THE UNIVERSITY OF
WARWICK

NAME: **Rebecca Williams**

YEAR: **Year 4** INTAKE YEAR**: 2009**

TYPE OF PROJECT: **CH401 Research Project**

TITLE: **Design and synthesis of novel star glycoclusters for specific bacterial toxin inhibition**

SUPERVISOR: **Dr. Matthew Gibson**

NUMBER OF WORDS: **8067**

Acknowledgements

Firstly I would like to thank Dr. Matthew Gibson for allowing me to carry out such an enjoyable and interesting research project and for always finding the time to provide me with invaluable advice and support. I would also like to thank Tom Congdon for being a fantastic supervisor and for his constant help, enthusiasm and vast practical knowledge. In addition, thank you to Lucienne Otten and Sarah-Jane Richards for their help with the plate reader and docking experiments. Finally, thank you to all the members of the Gibson research group that have made this project such an enjoyable time period and for creating such a warm and friendly environment.

Abstract

Carbohydrates mediate a multitude of biological processes including the control of host-pathogen interactions. Exploiting these interactions can provide a novel approach towards anti-infective therapies. Glycopolymers have the potential to mimic natural host carbohydrates and bind pathogen lectins, inhibiting their action. This work describes the synthesis of novel star glycoclusters aiming to inhibit the activity of cholera toxin subunit B.

Mono-, di- and trisaccharides were methacrylated and a Michael-type post-polymerisation technique was employed for the addition of thio-galactose across each double bond. These products were synthesised in good yield (from 73-95%) and their inhibition activity was measured using a fluorescence assay. All three of the final glycoclusters showed inhibition activity against cholera toxin B. However the observed trend showed an unexpected decrease in activity with polymer size. This provoked further investigation into the difference in binding modes between cholera toxin B and PNA lectins. The increased inhibition activity with polymer size observed when tested against PNA was postulated to be due to the differences in binding site. Although the results of this work are very suggestive, further work must be done to decipher the capacity of the different binding sites in order to aid the development of even more efficient inhibitors.

Abbreviations

PBS	Phosphate-buffered saline
PNA	Peanut agglutinin
FITC	Fluorescein isothiocyanate
GM1	Monosialotetrahexosylganglioside
CT	Cholera toxin
WHO	World Health Organisation
ET	Enterotoxin
ST	Shiga Toxin
TB	Tuberculosis
QCM	Quartz crystal microbalance
GPC	Gel permeation chromatography
SEC	Size exclusion chromatography
TEA	Triethylamine
BSA	Bovine serum albumin
HSA	Human serum albumin
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
ROP	Ring opening polymerisation
ROMP	Ring opening metathesis polymerisation
ATRP	Atom transfer radical polymerisation
TMEDA	Tetramethylethylenediamine
Cer	Ceramide
Gal	Galactose
Glc	Glucose
Cel	Cellobiose
Mal	Maltotriose
MeA	Methacrylate

MIC	Minimum inhibitory concentration
D_2O	Deuterium oxide
DCM	Dichloromethane
s	Singlet
d	Doublet
t	Triplet
q	Quartet
m	Multiplet
ppm	Parts per million
mmol	Millimoles
µmol	Micromoles
g	Grams
HPLC	High performance liquid chromatography
NMR	Nuclear magnetic resonance
ml	Millilitre
%	Percentage

List of Figures

List of Tables

Table of Contents

1. Introduction

1.1 The Glycocode

From bacteria to mammals, carbohydrates cover the surfaces of a wide range of cells. This 'sugar coating' is intimately involved in a variety of cellular functions such as cell adhesion, cell differentiation, cell signalling and immunological recognition.[3-5] Carbohydrates make up the 'glycocode' and can convey an astronomical level of information in the form of biological recognition systems.[6] This information is encoded through the position and configuration of glycosidic units and the arrangement of branching. For example four different monosaccharides can form 35,560 tetrasaccharides whereas four different amino acids or nucleotides may only form 24 tetramers. Even further diversity can be introduced through the functionalisation of hydroxyl groups, again promoting the idea that a huge number of oligosaccharides can be developed from a fairly small number of monosaccharides.[7]

Deciphering the importance of glycosylation within biological processes is extremely challenging for a number of reasons including the complexity of glycan structure, the multivalent nature of glycan recognition and the difficulty of carbohydrate analysis. These obstacles are being overcome by great improvements in the techniques used for the large-scale analysis of these interactions.[8] Simply deciphering the sequence of a specific oligosaccharide associated with a protein twenty years ago could take as long as one year. Now micromolar oligosaccharides can be fully characterised within 2 weeks using a techniques summarised in figure 1.[1] The development of molecular dynamics simulation techniques has also been an important advance towards the understanding of these highly important lectin-carbohydrate interactions.[9]

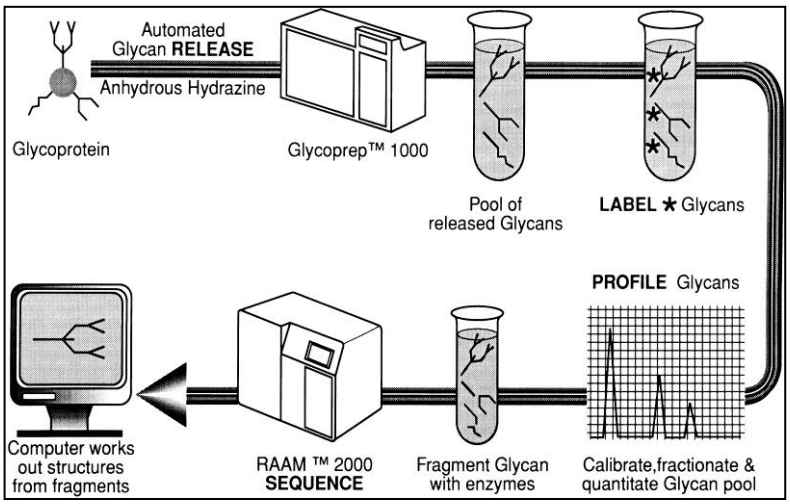


Figure 1: A general process for release, labelling, and sequencing of a glycan from a glycoprotein.[1]

1.2 Lectins

Lectins are defined as carbohydrate binding proteins capable of highly selective recognition of subtle variations in saccharide structure.[10] This allows them to act as decoders of information encoded by carbohydrates and are so selective that a lectin specific for galactose will not generally bind mannose.[7] Their individual interaction with mono- and oligosaccharides are often weak and non-covalent, however strong, highly specific bonds are required for recognition. Nature compensates for these low-affinity monovalent interactions by increasing valency through the arrangement of saccharides and their protein receptors to allow the formation of multiple protein-carbohydrate interactions.[11] This is often achieved through the assembly of individual protein subunits, each containing a single carbohydrate binding site, into multimers.[12] This therefore allows the simultaneous interaction of several individual carbohydrate ligands on the surface of a target cell.[13] This enhancement has been coined 'the cluster glycoside effect'.

Lectins have been identified in most organisms (viruses, plants, bacteria and animals) and represent a heterogenous group of oligomeric proteins that typically contain two or more carbohydrate combining sites and vary largely in size and structure.[14] Although the idea of lock and key style interactions implicated in specific reactions of the immune system were described in 1900, up until the late 1960s, carbohydrates were thought to serve only as energy sources.[15, 16] Lectins have since been shown to act as recognition determinants for the control of intracellular traffic, adhesion of infectious agents to host cells and the recruitment of leukocytes to sites of inflammation.[16] This along with the evidence suggesting their involvement in the development of malignant cancers and metastasis has resulted in significant research and investigation into lectin-carbohydrate interactions over recent years.

1.3 Carbohydrate binding proteins and their role in disease

Many pathogens require the use of carbohydrates to bind to sugar specific lectins for the expression of their pathogenicity.[17, 18] Some of these are summarised in the following table.

Pathogen	Receptor	Receptor specific saccharide
Cholera Toxin	GM1 Ganglioside	Gal(β1-3)GalNAc(β1-4) [Neu5Ac(α2-3)]Gal(β1-4)Glc-Cer
Ricin Toxin	IgA1	Gal(β1-3)
Shiga Toxin	Gb3 glycolipid	Gal(α1-4)Gal(β1-4)Glc(β1-1)Cer
HIV-1	DC-SIGN	Gal(β1-1)Cer
Legionella pneumophila	MBL	Man(α1-2)Man
Rotavirus	GA1 glycosphingolipid	Gal(β1-3)GalNAc(β1-4)Gal(β1-4)Glc(β1-1)Cer

Table 1: Pathogens that use carbohydrate-lectin interactions to infect host and their corresponding receptors.

1.3.1 Cholera

Cholera is an infection of the small intestine that is caused by the bacterium *Vibrio cholerae* which secretes the cholera toxin (CT). The World Health Organisation reported an annual figure of 589,584 cases across 58 countries in 2011. This is a huge 85% increase on the previous year and of these cases, 7,816 resulted in death. A further threat posed by this disease is its correlation with global warming. Studies have shown global warming is producing a much more favourable environment for the bacteria and Reyburn *et al.* predicted a 1°C increase in temperature over a period of four months could result in a two fold increase in the number of cases of cholera.[19]

The cholera toxin is a hexameric protein consisting of one enzymatically active single subunit A and 5 B sub units (AB_5) (figure 2). Infection is initiated when the B subunits mediate adhesion to the host cell through the simultaneous binding to carbohydrate moieties of five GM1 receptor gangliosides present on the surface of the intestinal epithelial cells.[5] This binding is essential for the internalisation and subsequent disease process of this AB_5 toxin.[20] Internalisation occurs *via* receptor mediated endocytosis upon which the toxic A subunit is cleaved resulting in the expression of cholera associated symptoms within the host.[4]

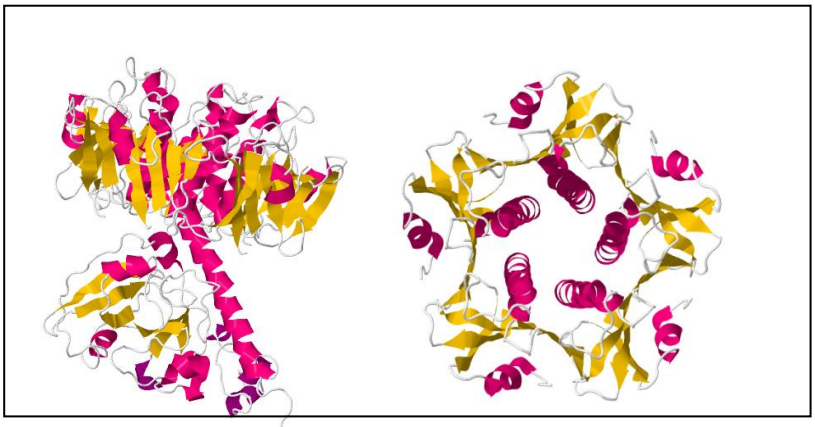


Figure 2: Cholera toxin AB_5 crystal structure, PDB ID- 1xtc (left). Cholera toxin B-pentamer, PDB ID- 1RCV (right).

1.3.2 Other diseases mediated *via* carbohydrate binding

The list of pathogens known to use carbohydrate-protein interactions to infect a host is extensive. *Escherichia coli* (E-coli) is a gram negative, rod-shaped bacterium capable of causing severe food-borne disease. E coli infections affect millions of people worldwide and although different strains vary in intensity, some are life threatening. The pathogenicity of these bacteria is attributed to the production of the heat labile enterotoxin (ET). ET is structurally analogous to the AB_5 cholera toxin and also works *via* B pentameter-mediated receptor binding, internalisation and activation of enzymatic A subunit.[21]

Shiga toxins are virulence factors produced by the bacteria *Shigella dysenteriae* and certain strains of *Escherichia coli*.[22] This family of toxins is categorised into two groups, shiga toxin 1 and 2 and all belong to the structural class AB_5. In order to reach their cytoplasmic target, Shiga toxins are endocytosed and transported by a retrograde pathway to the endoplasmic reticulum. Shiga toxins bind the carbohydrate moiety of the glycosphingolipid Gb3 in the extracellular leaflet of target cell plasma membranes.[23] However in order to exhibit the associated toxic effects on a cell, the A subunit must be enzymatically cleaved into two fragments, A1 and A2, normally held together by disulphide bridges between cysteine residues. The released A1 fragment is then translocated into the cytosol where it exerts its cytotoxic action. The toxic effects are brought about *via* inhibition of protein synthesis through catalytic inactivation of the 60S ribosomal subunit in target cells.[24] In addition to this it has also been reported that a long term effect of shiga toxin infection is the induction of apoptosis in affected cells.[25] Infections with Shiga-toxin producing bacteria are known to cause bloody diarrhoea as well as a more serious disease known as haemolytic uremic syndrome (HUS).

Symptoms of HUS include renal failure, thrombocytopenia (decreased blood platelet count) and microangiopathic hemolytic anemia (destruction of red blood cells).[26] Other more complex infections may also bring about neurological problems such as seizures and loss of consciousness.[27] There is currently very little in the way of treatment for these types of infections and maintenance of hydration is the primary option for some of the less serious symptoms.

The *Mycobacterium tuberculosis* is a small, aerobic, non-motile bacillus[28] and is the main cause of tuberculosis. This is a disease that affected 8.7 million people worldwide and caused 1.4 million deaths in 2011 (WHO data). Many host-pathogen interactions still remain unclear however it has been deduced that the C-terminal domain of the arabinosyltransferase *Mycobacterium tuberculosis* EmbC is a lectin-like carbohydrate binding module. This arabinosyltransferase enzyme (EmbC) is essential for the synthesis of arabinan, a polysaccharide vital within the cell envelope of the pathogen. Treatment of active TB normally involves the administration of a range of different antibiotics. However the understanding of this interaction has many possible implications for an alternative treatment through the inhibition of EmbC.[29]

1.4 Anti-adhesion therapy

The threat of bacterial antibiotic resistance is well known and has recently been widely reported in the media. It is also accepted that novel, alternative treatments must be explored in order to help combat this problem and prevent the regression of surgical medicine back to a period where bacterial infections were as life threatening as cancer and heart disease. Strategies that interfere with the pathogenicity of the bacteria rather than killing the bacteria itself may provide this all important alternative type of treatment. One possible example of this is the anti-adhesion strategy; a strategy targeting one of the earliest stages of bacterial infection.[30]

Adhesion of bacteria and viruses to the cell membrane is very often an initiation step for infection. This process is mediated by carbohydrate-protein interactions as described in some of the previous examples. Adhesion is an extremely important step as not only is it required for the progression of infection but it also confers species and tissue specificity.[31] The understanding of various cell surface carbohydrate-protein interactions led to the development of free carbohydrate structures aimed at interfering with pathogen attachment and hence prevention of cell adhesion and subsequent infection.[32] Any non-adhering bacteria can then be naturally removed by the body. As

the bacteria is removed rather than killed, this results in less selective pressure leading to resistance and any mutations resulting in a reduction of drug inhibition will confer a similar reduction in adhesive ability.

One major disadvantage of this type of therapy is that the majority of pathogens have more than one mechanism for adhesion. Therefore, for anti-adhesion therapy to be successful, it must disrupt all methods of pathogen attachment through either the use of multiple agents or a single agent able to exhibit a great range of anti-adhesive activity.[33]

Monovalent anti-adhesives have been developed for a number of different carbohydrate-protein interactions however their affinity for the target bacterial proteins is often low. Multivalent inhibitors have the potential to overcome this constraint *via* the mechanism described in figure 3 below.

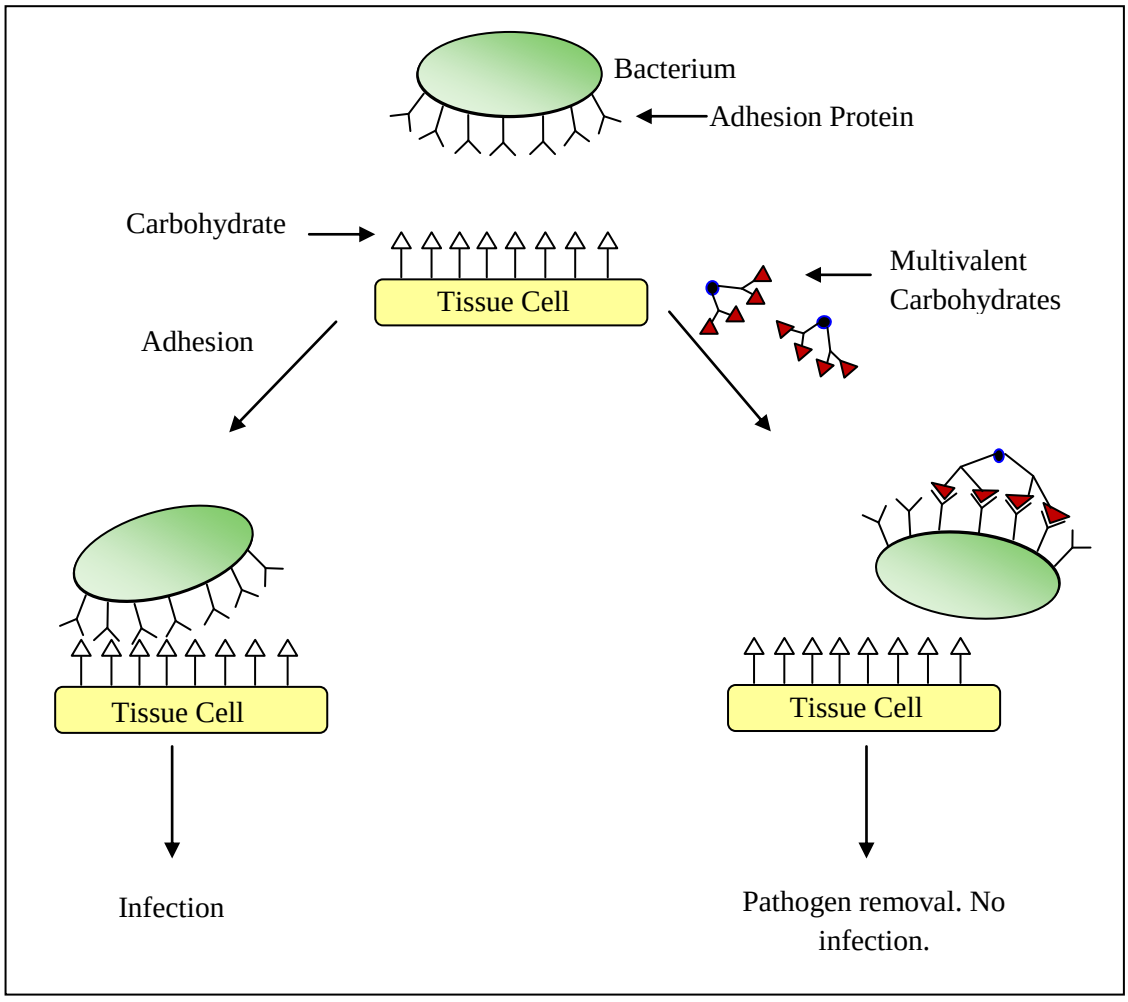


Figure 3: Schematic representation of bacterial adhesion (left) resulting in infection, contrasted with the use of multivalent inhibitors that bind the pathogen and aid removal from the body with no resultant infection (right).

Monovalent inhibitors provided the first evidence supporting the anti-adhesion theory and its potential for developing novel antibacterial agents. For example, early studies using various mannose containing structures for the inhibition of fimbriated E.Coli highlighted the importance of chemical modification upon the potency of an inhibitor. Methyl α-D-mannoside was used as the reference inhibitor and a relative increase of up to 30 times was observed for the lipophilic mannose derivative p-nitrophenyl α-D-mannoside.[34] These results led to a continuation of research into the potential of aryl mannosides[35] but ultimately the concentrations required for this level of potency were not clinically relevant. A further problem for monovalent inhibitors is caused by the high entropic cost of binding. Monovalent analogues of natural protein receptors will therefore find it very difficult to effectively compete with native membrane-bound receptors *in vivo*. Again high concentrations are required to overcome this issue.

Since a wide range of microbial pathogens and toxins use multivalent carbohydrate-protein interactions to assist adhesion, it is only logical that inhibitors of these interactions should adopt a similar structure. The method of using entropically favoured inhibitors to occupy the binding sites in a multivalent carbohydrate-binding protein has been applied to several members of the AB_5 structural family (ST, CT and ET).[36, 37]

The first of this family to be successfully inhibited by a multivalent compound was the type 1 fimbrial lectin of E. Coli.[38] Lindhorst *et al.* synthesised a range of multivalent mannose compounds; the most successful being a trivalent system, (figure 4). This structure is considered a fairly small trivalent structure and it was concluded that the inhibitory effects were a direct result of the third level of multivalency as larger systems showed little effect. Higher multivalent systems were designed by Nagahori *et al.* (up to 16 valent glycodendrimers) however the multivalency effect of these compounds were limited to one order of magnitude.[39] This is due to the structure of the fimbrial tips of this pathogen and in order to aid bridging even larger structures would be needed. Anomeric orientation also highly limited the effectiveness of these structures. It was therefore postulated that this specific pathogen has limited susceptiblity to the multivalent effects of inhibitors.

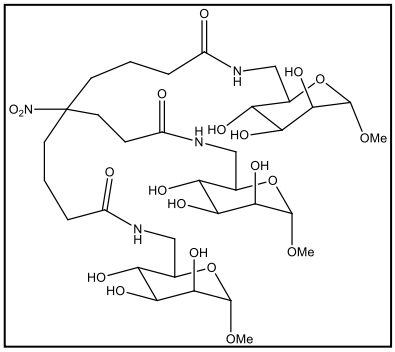


Figure 4: Trivalent mannose inhibitor developed by Lindhorst *et al.*

The gastric bacterium *Helicobacter Pylori* (H. pylori) colonises half of the world's population and is responsible for stomach ulcers and inflammation that may increase the risk of developing gastric cancers. The anti-adhesion target of H. pylori is considered complex as three different adhesive interactions have been identified. Despite this, the multivalent inhibitor 2,3-sialyllactose (Neu5Acα2,3Galb1,4Glc) has been shown to successfully inhibit adhesion of H. pylori and HSA conjugated 2,3 sialyllactose was found to be three orders of magnitude more potent than monovalent 2,3-sialyllactose on a per sugar basis.[40] This inhibitor was later tested on infected rhesus monkeys with some positive results with two out of six monkeys being completely cured under a specific regime.[41] However a major disadvantage of using such large protein-bound inhibitors is the likely immunogenicity effect. More recently, the inhibitory effects of porcine milk have been investigated which are thought to contain multivalent glycoproteins matching the appropriate carbohydrate profile.[42]

Effective multivalent inhibitors of Shiga Toxins have also been developed. Unlike CT and ET, ST-B subunits each contain three carbohydrate binding sites and hence the ST pentameter is capable of accommodating up to fifteen carbohydrate sequences.[43] Bast et al. conducted solid-phase binding experiments using site-directed mutants that revealed that two of the ST carbohydrate binding sites were more important than the third site positioned closest to the centre of the pentameter.[44] This provoked the design of divalent, bridged inhibitors from a Pk trisaccharide (αGal(1,4)βGal(1,4)βGlc) aiming to simultaneously inhibit sites 1 and 2 in a single ST-B subunit.[36] However, once again the concentrations of these inhibitors required for the desired potency was far too high for clinical applications. Therefore a complex 'starfish' inhibitor with a pentameric, bridged Pk trisaccharide structure was designed aiming to inhibit binding sites 1 and 2 on all 5 of the ST-B subunits (figure 5).[36] μM concentrations were shown to effectively neutralise ST toxicity for *in vitro* cell cultures. When 'starfish' was tested on mice, 90%

did not develop the symptoms associated with ST infection. The results of this study have extremely positive implications for the further design of soluble carbohydrate-based pharmaceuticals that demonstrate a dose range of 2-3 mg per kg of body weight.

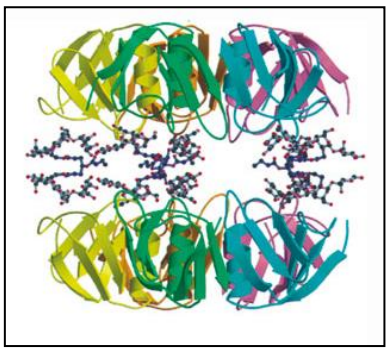


Figure 5: Crystal structure representation of Starfish ligand 'sandwiched' between two ST-B pentamers.

1.5 Synthetic Glycopolymers

1.5.1 Synthesis

Synthetic, sugar-containing macromolecules or 'glycopolymers' are becoming increasingly more attractive targets for the development of a wide range of biomedical and commercial applications. The diversity of the possible applications coupled with advances in polymerisation techniques has seen a huge increase in the level of research devoted to carbohydrate containing polymers.

The majority of glycopolymers are synthesised using polymerisable saccharide units and others are prepared *via* saccharide addition to a polymer backbone. There are very few limitations to the types of polymerisation techniques applicable to glycopolymers with examples of successful glycopolymer synthesis *via* ionic, ROP, ROMP, ATRP, living and controlled free radical polymerisation.[7] Free radical polymerisation is a widely used technique owing to its robust nature and high tolerance of solvent and monomer structure. However this technique results in a range of negative implications for the potential uses of the resulting glycopolymer. Low control of molecular weight, high polydispersity and poor end group functionality severely hinder the usefulness of a polymer particularly within the biomedical industry.

The first glycopolymer was synthesised by Horejsi *et al.* through the copolymerisation of acrylamide and allyl glycosides of different sugars.[45] The polymerisation was carried

out in water using TMEDA as a catalyst and ammonium persulfate as the initiator and the resulting glycopolymer showed mimicry activity towards other natural polysaccharides. Polyacrylamide derivatives remain popular within research due to the amide linkage conferring water solubility and stability towards hydrolysis.

Glycopolymers can be structurally tailored for a number of specific roles. Glycopolymers with pendant saccharides can display high affinities for proteins and amphiphilic polymers have uses as biomaterials. However, due to the highly precise nature of sugar-recognising proteins, particularly well defined glycopolymers are required to be synthetically and pharmacologically useful. Living polymerisation techniques are often employed to provide this control over molecular weight, end group functionality and low polydispersity.[46]

1.5.2 Post-polymerisation modification

The importance of precise control over chain length and carbohydrate density is often a key challenge when synthesising polymers for multivalent interactions. Direct polymerisation of two glycosylated monomers is unlikely to lead to identical degrees of polymerisation which may have serious implications on the desired carbohydrate-protein interactions.[47] Therefore the synthesis of glycopolymers *via* post-polymerisation modification of reactive polymer precursors is a very attractive alternative.[48] This technique is based upon the polymerisation of monomers with functional groups that are inert towards the polymerisation conditions but can be subsequently converted to a specific functional group. This method has the capacity to synthesise functional glycopolymers with excellent conversions, using mild reaction conditions tolerant of a wide range of functional groups.[2] The development of 'click' type chemistry has made these types of post-polymerisation reactions even more straightforward and a range of other techniques are summarised in figure 6.

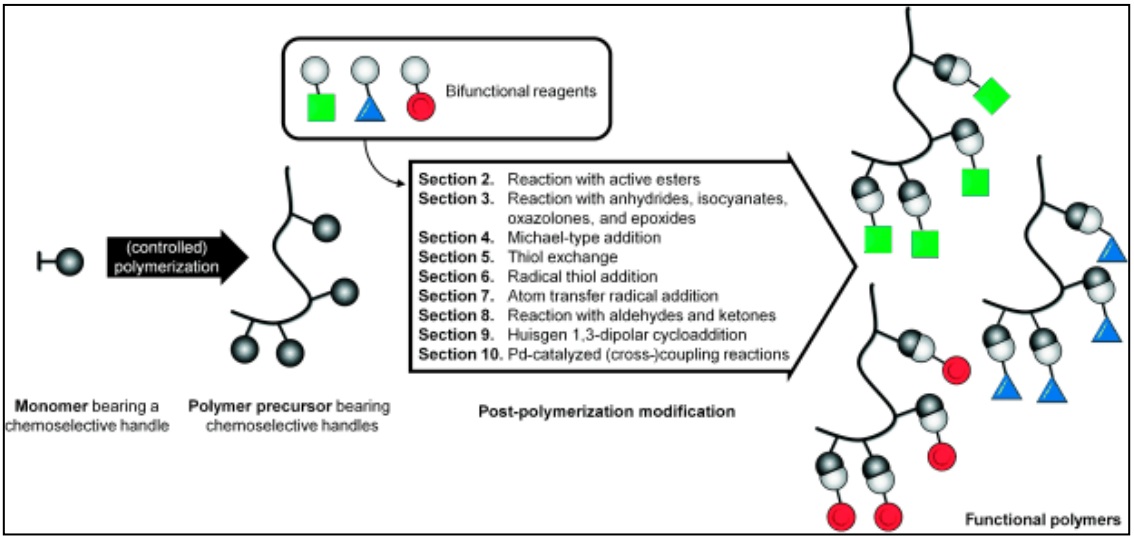


Figure 6: Synthesis of polymers by a range of post-polymerisation modification methods.[2]

An important route to post-polymerisation modification is the use of a Michael addition reaction between thiols and alkenes. These reactions proceed readily in aqueous solution at low temperatures.[49] Lou *et al.* have carried out successful post-polymerisation modification of acrylate-functionalised polyesters *via* Michael addition of thiol derivatives. They used this method to synthesise functional, ampiphilic poly(ε-caprolactone) under mild conditions without the need for protection/deprotection steps. The features of this reaction are highly beneficial for polymers with biomedical applications as the low temperatures prevent degradation of the polymer and there is no need for the use of toxic metal catalysts.[50]

The wide range of possible post-polymerisation techniques combined with the great range of potential saccharide architectures and polymer backbones means the prospects for functional glycopolymer synthesis is enormous.

1.6 Methods of assessing carbohydrate-protein interactions

Carbohydrate-protein interactions can be examined by a number of methods. X-ray crystallography is a highly useful method for observing the arrangement of the molecules involved, however quantitative measurements such as binding affinity cannot be drawn. Turbidimetry is a process that measures the decreased intensity of transmitted light due to the scattering of trapped particles inside and can be used to measure amount of absorbed light.[51] Another simple, cost-effective method for investigating these interactions is the use of QCM. QCM has the capacity to sensitively measure solution-surface interaction s with a wide detection range.[52] A unique feature of this technique is

the ability to also measure dissipation energy of the substrate bound to the surface (QCM-D).[53] However a much more effective measure of binding affinity can be achieved through the use of fluorescence assays. This relatively new technique uses a plate reader, only slightly larger than a desktop computer, to measure fluorescence of labelled proteins bound to a surface. Therefore, inhibitory action of different glycopolymers can be directly measured as a function of fluorescence. A high binding plate must be first functionalised with the pathogen receptor. The fluorescently labelled toxin and glycopolymer inhibitor can then be added at varying concentrations and incubated with the prepared plate. The lower the fluorescence value measured, the greater the inhibitory effect of the glycopolymer as unbound pathogen will be removed with washing. This technique is extremely useful as it can mimic the natural process of pathogen adhesion to cell-membrane carbohydrates to facilitate disease and infection.

2. Aims

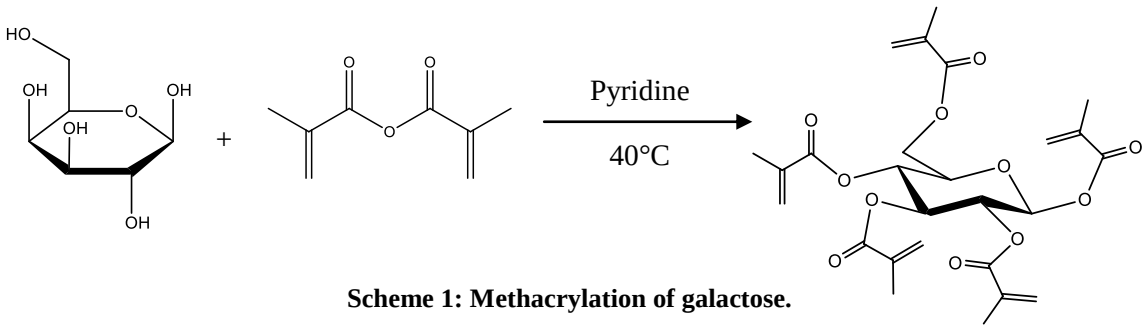
Interactions between lectins and carbohydrates mediate many cellular processes including cell adhesion and signalling. Glycopolymers have the potential to mimic the natural carbohydrate receptors of different bacterial toxins and hence inhibit their action. It is therefore proposed that inhibitors of these carbohydrate-protein interactions have the potential to form treatments for a range of life threatening diseases and infections.

This work aimed to synthesise several ‘star’ glycoclusters with different mono-,di- and trisaccharide cores. The cores will be methacrylated followed by a Michael addition of thio-galactose to the double bond. The inhibition activity of these glycoclusters will then be tested against fluorescently labelled cholera toxin B with GM1 functionalised plates.

3. Results and Discussion

3.1 Sugar Methacrylation

Since the pioneering work of Horejsi *et al.*,[45] *g*lycopolymers for use within the biomedical industry have been attracting significant attention. However synthetic methods for their production are often inefficient and the issue of polydispersity significantly reduces their applications.[54] Herein, the proposal of a two-step procedure for the methacrylation of different, mono-, di- and tri- saccharides is described as a precursor to glycosylated star polymers.



Scheme 1: Methacrylation of galactose.

This reaction sees the pyridine acting as both the solvent and the catalyst leading to the deprotonation of the sugar hydroxyl groups. The activated carbohydrate can then nucleophilically attack the methacrylic anhydride leading to the addition of a methacrylate group, (figure 7).

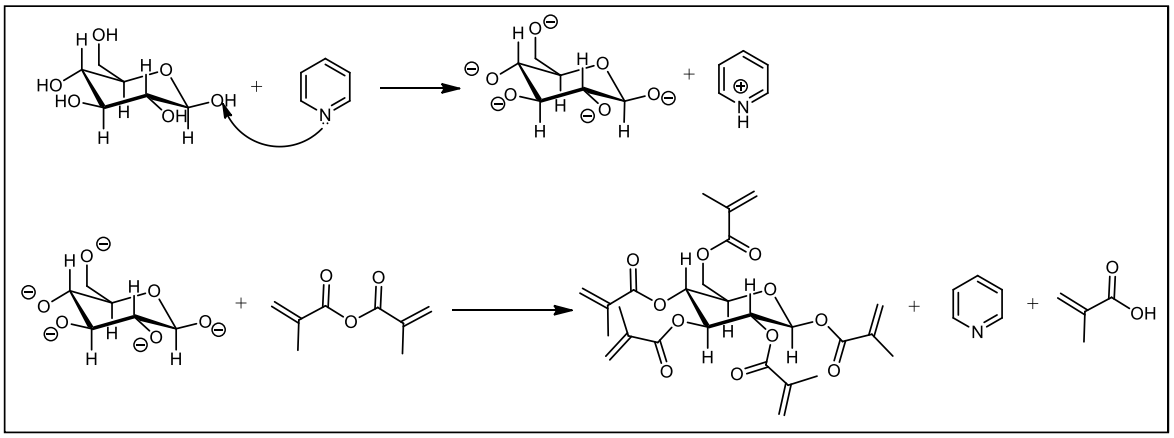


Figure 7: Reaction mechanism for the methacrylation of glucose.

In order to hydrolyse unreacted methacrylic anhydride, 0.1M NaOH was added and a colour change was observed between the two concentrated products. This reaction required optimisation of reaction conditions and concentration of the NaOH. ^{1}H NMR and TLC measurements indicated the presence of a methacrylic acid impurity and hence the need for purification. A further drawback of this procedure is the sensitivity of the methacrylated product. When left on the bench, the product often polymerised and turned solid. It was later found that storage at <10°C in a sealed container significantly reduced the prevalence of this issue.

3.1.1 Optimisation of purification

Numerous methods were explored attempting to remove the methacrylic acid adduct impurity, including the use of amber lite, DOWEX, recrystallisation methods and HPLC. The most successful procedure involved the use of a short column of silica eluted with 50:50 ethyl acetate/petroleum ether solvents.

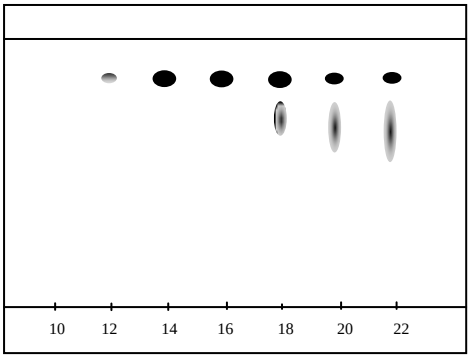


Figure 8: Reconstruction of TLC measurement following column of crude product.

Using the methacrylation of galactose as an example, in this case fractions 10-16 were collected, concentrated under vacuum and analysed by ^{1}H NMR. The NMR region where these impurities are visible is displayed in figure 9. The blue spectrum represents the crude product and the maroon spectrum represents the final methacrylated product following purification. This magnified spectrum highlights the large reduction in methacrylic impurities as a result of the column, (figure 9). Unfortunately this method did result in a significant loss of product however the purity of the resultant product was high.

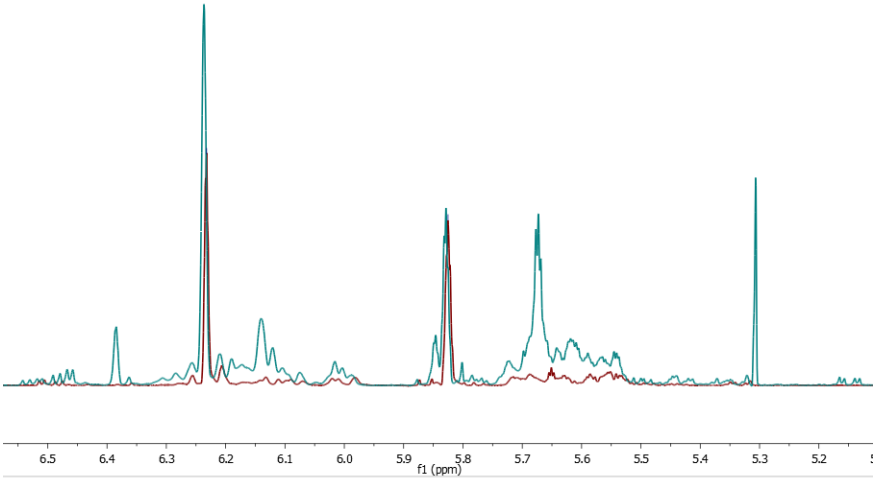


Figure 9: Magnified NMR methacrylate region, pre and post purification. Methacrylic impurities present before the column was carried out (shown in blue) and the maroon spectra highlights the removal of these impurities leaving only the predicted methacrylate peaks.

This procedure was successfully carried out for a range of different saccharides including glycerol, glucose, arabinose and mannose. However it was decided that further experimentation was to focus on galactose, cellobiose and maltotriose for the purpose of multivalency comparisons.

3.1.2 Characterisation

Characterisation of sugar containing polymers is very challenging. Anomeric hydrogens on both the core and the outer sugars make the assignment of individual protons extremely difficult, taking the Gal(Gal)$_5$ as an example, there are 64 possible anomeric combinations.

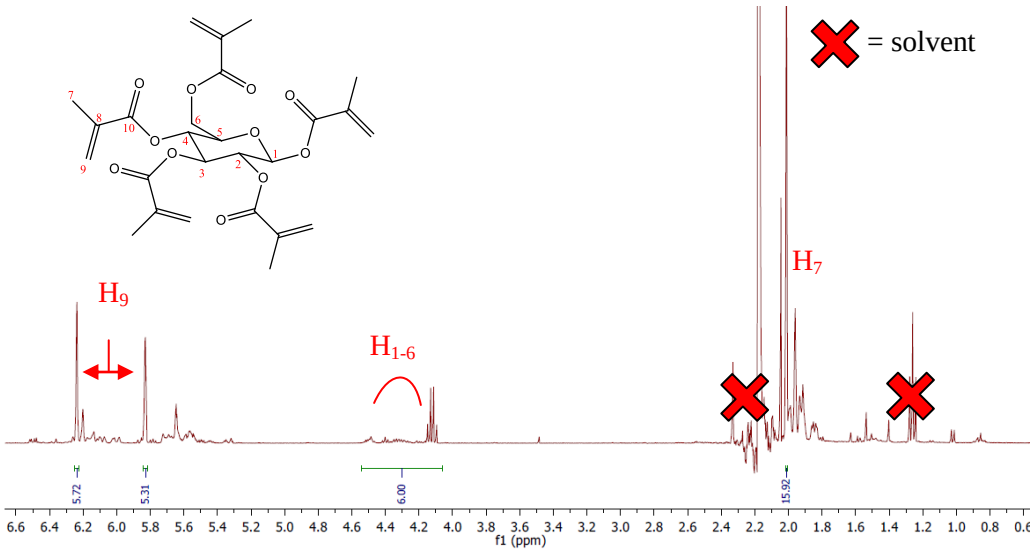
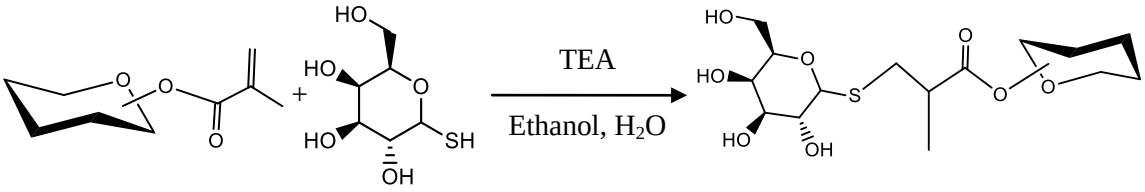


Figure 10: Assigned ^{1}H NMR for the purified Gal-MeA product.

Solvent peaks at 1.26ppm (petroleum ether, m), 2.17ppm (acetone, s) and 4.12ppm (ethyl acetate, q) have been omitted.

3.2 Glycosylation

With the successful synthesis of several different methacrylated sugar cores completed, post-functionalisation modifications were carried out. The alkene double bond from the methacrylate group provided an ideal platform for thiol-ene 'click' chemistry. Thiol functionalisation using a radical source and a range of different solvents were investigated including an interesting procedure proposed by the Haddleton research group involving the use of NMR to follow the reaction completion.[55] However differences in the solubilities of the reagents meant this trial was unsuccessful. The final optimised procedure involved the use of a strong base (triethylamine) to catalyse the Michael addition of thio-galactose in ethanol and water. Both the catalyst and reaction mixture were both degassed separately in order to ensure the reaction did not start with oxygen in the system, which has been shown to inhibit the reaction.[56]



Scheme 2: Glycosylation of sugar methacrylate *via* Michael addition.

As displayed in figure 11, the TEA abstracts a hydrogen atom from the thiol species to generate an anionic thiol. The thiol anion then adds onto the alkene in a 1,4 Michael addition. Addition occurs at the least hindered end of the bond to facilitate the formation of the more stable tertiary anion. This addition product can then abstract a hydrogen from the thiol to produce another thiol anion and the desired product.

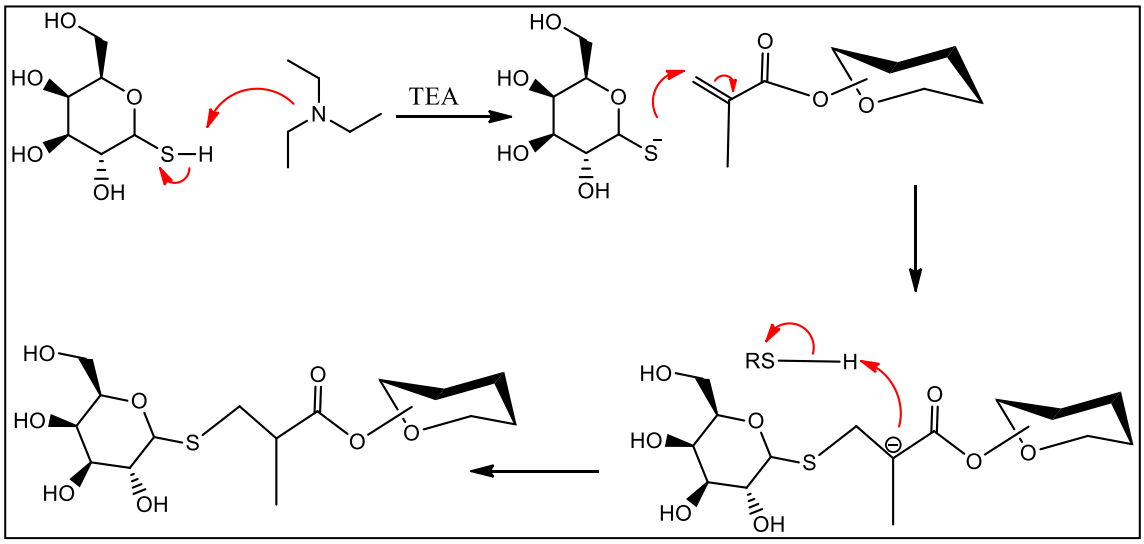


Figure 11: Reaction mechanism for the Michael addition of thio-galactose onto methacrylated saccharide core.

Initial spectral data obtained from the product of this reaction indicated high levels of impurity. Therefore purification of the crude product was carried out using dialysis, a simple yet highly effective method that can be carried out at temperatures that won't damage the polymer. Following dialysis, the product was freeze dried and the resultant sticky, off-white compound was analysed by ^{1}H and ^{13}C NMR. The ^{1}H NMR spectrum (figure 12) displays the expected spectrum for this product with only low levels of impurity. The peaks at 5.57 and 5.26 representing unreacted methacrylate functional groups are present however integrals show this is a very small percentage.

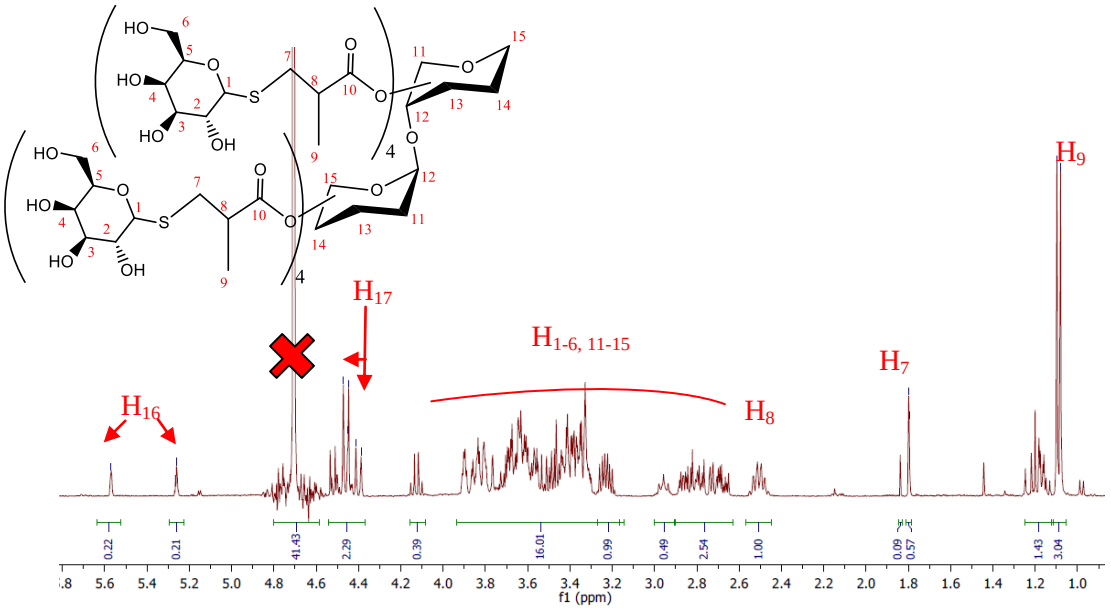


Figure 12: Assigned ^{1}H NMR spectrum for Cel(Gal)$_8$ in D_2O following dialysis.

H_{16} shown in the above spectrum represents unreacted CH_2 methacrylate peaks and H_{17} represents the presence of anomeric protons. See supplementary information for other labelled final product NMR. Both ^{1}H and ^{13}C spectra for the three final glycoclusters suggest high purity and successful synthesis. Using integration values for the methacrylate and thiol-CH_2 peaks, a percentage conversion for each reaction was obtained.

Gal(Gal)$_5$ - 88%

Cel(Gal)$_8$ - 73%

Mal(Gal)$_{11}$ >95% as shown by NMR.

These yields do not represent differences in reactive preference towards the Michael addition. It is much more likely that the yields represent an improvement in

experimental precision as the Cel$(Gal)_8$ was synthesised first, followed by Gal$(Gal)_5$ and Mal$(Gal)_{11}$.

Reaction conversion can also be observed using the disappearance of the methacrylate peak from the ^{13}C and ^{1}H NMR of the initial methacrylated sugar to the final glycosylated product.

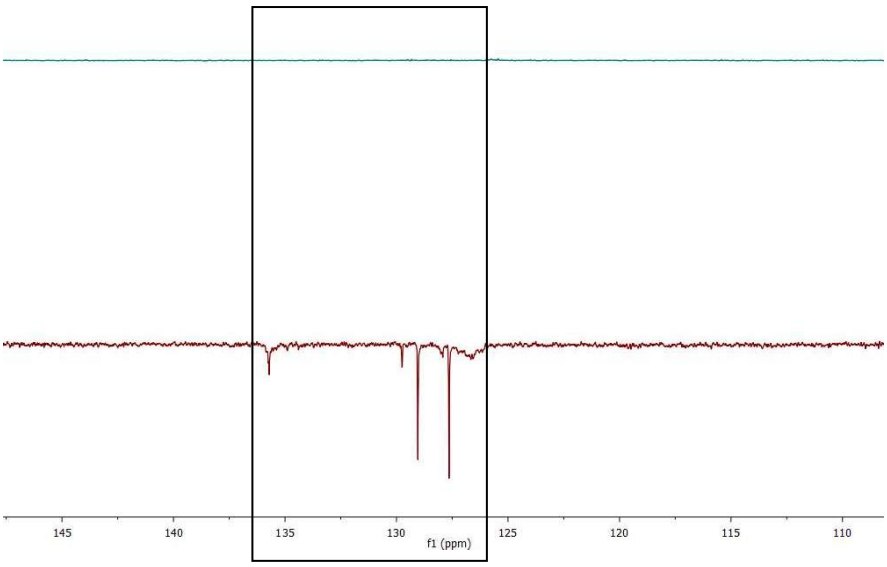


Figure 13: Stacked ^{13}C NMR showing presence of methacrylate peaks at 135.49 and 127.43 ppm for Mal-MeA (maroon) and absence of these peaks in the Mal$(Gal)_{11}$ spectra (blue).

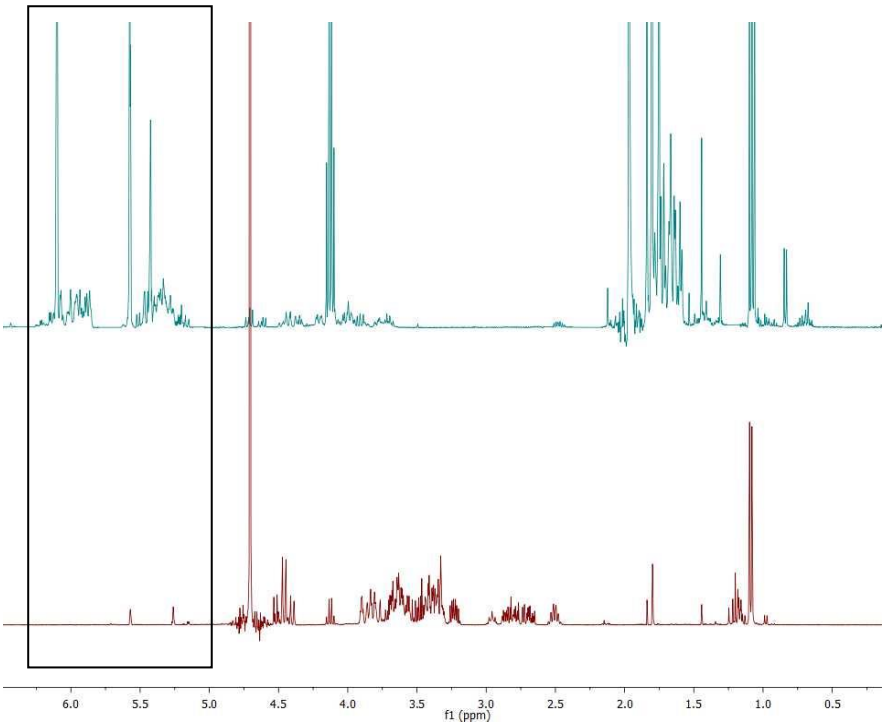


Figure 14: Stacked ^{1}H NMR showing presence of methacrylate peaks at 6.23 and 5.82 ppm for Cel-MeA (blue) and their reduction in the Cel$(Gal)_8$ spectra (maroon).

The methacrylated sugar and the final Michael addition product were analysed by size exclusion chromatography (THF-GPC) producing the following trace. Although exact molecular weights cannot be drawn from this measurement, there is a clear shift in molecular weight suggesting the thio-galactose modification was indeed successful. It is also important to note that while the trace for Cel(Gal)$_8$ (green) appears broad, this is due to the axis scale used and further GPC analysis of this product (figure 16) give a better representation of how narrow the peak actually is.

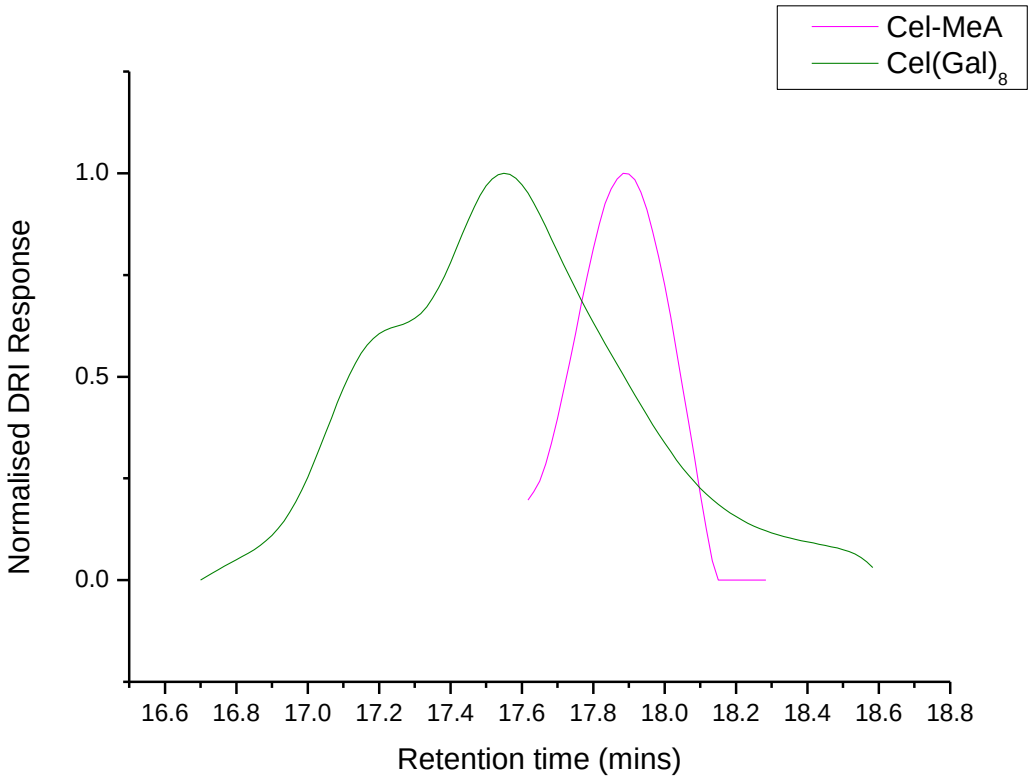


Figure 15: GPC analysis of Cel-MeA (pink) and Cel(Gal)$_8$ (green) following glycosylation of Cel-MeA.

Upon methacrylation and glycosylation of galactose, cellobiose and maltotriose, SEC analysis (DMF-GPC) was carried out on all three purified products to yield the following graph (figure 16). Again accurate conclusions regarding the exact molecular weights of these compounds cannot be drawn, but the expected shifts in molecular weights are clearly visible. However, it is also clear that the maltotriose polymer displays a multimodal trace. This suggests the presence of impurities or incomplete polymer stars. Conversely, both the ^{1}H and ^{13}C NMR for this product show the complete removal of methacrylate peaks and minimal impurities. Therefore NMR can be regarded as a more valid method for the investigation of these products over GPC.

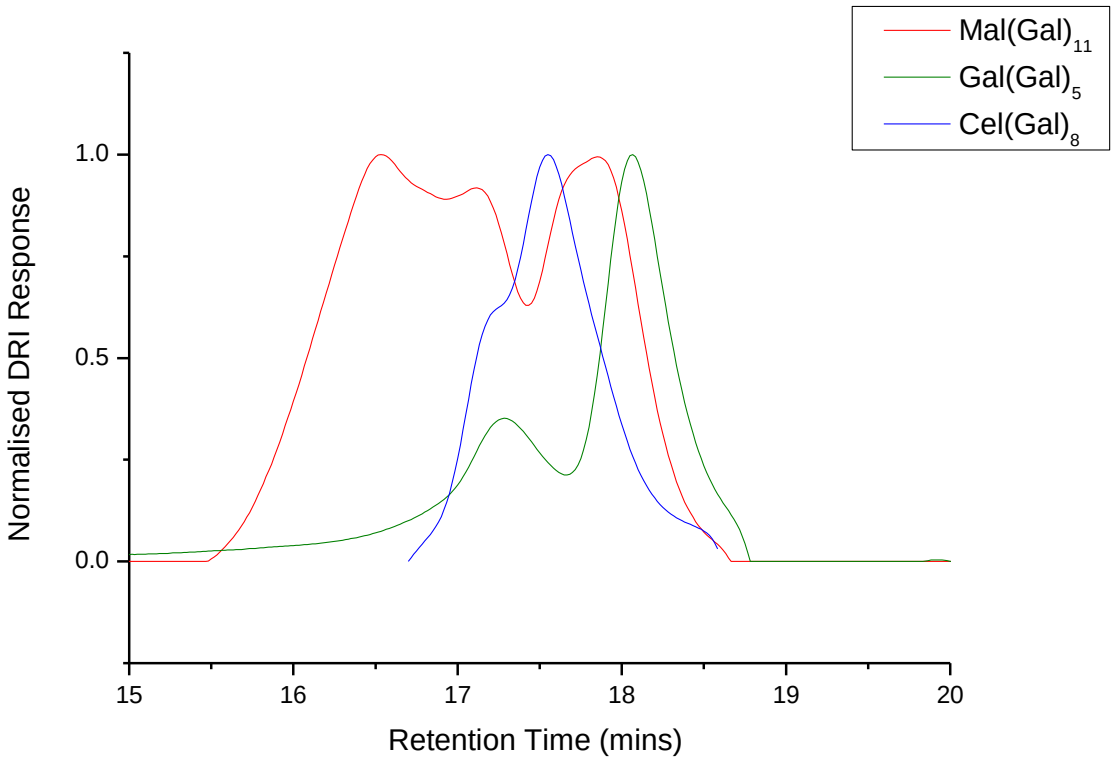


Figure 16: GPC analysis of all three final star glycoclusters.

Furthermore, an important feature of these star polymers is the feature of the ester linkage. This type of bond confers biodegradability and by-products of this degradation are simple carbohydrates. This is a very important consideration when designing glycopolymers for biomedical applications as toxicity of degradation products often restricts the usefulness of a drug.

3.3 Inhibition Studies

The ability of the synthesised star polymers to inhibit the binding of various lectins was measured using the BioTek Synergy HT Microplate reader. Fluorescence was measured at varying concentrations of all three sugars using an experimental design summarised in figure 17.

A GM1 functionalised surface was used to evaluate the level of glycocluster inhibition of both fluorescently labelled cholera toxin B and PNA (a). In the absence of the synthesised glycopolymers, the lectin binds to the GM1 surface and fluorescence is measured (b). In the presence of the glycopolymers, the glycopolymer binds the lectin, inhibiting surface binding and is therefore washed away prior to the reading, resulting in a reduced level of fluorescence being measured. Optimisation of this procedure was

required as initial attempts displayed a much weaker trend than expected. Using the procedure outlined by Liu *et al.*, 'blocking' of the surface using 1% BSA was carried out to yield much clearer results.[57] This 'blocking' technique is used to remove any non-functionalised sites on the high-binding plate that may bind to the toxin and result in inappropriate readings. This step does not in any way invalidate the data as vigorous washings were carried out in order to remove any non-bound BSA.

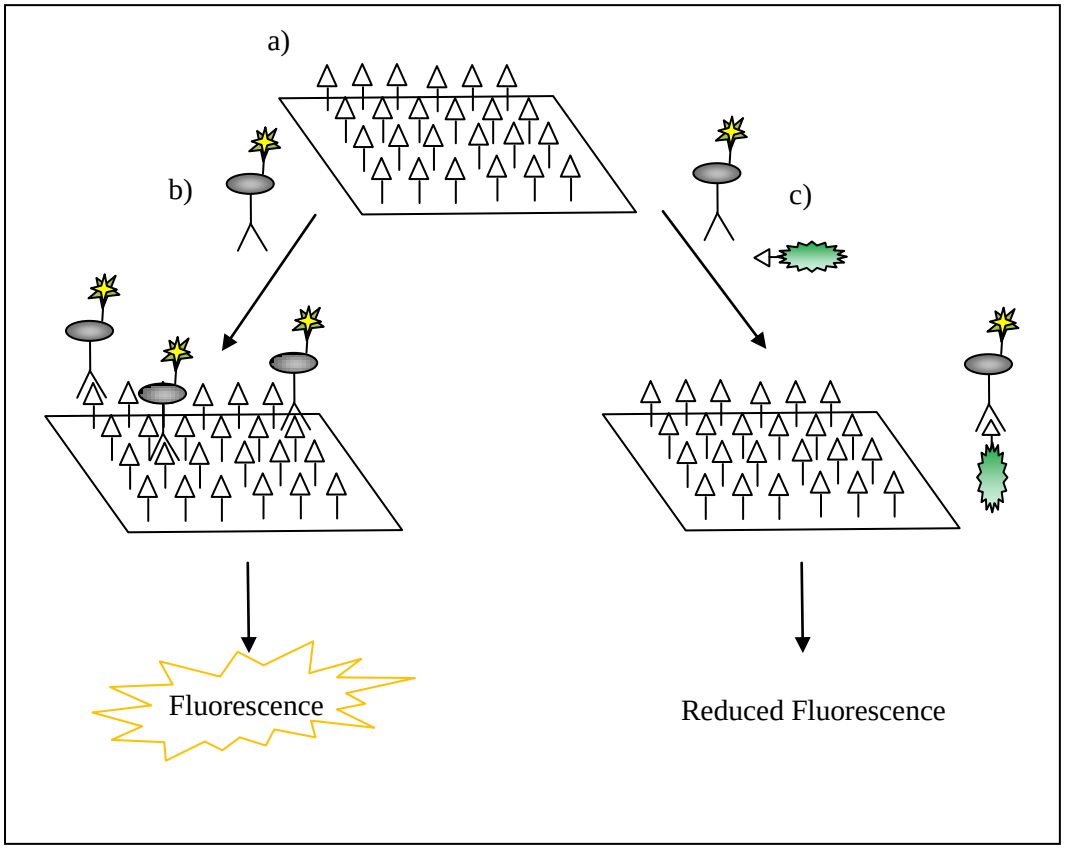


Figure 17: Experimental design of fluorescence assay to measure inhibitory effect of glycoclusters against different pathogens.

Initial inhibition studies focused on the cholera toxin as non-toxic fluorescently labelled B-subunits are readily available and are ideal for use within these assays. Furthermore, the successful inhibition of this toxin would have many positive implications for the future applications of these inhibitors. As previously described, cholera toxin B binds the carbohydrate sequence (Galβ1-3GalNAcβ1-4(Neu5Acα2-3)-Gal-β1-4Glc-ceramide) of five GM1 gangliosides. The potential of the synthesised glycoclusters to inhibit this toxin was tested using the fluorescence assay procedure described above.

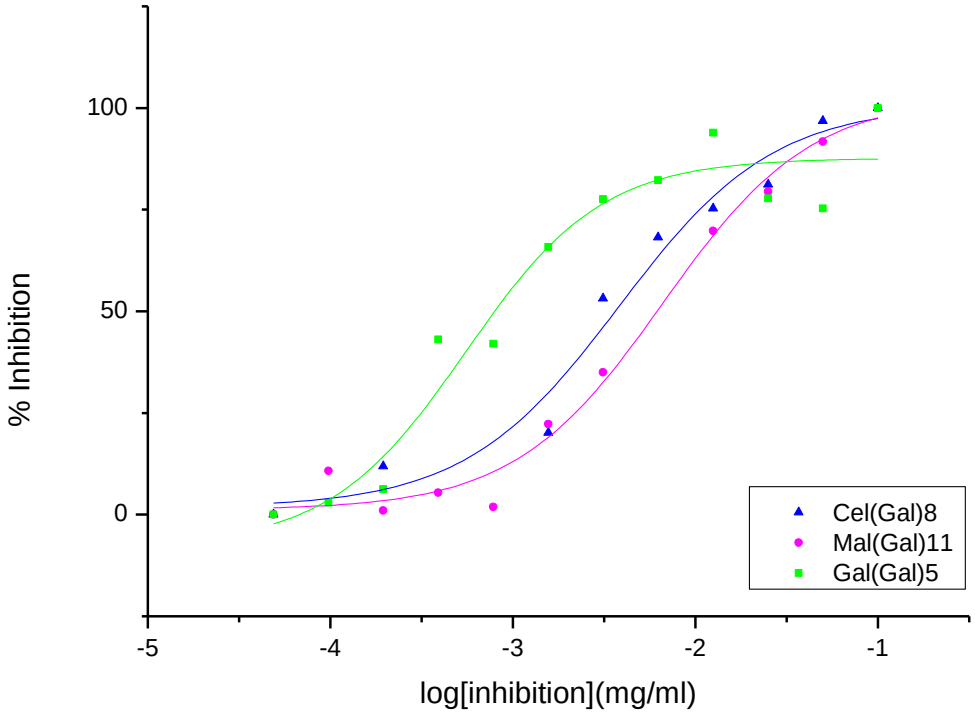


Figure 18: A graph showing the % inhibition exhibited by the three synthesised star glycoclusters as a function of mass concentration.

Inhibition curves from the fluorescence assay (figure 18) show all three glycoclusters showed inhibitory activity. This inhibition far surpassed that of monovalent galactose by approximately 100 times, which was also tested but the binding was too weak to display graphically. The sigmoidal shape is typical of this test and the increased activity of $Gal(Gal)_5$ above the other higher-valent glycoclusters is possibly due to the reason mass concentration was used, i.e. less moles of $Mal(Gal)_{11}$ present in the assay. Therefore assays using molar concentrations of the glycoclusters were carried out to yield the following results.

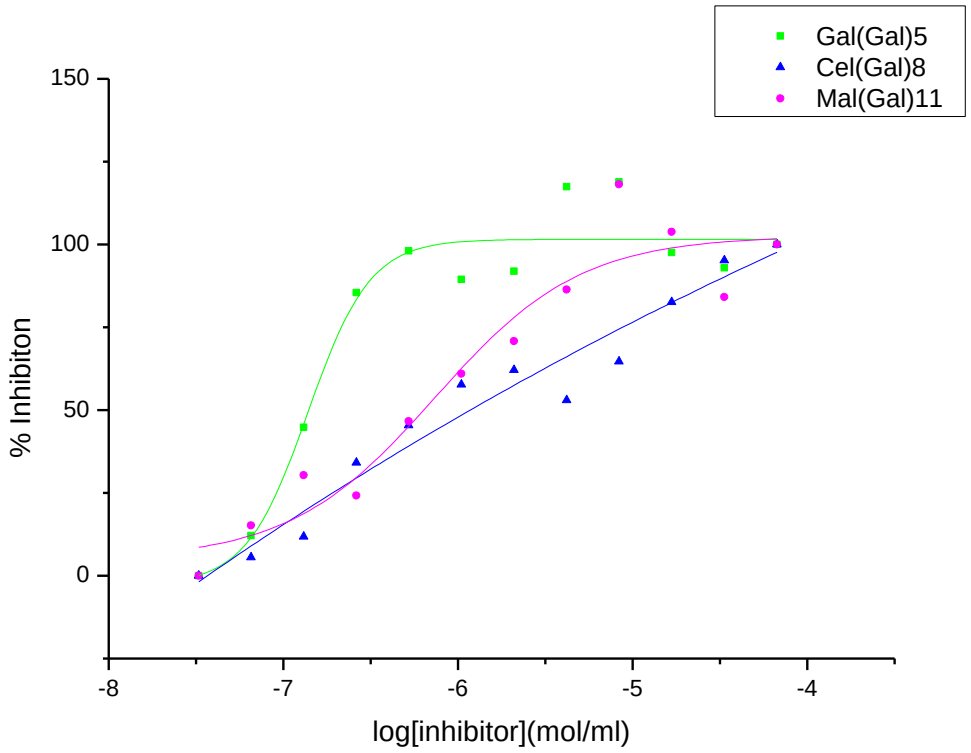


Figure 19: A graph showing the % inhibition exhibited by the three synthesised star glycoclusters as a function of molar concentration.

Despite the molar concentrations of each inhibitor being equal, $Gal(Gal)_5$ still showed greater inhibitory activity. Initially this was surprising as it was thought the higher valent cellobiose and maltotriose would show higher affinity for the cholera toxin resulting in increased inhibition. As the glycosylated galactose is responsible for the binding to the cholera lectin, the inhibitory activity of each versus the moles of galactose was also graphically analysed (figure 20).

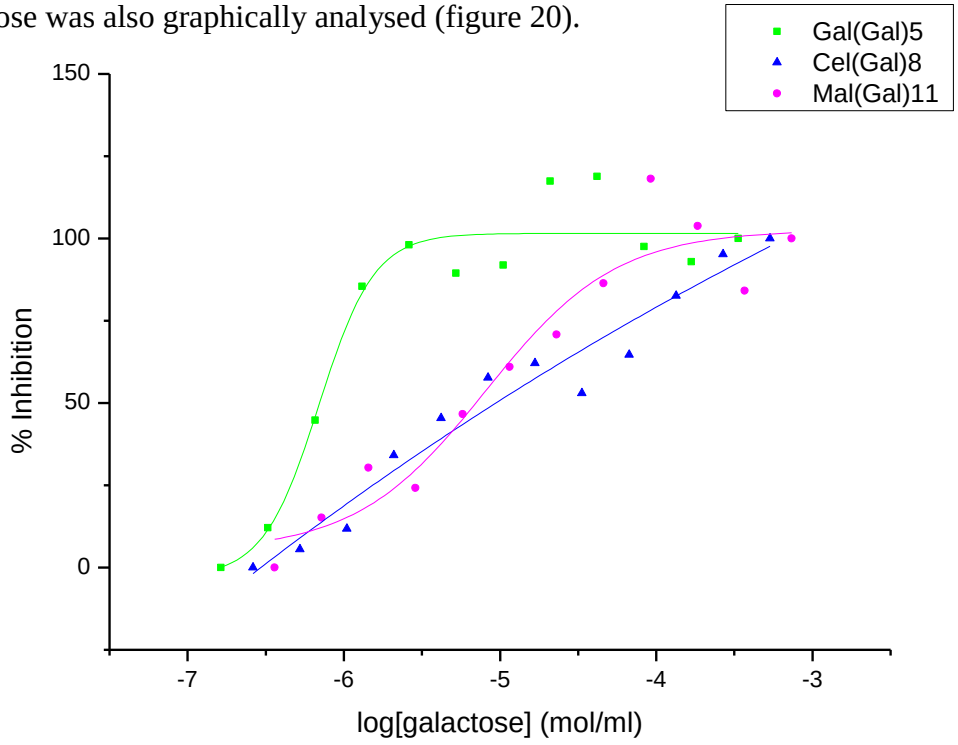


Figure 20: A graph showing the % inhibition exhibited by the three synthesised star glycoclusters as a function of galactose concentration.

The results of this graph confirmed the unexpected superlative activity of $Gal(Gal)_5$. This activity provoked further investigation into the actual binding site of the cholera toxin subunit. Previous work by Richards *et al*. found that increasing the galactose linker length increased activity.[48] Therefore it is possible that the long, narrow binding site cavity may be the reason for the observed trend. In other words, the increased hydrodynamic size of the synthesised glycoclusters may hinder the approach to the CT-B binding site and hence decrease their inhibitory effects. (See supplementary information for space fill representations.)

In order to test this theory, the inhibitory effects of the glycoclusters were tested on peanut agglutinin (PNA). PNA is a plant lectin protein derived from the fruits of Arachis hypogaea and binds the carbohydrate sequence Gal-(β-1-3)-GalNAc and therefore GM1. However, unlike CT-B, the PNA binding site is much more structurally open and accessible (figure 21).

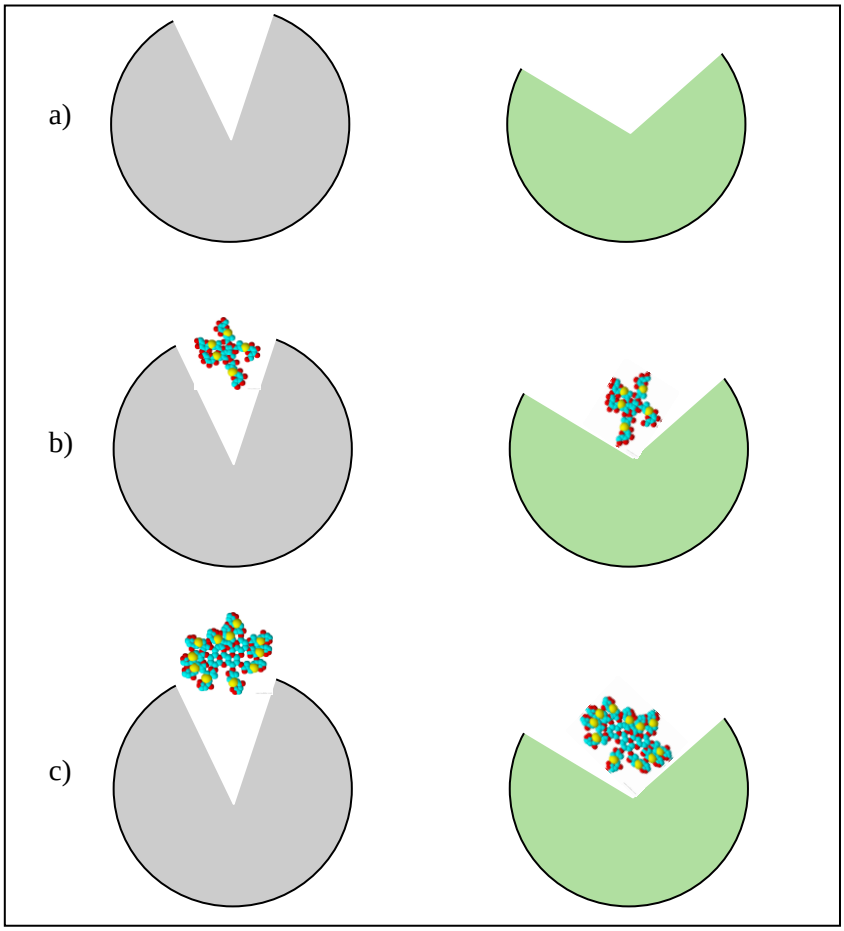


Figure 21: Cartoon representation of the relative binding sites of CT (left) and PNA (right) and their corresponding binding modes with the synthesised glycoclusters.

The previous cartoon-style representation describes the relative binding sites of CT-B (left) and PNA (right) (a). The smallest glycocluster Gal(Gal)$_5$ can easily access both sites to form lectin-carbohydrate interactions (b). However upon using a larger inhibitor e.g Mal(Gal)$_{11}$, the access to the CT-B receptors is hindered and there is only increased inhibition of PNA due to ease of access and increased carbohydrates for binding. This theory would account for the reduction in % inhibition upon increasing hydrodynamic size of the inhibitor.

The results for the inhibition assay of PNA were very interesting. The graph below representing mass concentrations of the inhibitors (0.1mg/ml, same as CT assay) shows how the activity of each is much more similar (figure 22). There are some data points highlighted on the graph for Gal(Gal)$_5$ that show higher activity than the larger glycoclusters, however considering the graph represents mass concentration, the results provide highly intriguing evidence towards the postulation that the increased size of the glycocluster reduces inhibition of CT-B.

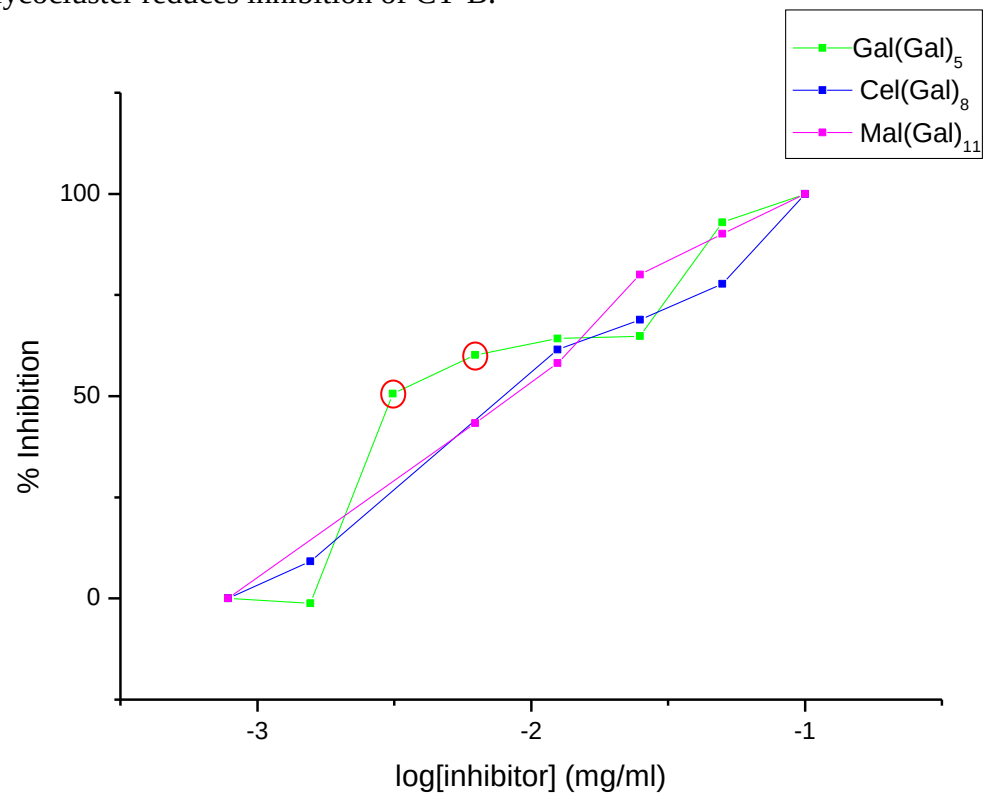


Figure 22: A graph showing the % inhibition of PNA exhibited by the three synthesised star glycoclusters as a function of mass concentration. (Possible anomalous results are circled.)

A further graph taking into consideration the molar concentrations of the inhibitors was plotted (figure 23). This graph shows a general trend of decreasing inhibition activity with decrease in size. This trend is as expected, the multivalency effect provided by the

larger glycoclusters increases binding affinity and hence inhibition. These findings also provide strong verification of the rationale for the previous CT results.

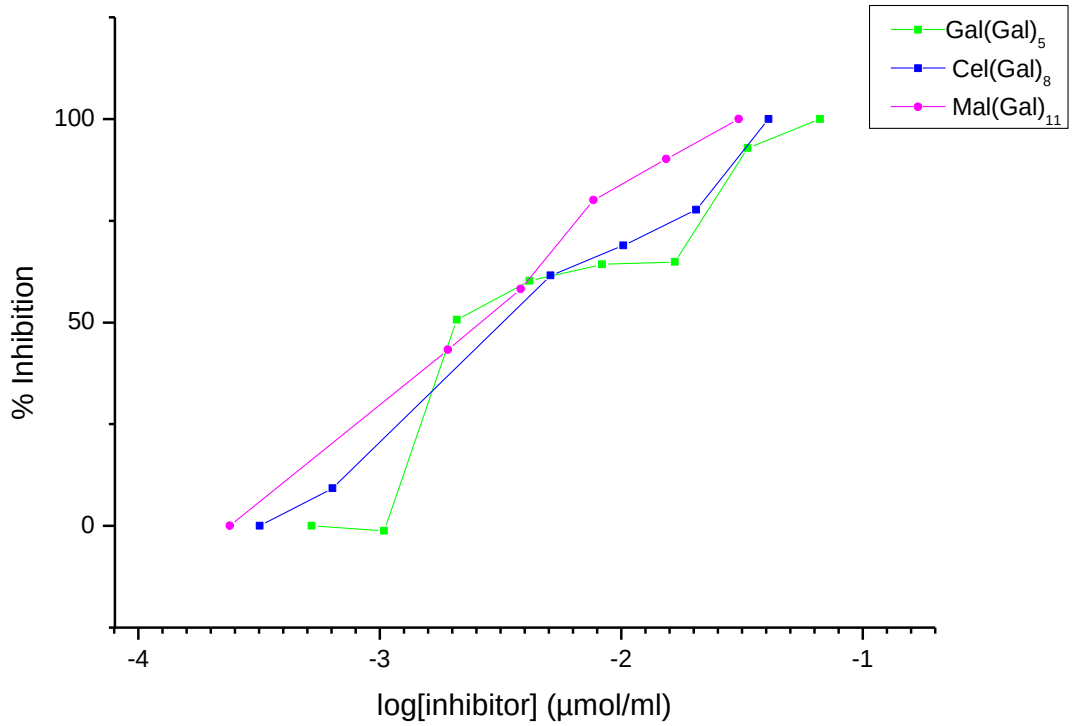


Figure 23: A graph showing the % inhibition of PNA exhibited by the three synthesised star glycoclusters as a function of molar concentration.

The MIC 50 values for the inhibition assays for both the cholera toxin and the PNA showed similar trends to the inhibiton assays. MIC 50 for CT-B is shown below.

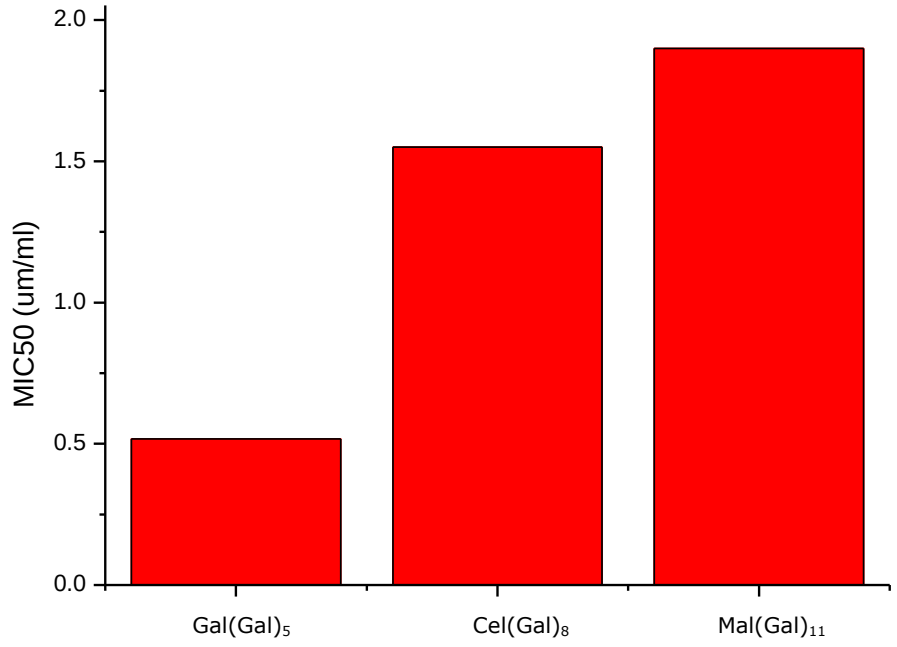


Figure 24: Bar chart representing MIC values for the three glycoclusters for the 50% inhibition of CT-B.

The MIC 50 is lowest for Gal$(Gal)_5$ and this concentration increases with increasing size of polymer. This supports the observed trend that the larger the polymer, the less effectively it inhibits cholera toxin activity promoting the idea of CT-B binding site restrictions.

On the other hand, the MIC 50 data for PNA inhibition showed the opposite relationship. The values for the PNA MIC 50 charts were taken using points on the line of best fit for the molar inhibition curve, (see supplementary information).

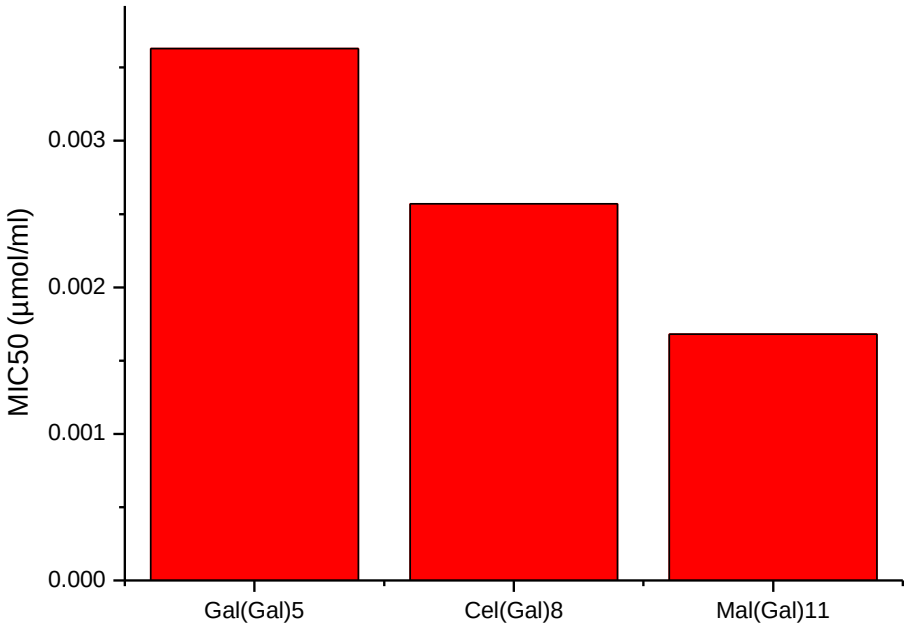


Figure 25: Bar chart representing MIC values for the three glycoclusters for the 50% inhibition of PNA.

The results gained from fluorescence assays indeed show clear trends and relationships however this technique is not without its disadvantages. Firstly the soaring cost of fluorescently labelled proteins and GM1 results in the availability of very small amounts of these substrates. Therefore with the level of optimisation required, experimenting with different concentrations (both molar and mass), repeats of the tests were not carried out. Therefore standard errors could not be applied to the results gained from these experiments. Furthermore, due to the heterogenous chemical nature of proteins, some may become labelled more efficiently than others. Therefore the fluorescence value is only indicative of the level of protein bound and cannot give an actual value. To overcome this calibration curves using well established standards must be formed.

The comparison of the binding assay results for CT-B and PNA were very supportive of the theory that the increased hydrodynamic size of the di- and trisaccharide derivatives hindered inhibition of CT-B. However, for reasons previously described, other methods should be explored to support this finding. In order to gain further evidence to support this hypothesis, molecular modelling docking software was use to visualise the binding interaction between the cholera toxin and the synthesised glycoclusters. Docking is a method that uses molecular modelling software to produce three dimensional images of the predicted binding mode. Software was initially used to convert the structures into the appropriate files for the docking software to utilise. The SwissDock program was then employed to generate the docking information and this was analysed using Chimera visualisation software. However unfortunately, due to the large size of the Cel$(Gal)_8$ and Mal$(Gal)_{11,}$ the files required to carry out the docking simply could not be constructed. Despite this, the docking analysis of the Gal$(Gal)_5$ bound to CT-B was quite revealing.

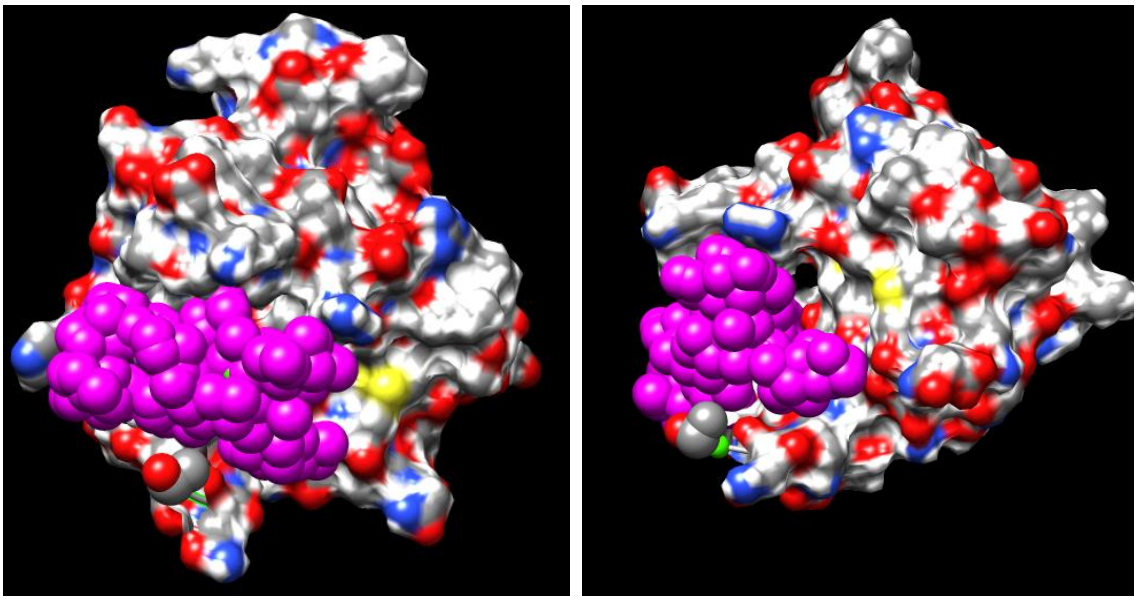


Figure 26: Hi-res image of docking prediction of CT-B and Gal$(Gal)_5$ binding mode created using Chimera.

These images show the synthesised Gal$(Gal)_5$ bound to the cholera toxin B subunit. It is clear the inhibitor is orientated in close proximity to the protein and the size appears to be a favourable fit for the binding crevice. The image on the right shows a side-view of the complex highlighting the depth of the binding site. Using these images it is possible to envisage how the much larger inhibitors would have a more hindered approach and less favourable fit with the binding site.

4. Conclusions

This work describes the two-part synthesis of novel, star glycoclusters. Methacrylated mono-, di- and trisaccharide cores were successfully synthesised and a simple method for their purification was achieved. Post-polymerisation modification was then carried out *via* a Michael-type thiol-ene addition. Thio-galactose was added to each of the methacrylate groups using TEA as a catalyst under very mild conditions. Both ^{13}C and ^{1}H NMR confirms the products synthesised were of high purity with high percentage conversions values.

Upon successful optimisation, synthesis and purification of these glycoclusters, fluorescence assays were carried out in order to assess their inhibitory activities. The initial results of the cholera toxin-B assay were not as expected as the increase in size and therefore valency of the glycoclusters did not increase inhibition of the toxin. Further repeats at varying concentrations suggested the same trend and so it was postulated that the binding site of the cholera toxin may be of a shape that preferably accommodates the smaller $Gal(Gal)_5$ over the larger inhibitors. In order to prove this theory, fluorescence assays using PNA were conducted. PNA was chosen as although it binds to galactose in the same way as CT-B, the binding site is much more open and accessible. The results of these assays showed that increasing inhibitor valency (Gal → Mal) and hence galactose concentration, increased inhibitory action. This trend not only highlights the effective inhibitory action of the synthesised glycoclusters and importance of the multivalency effect, but also provides interesting insight into the binding site of cholera-toxin-B. Docking analysis provided even more interesting evidence regarding the preferred carbohydrate orientation of the cholera toxin. All in all, these findings provide valid evidence for the great potential of glycopolymers for anti-adhesion therapies at low concentrations as well as an interesting platform for future work investigating the most effective structures for CT binding.

5. Future Work

This work provides an interesting platform for a wide range of future work and investigation. The synthesised glycoclusters are not monodisperse and this will hinder applications in biomedicine so optimisation of the synthesis is required.

Although the proposed binding theory was supported by both the inhibition assays and the docking study, molecular dynamics simulations should be carried out to provide further confirmation.

The synthesised glycoclusters successfully inhibited both CT-B and PNA. In order to find the most efficient use for these inhibitors, a range of structural analogues could be synthesised for example with different linker lengths between the core and outer sugars.

Further interesting work could see these glycoclusters tested against other pathogens that use galactose to mediate host adhesion e.g ricin or the attachment of these inhibitors to a protein to see if this improves the activity.

This study provides firm evidence that carbohydrate based therapeutics are worthy targets for the development of novel anti-infectives. However, there are many challenges that still must be overcome before this is a viable option for pharmaceutical companies such as the high cost assosciated with the mass-production of these types of drugs.

5. Experimental

5.1 Analytical Methods

^{1}H and ^{13}C NMR spectra were achieved using the Brüker DPX-400 spectrometer using deuterated solvents obtained from Aldrich. Chemical shifts were recorded as δ values in parts per million (ppm) and referenced to the solvent used. The Brüker DRX-500 spectrometer was used to obtain ^{1}H and ^{13}C NMR for all final products.

HPLC was used to analyse composition and purity of the step one synthesis products. A column of silica (~200 g) pre-eluted with 50:50 ethyl acteate/petroleum ether was used to collect approximately 24 fractions. These fractions were then analysed using TLC techniques and the appropriate fractions were collected and concentrated under vacuum.

Mass spectrometry was carried out on all methacrylated sugars using the LC-ICP-MS instrument.

GPC measurements were conducted using a Varian 390-LC MDS system equipped with a PL-AS RT/MT2 autosampler. This apparatus uses refractive index and viscometry detectors, 2 x PLgel 5μm mixed-D columns (300 x 7.5 mm), 1 x PLgel 5μm guard column (50 x 7.5mm) and a Shimadzu SPD-M20A diode array detector. The collected data was then analysed using Cirrus 3.3 software. Tetrahydrofuran and *N*,*N*-Dimethylformamide (including 2% triethylamine) was used as the eluent at a flow rate of 1 ml/min^{-1}.

Finally, all fluorescence measurements were achieved using a BioTek Synergy HT Microplate reader.

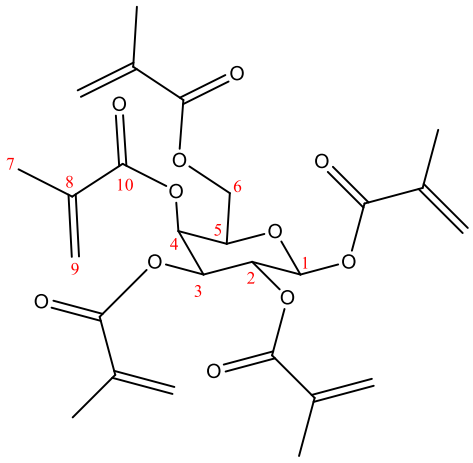
5.2 Materials

Compound	Manufacturer	CAS Number	Molecular Weight
Deuterated Solvents			
$CDCl_3$	Sigma-Aldrich	[865-49-6]	
D_2O	Sigma-Aldrich	[7789-20-0]	
Solvents			
Dichloromethane (DCM)	Fisher Scientific	[75-09-2]	
Ethanol	Fisher Scientific	[64-17-5]	
Pyridine	Sigma-Aldrich	[110-86-1]	
Ethyl Acetate	Fisher Scientific	[141-78-6]	

Petroleum Ether	Fisher Scientific	[8032-32-4]	
Reagents			
Methacrylic Anhydride	Sigma-Aldrich	[760-93-0]	154.16
Triethylamine (TEA)	Fisher Scientific	[121-44-8]	101.10 (p=0.726)
Sugars			
Glycerol	Fisher Scientific	[56-81-5]	92.09
L-(+)-Arabinose	Sigma-Aldrich	[5328-37-0]	150.13
α-D-Glucose	Sigma-Aldrich	[492-62-6]	180.2
D-(+)-Mannose	Alfa Aesar	[3458-28-4]	180.2
Galactose	Alfa Aesar	[59-23-4]	180.2
D-Cellobiose	Acros organics	[528-50-7]	342.3
Maltotriose	Sigma-Aldrich	[1109-28-0]	504.4
B-D-Thio-galactose	Carbosynth	[42891-22-5]	218.2
Inhibition Assays			
Cholera Toxin B	Sigma-aldrich	[9012-63-9]	
Peanut Agglutinin-FITC	Vector Labs Scientific	[26628-22-8]	
GM1	Carbosynth	[37758-47-7]	
HEPES	Sigma-Aldrich	[7365-45-9]	
BSA	Sigma-Aldrich	[9048-46-8]	
PBS	Sigma-Aldrich	-	

5.3 Experimental Procedures

5.3.1 General Procedure 1 (Methacrylation)



Galactose (5.067 g, 28.1 mmol) was dissolved in pyridine (150 ml, 1.85 mol) at 35°C. Methacrylic anhydride (25 ml, 0.169 mol) was added drop-wise with stirring. The solution was heated to 45°C and left to stir for 18 hours.

The methacrylated product was extracted using dichloromethane (50 ml) and the organic layer was washed five times using 3M hydrochloric acid (5 x 50 ml). The resulting product was concentrated under vacuum. This was then added to a 0.1M solution of sodium hydroxide (100 ml, 10 mmol) and stirred at room temperature for 15 hours. The product was then extracted with dichloromethane (50 ml) and concentrated under vacuum.

Purification of the methacrylated sugar was carried out using a column of silica (200 g) pre-eluted with ethyl acetate/petroleum ether 50:50 vol/vol. After elution using the same solvent mixture, the fractions were analysed by TLC and the appropriate fractions were collected and solvent removed under vacuum. This yielded a cream oil that was further analysed by ^{1}H NMR, ^{13}C NMR and mass spectroscopy.

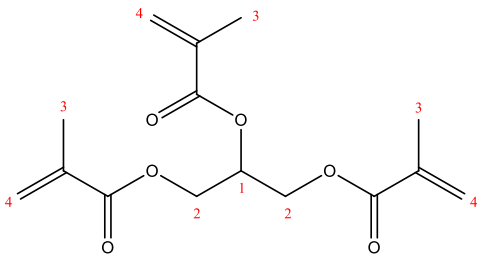
^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.20 (5H, m, H_{9a}), 5.80 (5H, m, H_{9b}), 4.49-4.06 (6H, m, $H_{1\text{-}6}$), 1.98 (15H, m, H_7).

^{13}C NMR ($CDCl_3$) 400MHz, δ_{ppm}; 171.22 (C_{10}), 135.83,129.08 ($C_{9,}$ C_{10}), 60.46-14.27 (6 ring carbons), 92.70 (small anomeric peak, others not visible).

MS (ESI) m/z = 475.0 [M^+], 927.4 [$2M^+$-Na^+]

The same above procedure was carried out for the methacrylation of the following sugars to produce the corresponding methacrylate:

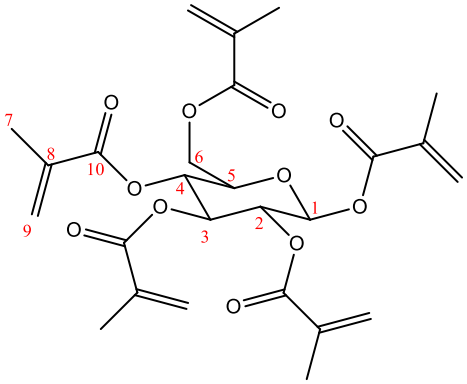
Glycerol Methacrylate:



Glycerol (1.103 g, 11.98 mmol), pyridine (10 ml) and methacrylic anhydride (6 ml, 40.54 mmol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.22 (3H, d, H_{4a}), 5.64 (3H, d, H_{4b}), 4.38 (1H, q, H_1), 1.95 (4H, d, H_2) 1.26 (9H, t, H_3).

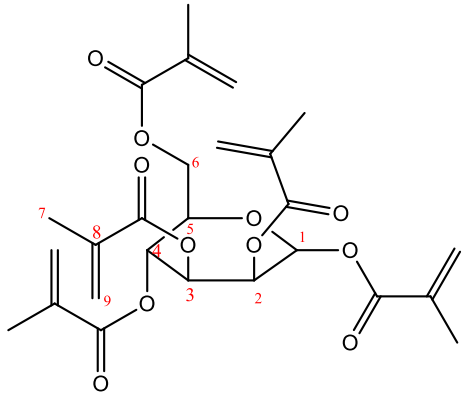
Glucose Methacrylate:



Glucose (1.061 g, 5.86 mmol), pyridine (10 ml), methacrylic anhydride (6 ml, 40.54 mmol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.21 (5H, d, H_{9a}), 5.80 (5H, d, H_{9b}), 4.58-3.65 (6H, m, $H_{1\text{-}6}$), 1.94 (15H, t, H_7).

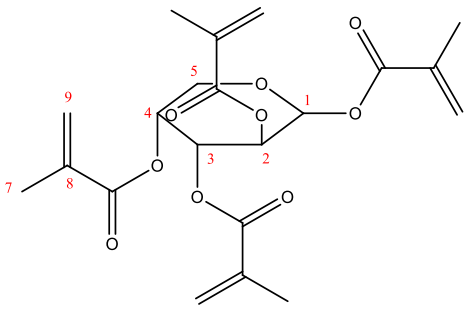
Mannose Methacrylate:



Mannose (5.006 g, 27.80 mmol), pyridine (150 ml, 1.85 mol), methacrylic anhydride (25 ml, 0.169 mol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.24 (5H, d, H_{9a}), 5.68 (5H, d, H_{9b}), 4.47-3.92 (6H, m, $H_{1\text{-}6}$), 2.04 (15H, t, H_7).

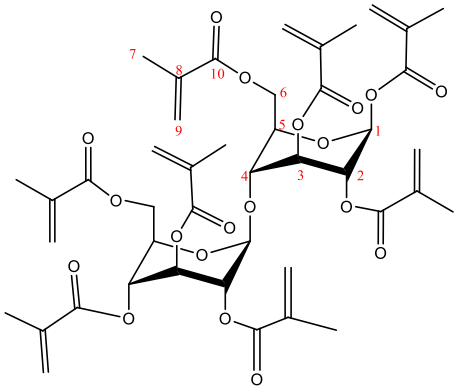
Arabinose Methacrylate:



Arabinose (1.018 g, 6.78 mmol), pyridine (10 ml), methacrylic anhydride (5 ml, 33.7 mmol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.25 (4H, m, H_{9a}), 5.68 (4H, d, H_{9b}), 4.52-3.61 (5H, m, $H_{1\text{-}5}$), 1.95 (12H, t, H_7).

Cellobiose Methacrylate:

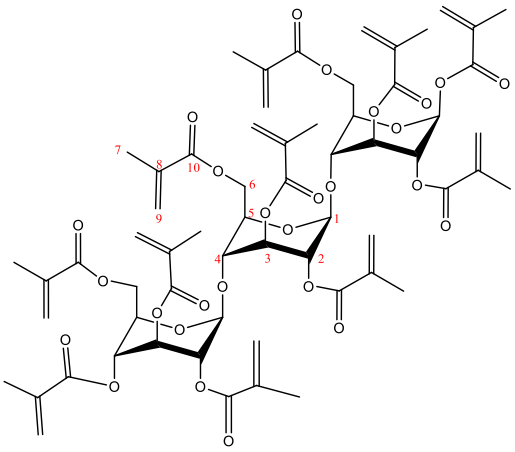


Cellobiose (1.004 g, 2.93 mmol), pyridine (20 ml), methacrylic anhydride (5 ml, 33.7 mmol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.23 (8H, m, H_{9a}), 5.82 (8H, q, H_{9b}), 4.76-4.65 (12H, m, $H_{1\text{-}6}$), 2.00 (24H, t, H_7).

^{13}C NMR ($CDCl_3$) 400MHz, δ_{ppm}; 172.56 (C_{10}), 135.73,128.01 ($C_{9,}$ C_{10}), 39.48-14.33 (12 ring carbons),101.88, 92.30, 81.88 (anomeric peaks, other not visible).

Maltotriose Methacrylate:

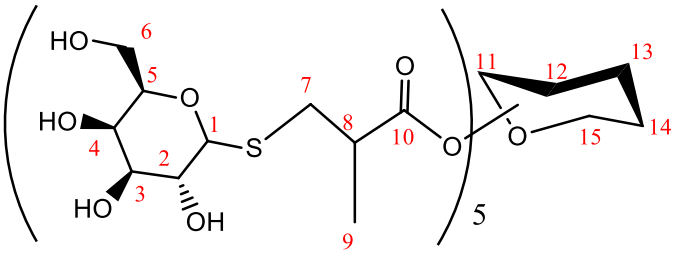


Maltotriose (0.33 g, 0.65 mmol), pyridine (10 ml), methacrylic anhydride (1.2 ml, 8.07mmol).

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 6.23 (11H, d, H_{9a}), 5.82 (11H, t, H_{9b}), 4.67-3.45 (18H, m, $H_{1\text{-}6}$), 2.01 (24H, dt, H_7).

^{13}C NMR ($CDCl_3$) 400MHz, δ_{ppm}; 171.96 (C_{10}), 135.49, 127.43 ($C_{9,}$ C_{10}), 53.23-14.41 (12 ring carbons),101.51 (anomeric peak, others not visible).

5.3.2 General Procedure 2 (Glycosylation)



Thio-galactose (97 mg, 0.5 mmol) was dissolved in ethanol (5 ml) and H_2O (1 ml) and galactose methacrylate (43 mg, 0.083 mmol) was added. This solution was degassed in nitrogen for 15 minutes. Catalytic triethylamine (0.603 µl, 0.0083 mmol) was also degassed in nitrogen and added to the solution. This solution was stirred for 20 hours.

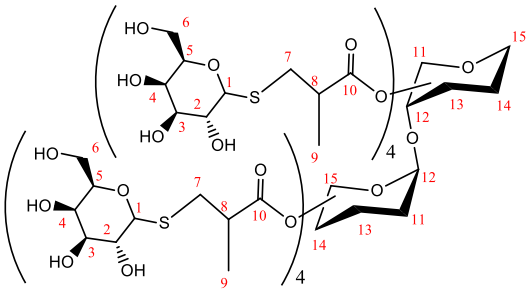
Ethanol removed using a rotary evaporator and the resulting product was dialysed in H_2O over 24 hours. Following dialysis the product was then freeze dried for 15 hours and ^{1}H NMR, ^{13}C NMR and mass spectroscopy were carried out.

^{1}H NMR ($CDCl_3$) 400MHz, $\delta_{ppm;}$ 4.40 (anomeric protons, d) 4.32 (anomeric protons, d), 3.88-3.48 (36H, m, $H_{1\text{-}6,\ 11\text{-}15}$),3.36 (10H, s, H_7), 3.21 (5H, q, H_8), 2.11 (15H, s, H_9).

^{13}C NMR ($CDCl_3$) 500MHz, δ_{ppm}; 215.3 (C_{10}), 120.3 (C_6 of core galactose) 86.1, 84.0 (anomeric thiol) 79.0-68.7 ($C_{1\text{-}5,\ 11\text{-}15}$), 61.1 ($C_7$), 46.7 ($C_6$), 30.2 ($C_8$), 11.5 ($C_9$).

The above procedure was repeated using sugar cores of cellobiose and maltotriose methacrylate.

Cellobiose

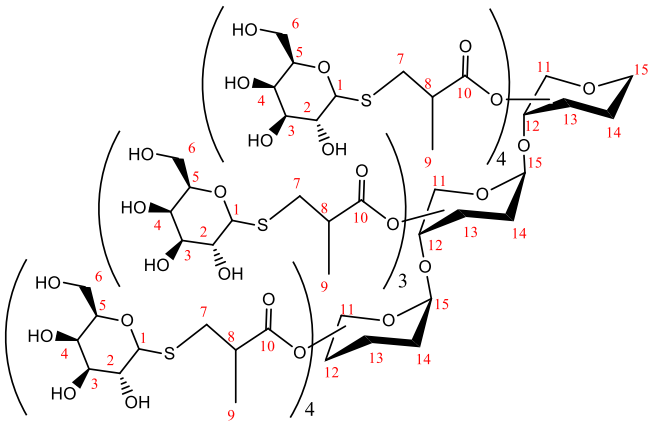


Cellobiose-methacrylate (0.142 g, 0.160 mmol), thiol-galactose (0.25 g, 1.30 mmol), ethanol (5 ml), water (1 ml), triethylamine (2.20 µl).

^{1}H NMR (D_2O) 400MHz, $\delta_{ppm;}$ 3.91-2.65 (60H, m, $H_{1\text{-}6,11\text{-}15}$), 3.14-3.10 (8H, dddd, H_8), 4.47,4.45 and 4.41,4.39 (anomeric core protons d,d), 1.84,1.80 (16H, d, H_7), 1.10,1.08 (24H, d, H_9).

^{13}C NMR ($CDCl_3$) 500MHz, δ_{ppm}; 184.3 (C_{10}), 142.4 (unreacted methacrylate), 120.2 (C_6 of core galactose) 102.6,102.5 (anomeric C of core sugar) 96.0,91.9 (anomeric thio-galactose) 86.4-86.9 ($C_{1\text{-}5,\ 11\text{-}15}$), 60.8 ($C_7$), 43.7 ($C_8$), 34.3 ($C_6$), 17.4 ($C_9$).

Maltotriose:



Maltotriose-methacrylate (0.060 g, 0.048 mmol), thio-galactose (0.101 g, 0.58 mmol), ethanol (5 ml), water (1 ml), triethylamine (0.660 µl).

^{1}H NMR (D_2O) 500MHz, $\delta_{ppm;}$ 3.72-3.34 (84H, m, $H_{1\text{-}6,11\text{-}15}$), 3.14-3.10 (11H, q, H_8), 4.32,4.30-4.25,4.23 (anomeric core protons d,d), 3.91,3.90-3.85,3.84 (anomeric thiol-gal protons, d,d) 3.50,3.51 (22H, d, H_7), 1.21,1.20,1.18 (33H, t, H_9).

^{13}C NMR ($CDCl_3$) 500MHz, δ_{ppm}; 215.4 (C_{10}), 103.8 (anomeric C on core sugar), 84.8-68.7 ($C_{1\text{-}5,\ 11\text{-}15}$), 61.1 ($C_7$), 46.7 ($C_6$), 30.2 ($C_8$), 11.5 ($C_9$).

All final % conversions were calculated using the measured ratio of methacrylate to corresponding CH_2 peaks. Unreacted methacrylate peaks display two singlets visible at around and 5.55 and 5.25ppm.

5.3.3 CT-B/ PNA-FITC GM1 Fluorescence Assay

Stock solutions of PBS (1 tablet per 200ml) and GM1 (0.1mg/ml in PBS) were made up. GM1 (150µl) was pipetted into each tray and incubated at 10°C for 1 hour. Unattached ganglioside was removed by washing with PBS followed by a 'blocking' procedure through the addition of BSA-PBS solution (1mg/ml). The wells were then incubated for 30 minutes at room temperature and then washed extensively with PBS. Stock solutions (1mg/ml in HEPES) of different star sugar polymers were diluted (0.5mg/ml) and added to the first column of wells (160 µl) of a low binding plate. Serial dilution across the wells was carried out using the final column as a control. CT-B (20µl, 0.3mg/ml) was added to each of the 96 wells and then incubated at 37°C for 30 minutes. The contents of the low binding plate were then transferred to the prepared high binding plate and incubated for a further 30 minutes at 37°C. The wells were then washed in PBS removing any unbound protein. The BioTek Synergy HT Microplate reader was then used to measure fluorescence at excitation/emission wavelengths of 485/528 nm with a sensitivity of 75 nm.

7. References

1. R. A. Dwek, *Chemical Reviews*, 1996, **96**, 683-720.
2. M. A. Gauthier, M. I. Gibson and H.-A. Klok, *Angewandte Chemie-International Edition*, 2009, **48**, 48-58.
3. R. S. Haltiwanger and J. B. Lowe, *Annual Review of Biochemistry*, 2004, **73**, 491-537.
4. B. D. Polizzotti and K. L. Kiick, *Biomacromolecules*, 2006, **7**, 483-490.
5. B. D. Polizzotti, R. Maheshwari, J. Vinkenborg and K. L. Kiick, *Macromolecules*, 2007, **40**, 7103-7110.
6. R. A. Laine, *Pure and Applied Chemistry*, 1997, **69**, 1867-1873.
7. V. Ladmiral, E. Melia and D. M. Haddleton, *European Polymer Journal*, 2004, **40**, 431-449.
8. K. T. Pilobello and L. K. Mahal, *Current Opinion in Chemical Biology*, 2007, **11**, 300-305.
9. Y. Cheong, G. Shim, D. Kang and Y. Kim, *Journal of Molecular Structure*, 1999, **475**, 219-232.
10. M. Ambrosi, N. R. Cameron and B. G. Davis, *Organic & Biomolecular Chemistry*, 2005, **3**, 1593-1608.
11. C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357-2364.
12. W. I. Weis and K. Drickamer, *Annual Review of Biochemistry*, 1996, **65**, 441-473.
13. G. Mulvey, P. I. Kitov, P. Marcato, D. R. Bundle and G. D. Armstrong, *Biochimie*, 2001, **83**, 841-847.
14. H. Lis and N. Sharon, *Chemical Reviews*, 1998, **98**, 637-674.
15. G. F. Gensini, A. A. Conti and D. Lippi, *J Infect*, 2007, **54**, 221-224.
16. N. Sharon and H. Lis, *Scientific American*, 1993, **268**, 82-89.
17. A. W. Paton, R. Morona and J. C. Paton, *Infection and Immunity*, 2001, **69**, 1967-1970.
18. R. J. Thomas, *Bioengineered bugs*, 2010, **1**, 17-30.
19. R. Reyburn, D. R. Kim, M. Emch, A. Khatib, L. von Seidlein and M. Ali, *American Journal of Tropical Medicine and Hygiene*, 2011, **84**, 862-869.
20. A. V. Pukin, H. M. Branderhorst, C. Sisu, C. A. G. M. Weijers, M. Gilbert, R. M. J. Liskamp, G. M. Visser, H. Zuilhof and R. J. Pieters, *Chembiochem*, 2007, **8**, 1500-1503.
21. L. de Haan and T. R. Hirst, *Molecular Membrane Biology*, 2004, **21**, 77-92.
22. K. Sandvig, *Toxicon*, 2001, **39**, 1629-1635.

23. A. A. Lindberg, J. E. Brown, N. Stromberg, M. Westlingryd, J. E. Schultz and K. A. Karlsson, *Journal of Biological Chemistry*, 1987, **262**, 1779-1785.

24. J. E. Brown, S. W. Rothman and B. P. Doctor, *Infection and Immunity*, 1980, **29**, 98-107.

25. V. L. Tesh, *Future Microbiology*, 2010, **5**, 431-453.

26. A. I. V. Serna and E. C. Boedeker, *Current Opinion in Gastroenterology*, 2008, **24**, 38-47.

27. S. Nathanson, T. Kwon, M. Elmaleh, M. Charbit, E. A. Launay, J. Harambat, M. Brun, B. Ranchin, F. Bandin, S. Cloarec, G. Bourdat-Michel, C. Pietrement, G. Champion, T. Ulinski and G. Deschenes, *Clinical Journal of the American Society of Nephrology*, 2010, **5**, 1218-1228.

28. S. D. Lawn and A. I. Zumla, *Lancet*, 2011, **378**, 57-72.

29. L. J. Alderwick, G. S. Lloyd, H. Ghadbane, J. W. May, A. Bhatt, L. Eggeling, K. Fuetterer and G. S. Besra, *Plos Pathogens*, 2011, **7**.

30. I. Ofek, D. L. Hasy and N. Sharon, *Fems Immunology and Medical Microbiology*, 2003, **38**, 181-191.

31. C. Bavington and C. Page, *Respiration*, 2005, **72**, 335-344.

32. R. J. Pieters, *Medicinal Research Reviews*, 2007, **27**, 796-816.

33. N. Sharon, *Biochimica Et Biophysica Acta-General Subjects*, 2006, **1760**, 527-537.

34. N. Firon, I. Ofek and N. Sharon, *Biochemical and Biophysical Research Communications*, 1982, **105**, 1426-1432.

35. N. Firon, S. Ashkenazi, D. Mirelman, I. Ofek and N. Sharon, *Infection and Immunity*, 1987, **55**, 472-476.

36. P. I. Kitov, J. M. Sadowska, G. Mulvey, G. D. Armstrong, H. Ling, N. S. Pannu, R. J. Read and D. R. Bundle, *Nature*, 2000, **403**, 669-672.

37. E. Fan, Z. Zhang, W. E. Minke, Z. Hou, C. L. M. J. Verlinde and W. G. J. Hol, *Journal of the American Chemical Society*, 2000, **122**, 2663-2664.

38. S. Kotter, U. Krallmann-Wenzel, S. Ehlers and T. K. Lindhorst, *Journal of the Chemical Society-Perkin Transactions 1*, 1998, 2193-2200.

39. N. Nagahori, R. T. Lee, S. Nishimura, D. Page, R. Roy and Y. C. Lee, *Chembiochem*, 2002, **3**, 836-844.

40. P. M. Simon, P. L. Goode, A. Mobasseri and D. Zopf, *Infection and Immunity*, 1997, **65**, 750-757.

41. J. V. Mysore, T. Wigginton, P. M. Simon, D. Zopf, L. M. Heman-Ackah and A. Dubois, *Gastroenterology*, 1999, **117**, 1316-1325.

42. A. Gustafsson, A. Hultberg, R. Sjostrom, I. Kacskovics, M. E. Breimer, T. Boren, L. Hammarstrom and J. Holgersson, *Glycobiology*, 2006, **16**, 1-10.

43. H. Ling, A. Boodhoo, B. Hazes, M. D. Cummings, G. D. Armstrong, J. L. Brunton and R. J. Read, *Biochemistry*, 1998, **37**, 1777-1788.

44. D. J. Bast, L. Banerjee, C. Clark, R. J. Read and J. L. Brunton, *Molecular Microbiology*, 1999, **32**, 953-960.

45. V. Horejsi, P. Smolek and J. Kocourek, *Biochimica et Biophysica Acta*, 1978, **538**, 292-298.

46. Y. Miura, *Polymer Journal*, 2012, **44**, 679-689.

47. M. G. Baek and R. Roy, *Macromolecular Bioscience*, 2001, **1**, 305-311.

48. S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angewandte Chemie-International Edition*, 2012, **51**, 7812-7816.

49. M. Friedman, J. F. Cavins and J. S. Wall, *Journal of the American Chemical Society*, 1965, **87**, 3672-3682.

50. J. Rieger, K. Van Butsele, P. Lecomte, C. Detrembleur, R. Jerome and C. Jerome, *Chemical Communications*, 2005, **0**, 274-276.

51. Y. Maynard, M. G. Scott, M. H. Nahm and J. H. Ladenson, *Clinical Chemistry*, 1986, **32**, 752-757.

52. K. A. Marx, *Biomacromolecules*, 2003, **4**, 1099-1120.

53. S. Slavin, A. H. Soeriyadi, L. Voorhaar, M. R. Whittaker, C. R. Becer, C. Boyer, T. P. Davis and D. M. Haddleton, *Soft Matter*, 2012, **8**, 118-128.

54. C. Xiao, C. Zhao, P. He, Z. Tang, X. Chen and X. Jing, *Macromolecular Rapid Communications*, 2010, **31**, 991-997.

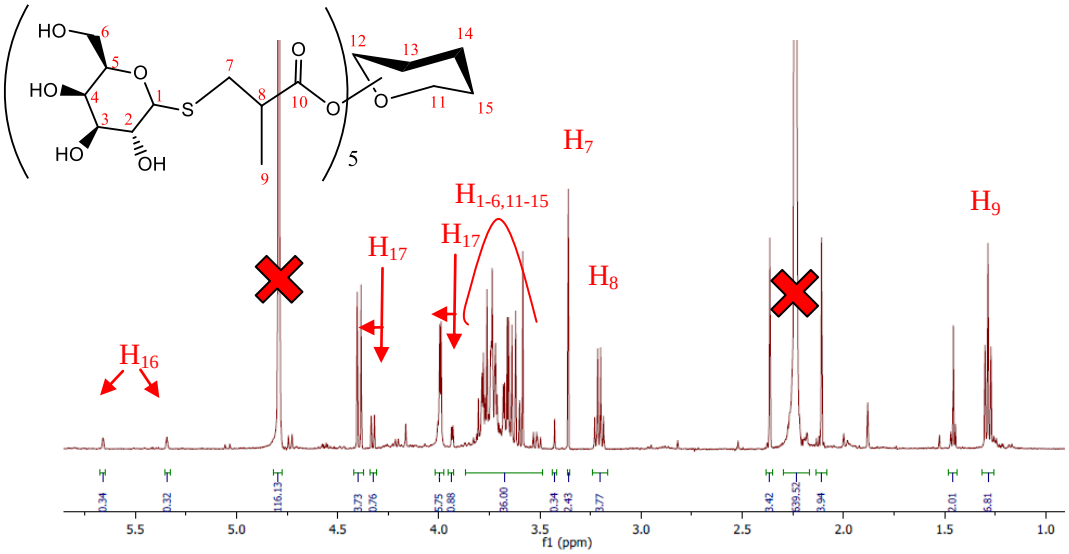
55. G.-Z. Li, R. K. Randev, A. H. Soeriyadi, G. Rees, C. Boyer, Z. Tong, T. P. Davis, C. R. Becer and D. M. Haddleton, *Polymer Chemistry*, 2010, **1**, 1196-1204.

56. L. M. Campos, K. L. Killops, R. Sakai, J. M. J. Paulusse, D. Damiron, E. Drockenmuller, B. W. Messmore and C. J. Hawker, *Macromolecules*, 2008, **41**, 7063-7070.

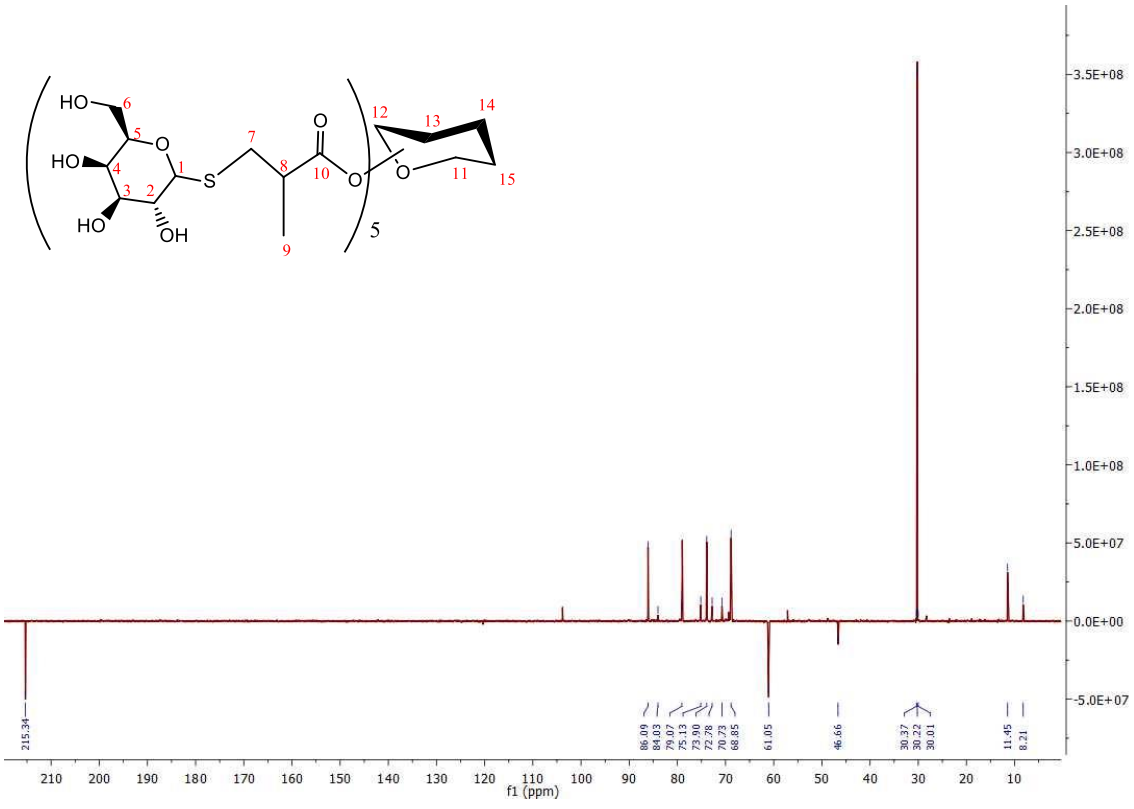
57. S. Liu and K. L. Kiick, *Macromolecules*, 2008, **41**, 764-772.

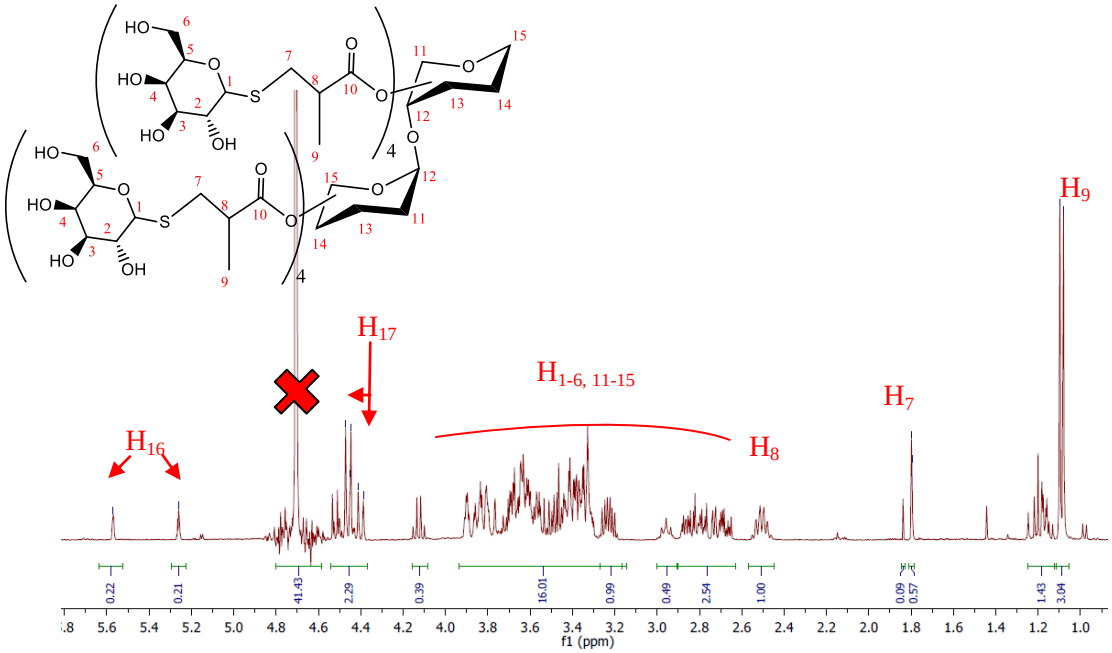
Supplementary Information

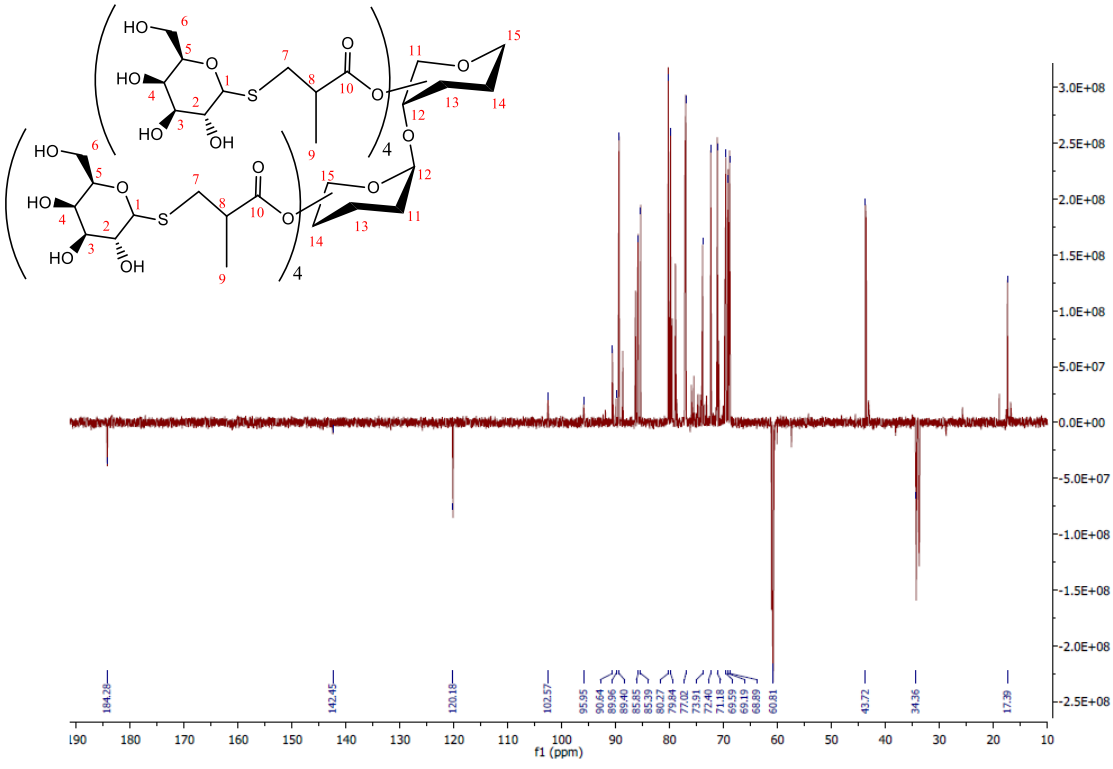
^{1}H and ^{13}C NMR of all final glycoclusters.

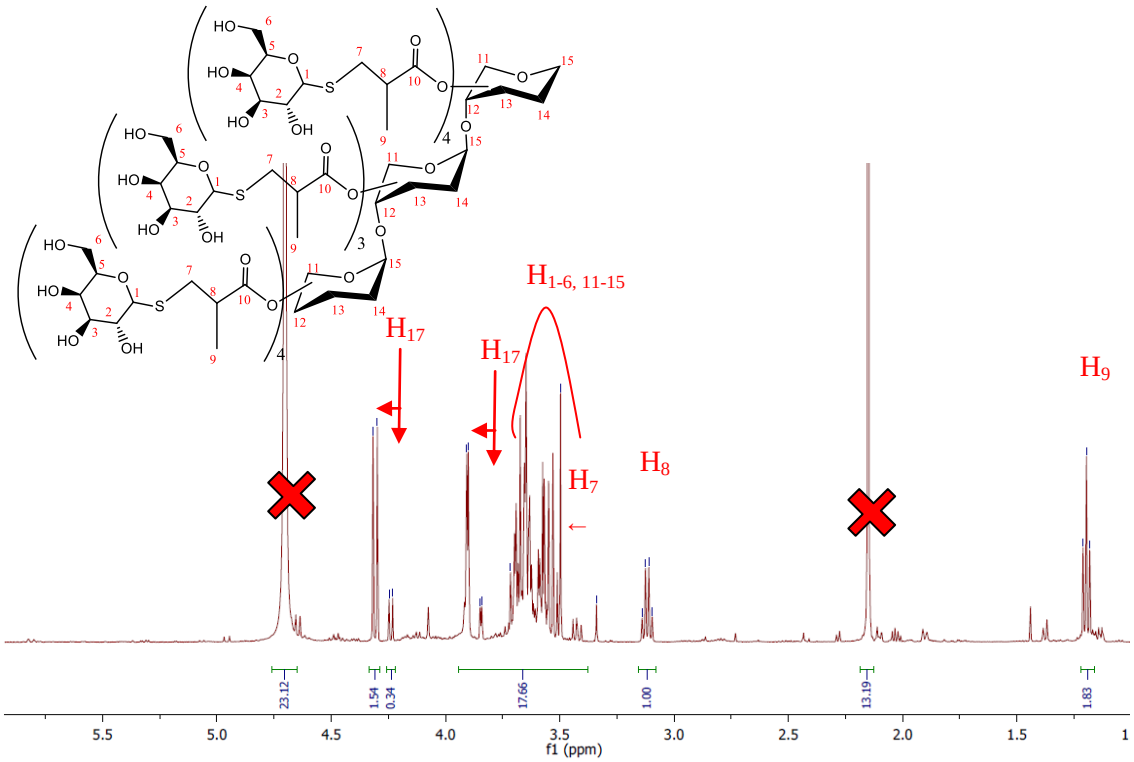


H_{16} represents unreacted methacrylate peaks and H_{17} relates to the anomeric protons of the core and outer sugars. Full assignments of carbon spectra are featured in the experimental section.

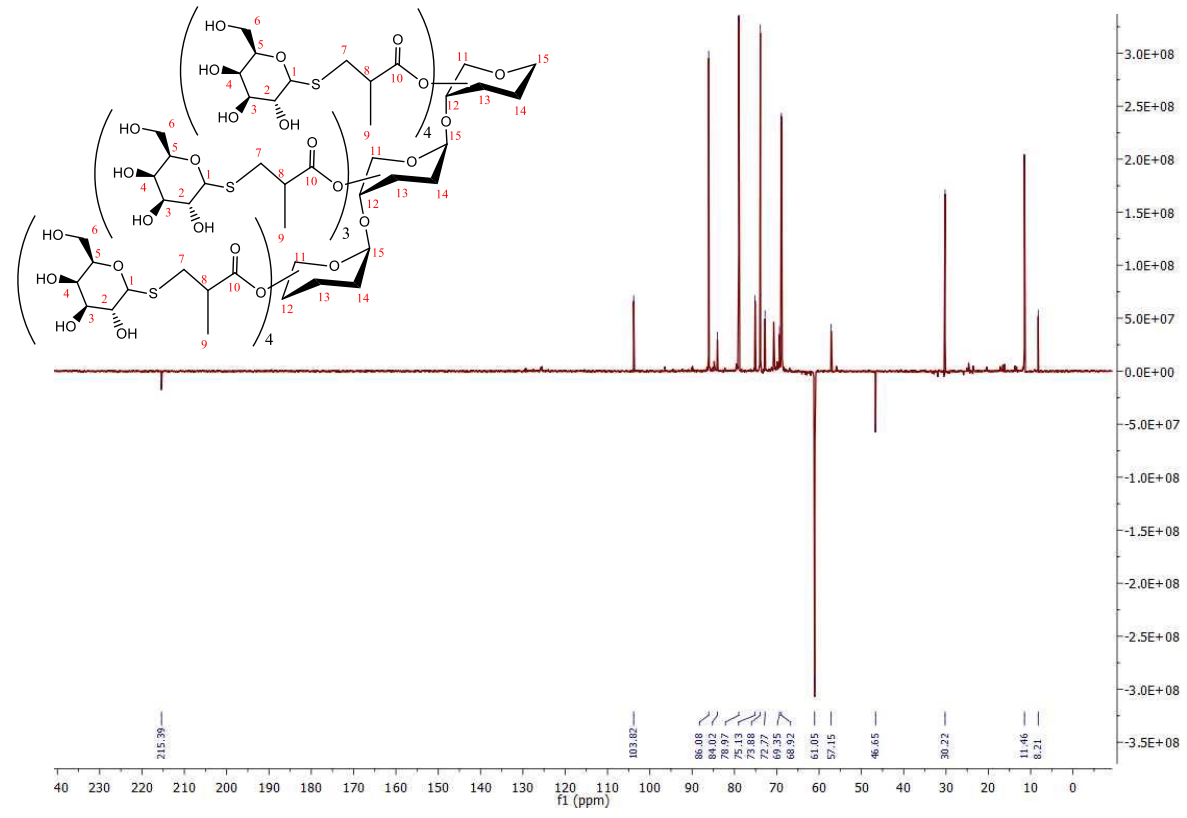



H9
H17
H1-6, 11-15
H7
H16
H8
f1 (ppm)


f1 (ppm)

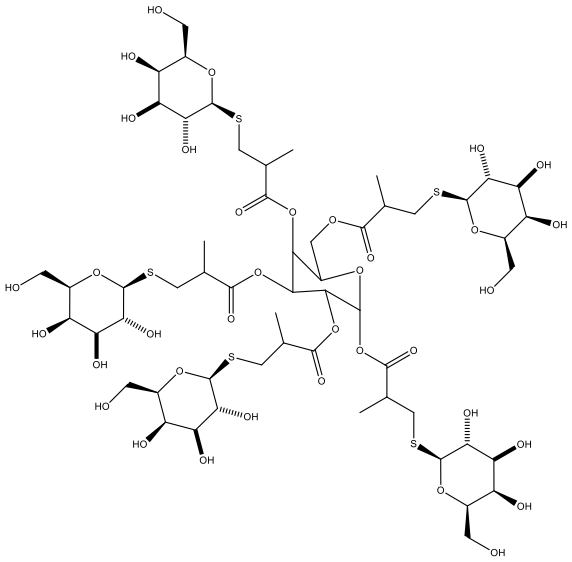


H_{17} represent anomeric hydrogen peaks fro thio-sugars and sugar core. H_7 is buried within the sugar multiplet however when zoomed in two peaks with the expected 7.5Hz splitting can be observed.



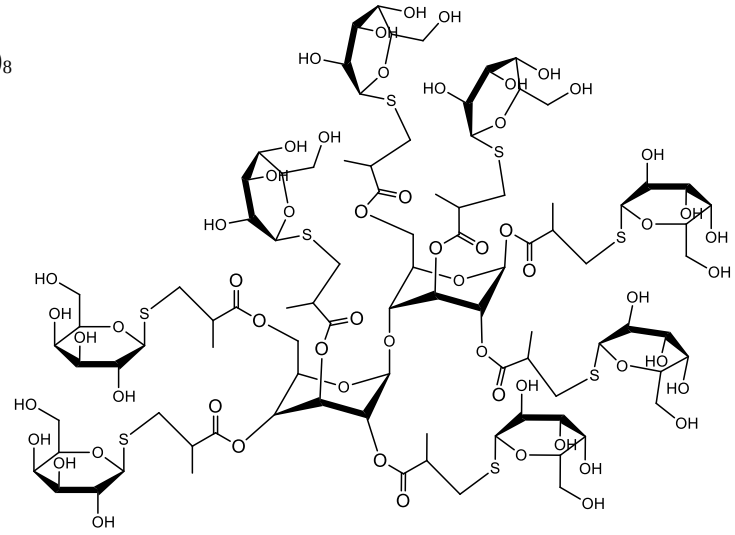
Synthesised glycoclusters represented by ChemDraw structures

$Gal(Gal)_5$



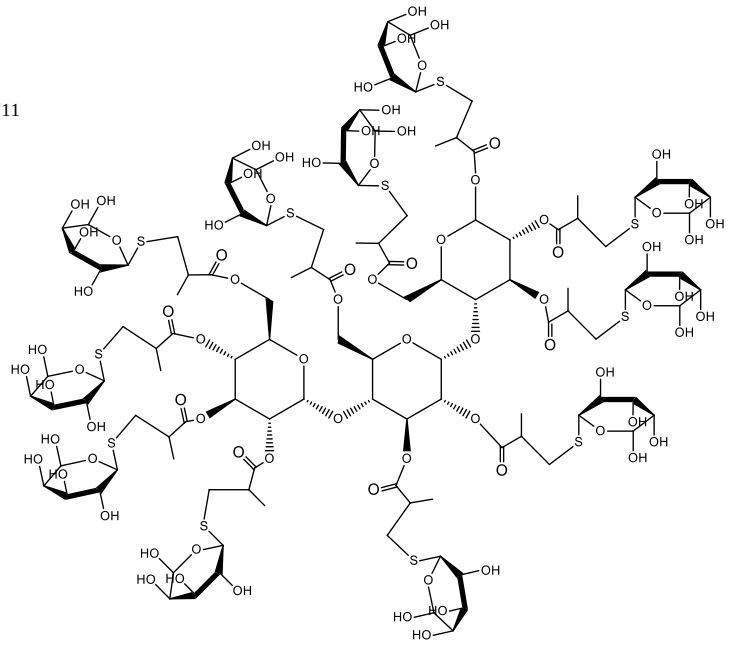
Exact Mass: 1500.40

$Cel(Gal)_8$



Exact Mass: 2454.65

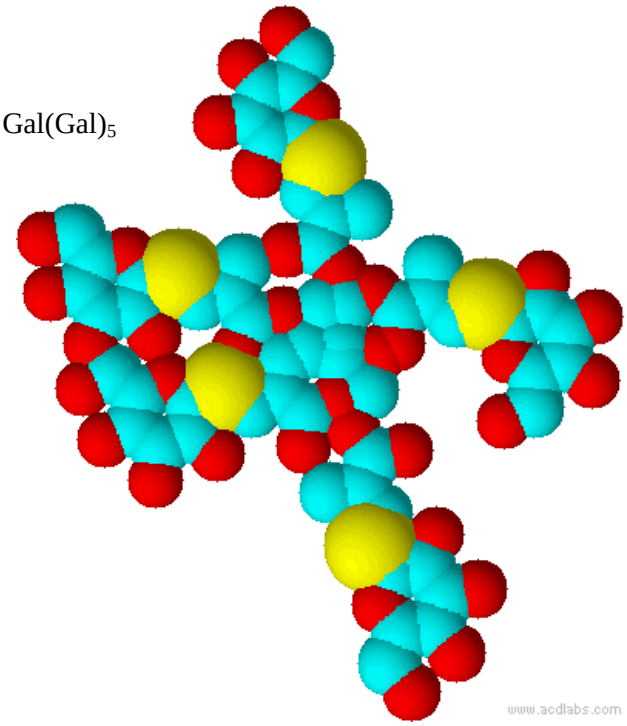
$Mal(Gal)_{11}$



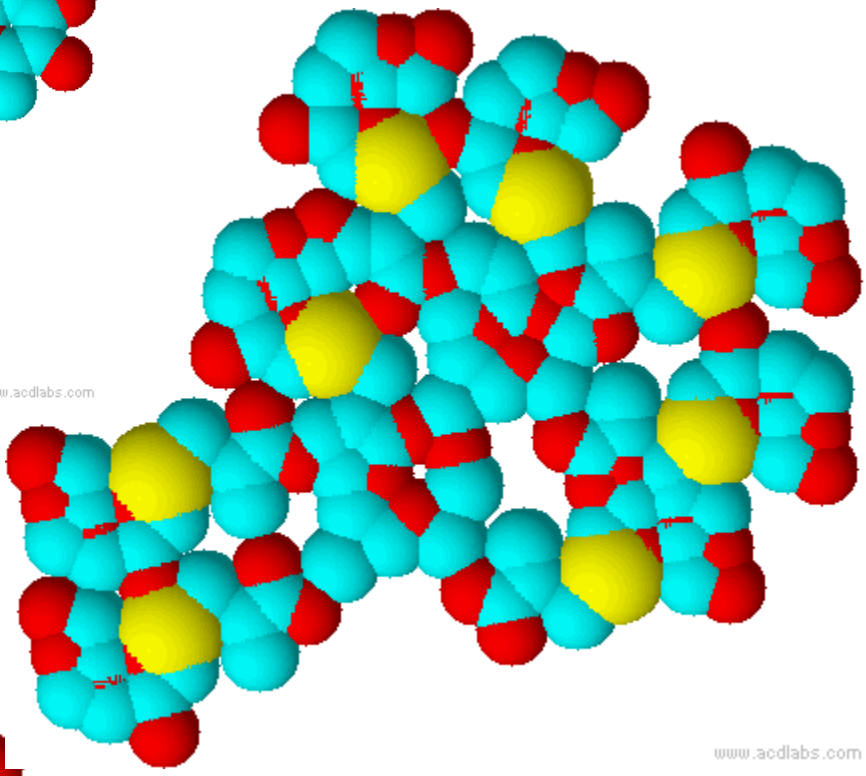
Exact Mass: 3254.73

Space fill representations of synthesised glycoclusters generated using Jmol 3D viewer.

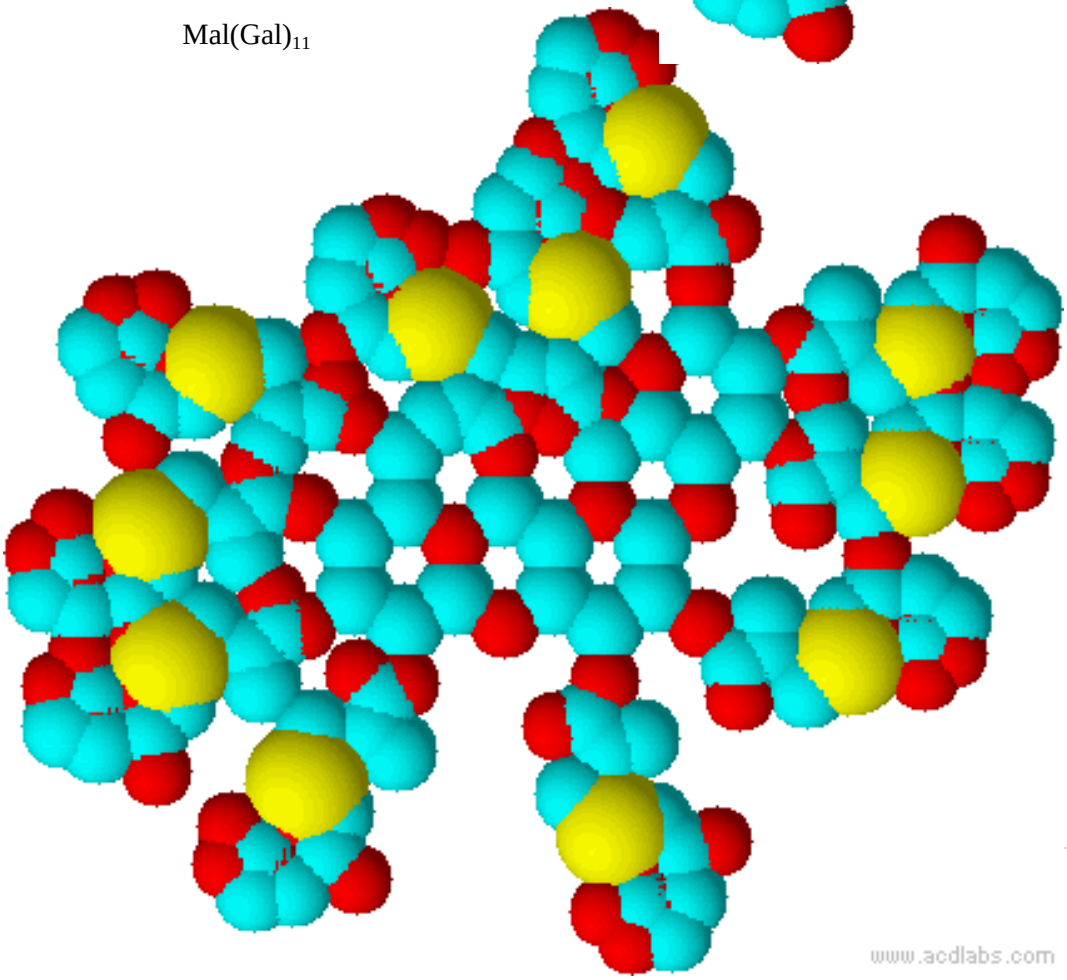
$Gal(Gal)_5$



$Cel(Gal)_8$



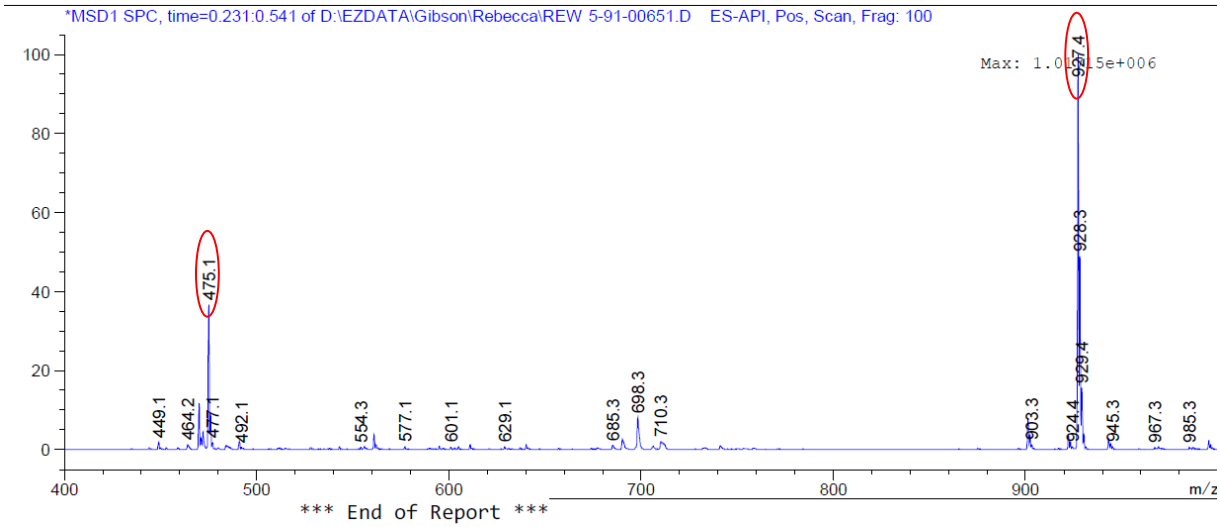
$Mal(Gal)_{11}$



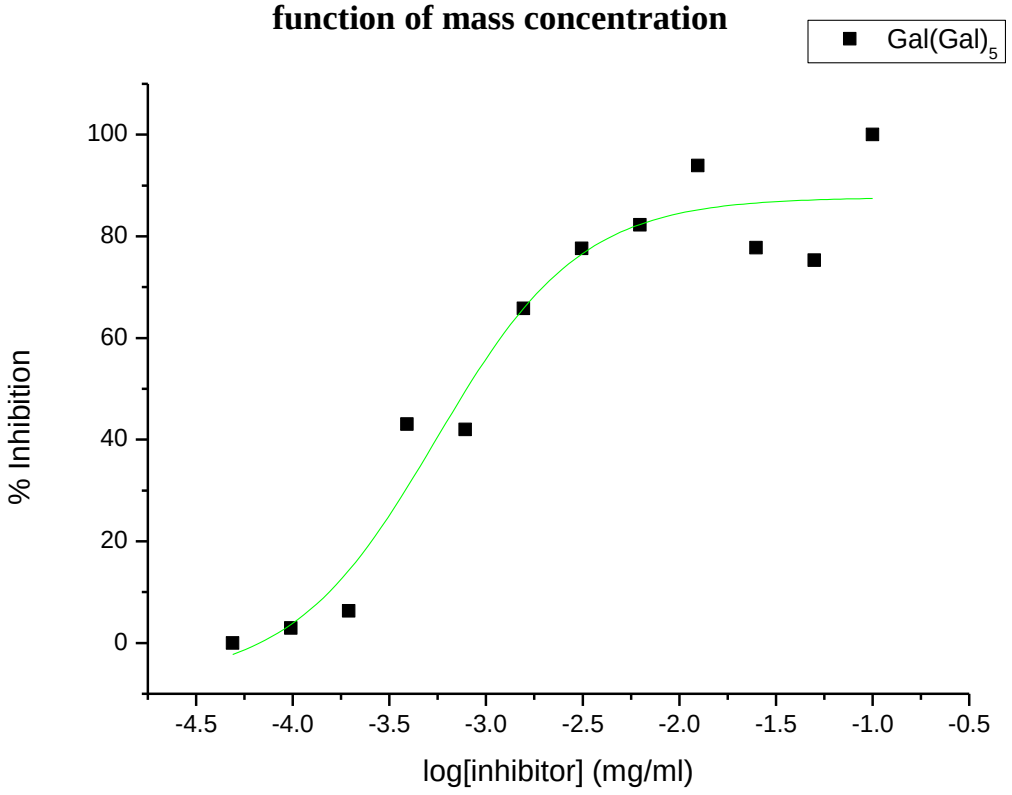
HEPES composition

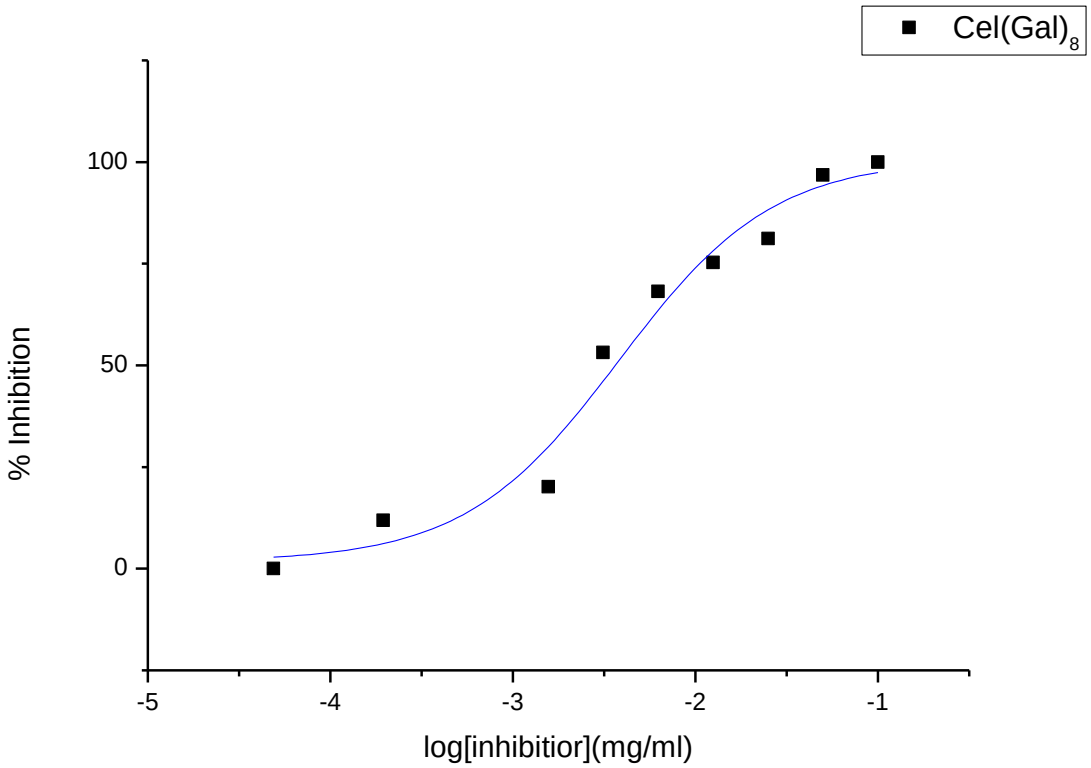
0.596g	HEPES
2.1915g	NaCl
0.003g	$CaCl_2$
250ml	Water

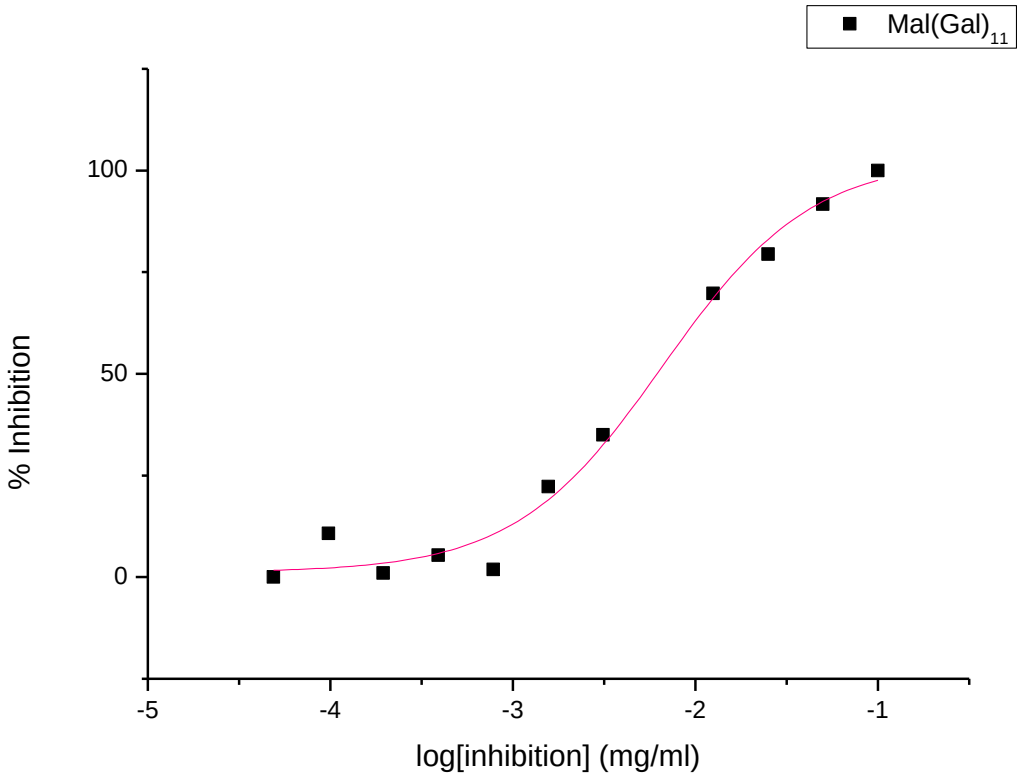
Mass spectrum of galactose methacrylate



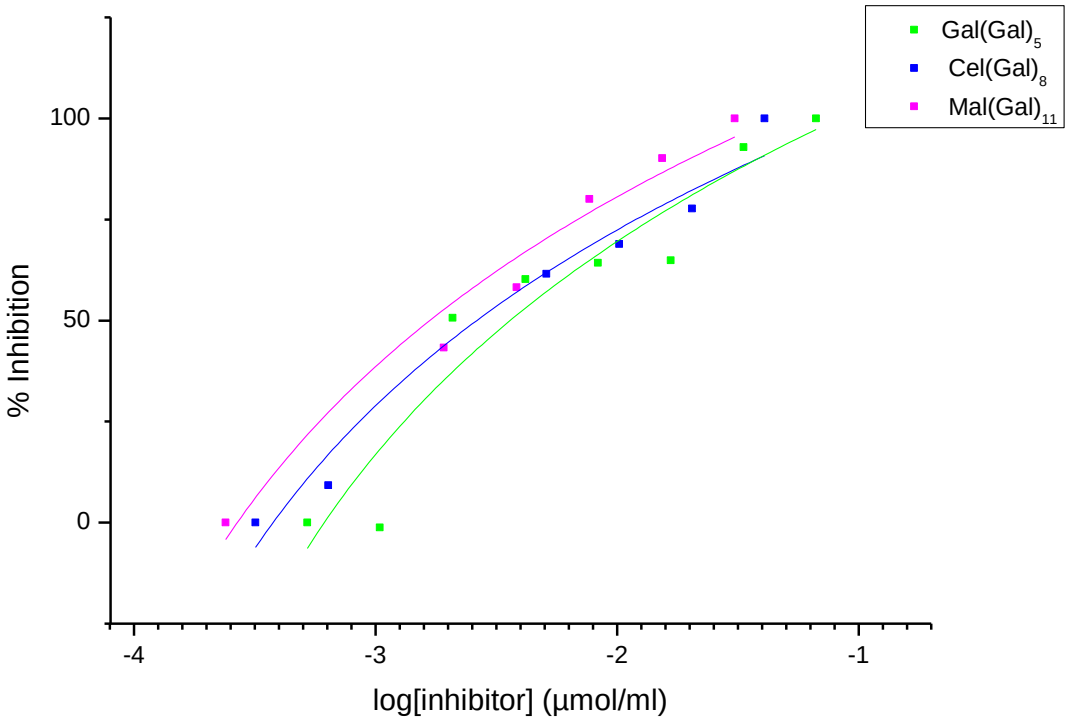
Individual inhibition curves for each glycocluster against CT-B as a function of mass concentration



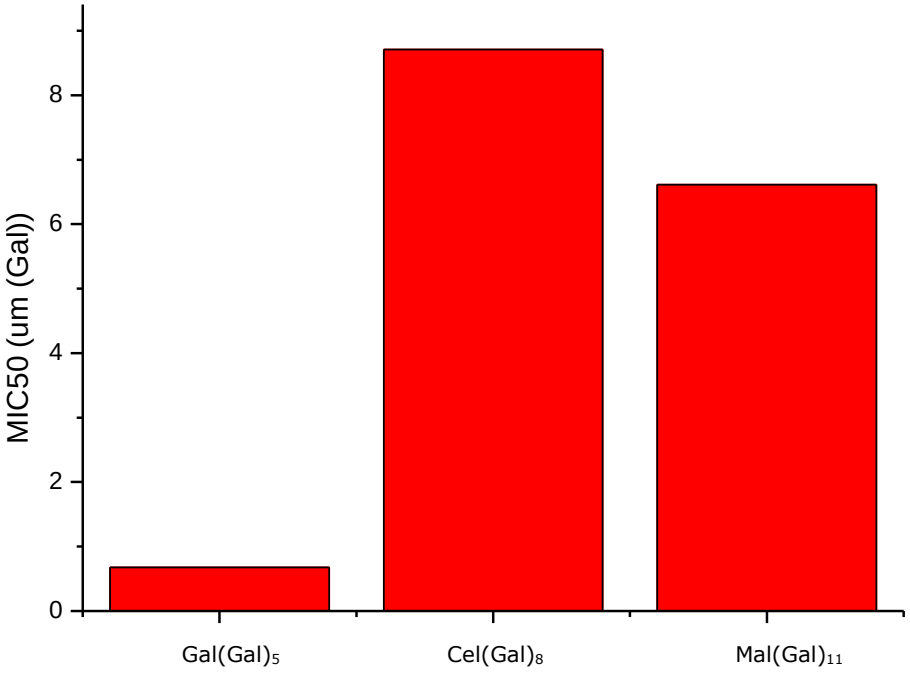


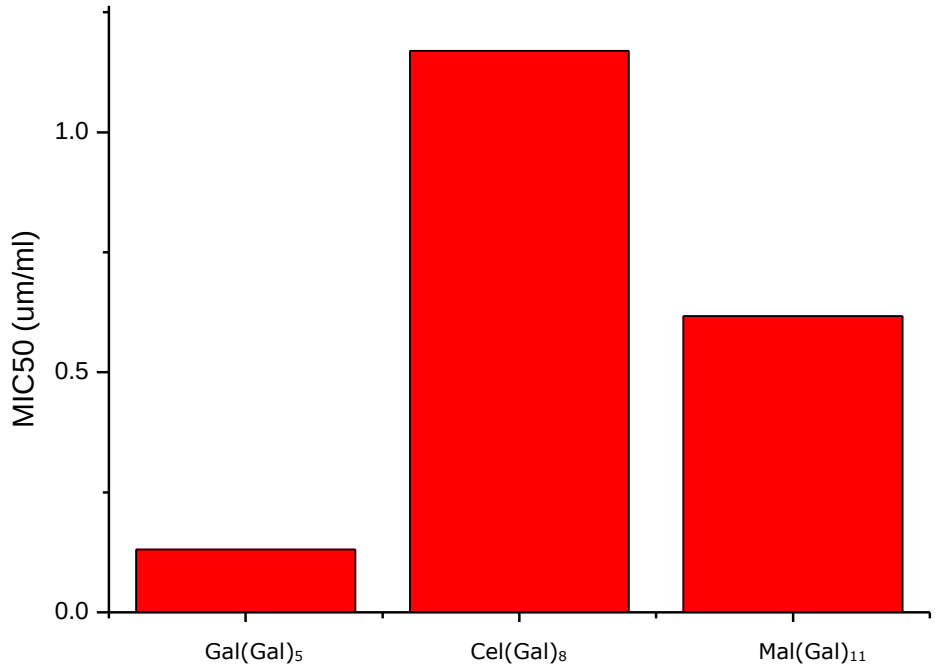


Combined inhibition curves for PNA assay using logarithm line of best fit.



MIC 50 bar charts derived from % inhibition vs. molar concentration and galactose concentration curves





Alternative views of CT-B bound to Gal(Gal)$_5$ generated using Chimera 1.7 imaging software