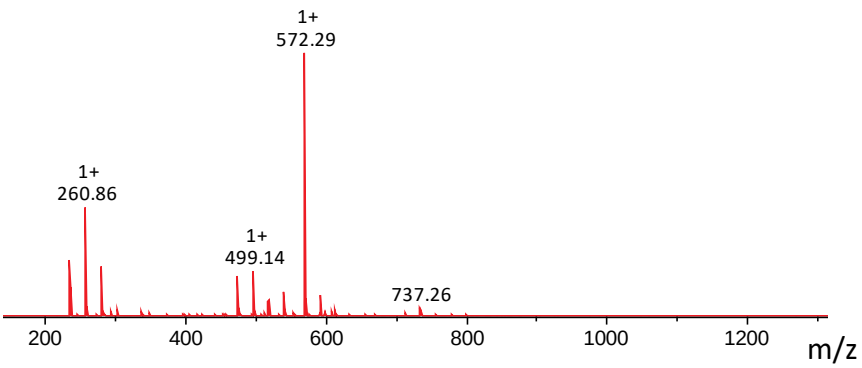
LC-MS: Found 1+ at 806.41.

LPNTAG 54



100 mg of preloaded glycine resin loading 0.79, 42 mg, 0.074 mmol, 93% yield.

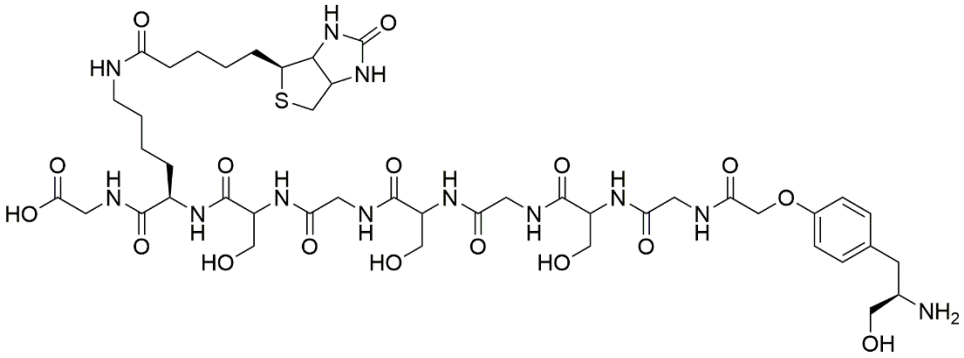




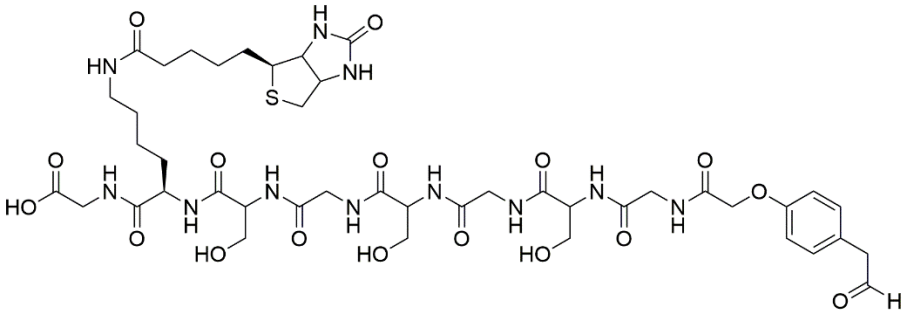
LC-MS: Found 1+ at 572.29.

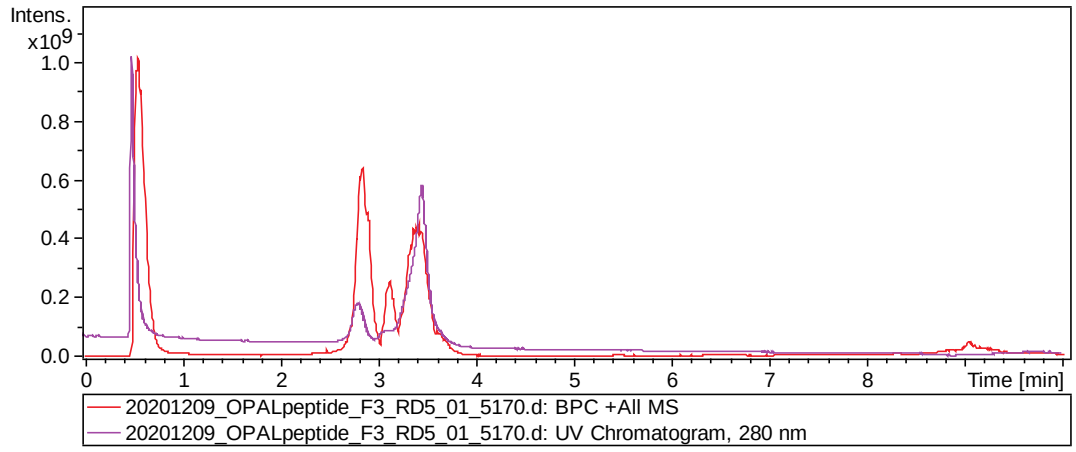
OPAL probe 37

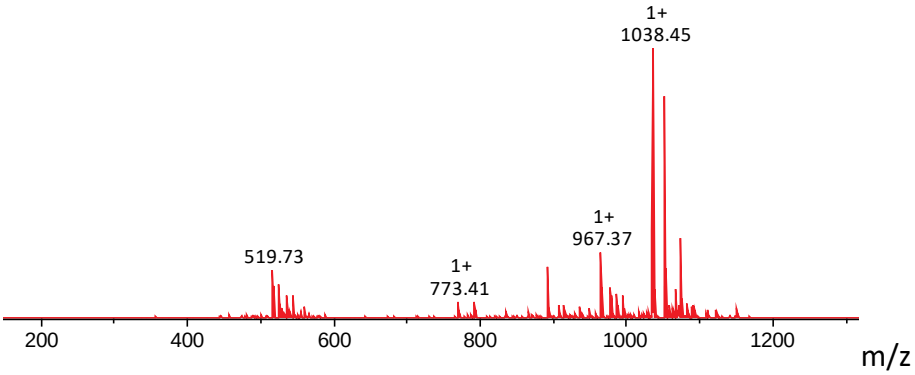
50 mg of preloaded glycine resin loading 0.79, 15.4 mg, 0.024 mmol, 61% yield.



This peptide was oxidised with sodium periodate and cleaned on a solid-phase extraction cartridge. 8.5 mg, 0.008 mmol, 34% yield. **LC-MS**: Found 1+ at 1038.45.



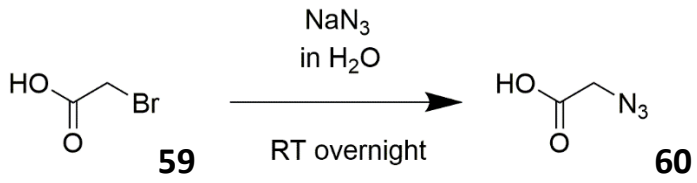




Chemical synthesis

NMR was performed using a Jeol ECS 400 instrument with Jeol Delta software and the results were analysed using MestReNova software.

2-azidoacetic acid [217]



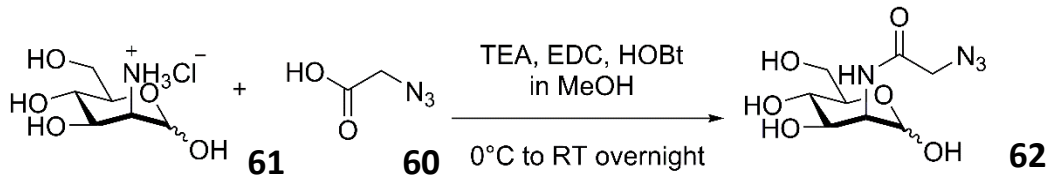
Following the procedure [218]: To a cooled (0°C) solution of 6.95g (100 mmol, 2 equiv) sodium azide in H_2O (30 mL), 7.15g (50 mmol, 1 equiv) bromoacetic acid **59** was added in parts over a 10 min period. The reaction was allowed to slowly warm to room temperature overnight and then acidified to pH 1 with conc. HCl. The product was extracted using Et_2O (5 x 10 mL), and the organic layers combined, dried over $MgSO_4$ and concentrated *in vacuo* affording 5.1g of azidoacetic acid **60** (98% yield).

HRESI-MS: Found $[M-H]^-$ 100.0155, $C_2H_2N_3O_2$ requires 100.0152.

NMR: 3.96 (s, 2H, CH_2) and 11.89 (s, 1H, COOH).

Synthesis of N-azidoacetyl-D-mannosamine AcManGN$_3$ 4 [68]

ManGN$_3$ 62



D-mannosamine hydrochloride **61** (50 mg, 0.232 mmol) was added to 5 mL dry methanol and stirred at RT until the sugar dissolved (10 min). Azidoacetic acid **60** (26 µL, 0.348 mmol) and triethylamine (TEA) (65 µL, 0.464 mmol) were added and the reaction mixture stirred for 10 min while being cooled to 0°C on ice. 1-HOBt (22 mg, 0.162 mmol) and EDC (31 mg, 0.162 mmol) were added and the reaction mixture stirred overnight (16 h), allowing it to slowly warm to RT as the ice melted. The mixture was concentrated to afford a crude orange syrup. The crude syrup was purified by flash chromatography (silica, 9:1 (v/v) DCM-methanol, R_F 0.18) to afford ManGN$_3$ **62** as a yellow foam (46.5 mg, 76%).

HRESI-MS: Found $[M+Na]^+$ 285.0818, $C_8H_{14}N_4NaO_6$ requires 285.0811.

IR (ν_{max}/cm^{-1}): 3272 (OH), 2105 (N_3).

$[\alpha]_D$ +3.0.

Major Diastereomer

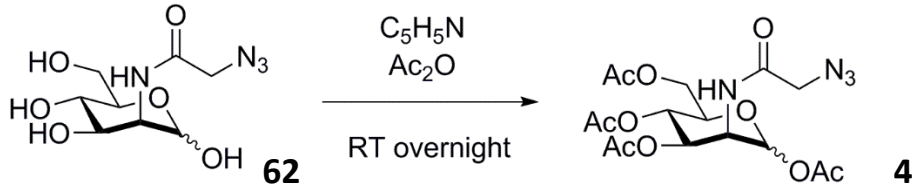
δ_H (400 MHz, $CDCl_3$); 5.24 (d, 1H, $J_{2,NH}$ 5.5 Hz, NH), 5.01 (d, 1H, $J_{1,2}$ 1.4 Hz, H-1), 4.28 (dd, 1H, $J_{2,3}$ 4.6 Hz, $J_{1,2}$ 1.4 Hz, H-2), 4.00 (dd, 1H, $J_{3,4}$ 9.6 Hz, $J_{2,3}$ 4.6 Hz, H-3), 3.92 (s, 2H, CH_2), 3.86-3.78 (m, H-6, H-6'), 3.75 (ddd, 1H, $J_{4,5}$ 9.6 Hz, $J_{5,6}$ 3.8 Hz, $J_{5,6'}$ 2.7 Hz, H-5), 3.54 (app t, 1H, $J_{3,4}$ 9.6 Hz, $J_{4,5}$ 9.6 Hz, H-4).

δ_C (101 MHz, MeOH-D_4); 171.0, 170.2 (C=O), 94.8, 94.8 (C-1), 78.5, 74.4, 73.6, 70.6, 69.5, 68.7, 68.3, 62.4 (C-3, C-4, C-5, C-6), 56.0, 55.3 (C-2), 52.8, 52.8 (CH_2-N_3).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 5.15 (d, 1H, $J_{2,NH}$ 1.5 Hz, NH), 4.90 (H-1 in solvent peak), 4.41 (dd, 1H, $J_{2,3}$ 4.3 Hz, $J_{1,2}$ 1.2 Hz, H-2), 3.90 (s, 2H, CH_2), 3.86-3.78 (m, H-6, H-6'), 3.66 (dd, 1H, $J_{3,4}$ 9.6 Hz, $J_{2,3}$ 4.3 Hz, H-3), 3.43 (app t, $J_{3,4}$ 9.6 Hz, $J_{4,5}$ 9.6 Hz, H-4), 3.24 (ddd, 1H, $J_{4,5}$ 9.6 Hz, $J_{5,6}$ 4.5 Hz, $J_{5,6'}$ 2.7 Hz, H-5).

AcManGN$_3$ 4



Acetic anhydride (1.4 mL, 12.7 mmol) was added to a solution of $ManGN_3$ **62** (46.5 mg, 0.177 mmol) in pyridine (1.4 mL) and stirred at RT for 14 h. The reaction mixture was concentrated by co-evaporating with toluene and dissolved in 4.67 mL ethyl acetate. The organic layer was washed with 1 M HCl (0.93 mL), aq. $NaHCO_3$ (0.93 mL), aq. NaCl (0.93 mL), dried (Na_2SO_4) and concentrated to afford a crude orange oil. The crude oil was purified by flash chromatography (silica, 7:3→3:2 (v/v) hexane-ethyl acetate, R_F 0.14) to afford $AcManGN_3$ **4** as a colourless foam (29 mg, 38%).

HRESI-MS: Found $[M+Na]^+$ 453.1220, $C_{16}H_{22}N_4NaO_{10}$ requires 453.1228.

IR (ν_{max}/cm^{-1}): 2108 (N_3), 1741 (C=O).

[α]$_D$ +16.4.

Major Diastereomer [219]

δ$_H$ (400 MHz, $CDCl_3$); 6.57 (d, 1H, $J_{2,NH}$ 9.3 Hz, NH), 6.05 (d, 1H, $J_{1,2}$ 1.8 Hz, H-1), 5.34 (dd, 1H, $J_{3,4}$ 10.1 Hz, $J_{2,3}$ 4.2 Hz, H-3), 5.22 (app t, 1H, $J_{3,4}$ 10.1 Hz, $J_{4,5}$ 10.1 Hz, H-4), 4.62 (ddd, 1H, $J_{2,NH}$ 9.3 Hz, $J_{2,3}$ 4.2 Hz, $J_{1,2}$ 1.8 Hz, H-2), 4.23 (dd, 1H, $J_{6,6'}$ 5.6 Hz , $J_{5,6}$ 4.1 Hz, H-6), 4.16-4.00 (m, 7H, H-6'), 4.07 (s, 2H, CH_2-N_3), 4.04 (ddd, 1H, $J_{4,5}$ 10.1 Hz, $J_{5,6}$ 4.1 Hz, $J_{5,6'}$ 2.4 Hz, H-5), 2.19 (s, 3H, C(O)CH_3), 2.12 (s, 3H, C(O)CH_3), 2.06 (s, 3H, C(O)CH_3), 2.00 (s, 3H, C(O)CH_3).

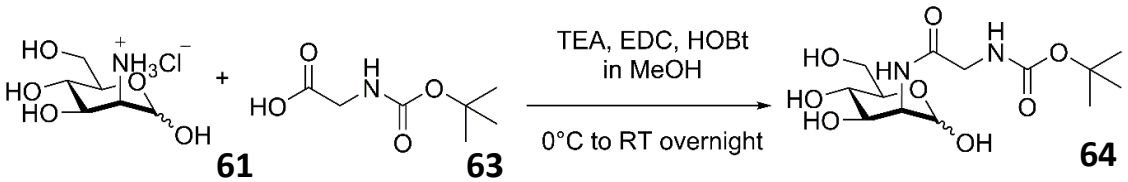
δ$_C$ (101 MHz, $CDCl_3$); 170.7, 170.3, 170.3, 169.7, 169.7, 168.5, 168.3, 167.5 (C(O)CH_3), 166.9, 166.9 (C(O)NH), 91.4, 90.4 (C-1), 73.5, 71.6, 70.4, 69.0, 65.2, 65.0, 61.9, 61.8, (C-3, C-4, C-5, C-6), 52.7 (C-2), 52.6, 49.9 (CH_2-N_3), 49.4 (C-2), 21.0, 20.9, 20.9, 20.8, 20.8, 20.8, 20.8, 20.7 (C(O)CH_3).

Minor Diastereomer

δ$_H$ (400 MHz, $CDCl_3$); 6.65 (d, 1H, $J_{2,NH}$ 9.0 Hz, NH), 5.89 (d, 1H, $J_{1,2}$ 1.6 Hz, H-1), 5.17 (app t, 1H, $J_{3,4}$ 9.8 Hz, $J_{4,5}$ 9.8 Hz, H-4), 5.06 (dd, 1H, $J_{3,4}$ 9.8 Hz, $J_{2,3}$ 3.9 Hz, H-3), 4.73 (ddd, 1H, $J_{2,NH}$ 9.0 Hz, $J_{2,3}$ 3.9 Hz, $J_{1,2}$ 1.6 Hz, H-2), 4.26 (dd, 1H, $J_{6,6'}$ 5.7 Hz , $J_{5,6}$ 4.5 Hz, H-6), 4.16-4.00 (m, 7H, H-6'), 4.04 (s, 2H, CH_2-N_3), 3.82 (ddd, 1H, $J_{4,5}$ 9.8 Hz, $J_{5,6}$ 4.5 Hz, $J_{5,6'}$ 2.5 Hz, H-5), 2.12 (s, 3H, C(O)CH_3), 2.12 (s, 3H, C(O)CH_3), 2.06 (s, 3H, C(O)CH_3), 2.01 (s, 3H, C(O)CH_3).

Synthesis of AcManGGN$_3$ 26

ManGBoc 64



61 **63** **64**

D-mannosamine hydrochloride **61** (100 mg, 0.464 mmol) was added to 20 mL dry methanol and stirred at RT until the sugar dissolved (10 min). Boc-Gly-OH **63** (122 mg, 0.696 mmol) and TEA (130 µL, 0.928 mmol) were added and the reaction mixture stirred for 10 min while being cooled to 0°C on ice. 1-HOBt (44 mg, 0.325 mmol) and EDC (62 mg, 0.325 mmol) were added and the reaction mixture stirred overnight (16 hr), allowing it to slowly warm to RT as the ice melted. The mixture was concentrated to

afford a crude yellow syrup. The crude syrup was purified by flash chromatography (silica, 9:1→7:1 (v/v) DCM-methanol, R_F 0.12) to afford ManGBoc **64** as a white foam (94 mg, 60%).

HRESI-MS: Found $[M+Na]^+$ 359.1430, $C_{13}H_{24}N_2NaO_8$ requires 359.1425.

IR (ν_{max}/cm^{-1}): 3365 (OH), 1686 (C=O).

$[\alpha]_D$ +2.3.

Major Diastereomer

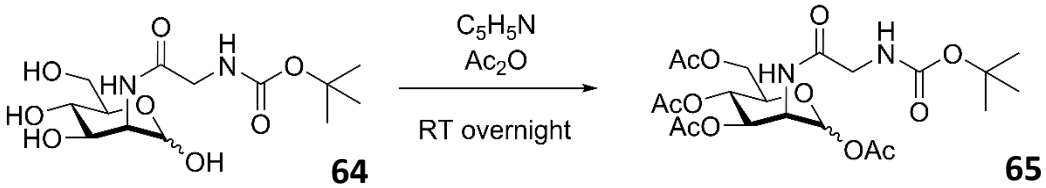
δ_H (400 MHz, $CDCl_3$); 5.20 (d, 1H, $J_{2,NHx}$ 5.1 Hz, NH_x), 5.00 (d, 1H, $J_{1,2}$ 1.3 Hz, H-1), 4.25 (dd, 1H, $J_{2,3}$ 4.6 Hz, $J_{1,2}$ 1.3 Hz, H-2), 3.96 (dd, 1H, $J_{3,4}$ 9.4 Hz, $J_{2,3}$ 4.6 Hz, H-3), 3.80-3.66 (m, 2H, H-6, H-6'), 3.76 (ddd, 1H, $J_{4,5}$ 9.4 Hz, $J_{5,6}$ 3.2 Hz, $J_{5,6'}$ 1.9 Hz, H-5), 3.51 (app t, 1H, $J_{3,4}$ 9.4 Hz $J_{4,5}$ 9.4 Hz, H-4), 1.42 (s, 9H, CH_3), 1.29 (t, 2H, $J_{CH2,NHy}$ 7.3 Hz, CH_2).

δ_C (101 MHz, $CDCl_3$); 174.1, 174.0, 173.1, 173.0 (C=O), 95.1, 94.9 (C-1), 80.9, 78.6, 74.7, 73.7, 70.8, 68.9 (C-3, C-4, C-5), 62.5, 62.4 (C-6), 56.2, 55.4 (C-2), 55.3, 55.0 (CH_2), 28.9 ($\underline{C}(CH_3)_3$), 9.4 ($(CH_3)_3$).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 5.24 (d, 1H, $J_{2,NH}$ 1.1 Hz, NH_x), 4.84 (H-1 in solvent peak), 4.37 (dd, 1H, $J_{2,3}$ 4.1 Hz, $J_{1,2}$ 1.0 Hz, H-2), 3.80-3.66 (m, 2H, H-6, H-6'), 3.62 (dd, 1H, $J_{3,4}$ 9.6 Hz, $J_{2,3}$ 4.1 Hz, H-3), 3.41 (app t, 1H, $J_{3,4}$ 9.6 Hz, $J_{4,5}$ 9.6 Hz, H-4), 3.23 (ddd, 1H, $J_{4,5}$ 9.6 Hz, $J_{5,6}$ 4.5 Hz, $J_{5,6'}$ 2.3 Hz, H-5), 1.42 (s, 9H, CH_3), 1.29 (t, 2H, $J_{CH2,NHy}$ 7.3 Hz, CH_2).

AcManGBoc 65



Acetic anhydride (2.2 mL, 23.3 mmol) was added to a solution of ManGBoc **64** (84 mg, 0.250 mmol) in pyridine (2.2 mL) and stirred at RT for 14 h. The reaction mixture was concentrated by co-evaporating with toluene to afford a yellow solid and dissolved in 7.35 mL ethyl acetate. The organic layer was washed with 1 M HCl (1.5 mL), aq. $NaHCO_3$ (1.5 mL), aq. NaCl (1.5 mL), dried (Na_2SO_4) and concentrated to afford a yellow oil. The

crude oil was purified by flash chromatography (silica, 7:3→3:2 (v/v) hexane-ethyl acetate, R_F 0.12) to afford AcManGBoc **65** as a yellow foam (95.4 mg, 76%).

HRESI-MS: Found $[M+Na]^+$ 527.1852, $C_{21}H_{32}N_2NaO_{12}$ requires 527.1848.

IR (ν_{max}/cm^{-1}): 1746 (C=O), 1219 (C-O).

$[\alpha]_D$ +14.4.

Major Diastereomer

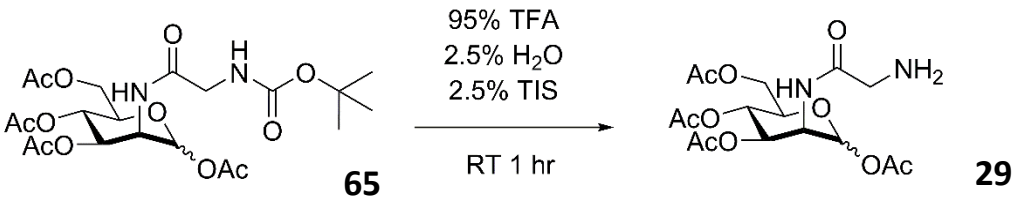
δ_H (400 MHz, $CDCl_3$); 6.90 (d, 1H, $J_{2,NHx}$ 9.2 Hz, NH_x), 6.22 (d, 1H, $J_{CH2,NHy}$ 5.5 Hz, NH_y), 6.00 (d, 1H, $J_{1,2}$ 1.4 Hz, H-1), 5.32 (dd, 1H, $J_{3,4}$ 10.1 Hz, $J_{2,3}$ 4.6 Hz, H-3), 5.17 (app t, 1H, $J_{3,4}$ 10.1 Hz, $J_{4,5}$ 10.1 Hz, H-4), 4.61 (ddd, 1H, $J_{2,NHx}$ 9.2 Hz, $J_{2,3}$ 4.6 Hz, $J_{1,2}$ 1.4 Hz, H-2), 4.54 (dd, 1H, $J_{6,6'}$ 5.8 Hz, $J_{5,6'}$ 2.5 Hz, H-6'), 4.23 (m, 1H, H-6), 4.01 (ddd, 1H, $J_{4,5}$ 10.1 Hz, $J_{5,6}$ 4.8 Hz, $J_{5,6'}$ 2.5 Hz, H-5), 3.82 Hz (d, 2H, $J_{CH2\text{-}NHy}$ 5.5 Hz, $C\underline{H}_2\text{-}NH_y$), 2.16 (s, 3H, $C(O)CH_3$), 2.10 (s, 3H, $C(O)CH_3$), 2.04 (s, 3H, $C(O)CH_3$), 1.99 (s, 3H, $C(O)CH_3$), 1.45 (s, 9H, CH_3).

δ_C (101 MHz, $CDCl_3$); 170.8, 170.8, 170.3, 170.2 ($\underline{C}(O)CH_3$), 169.7, 168.3 ($\underline{C}(O)NH$), 91.7, 90.6 (C-1), 73.5, 71.31, 70.3, 69.0, 65.6, 62.1 (C-3, C-4, C-5), 62.1, 60.5 (C-6), 49.6, 49.4 (C-2), 44.6, 44.6 (CH_2-NH), 28.4 ($\underline{C}(CH_3)_3$), 21.0, 20.8, 20.8 ($(CH_3)_3$).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 6.71 (d, 1H, $J_{2,NHx}$ 9.5 Hz, NH_x), 6.09 (d, 1H, $J_{CH2,NHy}$ 5.2 Hz, NH_y), 5.85 (d, 1H, $J_{1,2}$ 1.6 Hz, H-1), 5.11 (app t, 1H, $J_{3,4}$ 9.7 Hz, $J_{4,5}$ 9.7 Hz, H-4), 5.05 (dd, 1H, $J_{3,4}$ 9.7 Hz, $J_{2,3}$ 3.8 Hz, H-3), 4.75 (ddd, 1H, $J_{2,NHx}$ 9.5 Hz, $J_{2,3}$ 3.8 Hz, $J_{1,2}$ 1.6 Hz, H-2), 4.28 (dd, 1H, $J_{6,6'}$ 6.8 Hz, $J_{5,6}$ 2.5 Hz, H-6'), 4.11 (m, 1H, H-6), 3.85 Hz (d, 2H, $J_{CH2,NHy}$ 5.2 Hz, $C\underline{H}_2$-NH), 3.78 (ddd, 1H, $J_{4,5}$ 9.7 Hz, $J_{5,6}$ 5.3 Hz, $J_{5,6'}$ 2.5 Hz, H-5), 2.16 (s, 3H, $C(O)CH_3$), 2.09 (s, 3H, $C(O)CH_3$), 2.03 (s, 3H, $C(O)CH_3$), 2.00 (s, 3H, $C(O)CH_3$), 1.45 (s, 9H, CH_3).

AcManG 29



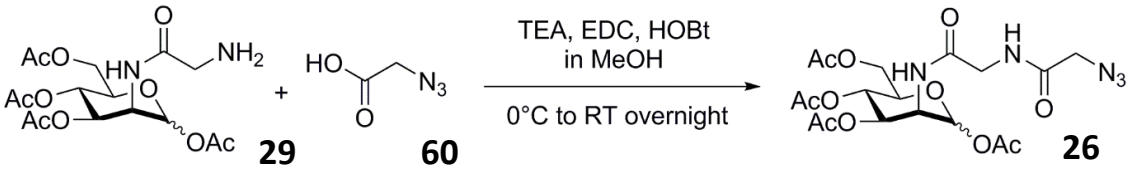
65 **29**

2 mL of cleavage cocktail containing 95% TFA, 2.5% H_2O, 2.5% TIS was added to AcManGBoc and the reaction stirred at RT for 1 h. The cleavage cocktail was evaporated under reduced pressure and AcManG **29** carried to the next step without purification.

HRESI-MS: Found $[M+Na]^+$ 427.1307, $C_{16}H_{24}N_2NaO_{10}$ requires 427.1323.

IR (ν_{max}/cm^{-1}): 1672 (C=O).

AcManGGN$_3$ 26



29 **60** **26**

AcManG **29** (38 mg, 0.094 mmol) was added to 2 mL dry methanol and stirred at RT until the sugar dissolved (10 min). Azidoacetic acid **60** (10.5 µL, 0.141 mmol) and TEA (26.5 µL, 0.188 mmol) were added and the reaction mixture stirred for 10 min while being cooled to 0°C on ice. 1-HOBt (9 mg, 0.0658 mmol) and EDC (13 mg, 0.0658 mmol) were added and the reaction mixture stirred overnight (16 h), allowing it to slowly warm to RT as the ice melted. The reaction mixture was concentrated by rotary evaporation to a crude orange oil and purified by flash chromatography (silica, 9:1 (v/v) DCM-methanol, R_F 0.6) to afford AcManGGN$_3$ **26** as an off-white foam (25 mg, 55%).

HRESI-MS: Found $[M+Na]^+$ 510.1426, $C_{18}H_{25}N_5NaO_{11}$ requires 510.1443.

IR (ν_{max}/cm^{-1}): 2108 (N_3), 1741 (C=O).

$[\alpha]_D$ +7.2.

Major Diastereomer

δ_H (400 MHz, $CDCl_3$); 7.08 (d, 1H, $J_{2,NH}$ 9.5 Hz, NH), 6.04 (d, 1H, $J_{1,2}$ 1.8 Hz, H-1), 5.31 (dd, 1H, $J_{2,3}$ 4.4 Hz, $J_{3,4}$ 9.7 Hz, H-3), 5.17 (app t, 1H, $J_{3,4}$ 9.7 Hz, $J_{4,5}$ 9.7 Hz, H-4), 4.62 (ddd, 1H, $J_{2,NH}$ 9.5 Hz, $J_{2,3}$ 4.4 Hz, $J_{1,2}$ 1.8 Hz, H-2), 4.29 (dd, 1H, $J_{6,6'}$ 5.5 Hz, $J_{5,6'}$ 2.8 Hz, H-6'), 4.22-4.11 (m, 1H, H-6), 4.07-4.02 (m, 2H, CH_2-N_3), 4.05 (ddd, 1H, $J_{4,5}$ 9.7 Hz, $J_{5,6}$ 6.7 Hz, $J_{5,6'}$ 2.8 Hz, H-5), 3.92 (d, 2H, $J_{CHx,NH}$ 5.3 Hz, CH_2-NH), 2.18 (s, 3H, C(O)CH_3), 2.12 (s, 3H, C(O)CH_3), 2.05 (s, 3H, C(O)CH_3), 1.99 (s, 3H, C(O)CH_3).

δ_C (101 MHz, $CDCl_3$); 171.1, 170.9, 170.2, 170.2, 169.8, 169.3, 169.0, 168.6 (C(O)CH_3), 168.4, 168.3, 168.2, 168.0 (C(O)NH), 91.7, 87.3 (C-1), 73.8, 71.4, 70.4, 69.2, 65.6, 65.5,

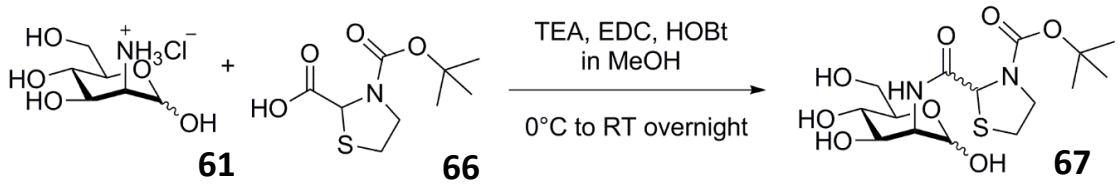
62.1, 62.1 (C-3, C-4, C-5, C-6), 52.5 (C-2), 52.5 (CH_2-N_3), 49.8 (C-2), 49.6 (CH_2-N_3), 43.6, 43.6 (CH_2-NH), 21.0, 21.0, 20.9, 20.9, 20.8, 20.8, 20.8, 20.8 (C(O)$\underline{C}H_3$).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 6.82 (d, 1H, $J_{2,NH}$ 9.4 Hz, NH), 5.86 (d, 1H, $J_{1,2}$ 1.8 Hz, H-1), 5.11 (app t, 1H, $J_{3,4}$ 9.5 Hz, $J_{4,5}$ 9.5 Hz, H-4), 5.05 (dd, 1H, $J_{3,4}$ 9.5 Hz, $J_{2,3}$ 3.9 Hz, H-3), 4.76 (ddd, 1H, $J_{2,NH}$ 9.4 Hz, $J_{2,3}$ 3.9 Hz, $J_{1,2}$ 1.8 Hz, H-2), 4.33 (dd, 1H, $J_{6,6'}$ 5.3 Hz, $J_{5,6'}$ 2.3 Hz, H-6'), 4.22-4.11 (m, 1H, H-6), 3.96 (d, 2H, $J_{CHy,NH}$ 5.3 Hz, CH_2-N_3), 3.88 (d, 2H, $J_{CHx,NH}$ 5.3 Hz, CH_2-NH), 3.81 (ddd, 1H, $J_{4,5}$ 9.5 Hz, $J_{5,6}$ 5.3 Hz, $J_{5,6'}$ 2.3 Hz, H-5), 2.12 (s, 3H, C(O)CH_3), 2.11 (s, 3H, C(O)CH_3), 2.06 (s, 3H, C(O)CH_3), 2.01 (s, 3H, C(O)CH_3).

Synthesis of AcManThz 27

ManThzBoc



The reaction shown followed a protocol from [68]. D-mannosamine hydrochloride **61** (200 mg, 0.928 mmol) was added to 20 mL dry methanol and stirred at RT until the sugar dissolved (10 min). Boc-thiazolidine **66** (325 mg, 1.391 mmol, kindly provided by Robin Brabham) and trimethylamine (TEA) (260 µl, 1.855 mmol) were added. The reaction was stirred and cooled to 0°C for 10 min, then 1-hydroxybenzotriazole (1-HOBt) (88 mg, 0.649 mmol) and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (125 mg, 0.649 mmol) were added. The reaction was stirred overnight, slowly warming to RT as the ice in the ice bath melted.

Completion of the reaction was checked using TLC with product detected by sugar stain (5% sulphuric acid in ethanol). The mixture was concentrated to afford a crude syrup and purified using flash chromatography (silica, 9:1 (v/v) DCM-methanol, R_F = 0.37) to afford ManThzBoc **67** as a white foam (190 mg, 52%).

HRESI-MS: Found $[M+Na]^+$ 417.1300, $C_{15}H_{26}N_2NaO_8S$ requires 417.1302.

IR (ν_{max}/cm^{-1}): 2978 (OH), 1691 (C=O).

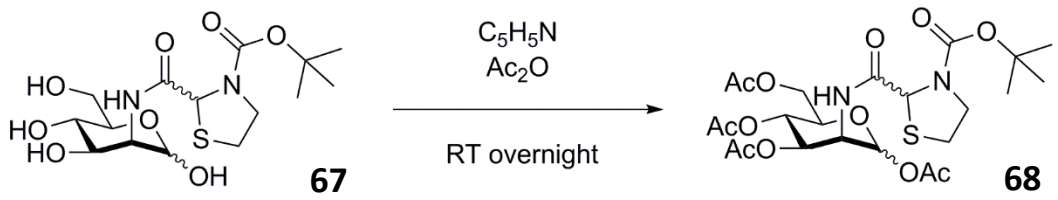
$[\alpha]_D$ +3.3.

Major Diastereomer

δ_H (400 MHz, MeOH-D_4); 5.07 (s, 1H, S-CH-N), 5.00 (d, 1H, $J_{1,2}$ 0.8 Hz, H-1), 4.20 (dd, 1H, $J_{2,3}$ 4.6 Hz, $J_{1,2}$ 1.4 Hz, H-2), 4.02 (dd, 1H, $J_{3,4}$ 9.7 Hz, $J_{2,3}$ 4.7 Hz, H-3), 3.90-3.72 (m, 5H, H-5, H-6, H-6', SCH_2) 3.54 (app t, 1H, $J_{3,4}$ 9.4 Hz, $J_{4,5}$ 9.4 Hz, H-4), 3.01 (m, 2H, NCH_2), 1.44 (s, 9H, CH_3).

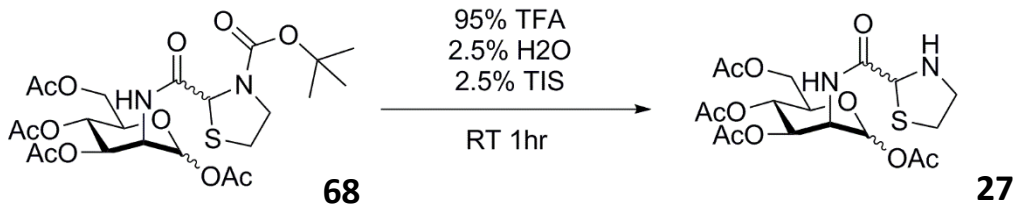
δ_C (101 MHz, MeOH-D_4); 176.2, 176.1 (O=C-O), 155.0, 155.0 (N-C=O), 95.0, 94.7 (C-1), 82.7, 82.3, 82.2, 82.1, 62.3, 61.2, (C-3, C-4, C-5, C-6), 55.7, 55.0 (C-2), 51.2, 51.1, (N-CH_2), 31.4, 30.3, (S-C-N), 28.8, 28.7, (S-CH_2), 28.7, 28.7 ($\underline{C}(CH_3)_3$), 9.3 ((CH_3)$_3$).

AcManThzBoc 67



A solution of ManThzBoc **67** (190 mg, 0.482 mmol), dry pyridine (3.8 mL) and acetic anhydride (3.8 mL) was stirred overnight at RT. The pyridine was removed by co-evaporating with toluene and the remaining residue dissolved in ethyl acetate (12.7 mL) and transferred to a separation funnel. The mixture was washed successively with 1 M aqueous HCl (3.8 mL), saturated aqueous $NaHCO_3$ (3.8 mL), saturated aqueous NaCl (3.8 mL), dried with anhydrous sodium sulfate and evaporated to form AcManThzBoc **68** as a yellow foam, which was taken to the next step without purification.

AcManThz 27



A solution of 5 mL cleavage cocktail containing 95% TFA, 2.5% H_2O, 2.5% TIS was added to AcManThzBoc **68** (183.4 mg) and the reaction stirred at RT for 1 h. The reaction mixture was purified using flash chromatography (silica, 9:1 (v/v) DCM-methanol, R_F = 0.7) to afford AcManThz **27** as a yellow foam (37.1 mg, 17% over two steps).

HRESI-MS: Found $[M+Na]^+$ 485.1199, $C_{18}H_{26}N_2NaO_{10}S$ requires 485.1200.

IR (v_{max}/cm^{-1}): 1743 (C=O), 1674 (C=O).

$[\alpha]_D$ +19.76.

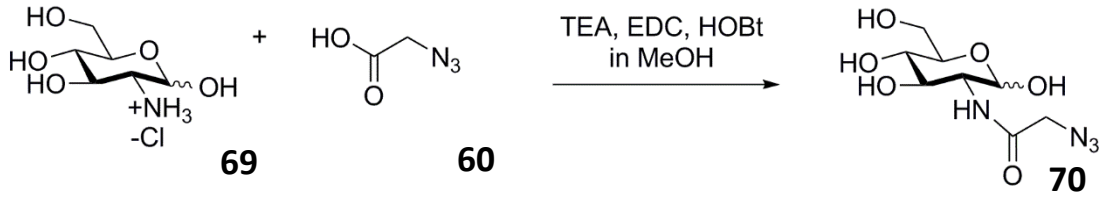
Major Diastereomer

δ_H (400 MHz, $CDCl_3$); 6.06 (s, 1H, H-1), 5.88 (s, 1H, S-C<u>H</u>-N), 5.30-5.00 (m, 2H, H-3, H-4), 4.59 (m, 1H, H-2), 4.30-4.15 (m, 2H, H-5, H-6), 4.12-3.65 (m, 2H, H-6', SCH_2), 3.23 (m, 2H, NCH_2), 2.17 (s, 3H, $C(O)CH_3$), 2.09 (s, 3H, $C(O)CH_3$), 2.07 (s, 3H, $C(O)CH_3$), 1.98 (s, 3H, $C(O)CH_3$).

δ_C (101 MHz, $CDCl_3$); 170.9, 170.8, 170.5, 170.4, 169.8, 169.3, 169.0, 168.8, (<u>C</u>$(O)CH_3$), 168.6, 168.5, (C(O)NH), 91.4, 90.6 (C-1), 70.5, 70.4, 69.2, 66.4, 65.6, 65.5, 62.3, 62.4 (NH-<u>C</u>H_2), 60.7, (C-3, C-4, C-5, C-6), 58.6 (NH-<u>C</u>H_2), 53.6, 50.0 (C-2), 32.0, 29.8 (S-CH_2), 29.7, 29.4 (S-C-N), 21.0, 21.0, 20.8, 20.8, 20.7, 20.7, 20.6, 20.6 (C(O)<u>C</u>H_3).

Synthesis of N-azidoacetyl-D-glucosamine $AcGlcGN_3$ 6 [68]

$GlcGN_3$ 70



69 **60** **70**

D-glucosamine hydrochloride **69** (200 mg, 0.928 mmol) was added to 20 mL dry methanol and stirred at RT until the sugar dissolved (10 min). Azidoacetic acid **60** (208 µL, 2.783 mmol) and TEA (260 µL, 1.855 mmol) were added and the reaction mixture stirred for 10 min while being cooled to 0°C on ice. 1-HOBt (87.7 mg, 0.649 mmol) and EDC (124.5 mg, 0.649 mmol) were added and the reaction mixture stirred overnight (16 hr), allowing it to slowly warm to RT as the ice melted. The mixture was concentrated to afford a crude syrup. The crude syrup was purified by flash chromatography (silica, 9:1 (v/v) DCM-methanol) to afford $GlcGN_3$ **70** (197 mg, 81%) as a pale yellow foam.

HRESI-MS: Found $[M+Na]^+$ 285.0806, $C_8H_{14}N_4NaO_6$ requires 285.0806.

IR (v_{max}/cm^{-1}): 2980 (OH), 2104 (N_3).

$[\alpha]_D$ +13.9.

Major Diastereomer

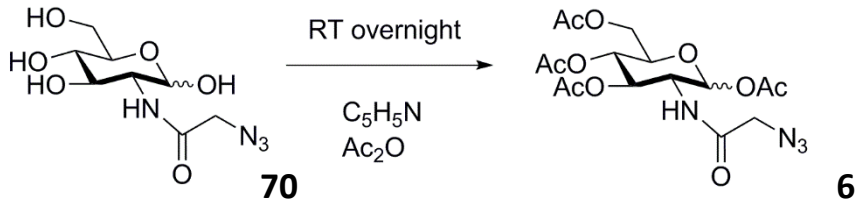
δ_H (400 MHz, $CDCl_3$); 5.45 (d, 1H, $J_{2,NH}$ 1.9 Hz, NH), 5.06 (d, 1H, $J_{1,2}$ 3.4 Hz, H-1), 3.88 (d, 1H, $J_{3,4}$ 4.7 Hz, H-3), 3.82 (dd, $J_{2,3}$ 10.8 Hz, $J_{1,2}$ 3.4 Hz, H-2), 3.79-3.69 (m, H-6, H-6'), 3.71 (s, 2H, CH_2), 3.66 (ddd, 1H, $J_{4,5}$ 10.1 Hz, $J_{5,6}$ 4.4, $J_{5,6'}$ 1.6 Hz, H-5), 3.62 (t, 1H, $J_{3,4}$ 5.8 Hz, $J_{4,5}$ 5.8 Hz, H-4).

δ_C (101 MHz, MeOH-D_4); 174.3, 170.6 (C=O), 92.7, 89.7 (C-1), 76.0, 74.8, 73.3, 72.8, 72.5, 72.3, 62.9, 61.0 (C-3, C-4, C-5, C-6), 56.1, 53.3 (C-2), 53.0, 52.8 (CH_2-N_3).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 5.86 (d, 1H, $J_{2,NH}$ 1.7 Hz, NH), 4.60 (d, 1H, $J_{1,2}$ 8.3 Hz, H-1), 3.79-3.69 (m, H-6, H-6'), 3.72 (s, 2H, CH_2), 3.71 (d, 1H, $J_{2,3}$ 8.6 Hz, H-3), 3.57 (dd, 1H, $J_{1,2}$ 8.3 Hz, $J_{2,NH}$ 1.7 Hz, H-2), 3.46 (t, 1H, $J_{3,4}$ 8.6 Hz, $J_{4,5}$ 8.6 Hz, H-4), 3.34 (ddd, 1H, $J_{4,5}$ 8.6 Hz, $J_{5,6}$ 2.6 Hz, $J_{5,6'}$ 1.1 Hz, H-5).

$AcGlcGN_3$ 6



70 **6**

Acetic anhydride (5 mL, 45.4 mmol) was added to a solution of $GlcGN_3$ (170 mg, 0.648 mmol) in pyridine (5 mL) and stirred at RT for 14 h. The reaction mixture was concentrated by co-evaporating with toluene and dissolved in 17 mL ethyl acetate. The organic layer was washed with 1 M HCl (3.4 mL), aq. $NaHCO_3$ (3.4 mL), aq. NaCl (3.4 mL), dried (Na_2SO_4) and concentrated to afford a crude orange oil. The crude oil was purified by flash chromatography (silica, 7:3→1:1 (v/v) hexane-ethyl acetate, R_F 0.25) to afford $AcGlcGN_3$ as a white foam (14.7 mg, 5%).

HRESI-MS: Found $[M+Na]^+$ 453.1228, $C_{16}H_{22}N_4NaO_{10}$ requires 453.1228.

IR (ν_{max}/cm^{-1}): 2106 (N_3), 1743 (C=O).

$[\alpha]_D$ +20.3.

Major Diastereomer (α) [220]

δ_H (400 MHz, $CDCl_3$); 6.45 (d, 1H, $J_{2,NH}$ 8.9 Hz, NH), 6.19 (d, 1H, $J_{1,2}$ 3.7 Hz, H-1), 5.25 (dd, 1H, $J_{2,3}$ 10.8 Hz, $J_{3,4}$ 9.7 Hz, H-3), 5.21 (app t, 1H, $J_{3,4}$ 9.7 Hz, $J_{4,5}$ 9.7 Hz, H-4), 4.44 (ddd,

1H, $J_{2,3}$ 10.8 Hz, $J_{2,NH}$ 8.9 Hz, $J_{1,2}$ 3.7 Hz, H-2), 4.29 (dd, 1H, $J_{6,6'}$ 7.9 Hz, $J_{5,6}$ 4.2 Hz, H-6), 4.09 (dd, 1H, $J_{6,6'}$ 7.9 Hz, $J_{5,6'}$ 2.3 Hz, H-6'), 4.02 (ddd, 1H, $J_{4,5}$ 9.7 Hz, $J_{5,6}$ 4.2 Hz, $J_{5,6'}$ 2.3 Hz, H-5), 3.94 (s, 2H, CH_2-N_3), 2.20 (s, 3H, C(O)CH_3), 2.09 (s, 3H, C(O)CH_3), 2.05 (s, 3H, C(O)CH_3), 2.05 s, (3H, C(O)CH_3).

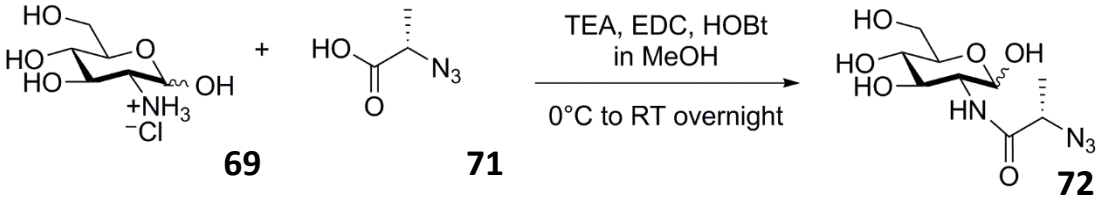
δ_C (101 MHz, $CDCl_3$); 171.7, 171.0, 170.8, 170.8, 169.5, 169.4, 169.3, 168.8 (C(O)CH_3), 167.2, 167.0 (C(O)NH), 92.3, 90.4 (C-1), 73.0, 72.3, 70.4, 69.9, 67.8, 67.5, 61.7, 61.6 (C-3, C-4, C-5, C-6), 53.3 (C-2), 52.7, 52.5 (CH_2-N_3), 51.3 (C-2), 21.0, 21.0, 20.9, 20.8, 20.8, 20.7, 20.7, 20.7 (C(O)CH_3).

Minor Diastereomer (β)

δ_H (400 MHz, $CDCl_3$); 6.52 (d, 1H, $J_{2,NH}$ 9.3 Hz, NH), 5.79 (d, 1H, $J_{1,2}$ 8.7 Hz, H-1), 5.29 (dd, 1H, $J_{2,3}$ 10.7 Hz, $J_{3,4}$ 9.6 Hz, H-3), 5.13 (app t, 1H, $J_{3,4}$ 9.6 Hz, $J_{4,5}$ 9.6 Hz, H-4), 4.27-4.18 (m, 2H, H-2, H-6), 4.11 (dd, 1H, $J_{6,6'}$ 12.7 Hz, $J_{5,6'}$ 2.3 Hz, H-6'), 3.92 (s, 2H, CH_2-N_3), 3.83 (ddd, 1H, $J_{4,5}$ 9.6 Hz, $J_{5,6}$ 4.6 Hz, $J_{5,6'}$ 2.3 Hz, H-5), 2.11 (s, 3H, C(O)CH_3), 2.09 (s, 3H, C(O)CH_3), 2.04 (s, 3H, C(O)CH_3), 2.04 (s, 3H, C(O)CH_3).

Synthesis of $AcGlcAN_3$ 28

$GlcAN_3$ 72



69 **71** **72**

Glc-HCl **69** (50 mg, 0.232 mmol) was stirred in 5 mL dry methanol until the sugar dissolved (10 min). *S*-2-azidopropionic acid **71** (140 mg, 0.653 mmol) and TEA (65 µl, 0.464 mmol) were added and the reaction mixture stirred for 10 min while being cooled to 0°C on ice. 1-HOBt (22 mg, 0.162 mmol) and EDC (31 mg, 0.162 mmol) were added and the reaction mixture stirred overnight (16 hr). The mixture was concentrated to afford a crude syrup. This was purified using flash chromatography (silica, 9:1 (v/v) DCM-methanol, R_F 0.38) to afford $GlcAN_3$ **72** as a pale yellow foam (52.3 mg, 81%).

HRESI-MS: Found $[M+Na]^+$ 467.1387, $C_{17}H_{24}N_4NaO_{10}$ requires 467.1385.

IR (ν_{max}/cm^{-1}): 2934 (OH), 2102 (N_3).

$[\alpha]_D$ +53.8.

Major Diastereomer

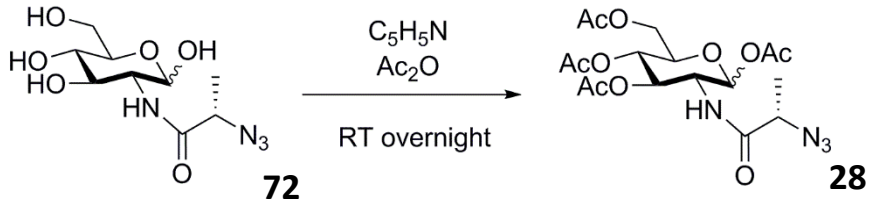
δ_H (400 MHz, $CDCl_3$); 5.02 (d, 1H, $J_{1,2}$ 3.4 Hz, H-1), 3.82-3.48 (m, H-6 and H-6'), 3.80 (app t, 1H, $J_{3,4}$ 7.0 Hz, $J_{4,5}$ 7.0 Hz, H-4), 3.76-3.69 (m, H-2), 3.74 (s, 1H, CH-N_3), 3.30 (dd, 1H, $J_{3,4}$ 8.9 Hz, $J_{2,3}$ 3.1 Hz, H-3), 3.00-2.93 (m, H-5), 1.28 (s, 3H, CH_3).

δ_C (101 MHz, MeOH-D_4); 177.6, 172.3 (C=O), 91.3, 89.5 (C-1), 78.1, 73.5, 73.3, 72.8, 72.5, 71.9, 60.9, 59.5, 56.5, 56.1 (C-2, C-3, C-4, C-5, C-6), 51.7, 26.1, 25.6 (CH-N_3), 18.1, 17.6 (CH_3).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 5.26 (d, 1H, $J_{1,2}$ 3.5 Hz, H-1), 3.82-3.48 (m, H-6 and H-6'), 3.78 (dd, 1H, $J_{3,4}$ 7.1 Hz, $J_{2,3}$ 1.9 Hz, H-3), 3.75 (s, 1H, CH-N_3), 3.68 (app t, 1H, $J_{3,4}$ 5.3 Hz, $J_{4,5}$ 5.3 Hz, H-4), 3.27-3.22 (m, H-5), 2.96 (ddd, 1H, $J_{2,3}$ 14.9 Hz, $J_{2,NH}$ 6.6 Hz, $J_{1,2}$ 3.9 Hz, H-2), 1.30 (s, 3H, CH_3).

$AcGlcAN_3$ 28



72 **28**

Acetic anhydride (2 mL, mmol) was added to a solution of $GlcAN_3$ (106.7 mg, 0.386 mmol) in pyridine (2 mL) and stirred at RT for 14 h. The reaction mixture was concentrated by co-evaporating with toluene and dissolved in 6 mL ethyl acetate. The organic layer was washed with 1 M HCl (1.2 mL), aq. $NaHCO_3$ (1.2 mL) aq. NaCl (1.2 mL), dried, and concentrated to afford a crude orange syrup. This was purified by flash chromatography (silica, 7:3 (v/v) hexane-ethyl acetate, R_F 0.80) to afford $AcGlcAN_3$ as a yellow foam (28.5 mg, 17%).

HRESI-MS: Found $[M+Na]^+$ 467.1387, $C_{17}H_{24}N_4NaO_{10}$ requires 467.1385.

IR (ν_{max}/cm^{-1}): 2106 (N_3), 1743 (C=O).

$[\alpha]_D$ +65.1.

Major Diastereomer

δ_H (400 MHz, $CDCl_3$); 6.49 (d, 1H, $J_{2,NH}$ 8.7 Hz, NH), 6.21 (d, 1H, $J_{1,2}$ 3.7 Hz, H-1), 5.27 (dd, 1H, $J_{2,3}$ 10.8 Hz, $J_{3,4}$ 9.5 Hz, H-3), 5.20 (app t, 1H, $J_{3,4}$ 9.5 Hz, $J_{4,5}$ 9.5 Hz, H-4), 4.38 (ddd, 1H, $J_{2,3}$ 10.8 Hz, $J_{2,NH}$ 8.7 Hz, $J_{1,2}$ 3.7 Hz, H-2), 4.27 (dd, 1H, $J_{5,6}$ 5.9 Hz, $J_{6,6'}$ 1.5 Hz, H-6), 4.18 (dd, 1H, $J_{5,6'}$ 10.3 Hz, $J_{6,6'}$ 1.5 Hz, H-6'), 4.02-3.97 (m, 1H, H-5), 3.77-3.67 (m, 1H, CH-N_3), 2.19 (s, 3H, C(O)CH_3), 2.08 (s, 3H, C(O)CH_3), 2.04 (s, 3H, C(O)CH_3), 2.03 (s, 3H, C(O)CH_3), 1.53 (s, 1H, CH_3-CH-N_3).

δ_C (101 MHz, $CDCl_3$); 171.7, 170.9, 170.8, 170.6, 170.3, 169.4, 169.4, 169.3 ($\underline{C}$(O)CH_3), 168.9, 168.8 (C(O)NH), 92.5, 90.3 (C-1), 73.0, 70.4, 69.9, 67.8, 67.5, 61.6, 59.4, 59.2 (C-3, C-4, C-5, C-6), 51.5, 48.3 (C-2), 25.5, 24.9 (CH-N_3), 21.0, 20.9, 20.9, 20.8, 20.8, 20.8, 20.7, 20.7 (C(O)$\underline{C}H_3$), 17.4, 17.3 ($\underline{C}H_3$-CH-N_3).

Minor Diastereomer

δ_H (400 MHz, $CDCl_3$); 6.56 (d, 1H, $J_{2,NH}$ 9.3 Hz, NH), 5.78 (d, 1H, $J_{1,2}$ 8.7 Hz, H-1), 5.26 (dd, 1H, $J_{3,4}$ 9.7 Hz, $J_{2,3}$ 2.6 Hz, H-3), 5.13 (app t, 1H, $J_{3,4}$ 9.7 Hz, $J_{4,5}$ 9.7 Hz, H-4), 4.30-4.23 (m, 1H, H-2, H-6), 4.11 (dd, 1H, $J_{5,6'}$ 12.5 Hz, $J_{6,6'}$ 2.3 Hz, H-6'), 3.81 (ddd, 1H, $J_{4,5}$ 9.7 Hz, $J_{5,6}$ 4.6 Hz, $J_{6,6'}$ 2.3 Hz, H-5), 3.77-3.67 (m, 1H, CH-N_3), 2.20 (s, 3H, C(O)CH_3), 2.08 (s, 3H, C(O)CH_3), 2.04 (s, 3H, C(O)CH_3), 2.02 (s, 3H, C(O)CH_3), 1.51 (s, 1H, CH_3-CH-N_3).

Microbiology protocols

Growth media recipes

Lysogeny broth (LB): NaCl 1% (w/v), Tryptone 1% (w/v), Yeast Extract 0.5% (w/v) in dH_2O, autoclaved.

Lysogeny broth peptone (LBP): NaCl 1% (w/v), Peptone 1% (w/v), Yeast Extract 0.5% (w/v) in dH_2O, autoclaved.

Lysogeny broth agar (LB agar): NaCl 1% (w/v), Tryptone 1% (w/v), Yeast Extract 0.5% (w/v), Agar 1.5% (w/v) in dH_2O, autoclaved. Where used, kanamycin was added to 50 µg / mL).

Super Optimal Broth with Catabolite repression (SOC): Tryptone 2% (w/v), Yeast Extract 0.5% (w/v), KCl 2.5 mM, NaCl 10 mM, $MgCl_2$ 10 mM, $MgSO_4$ 10 mM, Glucose 20 mM in dH_2O, autoclaved.

Sortase transformation into *E. coli*

The three commercially available sortase genes used (each one a gift from Hidde Ploegh) were pET28A Spy SrtA (Addgene plasmid #51139), pET30b 5M Sau SrtA (Addgene plasmid #51140) and pET30b 7M Sau SrtA (Addgene plasmid #51141). These arrived as a bacterial stab in DH5α cells. These were spread onto LB_{Kan} agar plates and incubated overnight. Colonies were used to inoculate 10 mL LB and grown for 17 h and the DNA extracted by small-scale plasmid DNA isolation (Miniprep). The QIAprep miniprep was performed according to manufacturer's instructions [221]. Sortase genes were verified by sequencing by GATC Biotech.

The *E. coli* strain BL21 (DE3) was chosen for protein production. 50 µL frozen electrocompetent [222] *E. coli* of the desired strain was thawed on ice, 1 µL of the desired plasmid (100 ng μL^{-1}) added, and the cells allowed to sit for 5 min on ice. Cells were transferred to a pre-chilled electroporation cuvette, shocked with 2.5 kilovolts using a BioRad Gene Pulse and immediately resuspended in 1 mL of SOC medium. The cells were subsequently incubated for 1 hour at 37°C, 180 rpm, then centrifuged for 3 min at 8000 rpm. 800 µL supernatant was removed and the cells resuspended in the remaining 200 µL, then spread onto LB_{Kan} agar plates and incubated overnight at 37°C.

GRASP Cloning

The Prozomigo database was trawled for sortase enzymes and the sequences of 28 desired enzymes ordered from Twist Bioscience. Synthetic DNA sequences containing these inserts were cloned into Top10 *E. coli* using the Genomics-based related activity screening protocol (GRASP). The restriction enzymes used were XhoI and NdeI (New England Biolabs, NEB), the plasmid vector pET28a (EMD Biosciences), and the buffer CutSmart (NEB). Inserts were digested for 2 h at 37°C, plasmid for 4 h at 37°C. Reactions were cleaned up with a Gelpure kit (NZYTech) according to manufacturer's instructions. Briefly, 5 volumes of binding buffer was added to the reaction mixture. This mixture was added to the spin column and the column centrifuged at 12,000 g for 1 min. The reaction mixture was again added to the spin column and centrifuged to maximise DNA binding. The flowthrough was discarded and 600 µL added to the column. This was centrifuged, the flowthrough discarded and the column centrifuged and then left for 20 min to completely remove residual ethanol. 50 µL elution buffer preheated to 55°C was added to the column, left for 1 min and the column centrifuged to elute the DNA.

Ligation reactions were performed with T4 DNA Ligase (NEB) overnight at 16°C, with insert: vector at a ratio of 2:1 or 4:1. The ligation reaction mixture was transformed into chemically competent Top10 *E. coli*. 50 µL frozen *E. coli* was thawed on ice for 10 min, 10 µL of ligation mixture added and the cells left on ice for 30 min. The cells were heat-shocked at 42°C for 2 min and placed on ice for 5 min to recover. 250 µL LBP was added and the cells incubated at 37°C, 180 rpm for 20 min then plated onto LB agar plates with the appropriate antibiotic and incubated overnight at 37°C.

To obtain plasmid from the transformed *E. coli*, 48 colonies were picked using a sterile tip and transferred to a 96-well LGC plate. The plate was incubated at 37°C overnight and sent to LGC for DNA extraction and plasmid sequencing. 25 µL of each plasmid was delivered to Prozomix, and sortase inserts were verified by sequencing. 22 of the initial 28 enzyme sequences had at least one plasmid with a verified insert.

Plasmid for each verified sequence was transformed into chemically competent BL21 (DE3) *E. coli*. 10 µL frozen *E. coli* was thawed on ice for 10 min, 0.5 µL of plasmid added and the cells left on ice for 30 min. The cells were heat-shocked at 42°C for 2 min and placed on ice for 5 min to recover. 100 µL LBP was added and the cells incubated at 37°C, 180 rpm for 60 min then plated onto LB_{Kan} agar plates and incubated overnight at 37°C.

Protein Expression

Buffer recipes

To make up 1 L, with pH adjusted using concentrated HCl or concentrated NaOH:

Tris I (pH 7.8): 25 mM Tris-HCl (3.02 g), 1.25 M NaCl (73.05 g), 25 mM $CaCl_2$ (2.78 g).

Tris II (pH 7.8): 25 mM Tris-HCl (3.02 g), 1.25 M NaCl (73.05 g), 2.5 mM EDTA (0.73 g).

Tris III (pH 7.2): 50 mM Tris-HCl (6.04 g), 150 mM NaCl (8.77 g), 10% v/v glycerol.

Tris IV (pH 7.2): 50 mM Tris-HCl (6.04 g), 150 mM NaCl (8.77 g), 10% v/v glycerol, 500 mM imidazole (34.04 g).

HEPES (pH 7.5): 50mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$.

Start buffer (pH 7.5): 10 mM HEPES, 500 mM NaCl, 10 mM imidazole.

Wash buffer (pH 7.5): 10 mM HEPES, 500 mM NaCl, 50 mM imidazole.

Elution buffer (pH 6.75): 10 mM HEPES, 500 mM NaCl, 500 mM imidazole.

gMBP Expression

A single colony of *E. coli* BL21 (DE3) transformed with gMBP TevOse (Amp) (a gift from Prof. Kurt Drickamer) was used to inoculate 5 mL LB_{Amp} and grown overnight at 37°C, 180 rpm. 1 mL of starter culture was used to inoculate 100 mL LB_{Amp} and grown overnight at 30°C. 20 mL of this was added to 1 L LB_{Amp} and incubated at 25°C for 3.5 h. Protein expression was induced by addition of IPTG (to a final concentration of 40 µM) and $CaCl_2$ (to a final concentration of 0.1 M), and culture was grown overnight at 25°C. Cells were harvested by centrifugation at 10,000 g for 15 min.

Sepharose Column Preparation

10 mL Sepharose 6B was washed with H_2O (3 x 10 mL) on a glass sinter funnel under vacuum, then put in a 50 ml falcon tube with 10 ml 0.5 M Na_2CO_3 buffer at pH 11 with 1 mL divinyl sulfone added. The mixture was spun on a rotator tube spinner for 70 min at room temperature, then returned to a glass sinter funnel and washed with H_2O (5 x 10 mL). Resin was resuspended in 10 mL 20% w/v sugar (either mannose or lactose) in 0.5 M Na_2CO_3 buffer pH 10 and rotated overnight. The mixture was again washed with H_2O (5 x 10 mL), then resuspended in 10 mL 0.5 M $NaHCO_3$ buffer at pH 8.5 and 200 µL 2-mercaptoethanol added. This was rotated for 2 h, washed in H_2O (5 x 10 mL) and the resin packed in a column with Tris I buffer.

gMBP Purification

This followed an earlier protocol from the Fascione group [223]. gMBP was purified by resuspending in 40 mL Tris I buffer with DNase and protease inhibitor tablet added, then lysed by sonication on ice using Soniprep 150 for 6 x 30 s cycles, with 30 s rest in between each cycle. Lysate was cleared by centrifugation at 35,000 g for 45 min and supernatant filtered through a 0.45 µm filter. Column purification used a 10 ml lactose-sepharose 6B affinity column for gMBP. The column was equilibrated with 5 CVs of Tris I buffer prior to supernatant loading. Bound protein was washed with 5 CVs Tris I buffer and eluted with 5 CVs Tris II buffer. Eluted protein was collected in fractions of 2 mL and checked for the protein of interest using SDS-PAGE. Fractions containing protein were pooled, dialysed 3 times in 100-fold Tris I buffer, concentrated, aliquoted and flash-frozen in liquid nitrogen to be stored at -80°C.

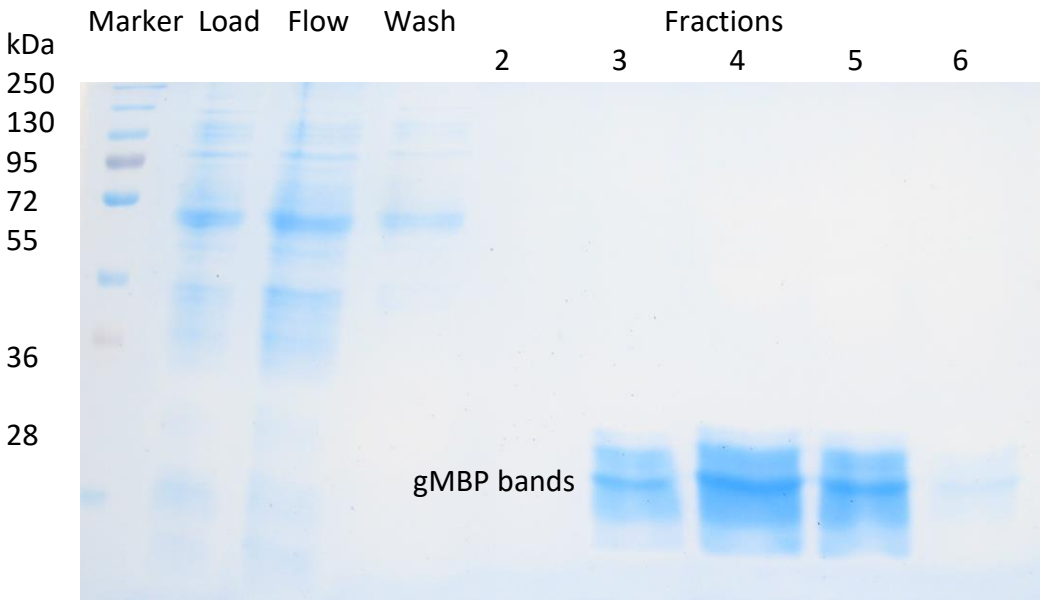


Figure 7.1: gMBP purification – fractions 3-6 were dialysed

Commercial Sortase Expression

Protein expression used a protocol adapted from the Liu group [160]. A single colony of *E. coli* BL21 (DE3) transformed with one of three sortase plasmids was used to inoculate 5 mL LB_{Kan} and grown overnight. 1 mL starter culture was added to 1 L LB_{Kan}, incubated at 37°C until OD_{600} reached 0.5 – 0.8 (about 3-3.5 hours). IPTG was added to a final concentration of 0.4 mM to induce sortase production and grown for 3 hours at 30°C. Cells were harvested by centrifugation at 10,000 g for 10 min.

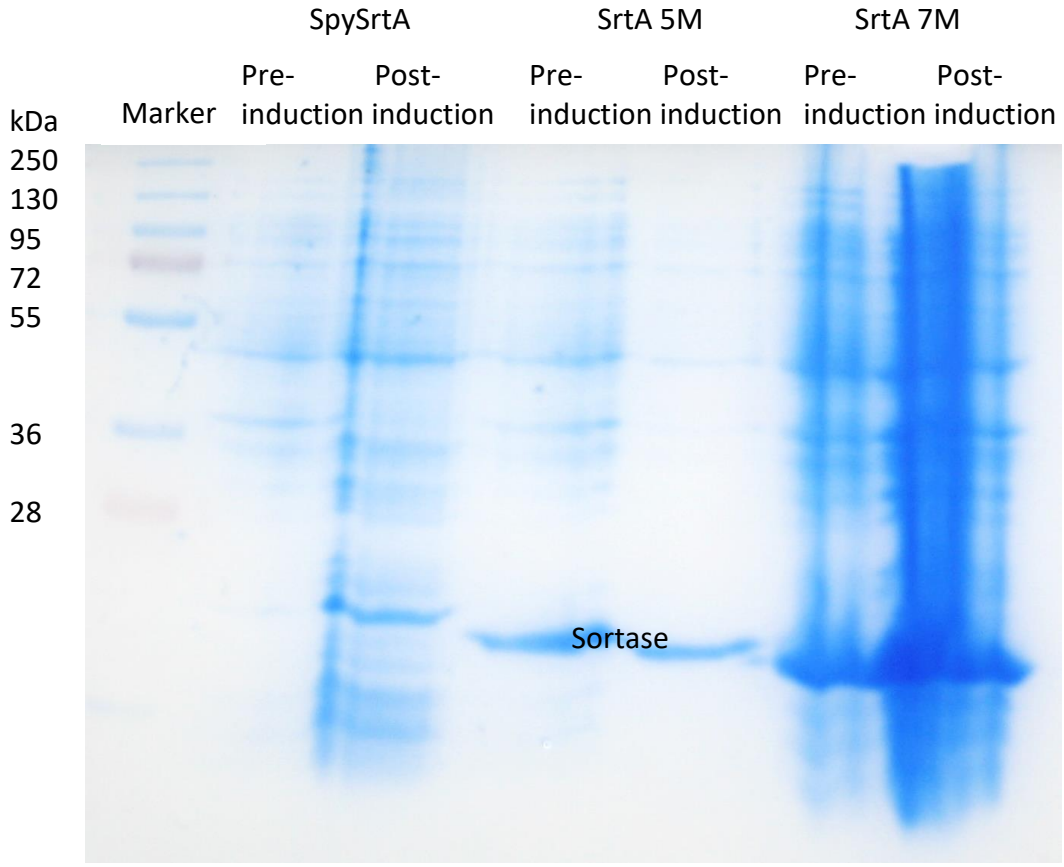


Figure 7.2: Successful sortase induction in *E. coli.*

Sortase Purification

Cell pellets were resuspended in 20 mL Tris III buffer with DNase and protease inhibitor tablet added, then lysed by sonication on ice using Soniprep 150 (6 x 30s cycles, 30s rest between each sonication). Lysate was cleared by centrifugation at 30,000 g for 45 min and supernatant filtered through 0.45 µm filter before column purification. Sortase was purified on an AKTA Start using a HiTrap Chelating High Performance column (1 mL or 5 mL) loaded with nickel. The column was pre-equilibrated with 5 CVs of Tris III buffer and the lysate loaded. The column was washed with 10 CVs of 8% Tris IV buffer in Tris III and protein elution with 10 CVs of Tris III – Tris IV gradient from 8% Tris IV – 100%. Depending on whether column size was 1 mL or 5 mL, fractions of 0.5 mL or 2 mL respectively were collected and checked for protein using SDS-PAGE. Clean protein fractions were pooled and dialysed in 100-fold HEPES buffer for 3 h at 4°C. The protein was concentrated, aliquoted, flash-frozen and stored at -80°C.

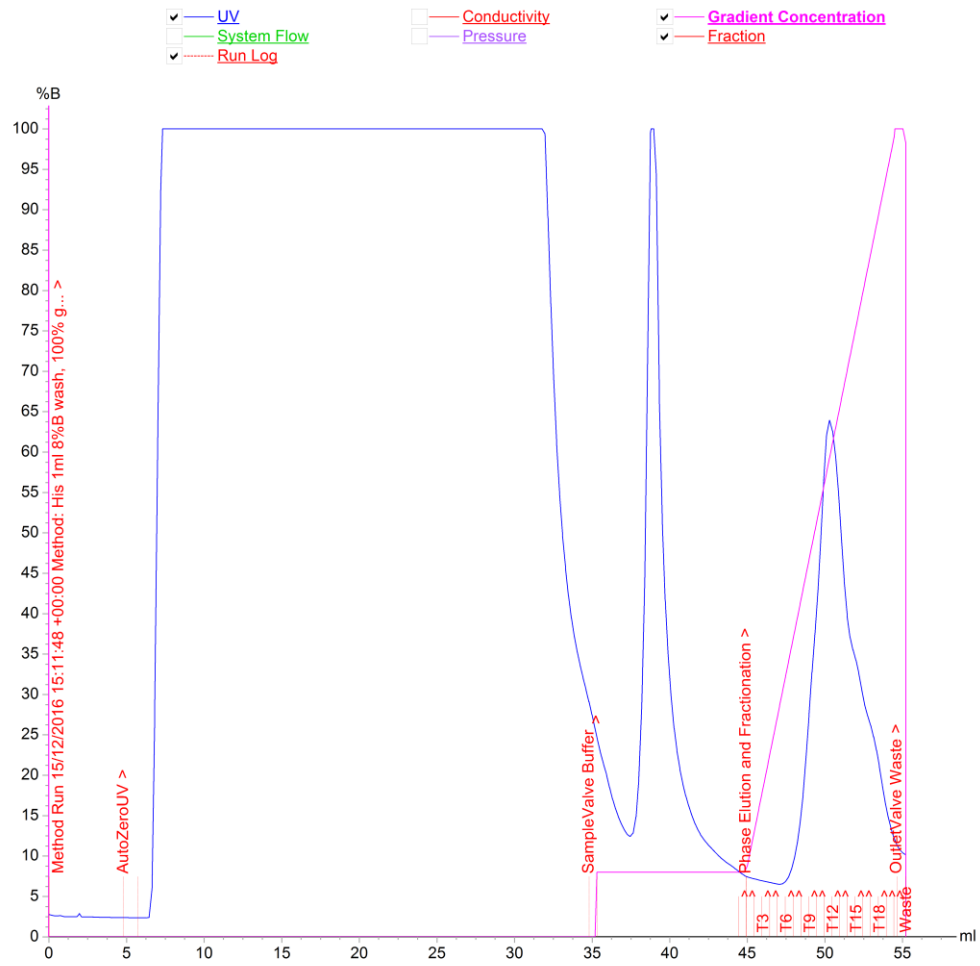


Figure 7.3: Spy SrtA purification on AKTA.

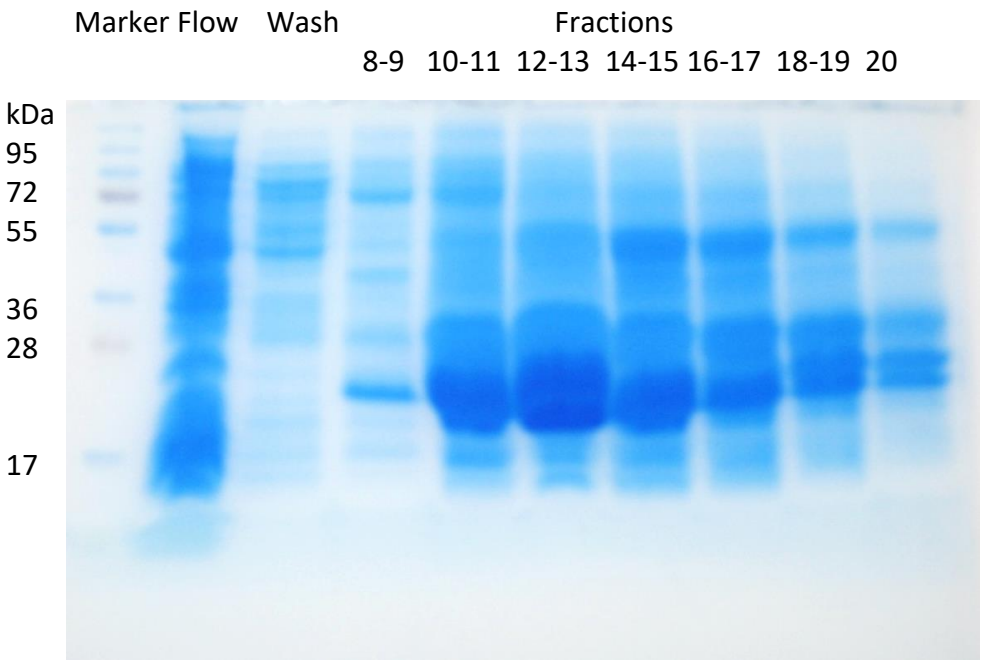


Figure 7.4: Spy SrtA purification, fractions 10-20 were dialysed.

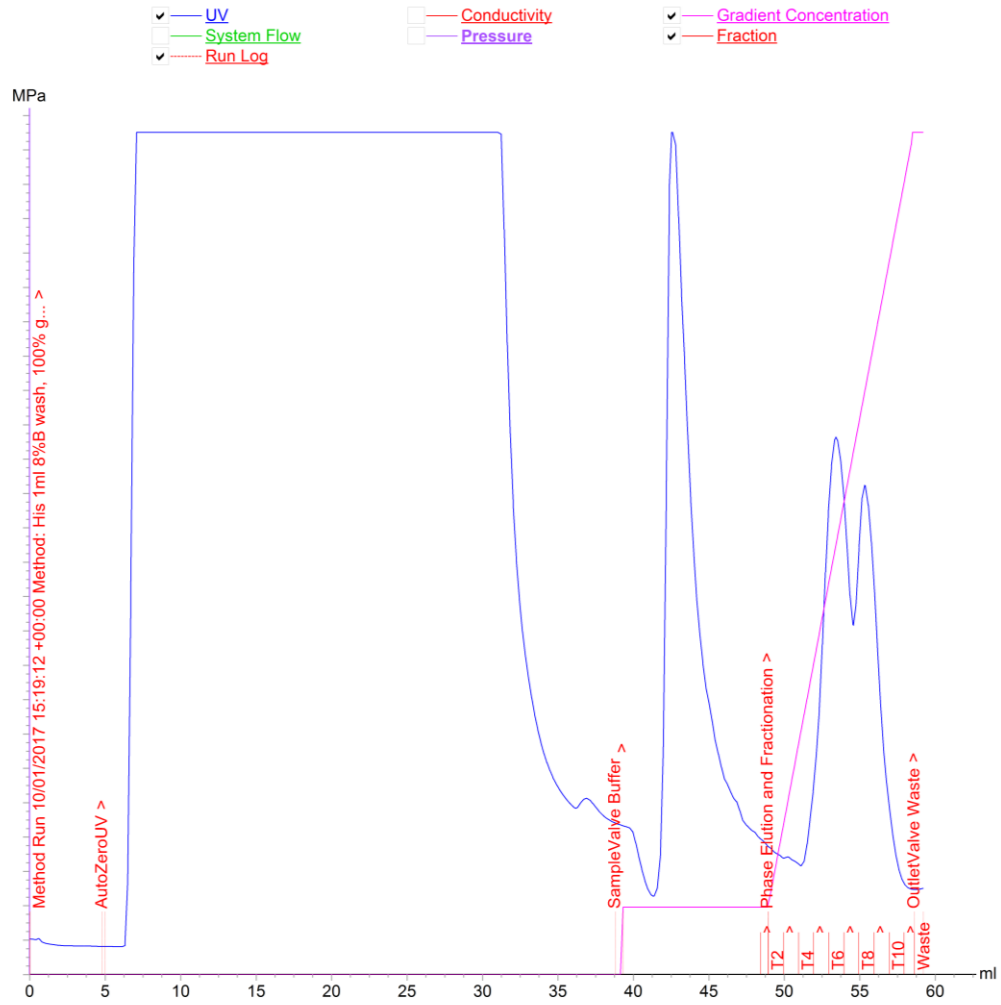


Figure 7.5: SrtA 5M purification on AKTA.

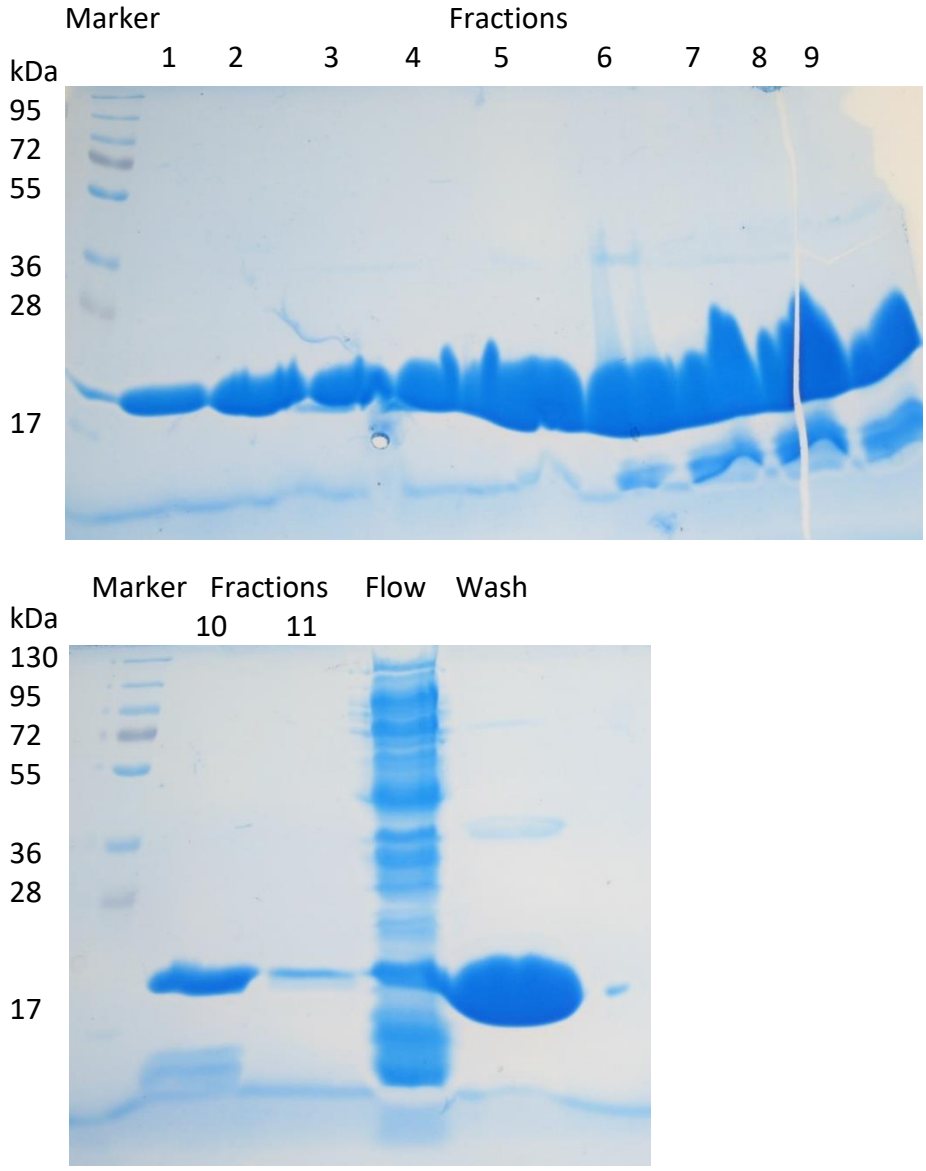


Figure 7.6: SrtA 5M purification, fractions 1-10 were dialysed.

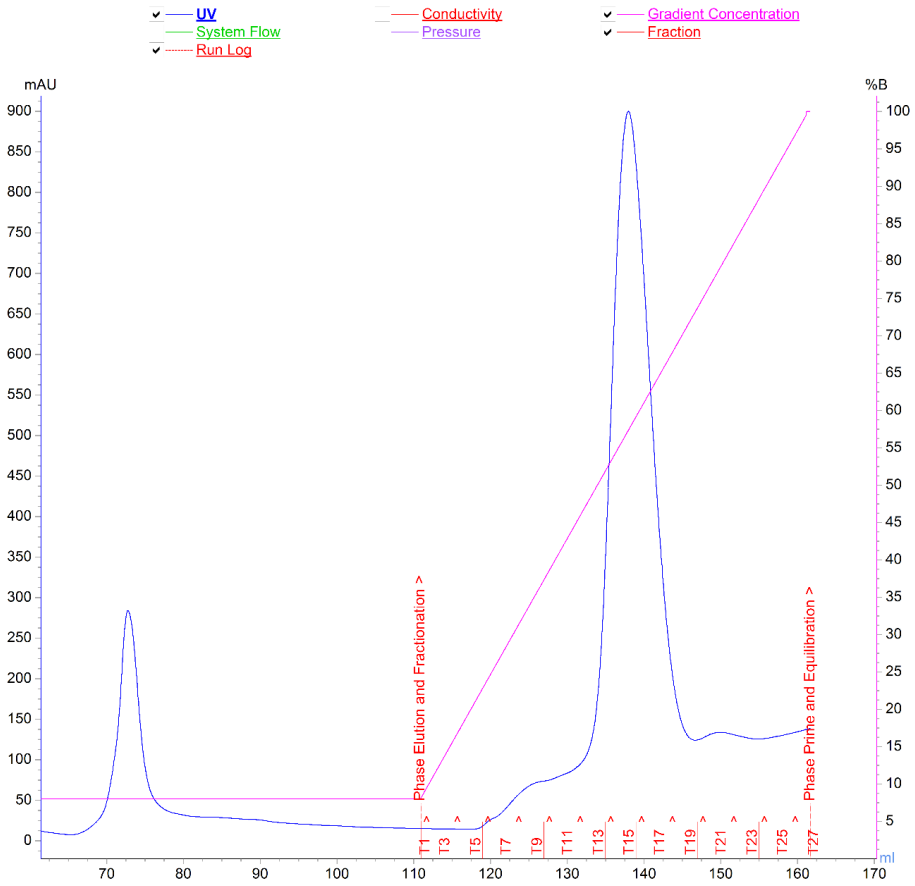


Figure 7.7: Sortase 7M purification on AKTA.

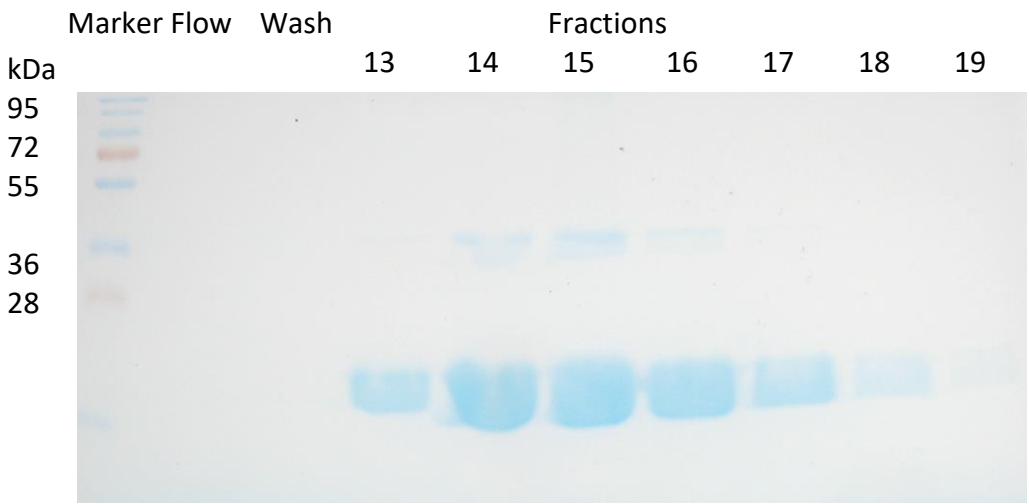


Figure 7.8: Sortase 7M purification, fractions 13-18 were dialysed.

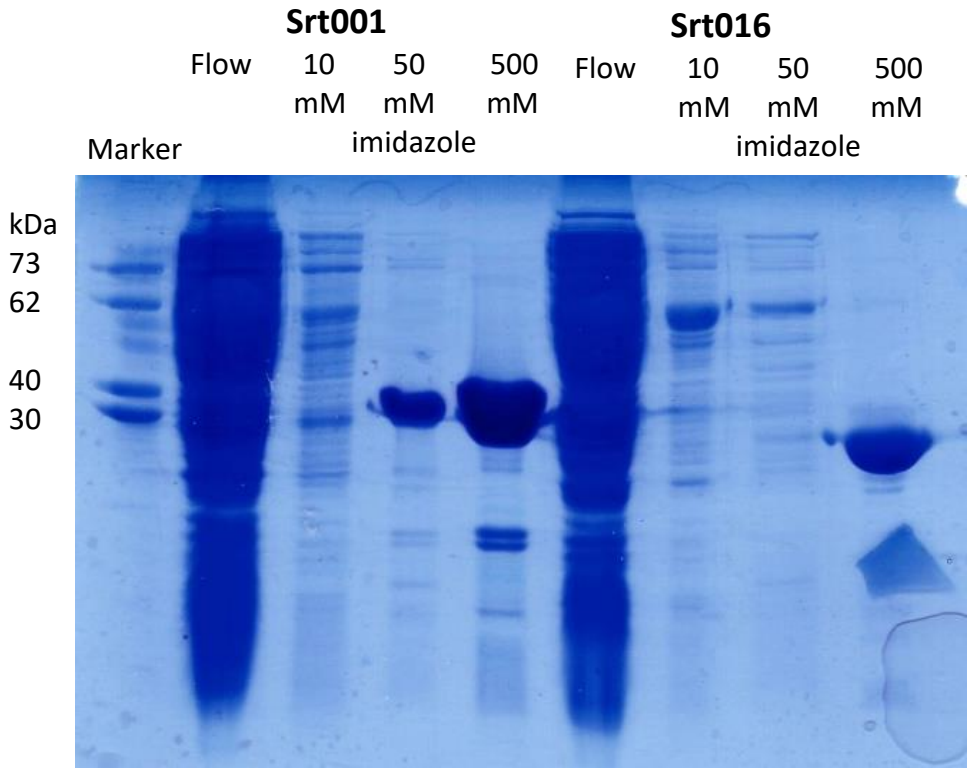
Sortase yields:

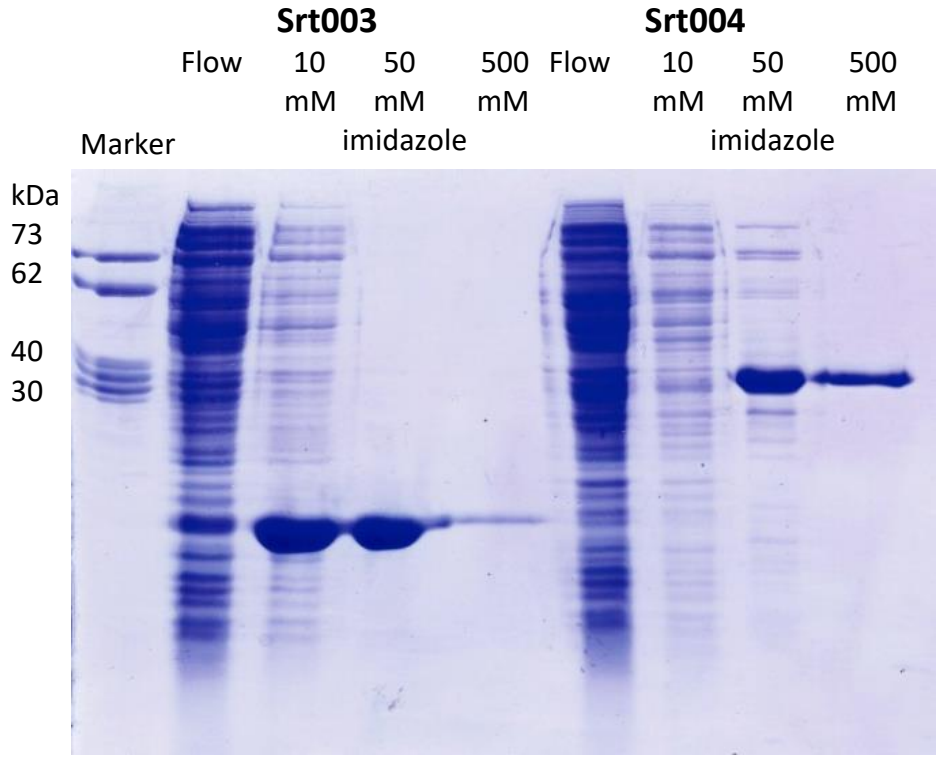
Sortase	Concentration (μM)	Volume (mL)
Spy SrtA	450	2.5
5M Sau SrtA	800	3
7M Sau SrtA	475	3.4

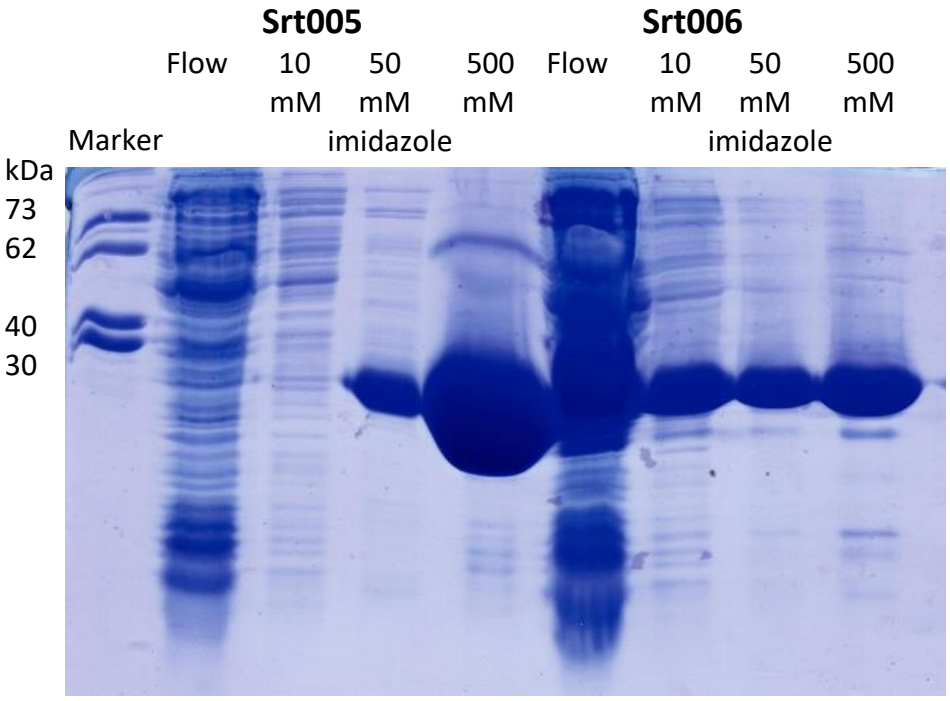
Prozomix Sortase Expression and Purification

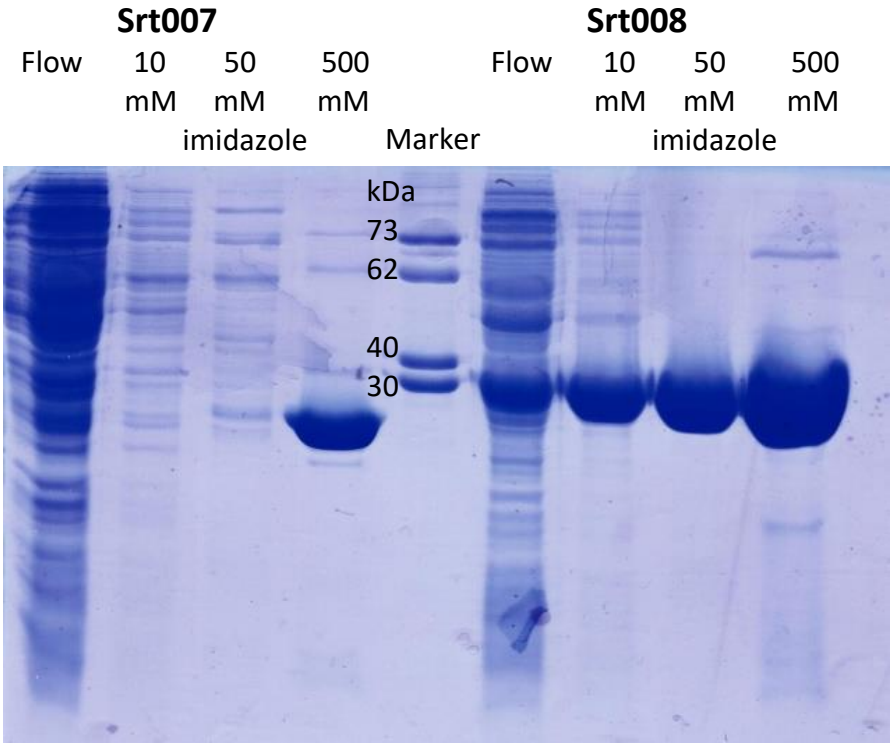
A single colony of *E. coli* BL21 (DE3) transformed with a new enzyme plasmid was used to inoculate 20 mL LB_{Kan} for starter culture and grown overnight at 37°C, 180 rpm. The starter culture was added to 750 mL LB_{Kan} in a 2 L baffled conical flask and the flask covered with kitchen paper and foil. Flasks were incubated at 25°C, 140 rpm for approximately 6 hours and the foil removed to improve aeration, leaving the paper covering on. Flasks were then grown overnight.

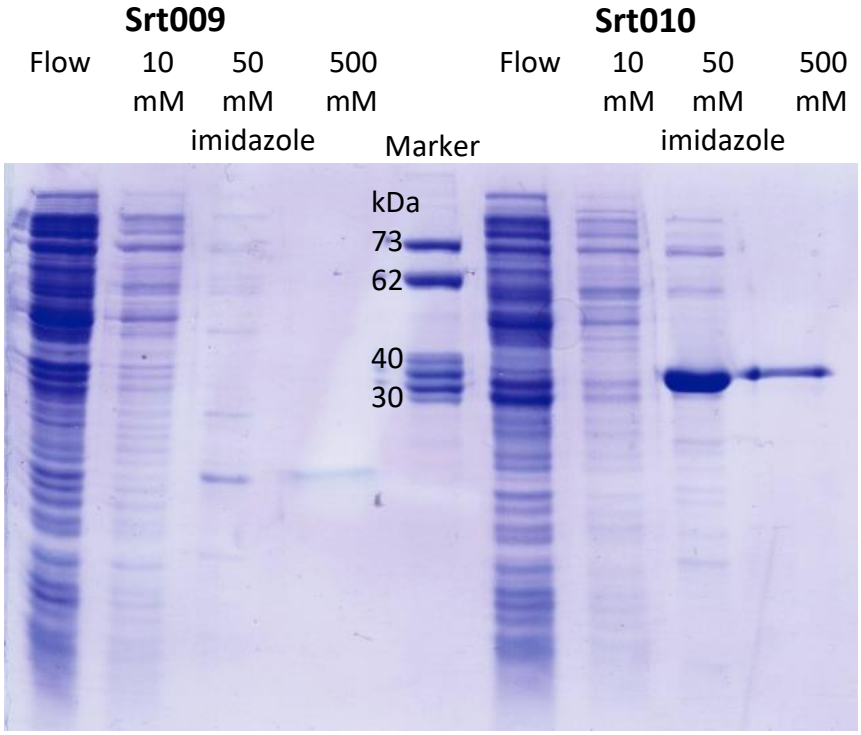
The cells were harvested by centrifugation in a Sorvall RC-5B refrigerated centrifuge at 7,500 rpm for 5 min. Cell pellets were weighed and 4x volume of Start buffer was added. The cells were resuspended and sonicated on ice for 1 min x 50 mL volume, for 4 rounds of sonication using Soniprep 150. The lysate was cleared by centrifugation at 11,000 rpm for 30 min and supernatant collected. This was purified on a gravity flow column packed with 8 mL HisPur Ni-NTA resin pre-equilibrated with 50 mL Start buffer. The lysate was loaded onto the column and washed with 50 mL Start buffer followed by 50 mL Wash buffer. The proteins were eluted with Elution buffer and fractions of 2 mL collected. Samples of 10 μL from each eluted fraction were added to 100 μL Bradford reagent (ThermoFisher) to check for protein and fractions that immediately turned the Bradford reagent bright blue were pooled. The volume of protein was measured, and 2x weight of ammonium sulfate was added and mixed before proteins were stored at 4°C. Gels for the purification of each sortase are shown below:

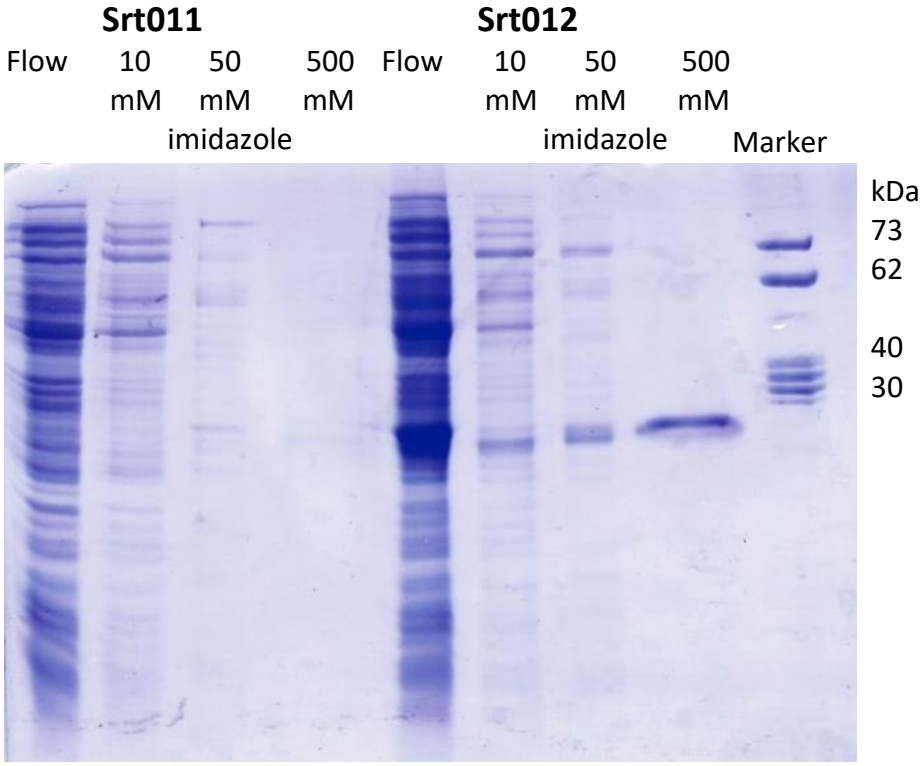

Srt001
Srt016
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
kDa
73
62
40
30

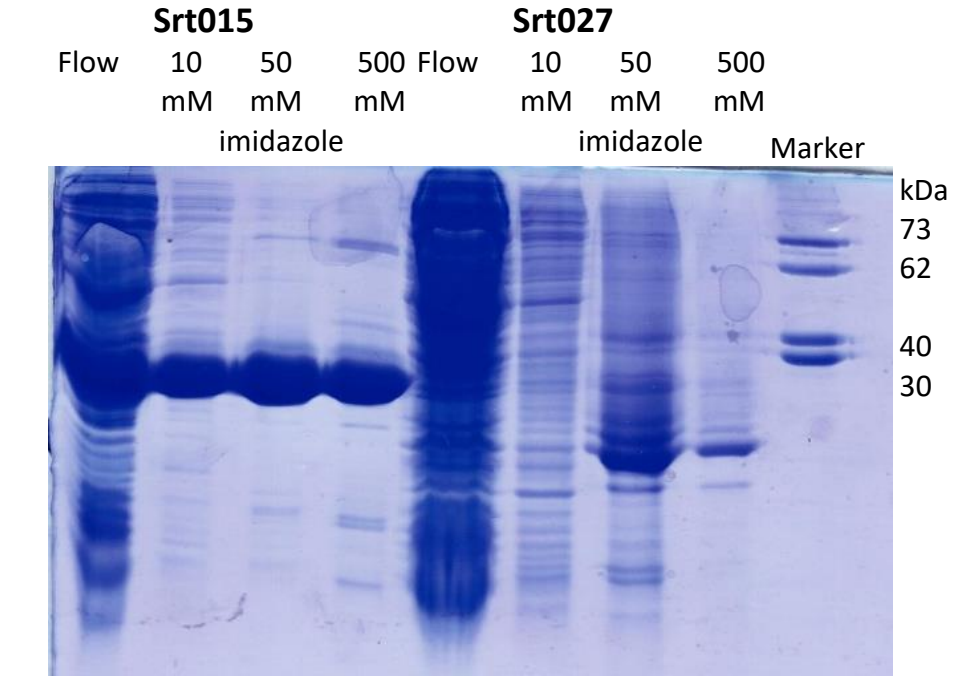

Srt003
Srt004
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
kDa
73
62
40
30

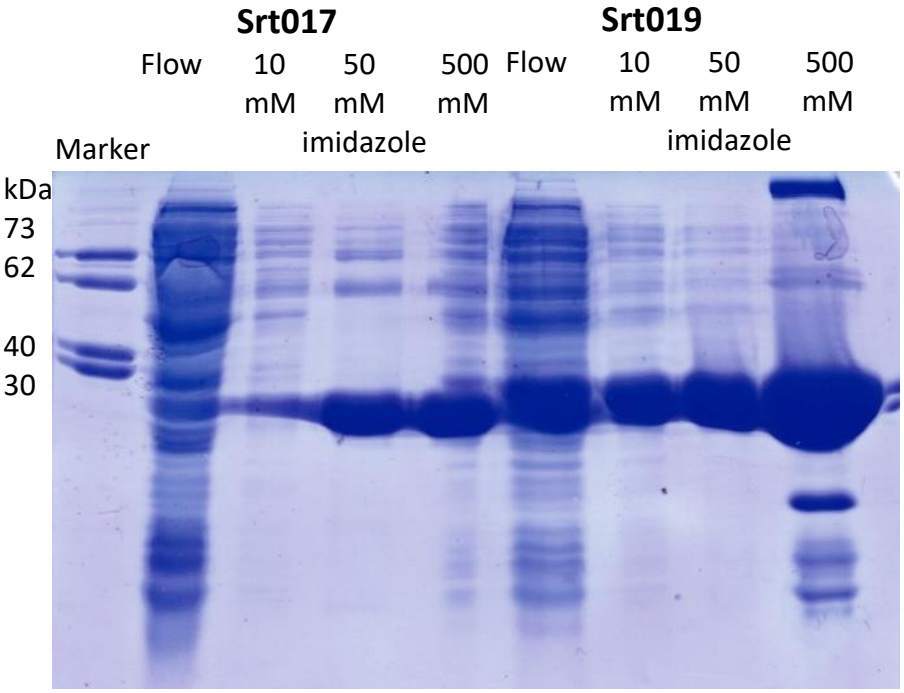

Srt005
Srt006
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
kDa
73
62
40
30

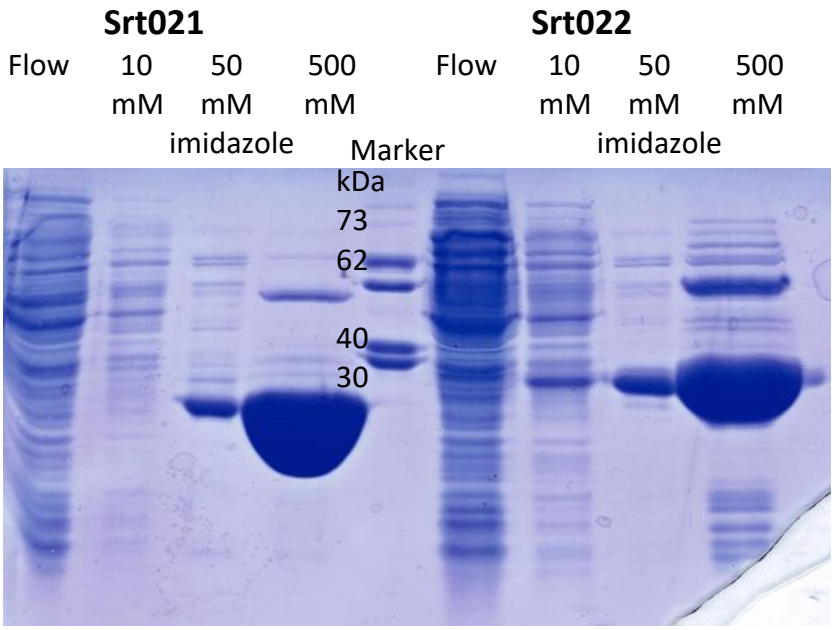

Srt007
Srt008
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
kDa
73
62
40
30

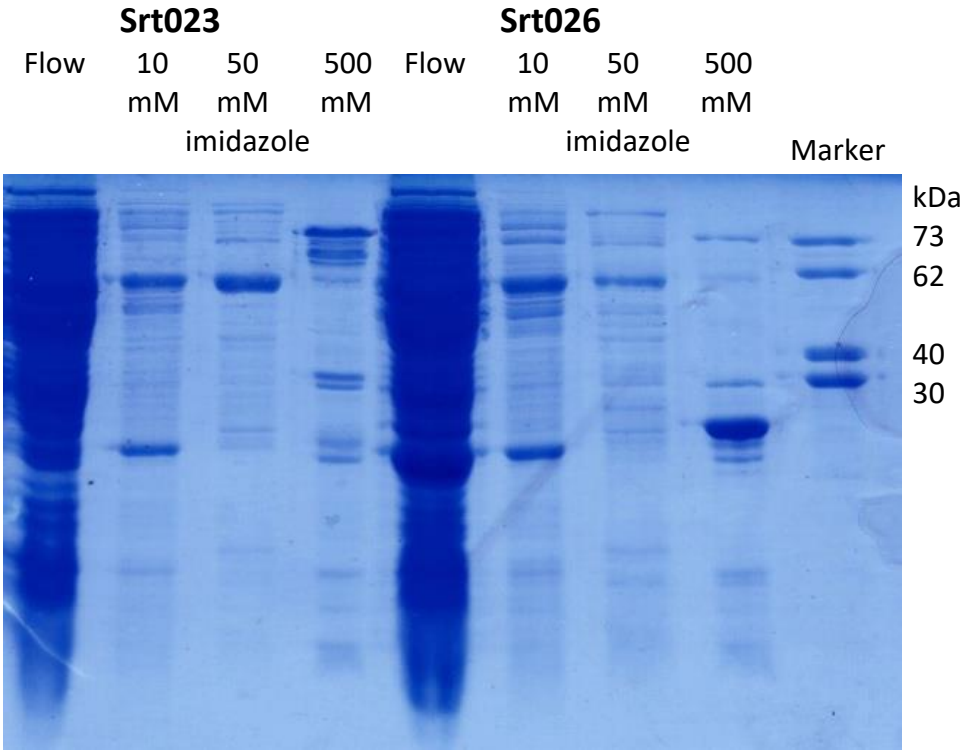

Srt009
Flow
10 mM imidazole
50 mM
500 mM
Marker
kDa
73
62
40
30
Srt010
Flow
10 mM
50 mM
500 mM
imidazole

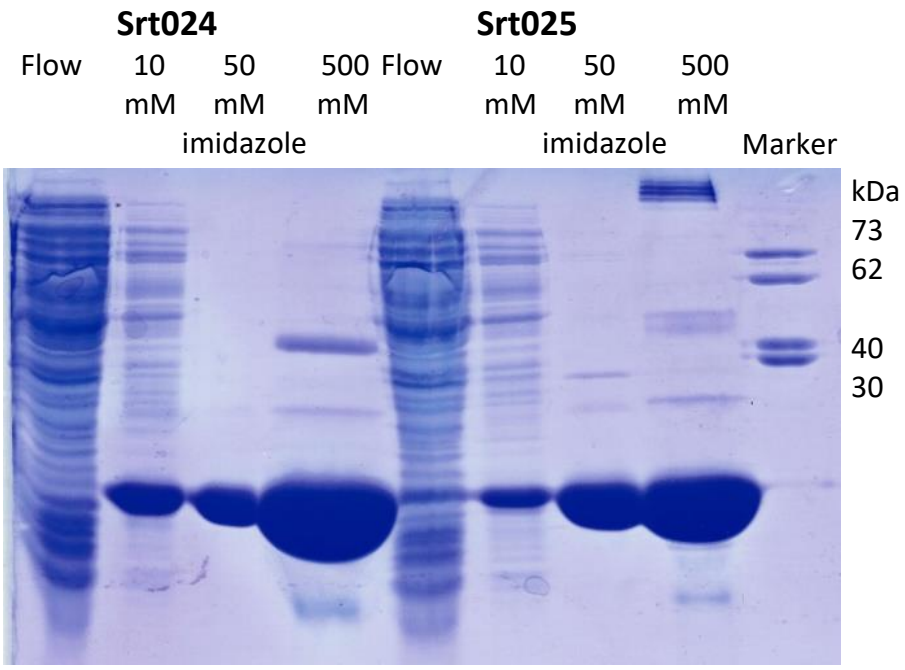

Srt011
Flow
10 mM
50 mM
500 mM
imidazole
Srt012
Flow
10 mM
50 mM
500 mM
imidazole
Marker
kDa
73
62
40
30


Srt015
Srt027
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
kDa
73
62
40
30


Srt017
Srt019
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Flow
10 mM imidazole
50 mM imidazole
500 mM imidazole
Marker
kDa
73
62
40
30


Srt021
Flow
10 mM imidazole
50 mM
500 mM
Marker
kDa
73
62
40
30
Srt022
Flow
10 mM imidazole
50 mM
500 mM


Srt023
Flow
10 mM imidazole
50 mM
500 mM
Srt026
Flow
10 mM imidazole
50 mM
500 mM
Marker
kDa
73
62
40
30



After the ammonium sulfate was found to interfere with the sortase reaction, the proteins were centrifuged at 17,000 *g* and the supernatant removed, then resuspended in the same volume of HEPES ligation buffer. The proteins were dialysed into three changes of 10x volume of HEPES ligation buffer at 4°C.

Sortase purification for crystals

Srt021 and Srt025 were further purified by size-exclusion chromatography for crystal trays. This used an AKTA FPLC with a Superdex 75 (3 kDa – 70kDa) column. The column was equilibrated with 1 CV of HEPES buffer, 10 mg of protein was loaded onto the column and the column run in HEPES buffer. 1 mL fractions were collected and the fractions shown with protein on the UV trace were checked with SDS-PAGE. Fractions 65-71 of Srt021 and 67-72 were pooled and concentrated by centrifuging at 4000 rpm with a 3 kDa MWCO. The proteins were concentrated to 10 mg / mL for the first batch, and 20 mg / mL for the second batch.

Each sortase had two crystal plates set up (kindly performed by Wendy Robinson). Each plate was a 96 well 2 drop plate containing either 1:1 or 1:2 ratio of protein to buffer in a sitting drop with a total volume of 300 nL. The commercial screens used were an INDEX screen, CSS 1 with bis-tris-propane at pH 6.0 and CSS 2 with bis-tris-propane at pH 8.0.

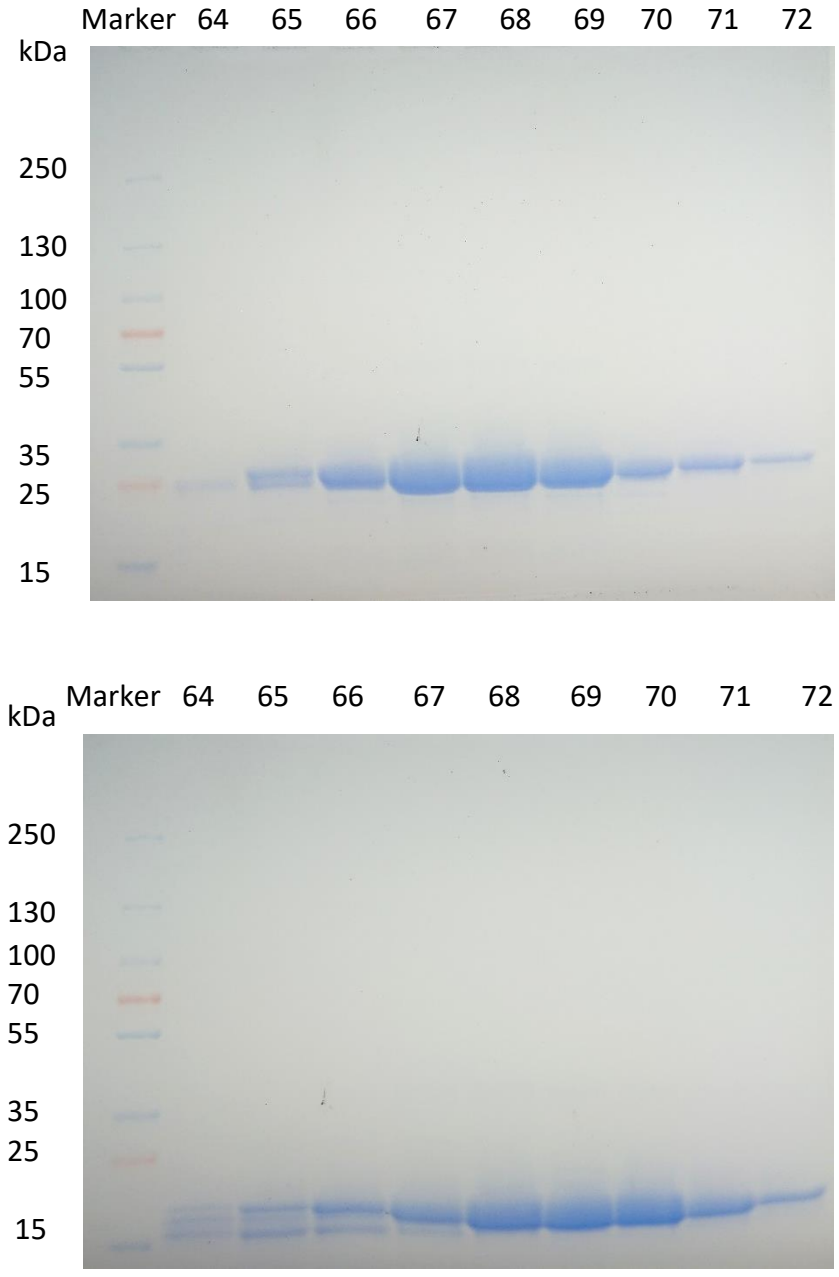


Figure 7.9: Size-exclusion purification of Srt021 (top) and Srt025 (bottom).

Sortase reactions

Sortase ligations were performed in HEPES ligation buffer.

	[stock], µM	Vol stock, µL	[reaction], µM
gMBP	240	25	60
Alkyne depsipeptide	2000	30	600
5M SrtA	800	0.75	6
HEPES	-	45	-

Alkyne-depsipeptide is hydrolysed over time, so 15 µl was added initially, and the other 15 µl added after 3 h. Samples were taken every hour for 6 hours. Samples were run on diluted in 50 µl 50% H_2O / 50% acetonitrile with 1% formic acid and analysed with liquid-chromatography mass-spectrometry (LC-MS).

Further sortase reactions were set up as follows in 100 µL HEPES buffer:

	[stock] µM	[reaction] µM	Mole eq.
SrtA 5M	800	60	1
GGG tripeptide	60000	600	10
LPETGG probe	9000	600	10

	[stock] µM	[reaction] µM	Mole eq.
Spy SrtA	450	60	1
AAA tripeptide	60000	600	10
LPETAA probe	2400	600	10

A negative control reaction mixture omitted the sortase, while a reaction to test for hydrolysis of the probe omitted the azidopeptide sugar. The reactions were repeated with tripeptide increased to 100 mole eq. Reactions were run for 3 h at 37°C and samples taken at hourly intervals and characterised on LC-MS.

Sortase reactions to test new sortases made at Prozomix were carried out in HEPES ligation buffer to a total volume of 50 or 100 µL. All hydrolysis reactions contained 60 µM sortase and 600 µM LPXTG / LPXTA / LPNTA probe. All ligation reactions contained

60 µM sortase and 600 µM LPXTG / LPXTA / LPNTA probe and 6000 µM nucleophile. Reactions were incubated at 37°C for 24 h. The sortase was precipitated out by addition of the same volume of ethanol and a 30 min incubation at -20°C, then centrifuged at 10,000 *g* to remove the protein pellet. Supernatant was diluted 1/3 into LC-MS solvent and run on an Accucore C18 column. Where the percentage of reacted peptide was calculated, the abundance of each peak was extracted from the mass spectra.

Srt021 and Srt025 were tested for hydrolysis and ligation of LPXTG and GGG under varying pH, temperature and buffers. These reactions were performed by Mohamed Abd El Bari. The reactions to test pH were performed in HEPES ligation buffer adjusted to pH 6, 7, 7.5 or 8 and incubated at 37°C for 24 h. Reactions to test temperature were performed in HEPES ligation buffer, pH 7.5 at 4°C, 25°C or 37°C for 24 h. For the reactions to test buffer, the sortases were dialysed once for 3 h at 4°C into the desired buffer before use in reactions at 37°C for 24 h.

Mammalian Cell Culture

Human Embryonic Kidney 293 (HEK293) cells were grown in Dulbecco's Modified Eagle Media (Gibco) supplemented with 10% (v/v) Fetal Bovine Serum, 1% (v/v) L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin. Cells were cultured in a humidified 5% CO_2 atmosphere at 37°C and passaged at 70-80% confluence. If cells were starved prior to feeding with sugars, the media was removed and fresh starvation media (DMEM supplemented with 1% (v/v) Fetal Bovine Serum, 1% (v/v) L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin) was added 24 h before seeding cells into well plates with sugar.

For cells that were assayed by flow cytometry, the cells were seeded in 6 well plates. 300 µL ethanol were added to each well of 6 well plates. Either 3 µL ethanol (control) or 0.3 µL, 1.5 µL or 3 µL of 100 mM stock solution of sugar in ethanol were added and mixed well by pipetting several times. The well plates were left uncovered for 15 min to completely evaporate the ethanol, then each well was seeded with 0.3 x 10^6 cells and 3 mL media. This gave final concentrations of 0 µM, 10 µM, 50 µM or 100 µM sugar. Each condition was tested in triplicate. The well plates were incubated for 72 h before performing the required assays.

For cells that were assayed by gel or western blot, cells were seeded in 10 cm tissue culture dishes. 1 mL ethanol were added to 10 cm tissue culture dishes (Falcon). Either 30 µL ethanol (control) or 30 µL of 100 mM stock solution of AcManThz **27** in ethanol were added and mixed well by pipetting several times. The tissue culture dishes were left uncovered for 15 min to completely evaporate the ethanol, then each dish was seeded with 2 x 10^6 cells and 10 mL media. This gave a final concentration of 300 µM AcManThz **27**, with each concentration tested in triplicate.

Cell Viability

Trypan Blue

To assess general cell growth and viability following treatment with sugars, culture media was aspirated and cells were washed with 500 µL PBS, then incubated with 250 µL 0.05% trypsin-EDTA (Gibco) for 5 min to release adherent cells. 750 µL media were added to neutralise the trypsin and the cells were mixed well by pipetting. 50 µL cell mix were added to 50 µL Trypan blue stain and the live (yellow) and dead (blue) cells counted using a haemocytometer.

MTS Assay

This assay based on the reduction of the 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2- (4-sulfophenyl)-2H-tetrazolium compound (MTS) [170] was used to assess potential toxicity linked to sugar treatments. 96 well plates were prepared by adding 20 µL ethanol to each well to cover the bottom, then adding 2 µL ethanol (control) or 0.2 µL, 1 µL, 2 µL, 4 µL or 6 µL of 10 mM stock solution of azidopeptide sugar in ethanol and mixing. This gave final concentrations of 10 µM, 50 µM, 100 µM, 200 µM and 300 µM respectively. Each sugar and concentration was tested in triplicate. The ethanol was allowed to evaporate for 15 min and each well was seeded with 5 x 10^3 cells and 200 µL media. After incubation at 37°C for 72 h 20 µL MTS reagent in electron coupling solutions (K300, BioVision) were added to each well and the plates incubated at 37°C 1 h. The A_{490} was measured using a BioRad iMark microplate reader and Microplate Manager 6 software.

Flow cytometry

Cells' incorporation of sugar derivatives was assessed using specific fluorescence labelling protocols described below combined with a Fluorescence-activated cell sorting

(FACS) analysis approach using a BD LSR Fortessa X-20 cell analyser. If DAPI stain was used, a 1 in 10,000 dilution of DAPI in PBS was added to samples for a final concentration of 1 in 20,000 immediately before running each sample on the Fortessa. FACS filters were set to the excitation maximum at 640 nm and emission maximum at 670 nm for streptavidin-647 (SA-647) detection, excitation maximum 488 nm and emission maximum 530 nm for fluorescein peptide detection, and excitation maximum 405 nm and emission maximum 450 nm for DAPI stain. At least 10 000 events were collected per sample. Data analysis was performed using FlowJo software from BD Biosciences.

Fluorescent Labelling of cells treated with sugars for flow cytometry analysis

Labelling of Mannosamine Derivatives by Click Chemistry

The media on cells treated for 72 h with mannosamine derivatives ($AcManGN_3$ **4**, $AcManGGN_3$ **26**) was aspirated and cells washed with 5 mL PBS and then incubated in 1.5 mL 10 mM EDTA for 5 min to gently release adherent cells. The EDTA was diluted by adding 3 mL PBS, or neutralised by adding 0.75 mL FBS. The cells were transferred to a flow cytometry vial and centrifuged at 200g, 4°C, for 3 min. The supernatant was poured off and cells resuspended in 200 µL flow cytometry buffer (PBS with 10 mM EDTA and 1% FBS, or PBS with 1% FBS if the cells were labelled at 37°C), then centrifuged at 200g, 4°C, for 3 min. This was repeated twice for a total of 3 washes. Cy5-DBCO dye was added to a final concentration of 10 µM and the cells left on ice in the dark for 1 h to allow the dye to react with exposed azides. Cells were centrifuged at 200g, 4°C, for 3 min, supernatant poured off and cells resuspended in 200 µL flow cytometry buffer. This was repeated twice for a total of 3 washes, and cells were detected by FACS.

Another method repeated the above protocol, but after cells were labelled with Cy5-DBCO dye and washed 3 times with 200 µL flow cytometry buffer, 200 µL FACS fixing buffer (PBS with 1% formaldehyde, 1% FBS and 0.05% azide) were added and the cells mixed and left in the fridge overnight. Cells were then centrifuged and washed 3 times in 200 µL flow cytometry buffer as above before FACS.

A third method labelled the cells with Cy5-DBCO first. The media was aspirated and the cells incubated in PBS with 10 µM Cy5-DBCO dye for 1 hour. The PBS was aspirated and the cells washed with PBS and gently detached by incubation with EDTA as above. The EDTA was diluted by adding 3 mL PBS and cells transferred to a flow cytometry vial,

centrifuged at 200g, 4°C, for 3 min and washed 3 times in 200 µL flow cytometry buffer as above.

Reduction of Mannosamine-Azide Derivatives to Amines

For cells that had been incubated in azide sugars for subsequent sortase ligation, 48 hours after HEK cells were seeded with media and sugars, 2DPBA in DMSO was added to a final concentration of 400 µM and the cells incubated for a further 24 hours before labelling with sortase. Where TCEP was used, TCEP was either added to 400 µM 24 hours before labelling as with 2DPBA, or added to 100 µM after the cells were harvested, and the cells incubated on ice for 15 min.

Fluorescent Labelling of Mannosamine-Amine Derivatives by Sortase Ligation

The media on cells treated for 72 h with mannosamine derivatives ($AcManGN_3$ **4**, $AcManGGN_3$ **26**) was aspirated and cells washed with 0.5 mL PBS and then incubated in 0.25 mL 5 mM EDTA for 5 min to gently release adherent cells. The EDTA was neutralised by adding 0.75 mL medium. The cells were transferred to a flow cytometry vial and centrifuged at 1000 rpm, 4°C, for 3 min. The supernatant was aspirated and cells resuspended in 200 µL PBS. The cells were centrifuged, supernatant aspirated and cells resuspended in flow cytometry buffer (PBS with 1% FBS) 2 times. The cells were again centrifuged, and resuspended in 100 µL sortase ligation buffer (50 mM HEPES, 150 mM NaCl, 5 mM $CaCl_2$ at pH 7.5). SrtA 5M was added to 20 µM and Fl-Gaba-LPETGG **30** probe added to 200 µM and the cells mixed and incubated for 1 h at 37°C. Cells were centrifuged at 1000 rpm, 4°C, for 3 min, supernatant poured off and cells resuspended in 200 µL flow cytometry buffer. This was repeated twice for a total of 3 washes. Fluorescent cells were detected using FACS.

Fluorescent Labelling of Mannosamine-Thz Derivatives by Palladium Decaging and FACS Detection

The media on cells treated for 72 h with AcManThz **27** was aspirated and cells washed with 1 mL PBS and then incubated in 250 µL 5 mM EDTA for 5 min to gently release adherent cells. The EDTA was neutralised with 0.75 mL PBS. Cells were transferred to a flow cytometry vial and centrifuged at 200g, 4°C, for 3 min. Supernatant was poured off and cells resuspended in 200 µL PBS, then centrifuged at 200g, 4°C, for 3 min. This was repeated twice for a total of 3 washes, with the cells resuspended in 200 µL PBS pH 7.4.

If a blocking step was used, cells were resuspended in 84.5 µL PBS and 12.5 µL Pro-tetrazole **36** and 2.9 µL phenylacetaldehyde was added to each sample and left at RT for 30 min. The cells were centrifuged 200g, 4°C, for 3 min and washed in 200 µL PBS three times.

Allylpalladium (II) chloride dimer **35** dissolved in DMSO (30 mM stock) was added to each sample to give a final concentration of 300 µM and left to react at RT for 30 min. Cells were centrifuged at 200g, 4°C, for 3 min and washed in 200 µL PBS three times, then resuspended in 84.5 µL PBS. OPAL probe **37** (2 µL of 10 mM stock) was added to a final concentration of 0.2 mM and the Pro-tetrazole catalyst **36** (12.5 µL of 200 mM stock) to a final concentration of 25 mM , and the reaction tubes incubated at RT for 30 min. Cells were centrifuged at 200g, 4°C, for 3 min and washed in 200 µL PBS three times, then resuspended in 100 µL PBS + 1% FBS with a 1 in 500 dilution of SA-647. Cells were centrifuged at 200g, 4°C, for 3 min and washed in 200 µL PBS three times, then analysed by FACS.

Western Blot analysis

Fluorescent Labelled Mannosamine-Thz Derivatives with palladium decaging

The media on cells treated for 72 h with AcManThz **27** was aspirated and cells washed with 0.5 mL PBS and then incubated in 0.25 mL 5 mM EDTA for 5 min to gently release adherent cells. The EDTA was neutralised by adding 0.75 mL medium. Cells were centrifuged at 200g, 4°C, for 3 min. Supernatant was poured off and cells resuspended in 200 µL PBS, then centrifuged at 200g, 4°C, for 3 min. This was repeated twice for a total of 3 washes.

Cells were resuspended in 100 µL lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 1% Igepal detergent (Sigma) and protease inhibitor) and left for 30 min on ice. The lysed cells were centrifuged at 7000 rpm for 10 min and the supernatant collected. The protein concentration of the lysate was measured at A_{280} using a Nanodrop spectrophotometer, and samples diluted with PBS to equal protein concentrations.

If a blocking step was used, the cells were resuspended in 73 µL PBS and 14.5 µL phenylacetaldehyde and 12.5 ul Pro-tetrazole **36** was added, and the samples incubated at RT for 30 min. The buffer was exchanged by centrifugation 10 kDa MWCO at 12,000 g for 10 min x 2 spins.

Allylpalladium (II) chloride dimer **35** dissolved in DMSO (30 mM stock) was added to a final concentration of 50 or 300 µM and the samples left at room temperature for 60 min. The samples were quenched with 10 ul 3-mercaptopropanoic acid. The samples were desalted with a PD G25 MiniTrap column using the spin protocol, and concentrated to < 50 µL by centrifuging at 12 000 g with a 3 kDa MWCO.

The OPAL-biotin probe **37** was added to the samples to a final concentration of 1 mM and the Pro-tetrazole catalyst **36** to a final concentration of 25 mM. The samples were left at RT for 60 min, then SDS loading buffer was added and the samples boiled at 95°C for 5 min before running on SDS-PAGE at 200 V for 45 min.

The western blot used a PVDF 0.2 µm membrane (ThermoFisher). The membrane was soaked in methanol for 30 s, then H_2O for 30 s, then in transfer buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS, 20% (v/v) methanol) for 10 min. The gel was also soaked in transfer buffer, for 5 min. The protein samples were transferred in a Bio-Rad Trans Blot Turbo at 25 V for 30 min. The membrane was incubated in blocking solution (PBS + 5% (w/v) skim milk powder + 0.1% Tween) overnight at 4°C. The membrane was incubated in blocking solution with 1/4000 dilution of streptavidin-horseradish peroxidase conjugate (ThermoFisher) for 1 h at 37°C. The membrane was washed for 5 min intervals in PBS with 0.1% Tween twice, then in PBS once. 1.5 mL each of substrate A and B from Western ECL Substrate kit (BioRad) was added and the membrane incubated at RT for 5 min. Bands were imaged using a Syngene G:Box Chemi XRQ with a Synoptics 4.0 MP camera and GeneSys software.

Fluorescent Labelling of Glucosamine Derivatives

Labelling and detection of intracellular glucosamine derivatives ($AcGlcGN_3$ **6**, $AcGlcAN_3$ **28**) used a gel labelling protocol. The media was aspirated and cells washed with 5 mL PBS, then incubated with 1.5 mL 0.05% trypsin-EDTA for 5 min to release adherent cells by cleaving cell surface proteins. 1.5 mL media were added to neutralise trypsin. The cells were transferred to centrifuge vials and centrifuged at 1000 rpm, 4°C, for 4 min. The supernatant was aspirated and the cells resuspended in 1 mL PBS, then centrifuged at 1000 rpm, 4°C, for 4 min. This was repeated for a total of 2 washes. The cells were resuspended in 100 µL PBS and centrifuged at 200 g for 5 min and the supernatant removed.

The cells were lysed by the addition of 100 µL lysis buffer (20 mM Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% Igepal detergent (Sigma)) and EDTA-free protease inhibitor cocktail to each vial, and the cells mixed and left for 30 min on ice. The cells were centrifuged at 6000 g for 5 min and the lysate collected. A solution of 20 mM HEPES with 1% SDS was added to the lysate to a total volume of 200 µL. The protein was prepared according to Invitrogen's protocol MP33368 [224]. Briefly, 600 µL methanol were added to the 200 µL samples and the samples vortexed. 150 µL chloroform were added, the samples vortexed, and finally 400 µL MilliQ water were added and the samples vortexed. The mixture was centrifuged for 5 min at 17,000 g and the upper aqueous phase removed, being careful not to disturb the protein layer at the interface. 450 µL methanol were added and the samples vortexed and centrifuged at 17,000 g to form a protein pellet. The supernatant was aspirated and the protein pellet air dried for 5 min and stored at -20°C.

Protein pellets were resolubilized in 50 µL of 1% SDS, 50 mM Tris-HCl pH 8 and labelled with the TAMRA alkyne using a copper-catalysed click protocol. TAMRA-alkyne was added to 25 µM, THPTA to 2.5 mM, $CuSO_4$ to 0.5 mM and sodium ascorbate to 5 mM. The reaction mixture was incubated at room temperature in the dark for 1 h.

Labelled protein samples were prepared for SDS-PAGE using protocol MP33368 as above. The dried protein pellets were resolubilized in 50 µL SDS-PAGE buffer by vortexing for 10 min followed by heating at 95°C for 5 min. The samples were run on SDS-PAGE (12% acrylamide) and imaged using a G:Box Chemi-XRQ gel doc, Syngene UV06 filter.

Appendix

Commercial sortase sequences

Spy SrtA sequence 5’ – 3’

ATGCGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGATCCGTCTTG
CAAGCACAAATGGCGGCTCAGCAACTTCCTGTTATAGGGGGCATTGCCATACCAGAGCTTGGC
ATTAATTTACCAATTTTTAAAGGTTTAGGAAATACTGAGCTTATTTATGGCGCAGGAACGATG
AAAGAAGAACAAGTTATGGGAGGAGAAAATAATTATTCTCTTGCCAGTCATCATATTTTTGGA
ATTACAGGTTCATCTCAAATGCTCTTTTCGCCGCTTGAAAGAGCACAAAATGGGATGTCCATCT
ATTTAACAGATAAAGAAAAAATTTACGAATACATCATAAAAGATGTTTTCACGGTAGCTCCTG
AACGCGTTGATGTTATCGATGATACAGCTGGTCTCAAAGAAGTGACTTTAGTGACTTGTACAG
ATATCGAAGCAACAGAACGTATTATTGTCAAAGGAGAACTAAAAACAGAATACGACTTTGATA
AAGCGCCCGCCGATGTATTGAAAGCTTTTAATCATTCTTATAACCAAGTATCTACCTAG

Protein sequence:

MRSSHHHHHH SSGLVPRGSV LQAQMAAQQL PVIGGIAIPE LGINLPIFKG LGNTELIYGA

GTMKEEQVMG GENNYSLASH HIFGITGSSQ MLFSPLERAQ NGMSIYLTDK EKIYEYIIKD

VFTVAPERVD VIDDTAGLKE VTLVTCTDIE ATERIIVKGE LKTEYDFDKA PADVLKAFNH

SYNQVST

Spy SrtA expressed in plasmid pET 28a, Kan^{R}. Plasmid 51139 from Addgene.

SrtA 5M sequence 5’ – 3’

ATGCAAGCTAAACCTCAAATTCCGAAAGATAAATCAAAAGTGGCAGGCTATATTGAAATTCCA
GATGCTGATATTAAAGAACCAGTATATCCAGGACCAGCAACACGCGAACAATTAAATAGAGG
TGTAAGCTTTGCAGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACACAC
TTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGT
GTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAGTATAAGAAACGTTAAGCC
AACAGCTGTAGAAGTTCTAGATGAACAAAAAGGTAAAGATAAACAATTAACATTAATTACTTG
TGATGATTACAATGAAGAGACAGGCGTTTGGGAAACACGTAAAATCTTTGTAGCTACAGAAG
TCAAACTCGAGCACCACCACCACCACCACTGA

Protein sequence (Pentamutant – original SrtA protein truncated with amino acids 2-59 removed and point mutations P94R, D160N, D165A, K190E, K196T underlined):

MQAKPQIPKD KSKVAGYIEI PDADIKEPVY PGPATREQLN RGVSFAEENE SLDDQNISIA

GHTFIDRPNY QFTNLKAAKK GSMVYFKVGN ETRKYKMTSI RNVKPTAVEV LDEQKGKDKQ

LTLITCDDYN EETGVWETRK IFVATEVKLE HHHHHH

SrtA 5M expressed in plasmid pET 30b, Kan^R. Plasmid 51140 from Addgene.

SrtA 7M sequence 5' – 3'

ATGCAAGCTAAACCTCAAATTCCGAAAGATAAATCAAAAGTGGCAGGCTATATTGAAATTCCA
GATGCTGATATTAAAGAACCAGTATATCCAGGACCAGCAACACGCGAACAATTAAATAGAGG
TGTAAGCTTTGCAAAAGAAAATCAATCACTAGATGATCAAAATATTTCAATTGCAGGACACAC
TTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGT
GTACTTTAAAGTTGGTAATGAAACACGTAAGTATAAAATGACAAGTATAAGAAACGTTAAGCC
AACAGCTGTAGAAGTTCTAGATGAACAAAAAGGTAAAGATAAACAATTAACATTAATTACTTG
TGATGATTACAATGAAGAGACAGGCGTTTGGGAAACACGTAAAATCTTTGTAGCTACAGAAG
TCAAACTCGAGCACCACCACCACCACCACTGA

Protein sequence (Heptamutant – original SrtA protein truncated with amino acids 2-59 removed and point mutations at P94R, E105K, E108Q, D160N, D165A, K190E, K196T underlined):

MQAKPQIPKD KSKVAGYIEI PDADIKEPVY PGPATREQLN RGVSFAKENQ SLDDQNISIA

GHTFIDRPNY QFTNLKAAKK GSMVYFKVGN ETRKYKMTSI RNVKPTAVEV LDEQKGKDKQ

LTLITCDDYN EETGVWETRK IFVATEVKLE HHHHHH

SrtA 7M expressed in plasmid pET 30b, Kan^R. Plasmid 51141 from Addgene.

Novel sortase sequences

These are the sequences of the sortase variants cloned and manufactured at Prozomix. The His tag and linker with thrombin cleavage site is underlined.

Srt001

MGSSHHHHHHSSGLVPRGSHMDWYISRHQGQVVDNYDKKAAQMSQKEINEALEKAQEYNEEL
LGNVVLTDPFDPAAVEKQNEDYDNLLNIGGDGVMGSVEIPSINVYLPIYHGTDSKSLEKGAGHLEN

SSLPIGGKGTHAIISAHTGLPSAKMFDDLTEVKEGDIFYIHVLNRTLAYEVNQIKVVLPENVSDLLIDK
NKDYVTLVTCTPYGINSHRLLVRGERIPYKEAKKKDNSDKKGSNMWKAY

Srt003

MGSSHHHHHHSSGLVPRGSHMCTIDIPKIGVYLPVQHGTGAETLERAVGHVVGTSLPVGGSGTH
AVLSAHSGMASAKLFSDIDQLAEGDMFYIHVLGEVLAYKVDAIHTVLPTDTSLLQIEDGKDQVTLV
TCTPFGVNTHRLLVRGHRVPYTPEQEATAAAEKPVASSWTQ

Srt004

MGSSHHHHHHSSGLVPRGSHMSNYLFENRADGIIDTVEKTADDADEEKYKEEIEAAQKYNAELAT
GHVVLKDPFVEEKLDEDAKEYNSLLNMADGGVMGFIKIPCIDVSLPIYHGTSAEILELGAGHLQGTS
LPIGGESTHSVITGHTGLSSAKLFTDLTELEEGDMFFLHVMGEKLAYKVGQISVILPEEMDKLTIENG
KDYCTLVTCTPYGVNTHRLLVRGERTEYTEDKEEEAGTEVKKTQSKWMEEYTKS

Srt005

MGSSHHHHHHSSGLVPRGSHMLEKAREYNEKLASSHVVLTDPFSEETNGGISEKEYYQLLNLDNT
GVMCSLEIPSINVDLPVYHGTSNSVLEKGVGHLEGTSLPVGGKDTHAVFTGHTGLNKAKLFTDLTE
LQKGDQFYIRVLDKILAYEVCQIDVVLPEDTSKLSVVDGQDLVTLVTCTPYGQNTHRLLVRGKRTKY
SPKAYEKERAKKKAGRSQWMR

Srt006

MGSSHHHHHHSSGLVPRGSHMTDYEQEVSHLSEEQENAMIEQAQEYNESLIGIGTIADPFSESNE
NQTEDDEYNKLLKIDDTGMMGYIDIPKLDVVLPVYHGTSEKVLQSGVGHLKNTSLPVGGESCHAV
LSGHRGLANAKIFTDLNKMEVGDVFYIKVLHHTFAYQVDQILTVLPSDIDSLQIEKGKDYVTLVTCT
PYAVNTHRLLVRGTRIPYEEAQKIDEEVGLHH

Srt007

MGSSHHHHHHSSGLVPRGSHMVSNYLYEKNGARVISSYDENAVRLSESEKQAMLEAARQYNREL
LGNIELLDPFSPLKKEVDARYQSLLNTNEAGMMGYIRIPKIDVELPIYHGTEERILQSGVGHFEGTSL
PVGGESSHTVLTGHRGLPSKLLFTDLDQMKEGDIFYLKILGETFAYKIDQILTVLPENTKALTIEPGKD
YATLVTCTPYAVNTHRLLVRGIRIPYEEAVRQVPDEKITPTLPFQVK

Srt008

MGSSHHHHHHSSGLVPRGSHMNQWNNHRQKQLISSYEDNLTQLTEAGDIDYAKELKKAQAYN
DALVPSILPDSFAVADAREEEDSAYMNCLNLTGDGMMGIVEIPKIAIKLPIYHGTSDEVLQQAAGH

LEGSSLPIGGESTHAVISAHRGLPSASLFTDLDQMKIGDHFLIHVLDNTLCYEVDQILVVEPEDTDAL
AVEEGEDLVTLLTCTPYGVNTQRLLVRGHRVDYVADEVAAEQTPLSGISLHT

Srt009

MGSSHHHHHHSSGLVPRGSHMILPDSFAIADAKEEKDRFYESCLNITGDGIMGMVEIPKIDVELPI
YHYTTEEVLKNAAGHLEGSSLPVGGKSTHSVISAHRGLPSAILFTDLDKMKKGDHFMIHVLDDILCY
EVDRISIVKPEDTSDLNVEKGKDLMTLLTCTPYGVNTERLLVRGHRVPYKEAYSSEKDSSISLE

Srt010

MGSSHHHHHHSSGLVPRGSHMNYLAQREQKEVIEEYAQTVEQSDKDKMARQWELAEEYNETLL
GDPVHDPFIPGTGYALPDNYESVLNVNKDGVMGYLKIPKIKVDLPIYHGTSEEVLEKGAGHVDVTA
LPIGGVNRHPVISAHRGLPSAELFTRLDELEKGDRFFLHILDKTLAYKVDQVRVIKPEELEQLQTYHD
KDYVTLLTCTPYGVNTHRLLVRGERIPYEVAEENGGEAFDDEKNQGMPQWVKE

Srt011

MGSSHHHHHHSSGLVPRGSHMINNYVSRSMVKEYTNRVAQMPSEKTQKMFEEATKYNNSLSN
NMIITDPFDEKAFQKIGANYEKTLNIDDNGLIGYIDVPKINVYLPIYHGTSEEILSKGAGHLQNTSLPV
GGASTHSVISAHSGFPGETFFDYLTDMKVGDEFYVHILDRTLKYEVDQIEVVLPSEINSLRIVDGEDL
VTLLTCTPYGINTHRLLVRGKRVDYDDTKYIETGESLAKFDNGYIFFLGYKIP

Srt012

MGSSHHHHHHSSGLVPRGSHMMINNSKFKDGMNDYTETVKKLDKKDNTKLFKSAKKYNHSLTQ
TSIITDPFDEEAYKAIGAHYNDVLNVDGKGLIGYVVVPRIDVNLPIYHGSSKKVLEKGAGHLQNTSM
PIGGKSTHAVISAHTGFPDQTFFDNLTDLVKGDIFYIKVLDKTLAYKVDQIKVVLPEDTNDLRIIPNED
HVTLLTCTPYGINTHRLLVRGVRTKYVPEEVAKNKLVKQAPFEKCFYFLGYKIP

Srt015

MGSSHHHHHHSSGLVPRGSHMDRKHQQNVIQTYQGNIEANSEEKLQEMLGEAERYNEMLWQ
TNGVLIGDIEQGILEEESYQSQLNLSGTGVMGTLSIPKINVDLPIYHGTEEEILANGVGHLQESSLPV
GGENTHCILTGHRGLPNAKLFTRLDEMEQGDLFFLTVCGEKLAYEVTKIEIVHPEDVEGLRIQAEKD
LVSLITCTPYGLNTKRLVVTGERIPYTEKQEQEIVPGSMSFRE

Srt016

MGSSHHHHHHSSGLVPRGSHMERQHQKDTVATYQGSIEEEDKSRIQDAIAKASEYNNMLFQTQ
GASIGDLQNGILSEENYENLLNLSGTGVMGSIEIPKINVDIPIYHGTSEEVLASGVGHFQDSSLPVGG

NNTRCILTGHRGLPNSKLFTRLDELEKEDLFFISTCGETLAYRITEIEVVEPEEAELLEILPEKDLCTLITC
TPYGINTQRLVITGERVPYEKAEYDSIERKLPSFRE

Srt017

MGSSHHHHHHSSGLVPRGSHMISTYESEVKQTDKSKVKEQIQSAQKYNDMLFQTRGASVGNIDT
EILSDENYESILNLTGKGIMGSIEIPKIGVDLPIYHGTSDDVLSNGVGHLQNSSFPVGGENTRTVLTG
HRGLPNAKLFTRLDELKKDDLFYIHVGNKTLAYQIYKIEIVKKEEAPDVLGIEEGKDLATLLTCTPYGV
NTHRLILTGKRVPYSKKKKEAIEPEMMSWRE

Srt019

MGSSHHHHHHSSGLVPRGSHMALQFESDRNLAATTATTAKEVAGWPYPQAEDKLTAARAYNKK
LAESGQPILGEAVDPFAAAQGGSQASGEDSASKKDKEYQSLLNTGNGVMGTIKVPKQSINLPFYH
GTSEEALASGAGHLYGTSLPVGGKSTHSVITGHRGLVEALMFTRLDEVKEGDFFYIEVMGETLGYK
VDRISVILPDDTSKLKIVPGEDRVTLMTCTPYGVNTHRLLISGHRVAIPMPAPEPNDVLDARN

Srt021

MGSSHHHHHHSSGLVPRGSHMVAKARAYNRRLAATPQVIGELSGEDGAVKGDFGFKSDREYQS
LLDFGDGIMATIEIPSIGVDLPVRHGADAYALDNGLGHLHGTSLPVGGTSTHSVITGHTGVADKAL
FTRLTELRKGDVFYVKVAAQTLAYKVTRIRTVDPDDLRSVRVEPGRDLVTLVTCTPIFLNTYRLLVTG
ERASMPGDAPYPEDAPKTSRRDMRP

Srt022

MGSSHHHHHHSSGLVPRGSHMSDRAHASVTTGYTEVVASMPAGERAAELDRARAYNARLGQG
SLTDPYSSSFSGQVDPAERADYLDQLSLTGSGTIAGIRVPSVGIALPVYHGTSEQTLTLGVGHLEGSA
LPVGGTGTNSVLTGHSGVPQARLFTDLHGVEAGDLVYLDVMGETFAYEIDRTDVVLPTETDLLQP
VKGEDLVTLVTCTPVGVNSHRLVVQAHRVEVVDEVLGAATVAGRDGAGF

Srt023

MGSSHHHHHHSSGLVPRGSHMREWNTQREVESVIHKFDETYSKGIDSDITSEPVKTSDKSDSDKN
DTADVTGQPANESDNKSETADDDKEKTDTSSSKNNTVTNTPKPTSHELASEGDDNAEAEQQTAG
NTDSTENTSEDSETALSENTDSDSDASDDEESATRPYQTLYGEMEKYNKDLTTNGQDIVDAWSYE
QQPLDLSSVDIDEDNPVIGYIEIPDMKIRLPLMLGASTKNLEKGAAVLSETSMPIGGKDTNCVIAGH
RGWEGSAYFQFIENMKKGSKVYITNPWETLVYECTSTQVIYPDDVQSILIQPGKDMVTLFTCHPYV
LGGGPYRYLVFCERVDTQKRKEADGILNPDATTPAPAEEADNVVVTEPEEVTSSDDSTSESTDEDL

SVENPAQNPGNNDDLSVDKEEQSTQDEAGTGNTSTDISPTPAKETTLDENLENDPEVKKAEERGK
RLLALEQ

Srt024

MGSSHHHHHHSSGLVPRGSHMTTTDLKQQWKKSLQTVDAKETTKPLATKGDGLLTIPSLNFEQVI
LEGASTDVLDRSIGHIKETGAPGKGNYALAGHRSFTKGLHFNKLPQLKTGADVIVTTKDHRYTYRM
ESSKLVKPTDLSVLDQDVKQPMITLITCDPPETATNRLIKQGVLIKTETLD

Srt025

MGSSHHHHHHSSGLVPRGSHMEELKQRWTQDLKAVDAKETKRPVATSGAGLLTIPSLDFEQVILE
GASTDILDQSIGHIKQTGSPGKGNYALAGHRSFTKGLHFNRLPELKKGAHVIVTTKSHRYTYKMMT
SQLVKPTDVSVLNQDVKQATITLITCDPPETATNRLIKQGTLIKTESL

Srt026

MGSSHHHHHHSSGLVPRGSHMDHIQGQAANYVATNHLSRQRKNEQKKKPSYNMKAVQPVSP
QSLANAYQHRRDYRAVGQIAIRDHNVLLNIYRGVGNVELNLGAGTMNQNQKMGEGNYALAGH
NMDDGRSFFSPLYTAKVRGNLPNGTTILLTDYKKVYYYKITSSRFISVYNLRLAWNNKEFKKKPVISL
FTCDWTGQGRLFIRGKYTGSQDYKGASKYVRSSFNF

Srt027

MGSSHHHHHHSSGLVPRGSHMDAQELPVVGGIAIPEVGINLPIFKGLGNTELTYGAGTMKEDQV
MGGENNYSLASHHVFGIAGASDMLFSPLDKAKEGMKIYLTDKNKVYTYVISEVKVVQPTEVAVVD
DTPGKSEVTLVTCTDAEATQRTIVKGELKSQVDFDKASSDIIEAFNKSYNQFQN

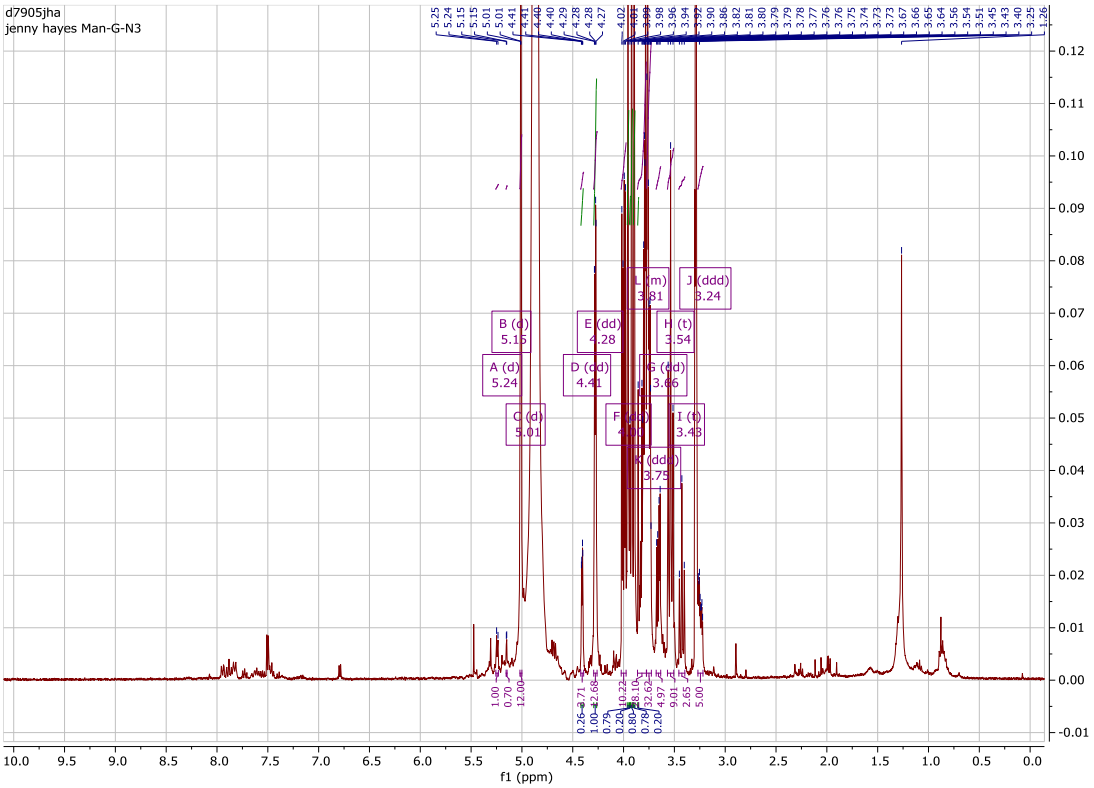
gMBP sequence

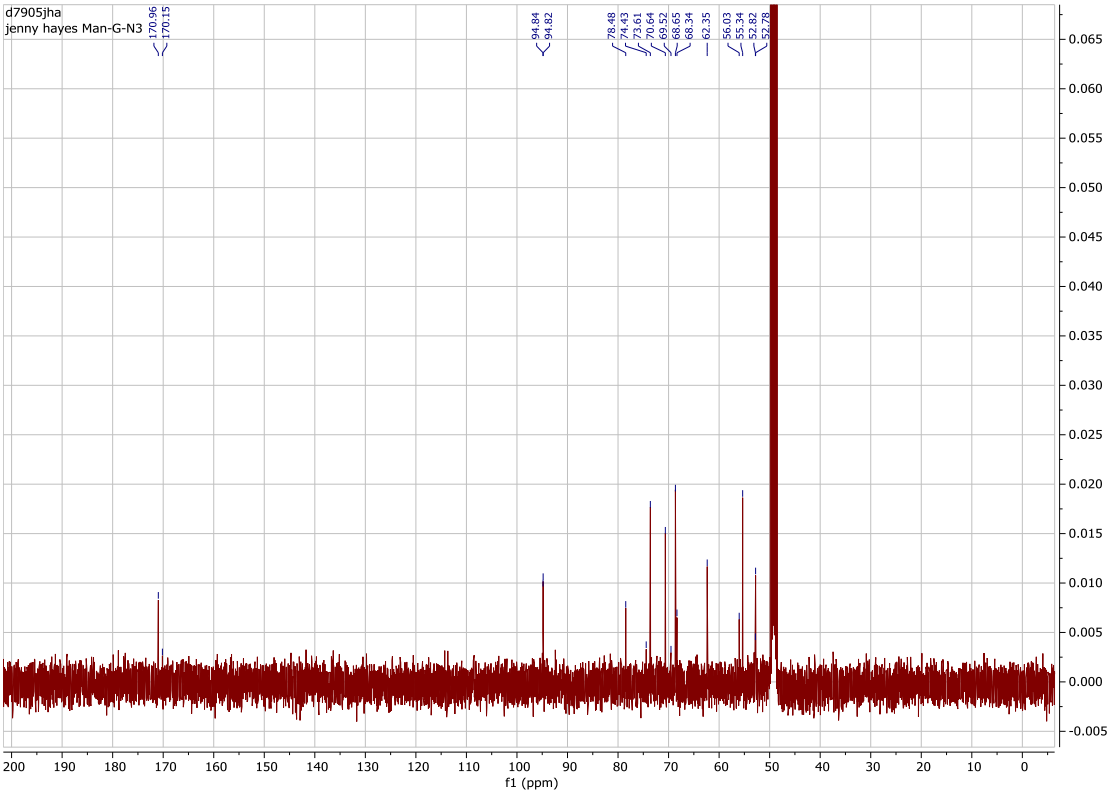
This is a modified version of hMBP-33 that binds galactose instead of mannose.

GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFLTNGEIMTFEKVKALCVKFQASVATPRN
AAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNEGQPDDWYGHGLGGGEDCVLLL
KNGQWNDVPCSTSHLAVCEFPI

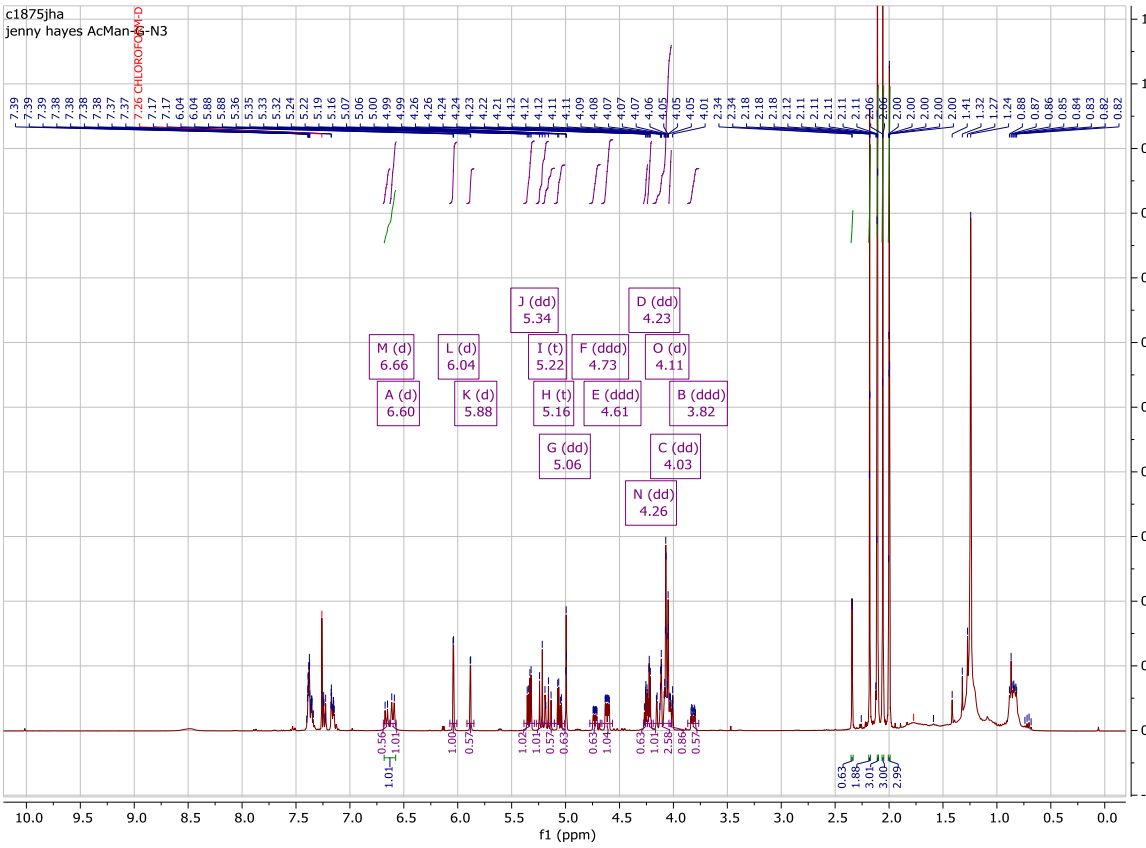
NMR Data

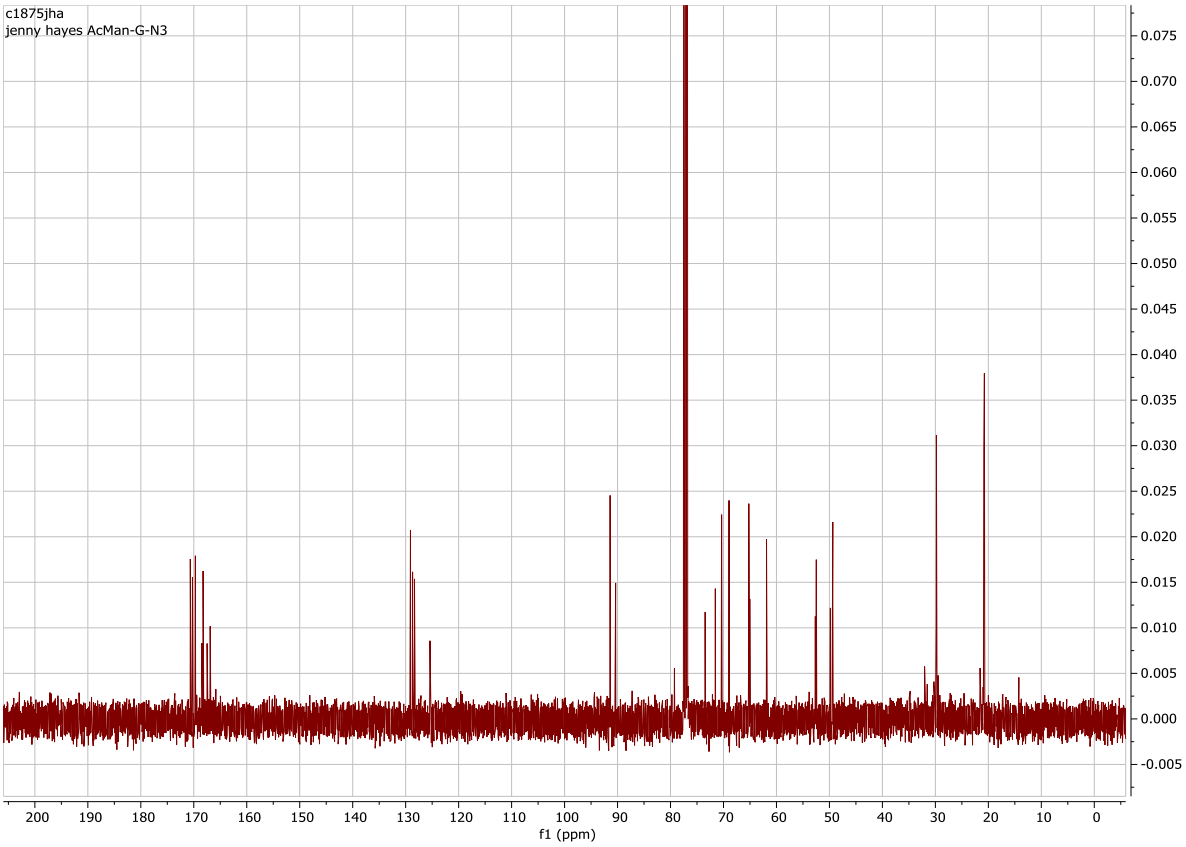
$ManGN_3$ **62**



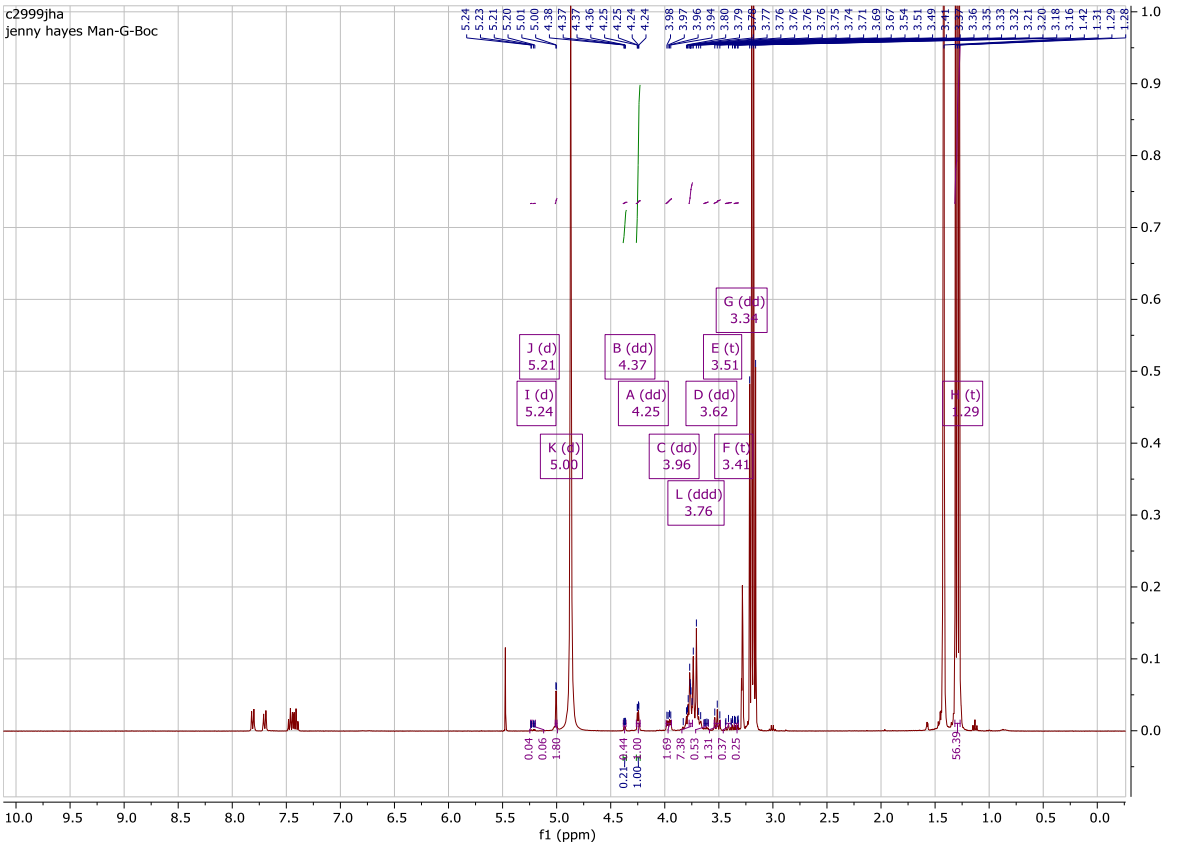


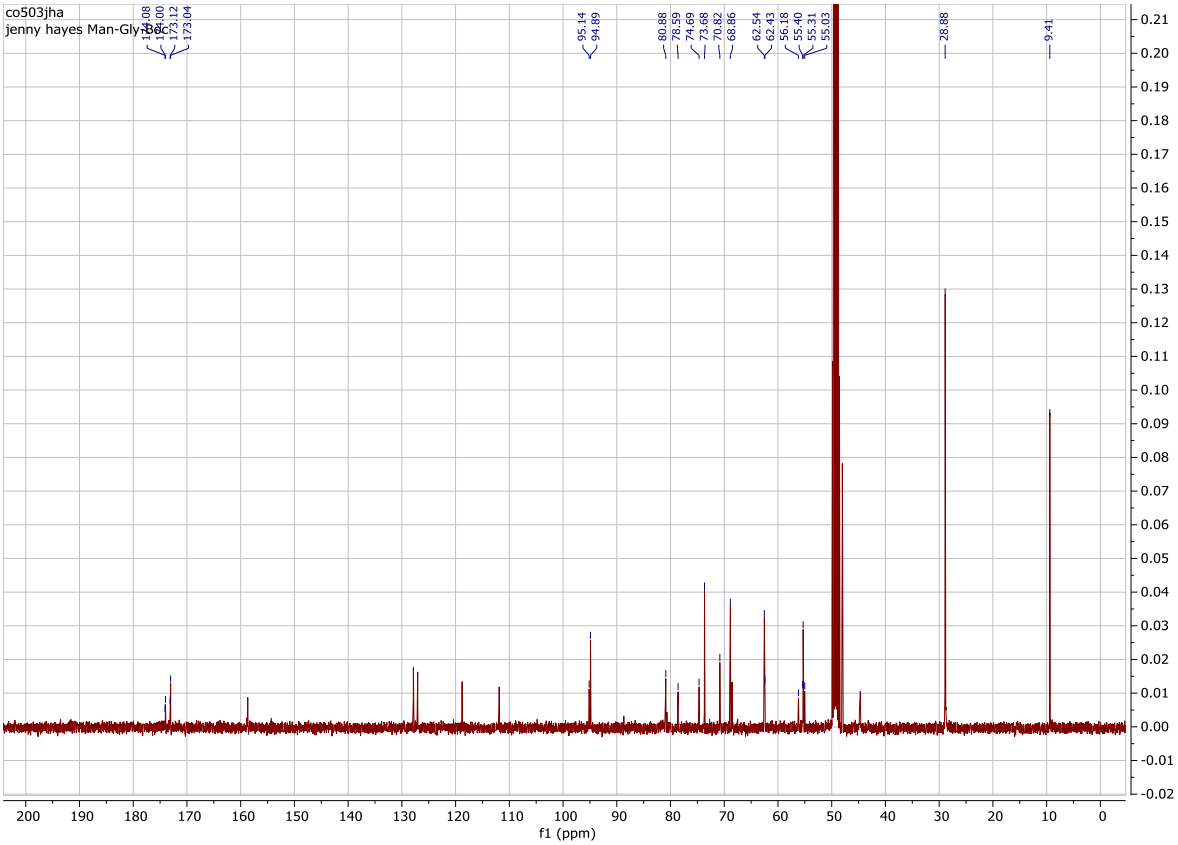
AcManGN$_3$ **4**


c1875jha
jenny hayes AcMan-G-N3
CHLOROFORM-D
J (dd) 5.34
D (dd) 4.23
M (d) 6.66
L (d) 6.04
I (t) 5.22
F (ddd) 4.73
O (d) 4.11
A (d) 6.60
K (d) 5.88
H (t) 5.16
E (ddd) 4.61
B (ddd) 3.82
G (dd) 5.06
C (dd) 4.03
N (dd) 4.26
f1 (ppm)

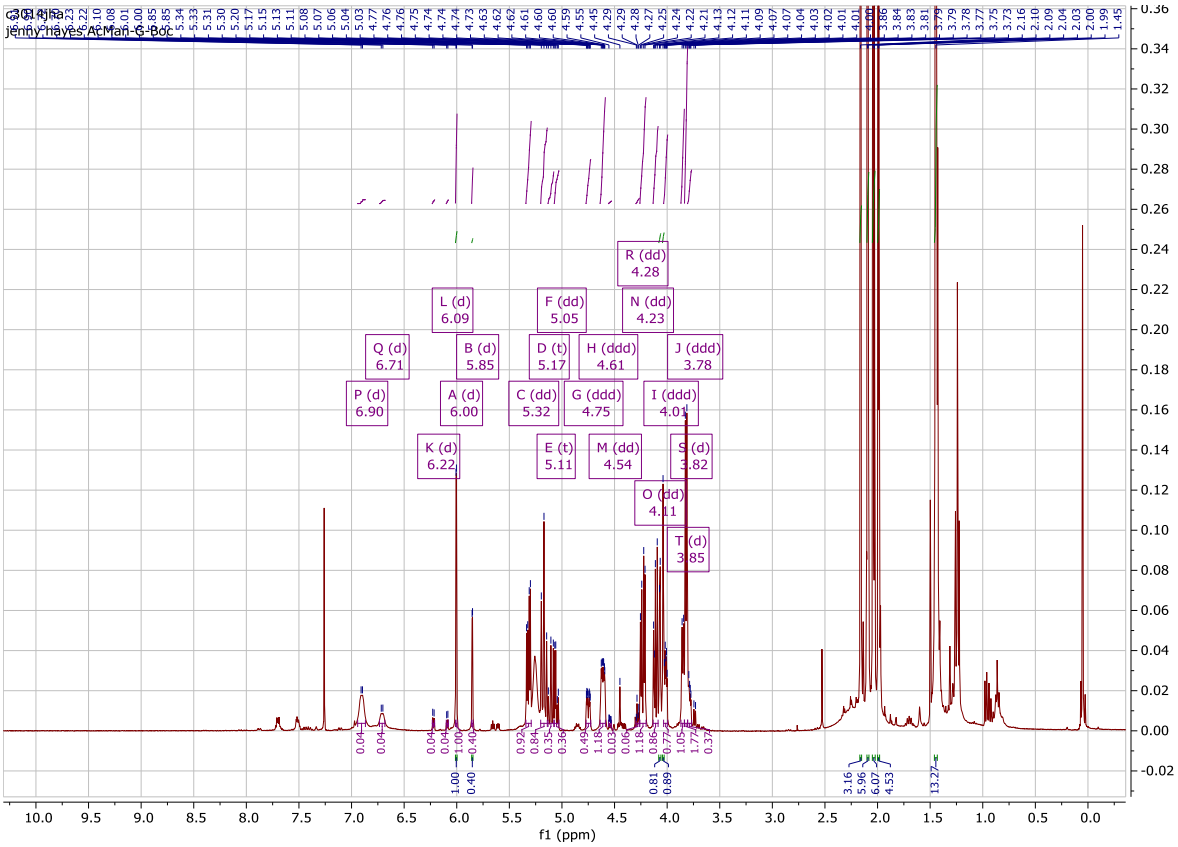

c1875jha
jenny hayes AcMan-G-N3
f1 (ppm)

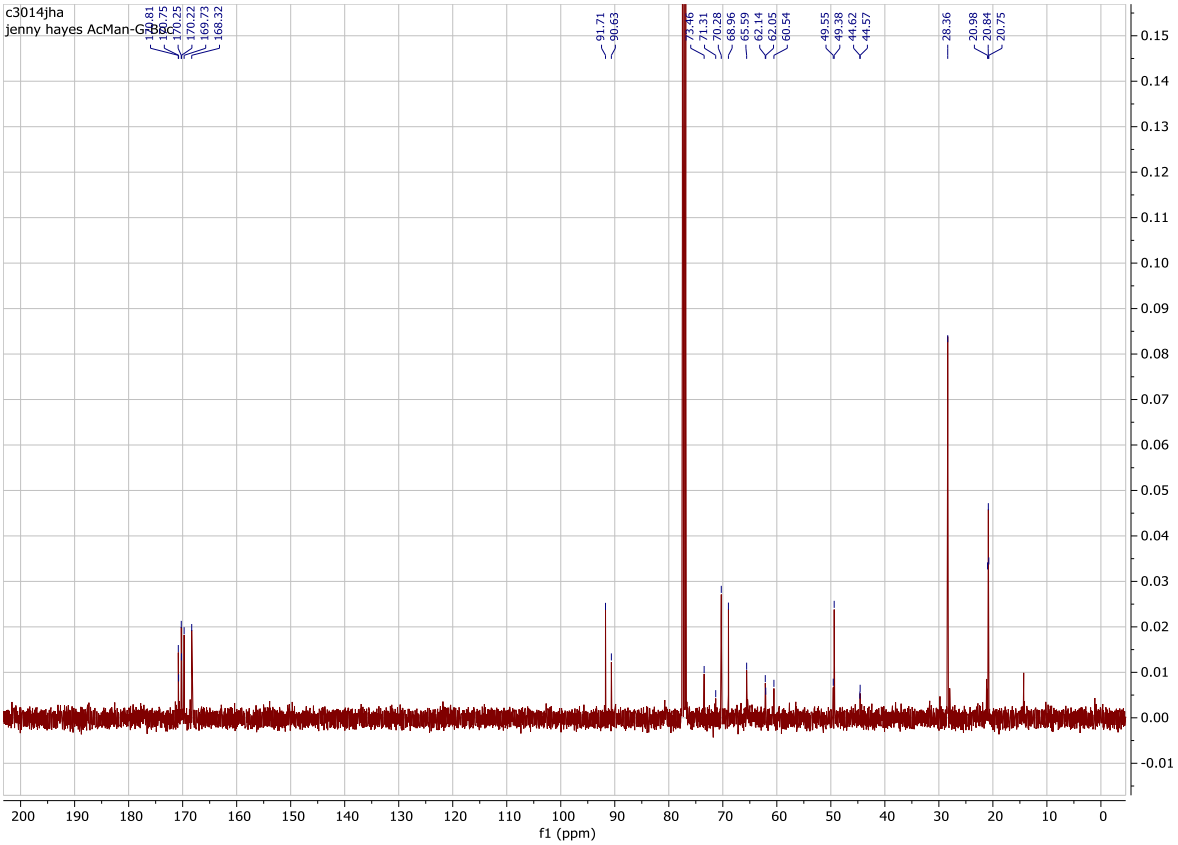
ManGBoc **64**


c2999jha
jenny hayes Man-G-Boc
J (d) 5.21
I (d) 5.24
K (d) 5.00
B (dd) 4.37
A (dd) 4.25
C (dd) 3.96
L (ddd) 3.76
D (dd) 3.62
E (t) 3.51
G (dd) 3.34
F (t) 3.41
H (t) 1.29
f1 (ppm)

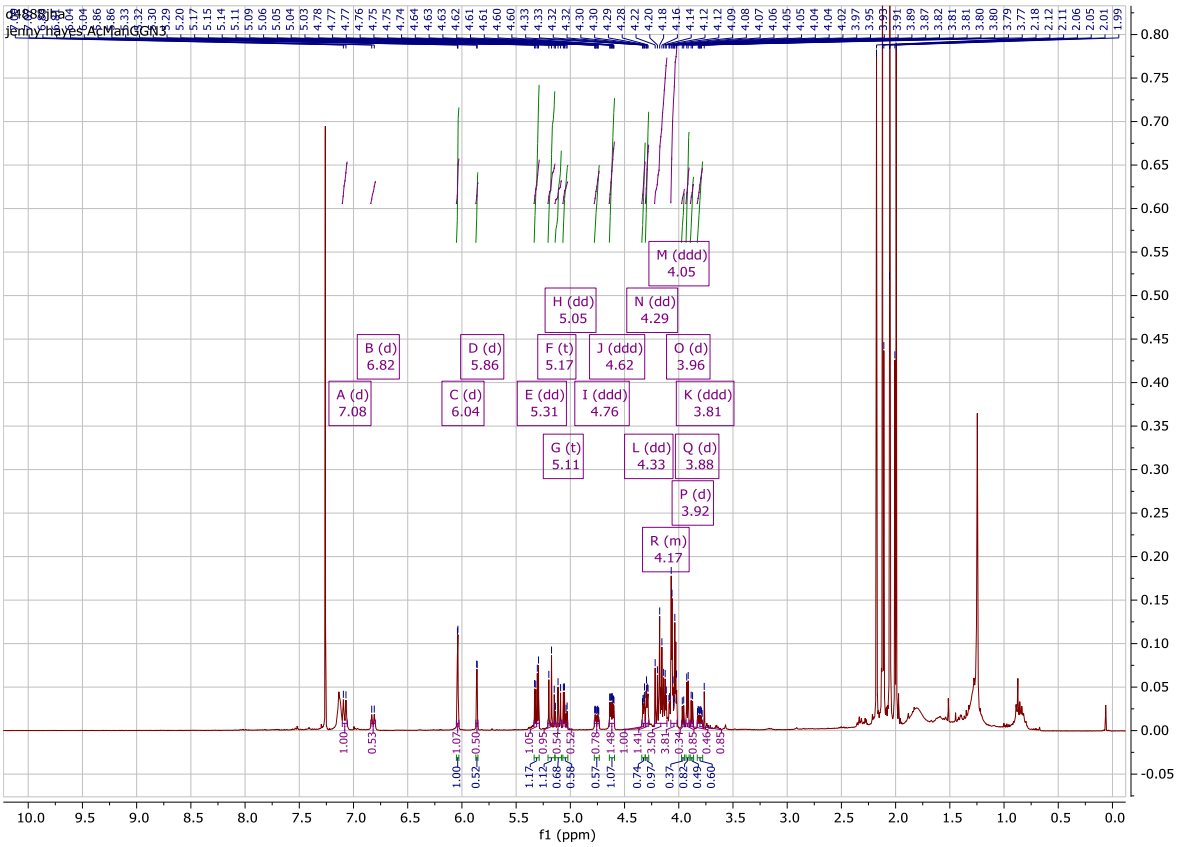

co503jha
jenny hayes Man-Gly-Boc
f1 (ppm)

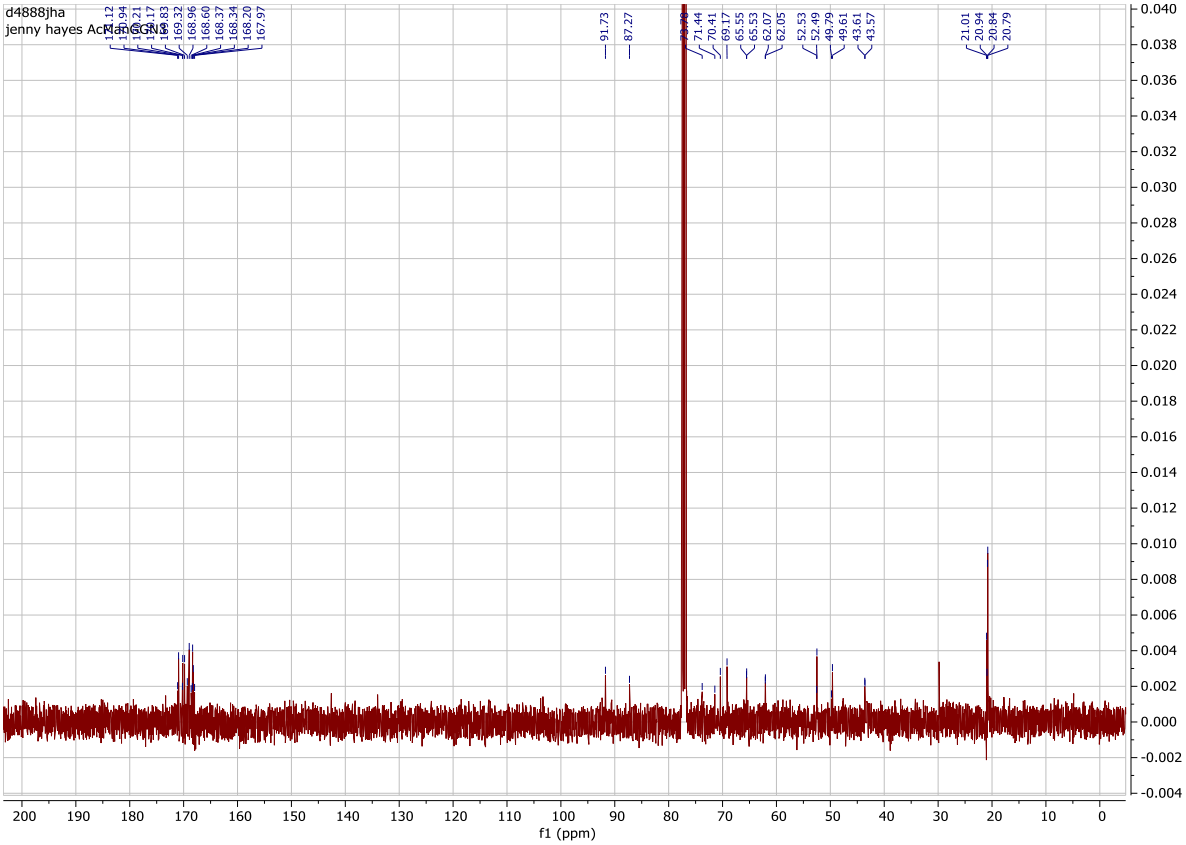
AcManGBoc **65**



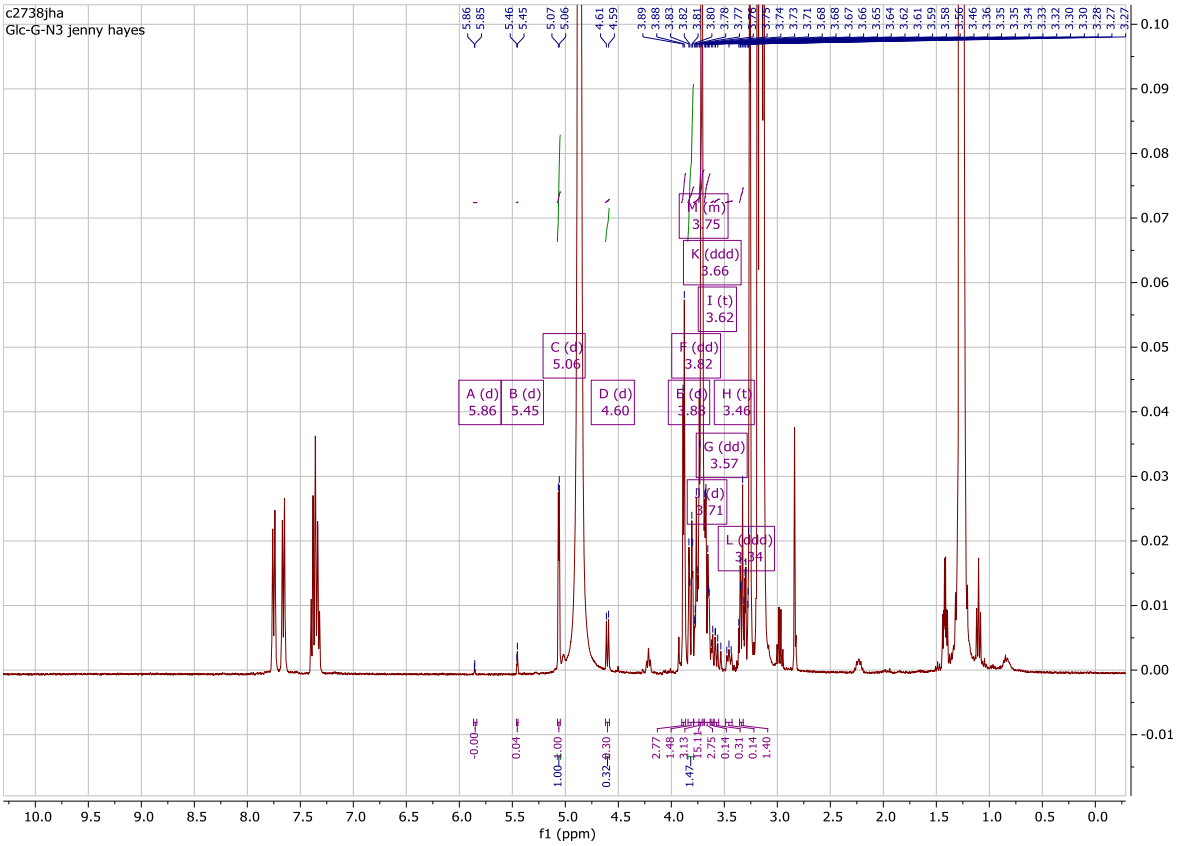


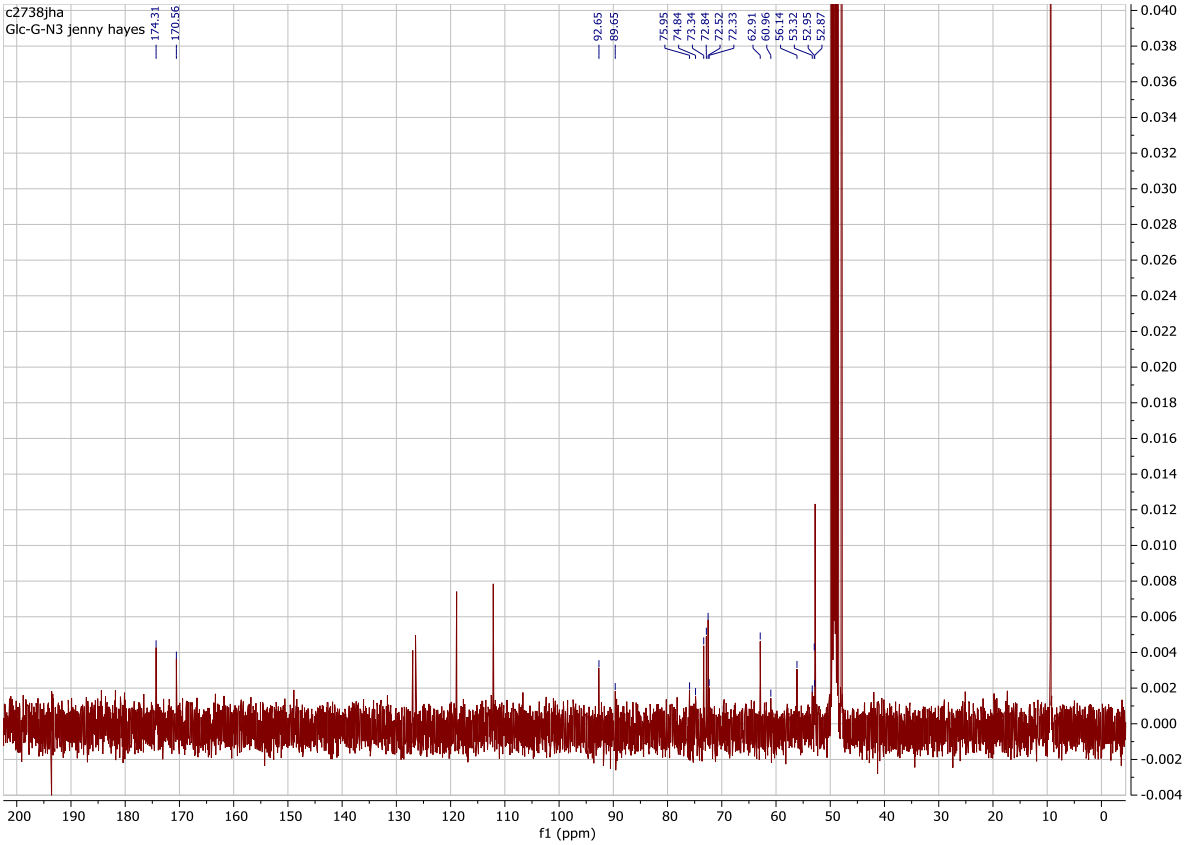
AcManGGN$_3$ **26**



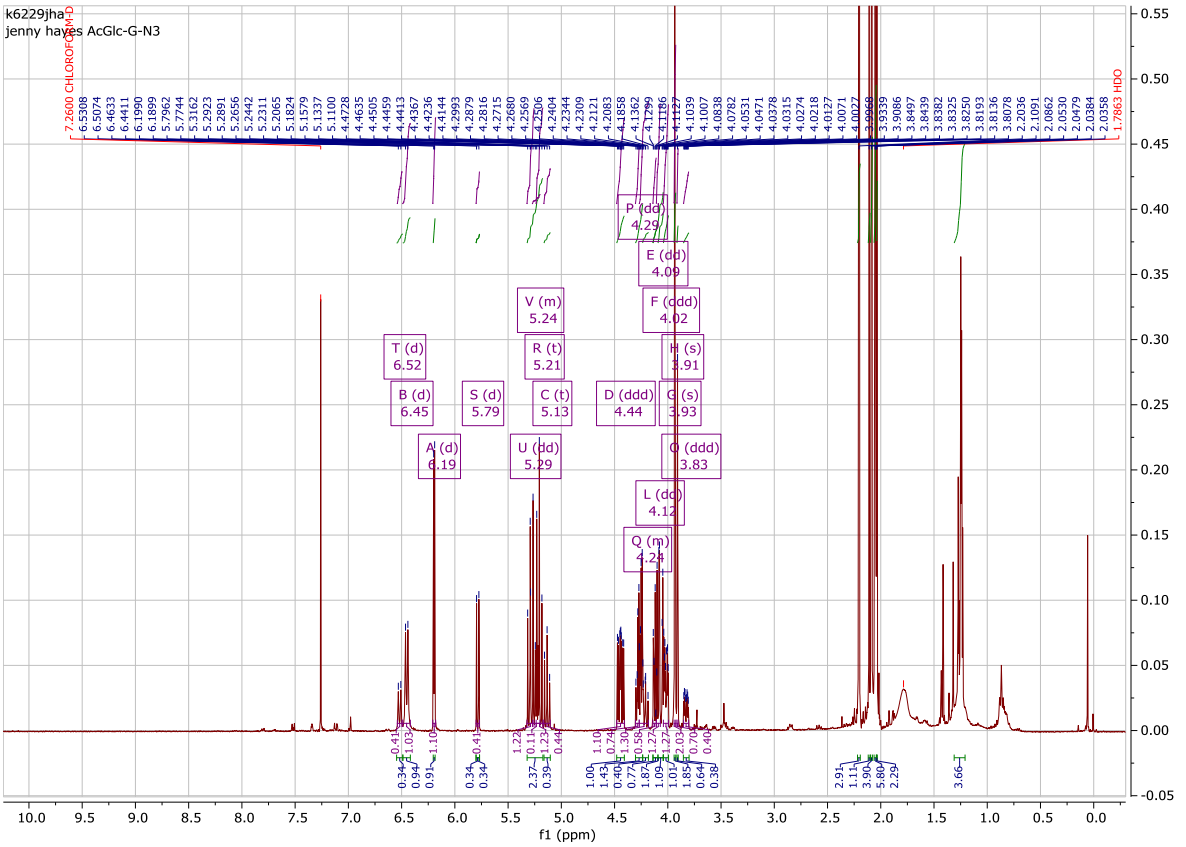


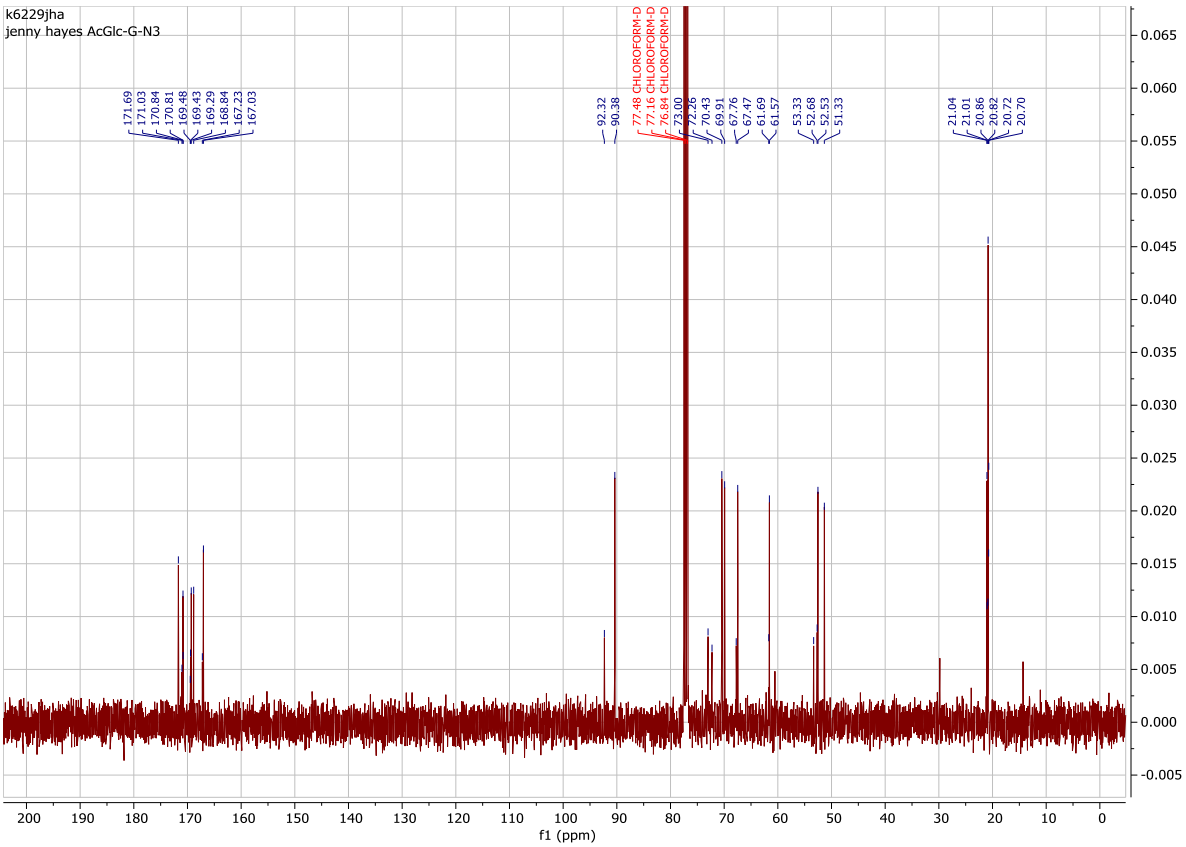
$GlcGN_3$ **70**


c2738jha
Glc-G-N3 jenny hayes
A (d) 5.86
B (d) 5.45
C (d) 5.06
D (d) 4.60
E (d) 3.88
F (dd) 3.82
G (dd) 3.57
H (t) 3.46
I (t) 3.62
K (ddd) 3.66
M (m) 3.75
L (ddd) 3.34
f1 (ppm)

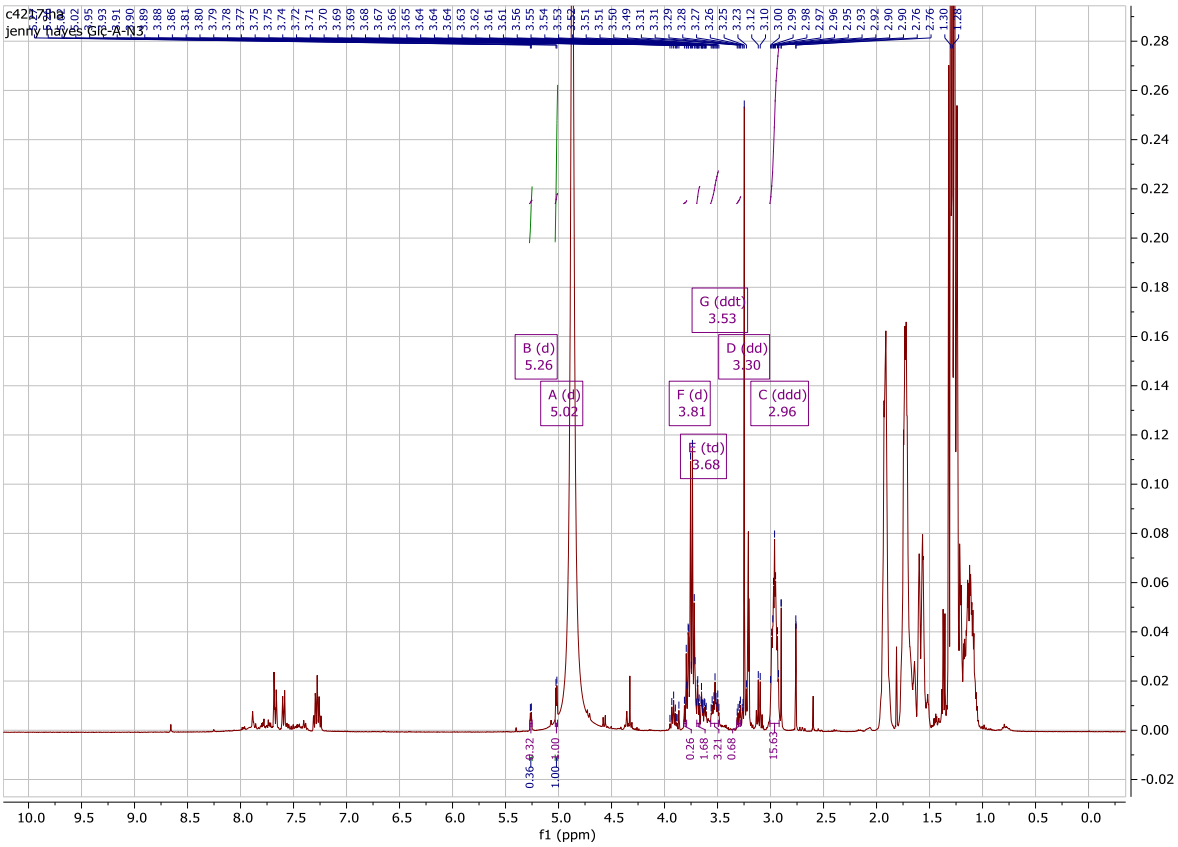

c2738jha
Glc-G-N3 jenny hayes
174.31
170.56
92.65
89.65
75.95
74.84
73.34
72.84
72.52
72.33
62.91
60.96
56.14
53.32
52.95
52.87
f1 (ppm)

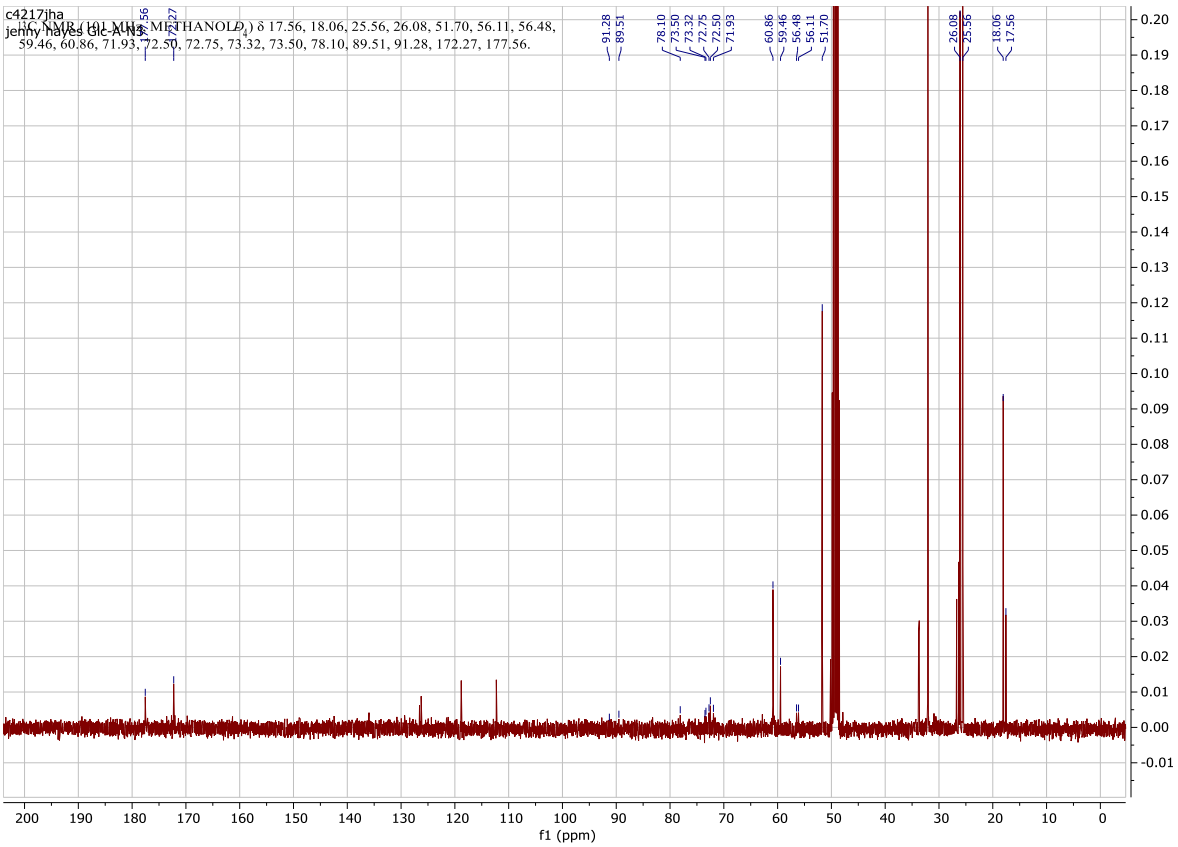
AcGlcGN$_3$ **6**


k6229jha
jenny hayes AcGlc-G-N3
P (dd) 4.29
E (dd) 4.09
V (m) 5.24
F (ddd) 4.02
T (d) 6.52
R (t) 5.21
H (s) 3.91
B (d) 6.45
S (d) 5.79
C (t) 5.13
D (ddd) 4.44
G (s) 3.93
A (d) 6.19
U (dd) 5.29
O (ddd) 3.83
L (dd) 4.12
Q (m) 4.24
f1 (ppm)

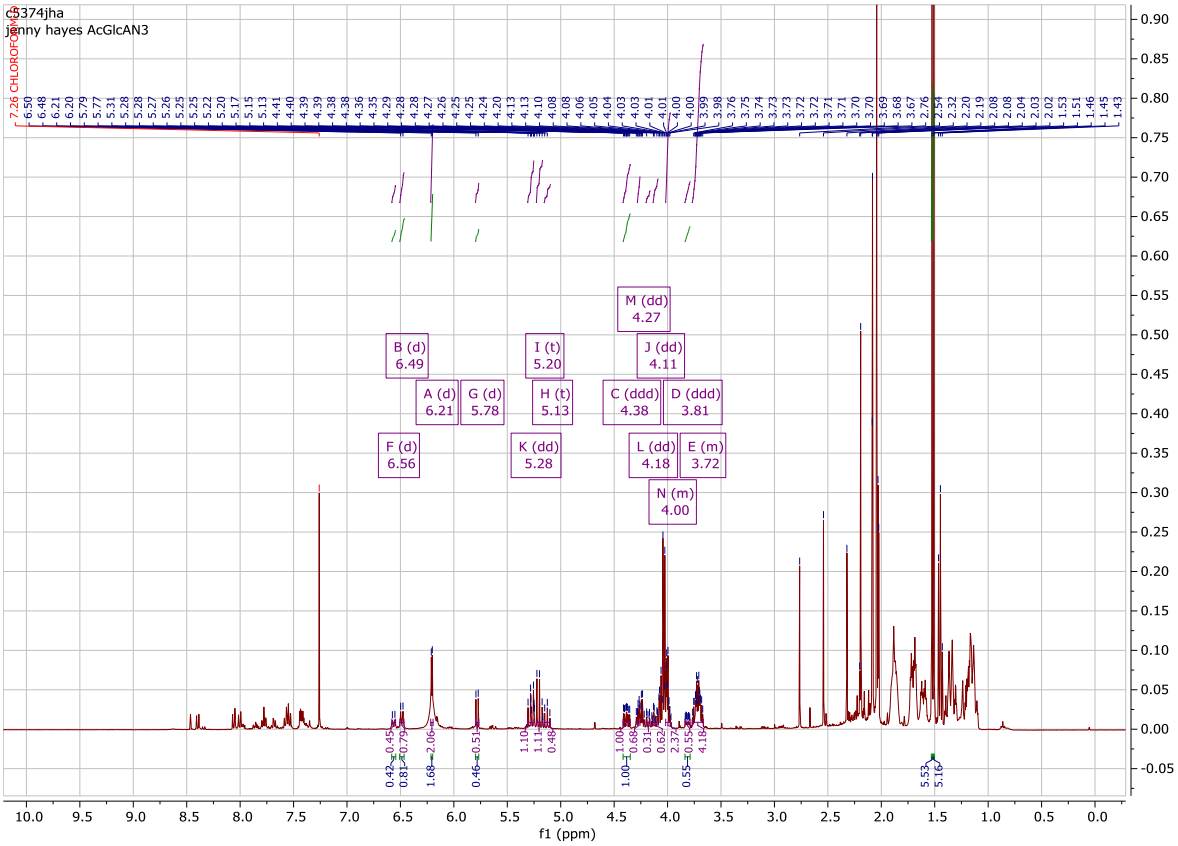

k6229jha
jenny hayes AcGlc-G-N3
f1 (ppm)

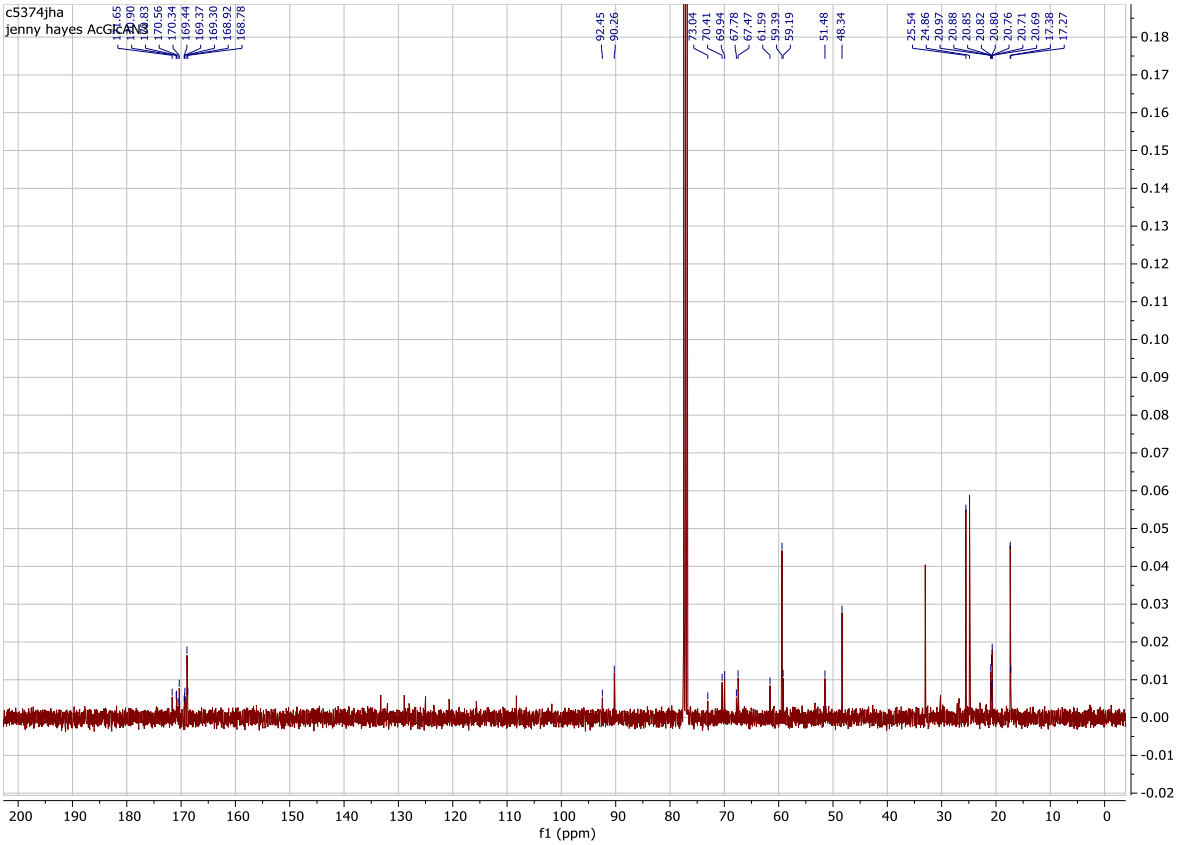
$GlcAN_3$ **72**


c4217jha
jenny hayes Glc-A-N3
B (d)
5.26
A (d)
5.02
G (ddt)
3.53
D (dd)
3.30
F (d)
3.81
C (ddd)
2.96
E (td)
3.68
f1 (ppm)


c4217jha
jenny hayes Glc-A-N3
13C NMR (101 MHz, METHANOL-d4) δ 17.56, 18.06, 25.56, 26.08, 51.70, 56.11, 56.48, 59.46, 60.86, 71.93, 72.50, 72.75, 73.32, 73.50, 78.10, 89.51, 91.28, 172.27, 177.56.
f1 (ppm)

$AcGlcAN_3$ **28**


c5374jha
jenny hayes AcGlcAN3
7.26 CHLOROFORM-D
B (d) 6.49
A (d) 6.21
G (d) 5.78
F (d) 6.56
I (t) 5.20
H (t) 5.13
K (dd) 5.28
M (dd) 4.27
J (dd) 4.11
C (ddd) 4.38
D (ddd) 3.81
L (dd) 4.18
E (m) 3.72
N (m) 4.00
f1 (ppm)


c5374jha
jenny hayes AcGlcAN3
f1 (ppm)

Abbreviations

2DPBA – 2-(diphenylphosphino) benzoic acid

9AzSia – 9-azido sialic acid

aa – amino acid

CASTing – Combinatorial active site saturation test

CHO – Chinese hamster ovary cells

CuAAC – Copper-catalysed azide-alkyne cycloaddition

CV – Column volume

Cy5 – Cyanine5

DAPI – 4',6-diamidino-2-phenylindole

DBCO – Dibenzocyclooctyne

DCM – Dichloromethane

DIPEA – *N,N*-Diisopropylethylamine

DMF – *N,N*-Dimethylformamide

EDC – 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride

EDTA – Ethylenediaminetetraacetic acid

E. coli – *Escherichia coli*

ESI-MS – Electrospray ionisation mass spectrometry

FACS – Fluorescence activated cell sorting

FBS – Fetal Bovine Serum

FITC – fluorescein isothiocyanate

FSC-A – Forward Scatter Area

Fuc – D-fucose

GABA – γ-aminobutyric acid

GAG – glycosaminoglycan

Gal – D-galactose

GalNAc – *N*-acetylgalactosamine

Glc – D-glucose

GlcNAc – *N*-acetylglucosamine

GLUT – Glucose transporter protein

GFP – Green fluorescent protein

gMBP – Galactose-binding protein

GRASP – Genomics-based related activity screening protocol

HCTU – 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate

HEK293 – Human Embryonic Kidney cells 293 (cell line)

HPLC – High performance liquid chromatography

IR – Infrared spectroscopy

Kdo – 3-deoxy-D-mannosoctulosonic acid

LC-MS – Liquid chromatography mass spectrometry

MALDI-MS/MS – Matrix-assisted laser desorption / ionization mass spectrometry

Man – D-mannose

ManBP – Mannose-binding protein

ManLev – *N*-levulinoyl mannosamine

ManNAc – *N*-acetylmannosamine

MFI – Mean fluorescence intensity

MOE – Metabolic oligosaccharide engineering

MRI – Magnetic resonance imaging

MTS –3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2- (4-sulfophenyl)-2H-tetrazolium inner salt

Neu5Ac – *N*-acetylneuraminic acid (sialic acid)

NMR – Nuclear magnetic resonance

OSTase – oligosaccharyltransferase

PBS – Phosphate-buffered saline solution

PCR – Polymerase chain reaction

PPh_3 – Triphenylphosphine

Pro-tetrazole – (*S*)-(-)-5-(2-Pyrrolidinyl)-1*H*-tetrazole

RT – room temperature

S. aureus – Staphylococcus aureus

S. pyogenes – Streptococcus pyogenes

SDS-PAGE – Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SGLT – sodium glucose symporter

SPAAC – Strain-promoted azide-alkyne cycloaddition

SPPS – Solid phase peptide synthesis

SSC-A – Side Scatter Area

TAMRA – 5-Carboxytetramethylrhodamine

TCEP – Tris(2-carboxyethyl)phosphine

TEA – Triethylamine

TFA – Trifluoroacetic acid

TIS – Triisopropylsilane

TLC – Thin layer chromatography

WT – Wild type

Xyl – D-xylose

References

1. Taylor, M.E. and K. Drickamer, *Introduction to Glycobiology*. 3rd ed. 2011: Oxford University Press.

2. Walsh, C.T., *Posttranslational Modification of Proteins*. 2006.

3. Ringenberg, M.A., S.M. Steenbergen, and E.R. Vimr, *The first committed step in the biosynthesis of sialic acid by Escherichia coli K1 does not involve a phosphorylated N-acetylmannosamine intermediate.* Mol. Microbiol., 2003. **50**(3): p. 961-75.

4. Aebi, M., *N-linked protein glycosylation in the ER.* Biochim. Biophys. Acta, 2013. **1833**(11): p. 2430-7.

5. Higel, F., et al., *N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins.* Eur. J. Pharm. Biopharm., 2016. **100**: p. 94-100.

6. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat. Rev. Mol. Cell Biol., 2012. **13**(7): p. 448-62.

7. Taylor, C.M., C.V. Karunaratne, and N. Xie, *Glycosides of hydroxyproline: Some recent, unusual discoveries.* Glycobiology, 2012. **22**(6): p. 757-767.

8. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins.* Nat. Methods, 2005. **2**(12): p. 905-9.

9. Zakeri, B., et al., *Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin.* Proc. Natl. Acad. Sci. U. S. A., 2012. **109**(12): p. E690-7.

10. A., V., et al., *Essentials of Glycobiology*. 2nd ed. 2009: Cold Spring Harbor Laboratory Press.

11. Saxon, E. and C.R. Bertozzi, *Chemical and biological strategies for engineering cell surface glycosylation.* Annu. Rev. Cell Dev. Biol., 2001. **17**: p. 1-23.

12. Dube, D.H. and C.R. Bertozzi, *Glycans in cancer and inflammation--potential for therapeutics and diagnostics.* Nat. Rev. Drug. Discov., 2005. **4**(6): p. 477-88.

13. Haun, R.S., et al., *Bioorthogonal labeling cell-surface proteins expressed in pancreatic cancer cells to identify potential diagnostic/therapeutic biomarkers.* Cancer Biol. Ther., 2015. **16**(10): p. 1557-1565.

14. Buck, C.A., M.C. Glick, and L. Warren, *A comparative study of glycoproteins from the surface of control and Rous sarcoma virus transformed hamster cells.* Biochemistry, 1970. **9**(23): p. 4567-76.

15. Bull, C., et al., *Targeting aberrant sialylation in cancer cells using a fluorinated sialic acid analog impairs adhesion, migration, and in vivo tumor growth.* Mol. Cancer Ther., 2013. **12**(10): p. 1935-46.

16. Munro, J.M., et al., *Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues.* Am. J. Pathol., 1992. **141**(6): p. 1397-408.

17. Nakamori, S., et al., *Increased expression of sialyl LewisX antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study.* Cancer Res., 1993. **53**(15): p. 3632-7.

18. Jorgensen, T., et al., *Up-regulation of the oligosaccharide sialyl LewisX: a new prognostic parameter in metastatic prostate cancer.* Cancer Res., 1995. **55**(9): p. 1817-9.

19. Kannagi, R., *Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression-The Warburg effect revisited.* Glycoconj. J., 2004. **20**(5): p. 353-64.

20. Munkley, J., *The Role of Sialyl-Tn in Cancer.* Int. J. Mol. Sci., 2016. **17**(3): p. 275.

21. Irie, A., et al., *The molecular basis for the absence of N-glycolylneuraminic acid in humans.* J. Biol. Chem., 1998. **273**(25): p. 15866-71.

22. Rich, S.M., et al., *The origin of malignant malaria.* Proc. Natl. Acad. Sci. U.S.A., 2009. **106**(35): p. 14902-7.

23. Nardy, A.F., et al., *The Sweet Side of Immune Evasion: Role of Glycans in the Mechanisms of Cancer Progression.* Front. Oncol., 2016. **6**: p. 54.

24. Samraj, A.N., et al., *A red meat-derived glycan promotes inflammation and cancer progression.* Proc. Natl. Acad. Sci. U.S.A., 2015. **112**(2): p. 542-7.

25. Fuster, M.M. and J.D. Esko, *The sweet and sour of cancer: glycans as novel therapeutic targets.* Nat. Rev. Cancer, 2005. **5**(7): p. 526-42.

26. Miyagi, T., et al., *Sialidase and malignancy: a minireview.* Glycoconj. J., 2004. **20**(3): p. 189-98.

27. Kato, T., et al., *Overexpression of lysosomal-type sialidase leads to suppression of metastasis associated with reversion of malignant phenotype in murine B16 melanoma cells.* Int. J. Cancer, 2001. **92**(6): p. 797-804.

28. Kakugawa, Y., et al., *Up-regulation of plasma membrane-associated ganglioside sialidase (Neu3) in human colon cancer and its involvement in apoptosis suppression.* Proc. Natl. Acad. Sci. U.S.A., 2002. **99**(16): p. 10718-23.

29. Granovsky, M., et al., *Suppression of tumor growth and metastasis in Mgat5-deficient mice.* Nat. Med., 2000. **6**(3): p. 306-12.

30. Kobata, A., *Altered glycosylation of surface glycoproteins in tumor cells and its clinical application.* Pigment Cell Res, 1989. **2**(4): p. 304-8.

31. Dennis, J.W., et al., *UDP-N-acetylglucosamine:alpha-6-D-mannoside beta1,6 N-acetylglucosaminyltransferase V (Mgat5) deficient mice.* Biochim. Biophys. Acta., 2002. **1573**(3): p. 414-22.

32. Dennis, J.W., M. Granovsky, and C.E. Warren, *Glycoprotein glycosylation and cancer progression.* Biochim. Biophys. Acta, 1999. **1473**(1): p. 21-34.

33. Tsuboi, S., et al., *Two opposing roles of O-glycans in tumor metastasis.* Trends Mol. Med., 2012. **18**(4): p. 224-32.

34. Hart, G.W., *Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins.* Annu. Rev. Biochem., 1997. **66**: p. 315-35.

35. Comer, F.I. and G.W. Hart, *O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate.* J. Biol. Chem., 2000. **275**(38): p. 29179-82.

36. Gao, Y., et al., *Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain.* J. Biol. Chem., 2001. **276**(13): p. 9838-45.

37. Arnold, C.S., et al., *The microtubule-associated protein tau is extensively modified with O-linked N-acetylglucosamine.* J. Biol. Chem., 1996. **271**(46): p. 28741-4.

38. Comstock, L.E. and D.L. Kasper, *Bacterial glycans: key mediators of diverse host immune responses.* Cell, 2006. **126**(5): p. 847-50.

39. Gautam, S., et al., *Wall teichoic acids prevent antibody binding to epitopes within the cell wall of Staphylococcus aureus.* ACS Chem. Biol., 2016. **11**(1): p. 25-30.

40. Cole, J.N., et al., *M protein and hyaluronic acid capsule are essential for in vivo selection of covRS mutations characteristic of invasive serotype M1T1 group A Streptococcus.* MBio., 2010. **1**(4).

41. Severi, E., D.W. Hood, and G.H. Thomas, *Sialic acid utilization by bacterial pathogens.* Microbiology, 2007. **153**(Pt 9): p. 2817-22.

42. Bouchet, V., et al., *Host-derived sialic acid is incorporated into Haemophilus influenzae lipopolysaccharide and is a major virulence factor in experimental otitis media.* Proc. Natl. Acad. Sci. U.S.A., 2003. **100**(15): p. 8898-903.

43. Coyne, M.J., et al., *Human symbionts use a host-like pathway for surface fucosylation.* Science, 2005. **307**(5716): p. 1778-81.

44. Ermert, D. and M. Laabei, *Catch Me if You Can: Streptococcus pyogenes Complement Evasion Strategies.* J. Innate Immun., 2018: p. 1-10.

45. Hentrich, K., et al., *Streptococcus pneumoniae Senses a Human-like Sialic Acid Profile via the Response Regulator CiaR.* Cell Host Microbe, 2016. **20**(3): p. 307-317.

46. Kim, H.K., et al., *Recurrent infections and immune evasion strategies of Staphylococcus aureus.* Curr. Opin. Microbiol., 2012. **15**(1): p. 92-9.

47. Jankowski, M.D., et al., *Sialic acid on avian erythrocytes.* Comp Biochem Physiol B Biochem Mol Biol, 2019. **238**: p. 110336.

48. Boriskin, Y.S., et al., *Arbidol: a broad-spectrum antiviral compound that blocks viral fusion.* Curr. Med. Chem., 2008. **15**(10): p. 997-1005.

49. Lew, W., X. Chen, and C.U. Kim, *Discovery and development of GS 4104 (oseltamivir): an orally active influenza neuraminidase inhibitor.* Curr. Med. Chem., 2000. **7**(6): p. 663-72.

50. Gray, E., J. Hogwood, and B. Mulloy, *The anticoagulant and antithrombotic mechanisms of heparin*, in *Handb. Exp. Pharmacol.* 2012. p. 43-61.

51. Goon, S. and C.R. Bertozzi, *Metabolic substrate engineering as a tool for glycobiology (Reprinted from Glycochemistry: Principles, Synthesis, and Applications, pg 641-674, 2001).* J. Carbohydr. Chem., 2002. **21**(7-9): p. 943-977.

52. Grammel, M. and H.C. Hang, *Chemical reporters for biological discovery.* Nat. Chem. Biol., 2013. **9**(8): p. 475-484.

53. Tuley, A., et al., *A genetically encoded aldehyde for rapid protein labelling.* Chem. Commun., 2014. **50**(56): p. 7424-6.

54. Prescher, J.A. and C.R. Bertozzi, *Chemistry in living systems.* Nat. Chem. Biol., 2005. **1**(1): p. 13-21.

55. Berg, J., J. Tymoczko, and L. Stryer, *Biochemistry*. 6th ed. 2006: W. H. Freeman.

56. Uldry, M., et al., *GLUT2 is a high affinity glucosamine transporter.* FEBS Lett, 2002. **524**(1-3): p. 199-203.

57. Wright, E.M., D.D. Loo, and B.A. Hirayama, *Biology of human sodium glucose transporters.* Physiol. Rev., 2011. **91**(2): p. 733-94.

58. Rodriguez, P., et al., *Redefining the facilitated transport of mannose in human cells: absence of a glucose-insensitive, high-affinity facilitated mannose transport system.* Biochemistry, 2005. **44**(1): p. 313-20.

59. Stasche, R., et al., *A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Molecular cloning and functional expression of UDP-N-acetyl-glucosamine 2-epimerase/N-acetylmannosamine kinase.* J. Biol. Chem., 1997. **272**(39): p. 24319-24.

60. Hang, H.C. and C.R. Bertozzi, *Chemoselective approaches to glycoprotein assembly.* Acc. Chem. Res., 2001. **34**(9): p. 727-36.

61. Saxon, E., et al., *Investigating cellular metabolism of synthetic azidosugars with the Staudinger ligation.* J. Am. Chem. Soc., 2002. **124**(50): p. 14893-902.

62. Wratil, P.R., R. Horstkorte, and W. Reutter, *Metabolic Glycoengineering with N-Acyl Side Chain Modified Mannosamines.* Angew. Chem. Int. Ed. Engl., 2016. **55**(33): p. 9482-512.

63. Tanner, M.E., *The enzymes of sialic acid biosynthesis.* Bioorg. Chem., 2005. **33**(3): p. 216-28.

64. Keppler, O.T., et al., *Biochemical engineering of the N-acyl side chain of sialic acid: biological implications.* Glycobiology, 2001. **11**(2): p. 11R-18R.

65. Oetke, C., et al., *Evidence for efficient uptake and incorporation of sialic acid by eukaryotic cells.* Eur. J. Biochem., 2001. **268**(16): p. 4553-61.

66. Mantey, L.R., et al., *Efficient biochemical engineering of cellular sialic acids using an unphysiological sialic acid precursor in cells lacking UDP-N-acetylglucosamine 2-epimerase.* FEBS Lett., 2001. **503**(1): p. 80-4.

67. Vocadlo, D.J., et al., *A chemical approach for identifying O-GlcNAc-modified proteins in cells.* Proc. Natl. Acad. Sci. U.S.A., 2003. **100**(16): p. 9116-21.

68. Laughlin, S.T. and C.R. Bertozzi, *Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation.* Nat. Protoc., 2007. **2**(11): p. 2930-2944.

69. Spaete, A.-K., et al., *Terminal Alkenes as Versatile Chemical Reporter Groups for Metabolic Oligosaccharide Engineering.* Chem. Eur. J., 2014. **20**(50): p. 16502-16508.

70. Kayser, H., et al., *New amino sugar analogues are incorporated at different rates into glycoproteins of mouse organs.* Experientia, 1993. **49**(10): p. 885-7.

71. Mahal, L.K., K.J. Yarema, and C.R. Bertozzi, *Engineering chemical reactivity on cell surfaces through oligosaccharide biosynthesis.* Science, 1997. **276**(5315): p. 1125-8.

72. DeChancie, J. and K.N. Houk, *The origins of femtomolar protein-ligand binding: hydrogen-bond cooperativity and desolvation energetics in the biotin-(strept)avidin binding site.* J. Am. Chem. Soc., 2007. **129**(17): p. 5419-29.

73. Saxon, E. and C.R. Bertozzi, *Cell surface engineering by a modified Staudinger reaction.* Science, 2000. **287**(5460): p. 2007-10.

74. Dumont, M., et al., *Plant cell wall imaging by metabolic click-mediated labelling of rhamnogalacturonan II using azido 3-deoxy-d-manno-oct-2-ulosonic acid.* Plant J., 2016. **85**(3): p. 437-447.

75. Agard, N.J., J.A. Prescher, and C.R. Bertozzi, *A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems.* J. Am. Chem. Soc., 2004. **126**(46): p. 15046-7.

76. Jones, M.B., et al., *Characterization of the cellular uptake and metabolic conversion of acetylated N-acetylmannosamine (ManNAc) analogues to sialic acids.* Biotechnol. Bioeng., 2004. **85**(4): p. 394-405.

77. Antonczak, A.K., Z. Simova, and E.M. Tippmann, *A critical examination of Escherichia coli esterase activity.* J. Biol. Chem., 2009. **284**(42): p. 28795-800.

78. Xiong, D.-C., et al., *Rapid probing of sialylated glycoproteins in vitro and in vivo via metabolic oligosaccharide engineering of a minimal cyclopropene reporter.* Org. Biomol. Chem., 2015. **13**(13): p. 3911-3917.

79. Yoon, H.I., et al., *Bioorthogonal Copper Free Click Chemistry for Labeling and Tracking of Chondrocytes In Vivo.* Bioconjugate Chem., 2016. **27**(4): p. 927-936.

80. Borrmann, A. and J.C.M. van Hest, *Bioorthogonal chemistry in living organisms.* Chem. Sci., 2014. **5**(6): p. 2123-2134.

81. Neves, A.A., et al., *Imaging sialylated tumor cell glycans in vivo.* FASEB J., 2011. **25**(8): p. 2528-2537, 10.1096/fj.10-178590.

82. Hanson, S.R., et al., *Tailored Glycoproteomics and Glycan Site Mapping Using Saccharide-Selective Bioorthogonal Probes.* J. Am. Chem. Soc., 2007. **129**(23): p. 7266-7267.

83. Jiang, H., et al., *Tracking Surface Glycans on Live Cancer Cells with Single-Molecule Sensitivity.* Angew. Chem., Int. Ed., 2015. **54**(6): p. 1765-1769.

84. Kang, K., et al., *Tissue-based metabolic labeling of polysialic acids in living primary hippocampal neurons.* Proc. Natl. Acad. Sci. U. S. A., 2015. **112**(3): p. E241-E248.

85. Boyce, M., et al., *Metabolic cross-talk allows labeling of O-linked β-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway.* Proc. Natl. Acad. Sci. U. S. A., 2011(Feb. 7 2011): p. 1-6, 6 pp.

86. Zaro, B.W., et al., *Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1.* Proc. Natl. Acad. Sci. U. S. A., 2011. **108**(20): p. 8146-51.

87. Hang, H.C. and C.R. Bertozzi, *Ketone isosteres of 2-N-acetamidosugars as substrates for metabolic cell surface engineering.* J. Am. Chem. Soc., 2001. **123**(6): p. 1242-3.

88. Moller, H., et al., *Glycan-specific metabolic oligosaccharide engineering of C7-substituted sialic acids.* Angew. Chem. Int. Ed. Engl., 2012. **51**(24): p. 5986-90.

89. Zaro, B.W., K.N. Chuh, and M.R. Pratt, *Chemical Reporter for Visualizing Metabolic Cross-Talk between Carbohydrate Metabolism and Protein Modification.* ACS Chem. Biol., 2014. **9**(9): p. 1991-1996.

90. Rostovtsev, V.V., et al., *A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.* Angew. Chem. Int. Ed. Engl., 2002. **41**(14): p. 2596-9.

91. Woo, C.M., et al., *Development of IsoTaG, a Chemical Glycoproteomics Technique for Profiling Intact N- and O-Glycopeptides from Whole Cell Proteomes.* J. Proteome Res., 2017. **16**(4): p. 1706-1718.

92. Spiciarich, D.R., et al., *Bioorthogonal Labeling of Human Prostate Cancer Tissue Slice Cultures for Glycoproteomics.* Angew. Chem. Int. Ed. Engl., 2017. **56**(31): p. 8992-8997.

93. Nazarova, L.A., et al., *Extracellular Toxoplasma gondii tachyzoites metabolize and incorporate unnatural sugars into cellular proteins.* Microbes Infect., 2016. **18**(3): p. 199-210.

94. Josa-Cullere, L., et al., *Diazo group as a new chemical reporter for bioorthogonal labelling of biomolecules.* RSC Adv., 2014. **4**(94): p. 52241-52244.

95. Benito-Alifonso, D., et al., *Imidazolium-tagged glycan probes for non-covalent labeling of live cells.* Chem. Commun., 2016. **52**(27): p. 4906-4909.

96. Chang, P.V., et al., *Imaging Cell Surface Glycans with Bioorthogonal Chemical Reporters.* J. Am. Chem. Soc., 2007. **129**(27): p. 8400-8401.

97. Cole, C.M., et al., *Fluorescent Live-Cell Imaging of Metabolically Incorporated Unnatural Cyclopropene-Mannosamine Derivatives.* ChemBioChem, 2013. **14**(2): p. 205-208.

98. Niederwieser, A., et al., *Two-color glycan labeling of live cells by a combination of Diels-Alder and click chemistry.* Angew. Chem. Int. Ed. Engl., 2013. **52**(15): p. 4265-8.

99. Gilormini, P.A., et al., *A sequential bioorthogonal dual strategy: ManNAl and SiaNAl as distinct tools to unravel sialic acid metabolic pathways.* Chem. Commun., 2016. **52**(11): p. 2318-2321.

100. Feng, L., et al., *Bifunctional unnatural sialic acids for dual metabolic labeling of cell-surface sialylated glycans.* J. Am. Chem. Soc., 2013. **135**(25): p. 9244-7.

101. Batt, A.R., et al., *Metabolic Chemical Reporters of Glycans Exhibit Cell-Type-Selective Metabolism and Glycoprotein Labeling.* Chembiochem, 2017. **18**(13): p. 1177-1182.

102. Chuh, K.N., et al., *Changes in Metabolic Chemical Reporter Structure Yield a Selective Probe of O-GlcNAc Modification.* J. Am. Chem. Soc., 2014. **136**(35): p. 12283-12295.

103. Zaro, B.W., et al., *The Small Molecule 2-Azido-2-deoxy-glucose Is a Metabolic Chemical Reporter of O-GlcNAc Modifications in Mammalian Cells, Revealing an Unexpected Promiscuity of O-GlcNAc Transferase.* ACS Chem. Biol., 2017. **12**(3): p. 787-794.

104. Darabedian, N., et al., *The Metabolic Chemical Reporter 6-Azido-6-deoxy-glucose Further Reveals the Substrate Promiscuity of O-GlcNAc Transferase and Catalyzes the Discovery of Intracellular Protein Modification by O-Glucose.* J. Am. Chem. Soc., 2018. **140**(23): p. 7092-7100.

105. Sun, Y., et al., *Mechanistic Investigation and Multiplexing of Liposome-Assisted Metabolic Glycan Labeling.* J. Am. Chem. Soc., 2018.

106. Xie, R., et al., *Cell-selective metabolic glycan labeling based on ligand-targeted liposomes.* J. Am. Chem. Soc., 2012. **134**(24): p. 9914-9917.

107. Xie, R., et al., *Targeted imaging and proteomic analysis of tumor-associated glycans in living animals.* Angew. Chem. Int. Ed. Engl., 2014. **53**(51): p. 14082-6.

108. Xie, R., et al., *In vivo metabolic labeling of sialoglycans in the mouse brain by using a liposome-assisted bioorthogonal reporter strategy.* Proc. Natl. Acad. Sci. U. S. A., 2016. **113**(19): p. 5173-5178.

109. Robinson, P.V., et al., *Live-Cell Labeling of Specific Protein Glycoforms by Proximity-Enhanced Bioorthogonal Ligation.* J. Am. Chem. Soc., 2015. **137**(33): p. 10452-10455.

110. Rabuka, D., et al., *A chemical reporter strategy to probe glycoprotein fucosylation.* J. Am. Chem. Soc., 2006. **128**(37): p. 12078-9.

111. Hsu, T.L., et al., *Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells.* Proc. Natl. Acad. Sci. U.S.A., 2007. **104**(8): p. 2614-9.

112. Wang, W., Y. Zhu, and X. Chen, *Selective Imaging of Gram-Negative and Gram-Positive Microbiotas in the Mouse Gut.* Biochemistry, 2017. **56**(30): p. 3889-3893.

113. Dumont, A., et al., *Click-Mediated Labeling of Bacterial Membranes through Metabolic Modification of the Lipopolysaccharide Inner Core.* Angewandte Chemie International Edition, 2012. **51**(13): p. 3143-3146.

114. Geva-Zatorsky, N., et al., *In vivo imaging and tracking of host-microbiota interactions via metabolic labeling of gut anaerobic bacteria.* Nat. Med., 2015. **21**(9): p. 1091-1100.

115. Sadamoto, R., et al., *Cell-wall engineering of living bacteria.* J. Am. Chem. Soc., 2002. **124**(31): p. 9018-9.

116. Foley, H.N., et al., *Bioorthogonal Chemical Reporters for Selective In Situ Probing of Mycomembrane Components in Mycobacteria.* Angew. Chem., Int. Ed., 2016. **55**(6): p. 2053-2057.

117. Gautam, S., et al., *Exterior design: strategies for redecorating the bacterial surface with small molecules.* Trends Biotechnol., 2013. **31**(4): p. 258-267.

118. Liu, F., et al., *The engineering of bacteria bearing azido-pseudaminic acid-modified flagella.* Chembiochem, 2009. **10**(8): p. 1317-20.

119. Mas Pons, J., et al., *Identification of Living Legionella pneumophila Using Species-Specific Metabolic Lipopolysaccharide Labeling.* Angew. Chem., Int. Ed., 2014. **53**(5): p. 1275-1278.

120. Collins, B.E., et al., *Conversion of cellular sialic acid expression from N-acetyl- to N-glycolylneuraminic acid using a synthetic precursor, N-glycolylmannosamine pentaacetate: inhibition of myelin-associated glycoprotein binding to neural cells.* Glycobiology, 2000. **10**(1): p. 11-20.

121. Herrmann, M., et al., *Consequences of a subtle sialic acid modification on the murine polyomavirus receptor.* J. Virol., 1997. **71**(8): p. 5922-31.

122. Keppler, O.T., et al., *Elongation of the N-acyl side chain of sialic acids in MDCK II cells inhibits influenza A virus infection.* Biochem. Biophys. Res. Commun., 1998. **253**(2): p. 437-42.

123. Memmel, E., et al., *Metabolic glycoengineering of Staphylococcus aureus reduces its adherence to human T24 bladder carcinoma cells.* Chem. Commun., 2013. **49**(66): p. 7301-7303.

124. Almaraz, R.T., et al., *Metabolic oligosaccharide engineering with N-Acyl functionalized ManNAc analogs: cytotoxicity, metabolic flux, and glycan-display considerations.* Biotechnol. Bioeng., 2012. **109**(4): p. 992-1006.

125. Paul, B., R.J. Bernacki, and W. Korytnyk, *Synthesis and biological activity of some 1-N-substituted 2-acetamido-2-deoxy-beta-D-glycopyranosylamine derivatives and related analogs.* Carbohydr. Res., 1980. **80**(1): p. 99-115.

126. Sharma, M., et al., *Fluorinated carbohydrates as potential plasma membrane modifiers. Synthesis of 3-deoxy-3-fluoro derivatives of 2-acetamido-2-deoxy-D-hexopyranoses.* Carbohydr. Res., 1993. **240**: p. 85-93.

127. Witte, C., et al., *Live-cell MRI with Xenon Hyper-CEST Biosensors Targeted to Metabolically Labeled Cell-Surface Glycans.* Angew. Chem., Int. Ed., 2015. **54**(9): p. 2806-2810.

128. Neves, A.A., et al., *Imaging Glycosylation In Vivo by Metabolic Labeling and Magnetic Resonance Imaging.* Angew. Chem., Int. Ed., 2016. **55**(4): p. 1286-1290.

129. Kieseier, B.C., *The mechanism of action of interferon-beta in relapsing multiple sclerosis.* CNS Drugs, 2011. **25**(6): p. 491-502.

130. Luchansky, S.J., et al., *Metabolic Functionalization of Recombinant Glycoproteins.* Biochemistry, 2004. **43**(38): p. 12358-12366.

131. Breidenbach, M.A., et al., *Targeted metabolic labeling of yeast N-glycans with unnatural sugars.* Proc. Natl. Acad. Sci. U.S.A., 2010. **107**(9): p. 3988-93.

132. Lee, J.H., et al., *Engineering novel cell surface receptors for virus-mediated gene transfer.* J. Biol. Chem., 1999. **274**(31): p. 21878-84.

133. Vanbeselaere, J., et al., *Alkynyl monosaccharide analogues as a tool for evaluating Golgi glycosylation efficiency: application to Congenital Disorders of Glycosylation (CDG).* Chem. Commun., 2013. **49**(96): p. 11293-5.

134. Chang, P.V., et al., *A strategy for the selective imaging of glycans using caged metabolic precursors.* J. Am. Chem. Soc., 2010. **132**(28): p. 9516-8.

135. Mazmanian, S.K., et al., *Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall.* Science, 1999. **285**(5428): p. 760-3.

136. Ton-That, H., et al., *Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif.* Proc. Natl. Acad. Sci. U.S.A., 1999. **96**(22): p. 12424-9.

137. Mazmanian, S.K., et al., *An iron-regulated sortase anchors a class of surface protein during Staphylococcus aureus pathogenesis.* Proc. Natl. Acad. Sci. U.S.A., 2002. **99**(4): p. 2293-8.

138. Ton-That, H. and O. Schneewind, *Assembly of pili on the surface of Corynebacterium diphtheriae.* Mol. Microbiol., 2003. **50**(4): p. 1429-38.

139. Spirig, T., E.M. Weiner, and R.T. Clubb, *Sortase enzymes in Gram-positive bacteria.* Mol. Microbiol., 2011. **82**(5): p. 1044-59.

140. Chang, C., et al., *Cell surface display of minor pilin adhesins in the form of a simple heterodimeric assembly in Corynebacterium diphtheriae.* Mol. Microbiol., 2011. **79**(5): p. 1236-47.

141. Di Girolamo, S., et al., *Characterization of the housekeeping sortase from the human pathogen Propionibacterium acnes: first investigation of a class F sortase.* Biochem. J., 2019. **476**(4): p. 665-682.

142. Popp, M., et al., *Sortagging: a versatile method for protein labeling.* Nat. Chem. Biol., 2007. **3**(11): p. 707-708.

143. Jacobitz, A.W., et al., *Sortase Transpeptidases: Structural Biology and Catalytic Mechanism.* Adv. Protein Chem. Struct Biol., 2017. **109**: p. 223-264.

144. Fottner, M., et al., *Site-specific ubiquitylation and SUMOylation using genetic-code expansion and sortase.* Nat. Chem. Biol., 2019. **15**(3): p. 276-284.

145. Williamson, D.J., et al., *Efficient N-terminal labeling of proteins by use of sortase.* Angew. Chem., Int. Ed., 2012. **51**(37): p. 9377-80.

146. Williamson, D.J., M.E. Webb, and W.B. Turnbull, *Depsipeptide substrates for sortase-mediated N-terminal protein ligation.* Nat. Protoc., 2014. **9**(2): p. 253-62.

147. Williamson, D.J., *Sortase-mediated modification and intracellular delivery of cholera toxin analogues*, in *School of Chemistry*. 2014, Leeds.

148. Zong, Y., et al., *Crystal structures of Staphylococcus aureus sortase A and its substrate complex.* J. Biol. Chem., 2004. **279**(30): p. 31383-9.

149. Ilangovan, U., et al., *Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus.* Proc. Natl. Acad. Sci. U. S. A., 2001. **98**(11): p. 6056-61.

150. Chen, I., B.M. Dorr, and D.R. Liu, *A general strategy for the evolution of bond-forming enzymes using yeast display.* Proc. Natl. Acad. Sci. U.S.A., 2011. **108**(28): p. 11399-404.

151. Zaccolo, M., et al., *An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues.* J. Mol. Biol., 1996. **255**(4): p. 589-603.

152. Sarpong, K. and R. Bose, *Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins.* Anal. Biochem., 2017. **521**: p. 55-58.

153. Freund, C. and D. Schwarzer, *Engineered Sortases in Peptide and Protein Chemistry.* Chembiochem, 2021. **22**(8): p. 1347-1356.

154. Chen, L., et al., *Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency.* Sci. Rep., 2016. **6**: p. 31899.

155. Hirakawa, H., S. Ishikawa, and T. Nagamune, *Design of Ca2+-independent Staphylococcus aureus sortase A mutants.* Biotechnol. Bioeng., 2012. **109**(12): p. 2955-61.

156. Hirakawa, H., S. Ishikawa, and T. Nagamune, *Ca2+ -independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of Escherichia coli.* Biotechnol. J., 2015. **10**(9): p. 1487-92.

157. Yamamura, Y., et al., *Enhancement of sortase A-mediated protein ligation by inducing a beta-hairpin structure around the ligation site.* Chem. Commun., 2011. **47**(16): p. 4742-4.

158. Zou, Z., et al., *Enhancing Robustness of Sortase A by Loop Engineering and Backbone Cyclization.* Chemistry, 2020. **26**(60): p. 13537.

159. Zou, Z., et al., *Directed sortase A evolution for efficient site-specific bioconjugations in organic co-solvents.* Chem. Commun. (Camb). 2018. **54**(81): p. 11467-11470.

160. Dorr, B.M., et al., *Reprogramming the specificity of sortase enzymes.* Proc. Natl. Acad. Sci. U.S.A., 2014. **111**(37): p. 13343-8.

161. Schmohl, L., et al., *Engineering sortase A by screening a second-generation library using phage display.* J Pept Sci, 2017. **23**(7-8): p. 631-635.

162. Guimaraes, C.P., et al., *Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions.* Nat. Protoc., 2013. **8**(9): p. 1787-99.

163. Chang, C., et al., *In vitro reconstitution of sortase-catalyzed pilus polymerization reveals structural elements involved in pilin cross-linking.* Proc. Natl. Acad. Sci. U. S. A., 2018. **115**(24): p. E5477-E5486.

164. Susmitha, A., K.M. Nampoothiri, and H. Bajaj, *Insights into the biochemical and functional characterization of sortase E transpeptidase of Corynebacterium glutamicum.* Biochem. J., 2019. **476**(24): p. 3835-3847.

165. Reetz, M.T., et al., *Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test.* Angew. Chem. Int. Ed., 2005. **44**(27): p. 4192-6.

166. Brabham, R.L., et al., *Palladium-unleashed proteins: gentle aldehyde decaging for site-selective protein modification.* Chem. Commun., 2018. **54**(12): p. 1501-1504.

167. Spears, R.J., et al., *Site-selective C-C modification of proteins at neutral pH using organocatalyst-mediated cross aldol ligations.* Chem. Sci., 2018.

168. Zeng, Y., et al., *High-efficiency labeling of sialylated glycoproteins on living cells.* Nat. Methods, 2009. **6**(3): p. 207-9.

169. Li, Z., et al., *Bioorthogonal chemistry for selective recognition, separation and killing bacteria over mammalian cells.* Chem. Commun., 2016. **52**(17): p. 3482-3485.

170. Buttke, T.M., J.A. McCubrey, and T.C. Owen, *Use of an aqueous soluble tetrazolium/formazan assay to measure viability and proliferation of lymphokine-dependent cell lines.* J. Immunol. Methods, 1993. **157**(1-2): p. 233-40.

171. Baskin, J.M., et al., *Copper-free click chemistry for dynamic in vivo imaging.* Proc. Natl. Acad. Sci. U.S.A., 2007. **104**(43): p. 16793-7.

172. Adan, A., et al., *Flow cytometry: basic principles and applications.* Crit. Rev. Biotechnol., 2017. **37**(2): p. 163-176.

173. Lo Conte, M., et al., *Multi-molecule reaction of serum albumin can occur through thiol-yne coupling.* Chem. Commun., 2011. **47**(39): p. 11086-8.

174. Ning, X., et al., *Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions.* Angew. Chem. Int. Ed. Engl., 2008. **47**(12): p. 2253-5.

175. Race, P.R., et al., *Crystal structure of Streptococcus pyogenes sortase A: implications for sortase mechanism.* J. Biol. Chem., 2009. **284**(11): p. 6924-33.

176. Hong, V., et al., *Labeling live cells by copper-catalyzed alkyne--azide click chemistry.* Bioconjug. Chem., 2010. **21**(10): p. 1912-6.

177. Kapuscinski, J., *DAPI: a DNA-specific fluorescent probe.* Biotech. Histochem., 1995. **70**(5): p. 220-33.

178. Hosseini, M.J., et al., *New mechanistic approach of inorganic palladium toxicity: impairment in mitochondrial electron transfer.* Metallomics, 2016. **8**(2): p. 252-9.

179. Naruse, N., et al., *Copper-Mediated Deprotection of Thiazolidine and Selenazolidine Derivatives Applied to Native Chemical Ligation.* J. Org. Chem., 2020. **85**(3): p. 1425-1433.

180. Susmitha, A., H. Bajaj, and K. Madhavan Nampoothiri, *The divergent roles of sortase in the biology of Gram-positive bacteria.* Cell Surf., 2021. **7**: p. 100055.

181. Ton-That, H., et al., *Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates.* J. Biol. Chem., 2000. **275**(13): p. 9876-81.

182. Duong, A., et al., *Aerial development in Streptomyces coelicolor requires sortase activity.* Mol. Microbiol., 2012. **83**(5): p. 992-1005.

183. Cossart, P. and R. Jonquieres, *Sortase, a universal target for therapeutic agents against gram-positive bacteria?* Proc. Natl. Acad. Sci. U.S.A., 2000. **97**(10): p. 5013-5.

184. Iqbal, H.A., Z. Feng, and S.F. Brady, *Biocatalysts and small molecule products from metagenomic studies.* Curr. Opin. Chem. Biol., 2012. **16**(1-2): p. 109-16.

185. Barnett, T.C. and J.R. Scott, *Differential recognition of surface proteins in Streptococcus pyogenes by two sortase gene homologs.* J. Bacteriol., 2002. **184**(8): p. 2181-91.

186. Kang, H.J., et al., *Crystal structure of Spy0129, a Streptococcus pyogenes class B sortase involved in pilus assembly.* PLoS One, 2011. **6**(1): p. e15969.

187. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.

188. Shetty, S.A., Zuffa, S., Bui, T. P. N., Aalvink, S., Smidt, H., De Vos, W. M., *Reclassification of Eubacterium hallii as Anaerobutyricum hallii gen. nov., comb. nov., and description of Anaerobutyricum soehngenii sp. nov., a butyrate and propionate-producing bacterium from infant faeces.* Int. J. Syst. Evol. Microbiol., 2018. **68**(12): p. 3741-3746.

189. Iino, T., et al., *Oscillibacter valericigenes gen. nov., sp. nov., a valerate-producing anaerobic bacterium isolated from the alimentary canal of a Japanese corbicula clam.* Int. J. Syst. Evol. Microbiol., 2007. **57**(Pt 8): p. 1840-1845.

190. Rajilic-Stojanovic, M. and W.M. de Vos, *The first 1000 cultured species of the human gastrointestinal microbiota.* FEMS Microbiol. Rev., 2014. **38**(5): p. 996-1047.

191. Liu, C., et al., *Reclassification of Clostridium coccoides, Ruminococcus hansenii, Ruminococcus hydrogenotrophicus, Ruminococcus luti, Ruminococcus productus and Ruminococcus schinkii as Blautia coccoides gen. nov., comb. nov., Blautia hansenii comb. nov., Blautia hydrogenotrophica comb. nov., Blautia luti comb. nov., Blautia producta comb. nov., Blautia schinkii comb. nov. and description of Blautia wexlerae sp. nov., isolated from human faeces.* Int J Syst Evol Microbiol, 2008. **58**(Pt 8): p. 1896-902.

192. Holdeman, L.V., and Moore, W. E. C., *New Genus, Coprococcus, Twelve New Species, and Emended Descriptions of Four Previously Described Species of Bacteria from Human Feces.* Int. J. Syst. Evol. Microbiol., 1974. **24**(2): p. 260-277.

193. Moore, W.E.C., Johnson, J. L., Holdeman, L. V., *Emendation of Bacteroidaceae and Butyrivibrio and Descriptions of Desulfomonas gen. nov. and Ten New Species in the Genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus.* Int. J. Syst. Evol. Microbiol., 1976. **26**(2): p. 238-252.

194. Winter, J., et al., *Mode of action of steroid desmolase and reductases synthesized by Clostridium "scindens" (formerly Clostridium strain 19).* J. Lipid. Res., 1984. **25**(10): p. 1124-31.

195. Seshadri, R., et al., *Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection.* Nat. Biotechnol., 2018. **36**(4): p. 359-367.

196. Sakata, S., et al., *Unification of Bifidobacterium infantis and Bifidobacterium suis as Bifidobacterium longum.* Int. J. Syst. Evol. Microbiol., 2002. **52**(Pt 6): p. 1945-1951.

197. Duranti, S., et al., *Evaluation of genetic diversity among strains of the human gut commensal Bifidobacterium adolescentis.* Sci. Rep., 2016. **6**: p. 23971.

198. Yumoto, I., et al., *Exiguobacterium oxidotolerans sp. nov., a novel alkaliphile exhibiting high catalase activity.* Int. J. Syst. Evol. Microbiol., 2004. **54**(Pt 6): p. 2013-2017.

199. Fruhling, A., et al., *Exiguobacterium undae sp. nov. and Exiguobacterium antarcticum sp. nov.* Int. J. Syst. Evol. Microbiol., 2002. **52**(Pt 4): p. 1171-1176.

200. Cato, E.P., Moore, W. E. C., Johnson, J. L., *Synonymy of Strains of "Lactobacillus acidophilus" Group A2 (Johnson et 81. 1980) with the Type Strain of Lactobacillus crispatus (Brygoo and Aladame 1953) Moore and Holdeman 1970* Int J Syst Evol Microbiol, 1983. **33**(2): p. 426-428.

201. Delorme, C., et al., *Genomics of Streptococcus salivarius, a major human commensal.* Infect. Genet. Evol., 2015. **33**: p. 381-92.

202. Dramsi, S., P. Trieu-Cuot, and H. Bierne, *Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria.* Res. Microbiol., 2005. **156**(3): p. 289-97.

203. Gaspar, A.H. and H. Ton-That, *Assembly of distinct pilus structures on the surface of Corynebacterium diphtheriae.* J. Bacteriol., 2006. **188**(4): p. 1526-33.

204. Mandlik, A., A. Das, and H. Ton-That, *The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria.* Proc. Natl. Acad. Sci. U. S. A., 2008. **105**(37): p. 14147-52.

205. Swaminathan, A., et al., *Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in Corynebacterium diphtheriae.* Mol. Microbiol., 2007. **66**(4): p. 961-74.

206. Marraffini, L.A. and O. Schneewind, *Sortase C-mediated anchoring of BasI to the cell wall envelope of Bacillus anthracis.* J. Bacteriol., 2007. **189**(17): p. 6425-36.

207. Marraffini, L.A. and O. Schneewind, *Targeting proteins to the cell wall of sporulating Bacillus anthracis.* Mol Microbiol, 2006. **62**(5): p. 1402-17.

208. Cozzi, R., et al., *Group B Streptococcus pilus sortase regulation: a single mutation in the lid region induces pilin protein polymerization in vitro.* FASEB J, 2013. **27**(8): p. 3144-54.

209. Popp, M.W., J.M. Antos, and H.L. Ploegh, *Site-specific protein labeling via sortase-mediated transpeptidation.* Curr. Protoc. Protein Sci., 2009. **Chapter 15**: p. Unit 15 3.

210. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat. Protoc., 2015. **10**(6): p. 845-58.

211. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.

212. Harvey, B.E. and P. Thomas, *Inhibition of CMP-sialic acid transport in human liver and colorectal cancer cell lines by a sialic acid nucleoside conjugate (KI-8110).* Biochem. Biophys. Res. Commun., 1993. **190**(2): p. 571-5.

213. Traynor, A.J., et al., *Inhibition of UDP-N-acetylglucosamine import into Golgi membranes by nucleoside monophosphates.* J. Med. Chem., 1996. **39**(15): p. 2894-9.

214. Hinderlich, S., et al., *A bifunctional enzyme catalyzes the first two steps in N-acetylneuraminic acid biosynthesis of rat liver. Purification and characterization of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase.* J. Biol. Chem., 1997. **272**(39): p. 24313-8.

215. Antos, J.M., et al., *Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity.* J. Am. Chem. Soc., 2009. **131**(31): p. 10800-1.

216. Bradshaw, W.J., et al., *Molecular features of the sortase enzyme family.* FEBS J., 2015. **282**(11): p. 2097-114.

217. Roiban, G.D., et al., *A general solid phase method for the synthesis of sequence independent peptidyl-fluoromethyl ketones.* Org. Biomol. Chem., 2012. **10**(23): p. 4516-23.

218. Srinivasan, R., et al., *High-throughput synthesis of azide libraries suitable for direct "click" chemistry and in situ screening.* Org. Biomol. Chem., 2009. **7**(9): p. 1821-8.

219. Wang, H., et al., *Targeted Ultrasound-Assisted Cancer-Selective Chemical Labeling and Subsequent Cancer Imaging using Click Chemistry.* Angew. Chem. Int. Ed. Engl, 2016. **55**(18): p. 5452-6.

220. Tomaszewska, J., K. Kowalkska, and K. Koroniak-Szejn, *Glucosamine- and galactosamine- based monosaccharides with highly fluorinated motifs.* J. Fluor Chem, 2016. **191**: p. 1-13.

221. Qiagen, *QIAprep Miniprep Handbook*. 2015.

222. Gonzales, M.F., et al., *Rapid protocol for preparation of electrocompetent Escherichia coli and Vibrio cholerae.* J. Vis. Exp, 2013(80).

223. Bucaite, G., *Sortase A modifications of proteins for SPR experiments*, in *Chemistry*. 2016, York.

224. Invitrogen, *Click-iT O-GlcNAc Enzymatic Labeling System*, in *MP33368*. 2007. p. 7.