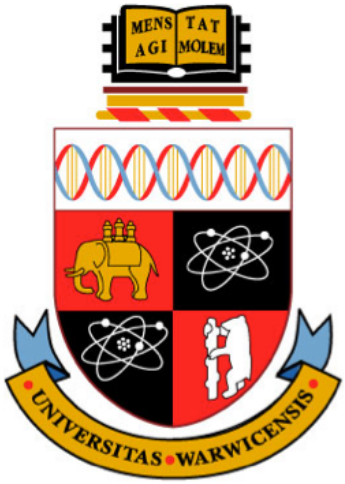


Pathogen Detection Based on Carbohydrate Adhesion

by

Lucienne Carlien Otten

Thesis

Submitted to The University of Warwick

for the degree of

Doctor of Philosophy

Systems Biology DTC

September 2015

THE UNIVERSITY OF WARWICK

Contents

List of Schemes and Figures

List of Tables

Abbreviations

3D7	*Plasmodium falciparum* 3D7
Ara	Arabinose
Biotin-NHS	(+)- Biotin *N*-hydroxysuccinimide ester
BSA	Bovine Serum Albumin
C. difficile	*Clostridium difficile*
cAMP	Cyclic adenosine monophosphate
CDC	Centre for Disease Control
Cel	Cellobiose
CFG	Consortium for Functional Glycomics
CFTR	Cystic fibrosis transmembrane conductance regulator
Con A	Concanavalin A
CTx	Cholera toxin
CTxB	Cholera toxin B subunit
DBA	*Dolichos biflorus* Agglutinin
Dex	Dextran

DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DSA	Drop shape analysis
E. coli	*Escherichia coli*
FDA	Food and Drug Administration
FITC	Fluorescein isothiocyanate
Gal	Galactose
GalNAc	*N*-Acetyl-D-Galactosamine
GI	Gastrointestinal
Glc	Glucose
GlcNAc	*N*-Acetyl-D-Glucosamine
Gly	Glyceraldehyde
GlycoAuNPs	Sugar functionalised gold nanoparticles
GM1	Monosialyltetrahexosylganglioside
GPI	Glycophosphatidylinositol
H. influenza	*Haemophilus influenza*

HEPES	2-[4-(2-hydroxyethyl)piperizan-1-yl]ethanesulfonic acid
HIV	Human immunodeficiency virus
HSB	Hue, Saturation, Brightness
IC_{50}	Half maximal binding concentration
Kd	Dissociation constant
Kd_{app}	Apparent dissociation constant
L-PHA	Plant lectin leukoagglutinin
L. casei	*Lactobacillus casei*
Lac	Lactose
M. marinum	*Mycobacterium marinum*
M. smegmatis	*Mycobacterium smegmatis*
M. tuberculosis	*Mycobacterium tuberculosis*
Man	Mannose
ManNAc	*N*-Acetyl-D-Mannosamine
MHC1	Major histocompatibility complex 1
MRI	Magnetic resonance imaging

N. americanus	*Necator americanus*
N. brasiliensis	*Nippostrongylus brasiliensis*
NHS	*N*-hydroxysuccinimide
P. aeruiginosa	*Pseudemonas aeruginosa*
P. falciparum	*Plasmodium falciparum*
P. knowlsei	*Plasmodium knowlsei*
P. malariae	*Plasmodium malariae*
P. ovale	*Plasmodium ovale*
P. putida	*Pseudemonas putida*
P. vivax	*Plasmodium vivax*
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PCs	Principle components
PfEMP1	*Plasmodium falciparum* erythrocyte membrane protein 1
Plasmodium spp.	*Plasmodium* species
PNA	Peanut agglutinin

pRBC	Trophozoite infected red blood cells
RBCs	Red blood cells
RCA	Ricin
RCA120	Non-toxic derivative of the ricin B subunit
RNA	Ribonucleic acid
S. aureus	*Staphylococcus aureus*
S. pneumoniae	*Streptococcus pneumoniae*
S. suis	*Streptococcus suis*
SBA	Soybean Agglutinin
SiNW	Silicon nanowires
Sp	Spacer
Strain 4	*Plasmodium falciparum* TH004-004
Strain 23	*Plasmodium falciparum* TH004-023
Strain 36	*Plasmodium falciparum* TH004-036
Strain 64	*Plasmodium falciparum* TH004-064

UV	Ultraviolet
V. cholerae	*Vibrio cholerae*
VBNC	Viable but not culturable
XPS	X-ray photoelectron spectroscopy

Acknowledgments

First of all I would like to thank Matt, for his endless patience, enthusiasm and optimism. For always putting a positive spin on even the worst results and for providing the positive outlook necessary to help me complete my thesis. I would also like to thank Stephen Edmondson for his assistance with the variable angle XPS, Liz Fullam and Alasdair Hubbard for their assistance with the microbiology work and Eva Caamaño-Gutiérrez, Matthew Phanchana and Ahmed Saif for their help with the Malaria work.

I would also like to thank all current and past members of the Gibson group. It has been a pleasure to work with all of you. Special thanks goes to Daniel Phillips, Caroline Biggs, Sarah-Jane Richards and Tom Congdon for all their support during my Masters project and throughout my PhD and for generally putting up with me in the office! Thanks also to Richard Lowery for all your support and friendship, for providing general entertainment in the office and for science balloons.

Without the funding and support from the BBSRC and the Systems Biology DTC this work would not have been possible. Being part of the Systems Biology DTC provided me with the skills necessary to help make this project my own and allowed me to make some amazing friends. Specifically Eva and Nikita, without whom my Masters year would not have been the enjoyable experience it was. Thank you both for your endless support, friendship and for all the days we just span on the chairs in the office.

I would also like to thank Nick Barker for his inspirational dedication to Outreach and for helping me to develop my science communication and teaching skills (and for

allowing me to make slime!). Thanks also to all those at the Brilliant Club for giving me the opportunity to teach elements of my thesis to students around the West Midlands. A special mention to the student who hand drew their presentation images, as they did not have access to PowerPoint and the person who rapped their presentation to music- you are both inspirational. Having students be enthusiastic about your research really helps after a bad day!

To all those at St John Ambulance, especially those of you that have waded through fields, sat in first aid posts, frozen at the pitch side and crewed ambulances with me, I thank you for; all the experiences you have given me, providing an endless source of procrastination (when needed), being a source of endless entertainment and for generally feigning interest in my science. You were literally 'the difference' when it came keeping me sane at times and I could not have done this without your support!

I would like to thank my friends and family for always being there to support me not just through my PhD but throughout my life. Thanks to Fiona for always being there and for providing emergency supplies of 'happy hippos' when times got tough, you truly are an inspiration to me and I will forever be thankful you chose me to live in your house all those years ago! I would also like to thank Aoife for her many motivational messages, Akshay for his general enthusiasm about all of my work and especially Chris for listening to my endless rants about science, work and life in general and for constantly reminding me why I do the things that I do.

Finally, heartfelt thanks to my mum, dad and my sisters Lysandra and Lesley for providing constant love and support without which this thesis would never have happened.

Declaration

The work presented in this thesis is entirely original and my own work, except where acknowledged below. I confirm that this thesis has not been submitted for another degree at another University.

The inhibitor design element of Chapter 2 was published as:

Jones, M.W., **Otten, L.,** Richards, S-J., Lowery, R., Phillips, D.J., Haddleton, D.M., Gibson, M.I. (2014) Glycopolymers with secondary binding motifs mimic glycan branching and display bacterial lectin selectivity in addition to affinity. *Chemical science*, **5**, 1611-1616.

Chapter 3:

All variable angle XPS was carried out by Stephen Edmondson, School of Materials Science at the University of Manchester.

Chapter 4 contains elements published as:

Otten, L., Gibson, M.I. (2015) Discrimination between lectins with similar specificities by ratiometric profiling of binding to glycosylated surfaces; a chemical 'tongue' approach. *RSC Advances*, **5**, 53911-53914.

Richards, S-J., **Otten, L.**, Gibson, M.I. (2016) Glycosylated gold nanoparticle libraries for label-free multiplexed lectin biosensing. *Journal of Materials Chemistry B*, DOI: 10.1039/C5TB01994J.

And is in preparation for:

Otten, L., Vlachou, D., Richards, S-J., Gibson, M.I. Glycan heterogeneity on gold nanoparticles increases lectin discrimination capacity in label-free multiplexed bioassays.

Chapter 5 has been published as:

Otten, L., Fullam, E., Gibson, M.I. (2016) Discrimination between bacterial species by ratiometric analysis of their carbohydrate binding profile. *Molecular Biosystems*, DOI: 10.1039/C5MB00720H.

Chapter 6 is in preparation for:

Caamaño-Gutiérrez, E., **Otten, L.**, Phanchana, M., Saif, A., Biagini, G., Gibson, M.I. The 'sweet tooth' of *Plasmodium falciparum*: a key for Artemisinin resistance classification.

Within this chapter all parasite work was completed at the Liverpool School of Tropical medicine with the assistance of Eva Caamaño-Gutiérrez, Matthew Phanchana and Ahmed Saif from the Department of Parasitology. I completed all data analysis and surface preparation.

Abstract

The rapid detection of pathogenic organisms to ensure appropriate administration of treatment remains a global healthcare challenge. This is becoming increasingly difficult, as identification of the organism alone is no longer enough, with the rise of drug resistance amongst many pathogens it is becoming increasingly important that both the pathogen and drug resistance are identified.

Currently, rapid identification can be achieved through a variety of techniques. However, many of these techniques are expensive, require extensive sample preparation, or highly trained personnel to run with results often not rapidly available. This leaves health care professionals to make point-of-care treatment decisions based on symptoms without any indication of drug resistance. The use of carbohydrate microarrays for pathogen detection has been identified as both a method for detection but also as a basis for identifying new drug targets. This exploits the initial protein-carbohydrate interaction that many pathogens utilise in the initial stages of infection. However, the use of microarrays is also challenging, as highly sensitive identification of pathogens often requires expensive or synthetically challenging oligosaccharides or coupling with a highly sensitive detection method thus limiting its point of care application.

Herein we describe the coupling of a facile surface chemistry for glycan addition with a powerful statistical algorithm to improve the sensitivity of a cheap monosaccharide functionalised surface without using expensive detection methodologies. This technique was then applied to the detection and identification of toxic lectins, bacterial samples and finally the life-stage specific detection of *Plasmodium falciparum* (one of the parasites responsible for human malaria). In this last case, drug resistance related to carbohydrate binding profile was also observed.

Chapter 1

1.2 Introduction

Carbohydrates in the glycocalyx, which surrounds almost all cells, mediate a multitude of biological processes through protein-carbohydrate interactions. The glycocalyx is also important for many diseases and can reflect disease states, such as cancer, through aberrant glycosylation patterns. Many pathogens also utilise proteins to interact with carbohydrates as an important step in pathogenicity.

Protein-carbohydrate interactions are exploited by *Vibrio cholera*e in order to mediate toxin binding thus causing cholera infection, which results in the death of around 100,000 people a year.[1] Malaria is estimated to kill one child in Africa every 8 minutes. It is caused by infection with parasites of the *Plasmodium spp.* and initial binding and internalisation of *Plasmodium falciparum* is mediated by a lectin-like protein. Protein-carbohydrate interactions are also crucial for infection with Influenza A and for the direct binding of pathogenic strains of *Escherichia coli* to give but a few examples.[2]

Many of the current detection methodologies for these pathogens are complex or too expensive for robust point-of-care diagnostics and as such, increasing the resolution of current cheap and facile methods without the use of expensive detection systems is an important area of research for point-of-care pathogen detection.

1.2.1 Carbohydrates

Carbohydrates were once described as a molecule "in search of a function"[3] and thought to be less information rich than proteins.[4] The importance of carbohydrate signalling has also been underestimated previously.[5] However, in the last 20 years their role in complex biological processes; such as cell-cell communication, immunity, cell differentiation and fertilisation, has come to light and it is now known that carbohydrates play a crucial role in many processes.[4] This has contributed to the delay in generating techniques for their study when compared to DNA and protein techniques and many challenges remain in the study of carbohydrates.

Unlike proteins and DNA, carbohydrate synthesis is not trivial, it is possible to remove carbohydrate chains from cells but most procedures require multiple chromatographic or purification steps. In nature, sugar chain synthesis requires multiple enzymes and highly complex synthesis pathways and whilst there has been some success in the production of synthetic oligosaccharides the process is often laborious involving many protection and deprotection steps.[6-10] Generation of the carbohydrate structures found in bacteria is further complicated by the presence of different glycosylation enzymes and their ability to utilise a larger library of monosaccharides than mammalian cells resulting in the production of even more complex structures.[5]

1.2.2 Glycan complexity

Carbohydrates are now regarded as being unsurpassed in their structural diversity, which makes them challenging to study.[5] Only ten monosaccharides are utilised in

mammalian cells but that is enough to potentially generate ten million different tetrasaccharides.[3]

The monosaccharides are themselves able to exist as different enantiomers known as D- and L- depending on the spatial configuration around the chiral carbon (Figure 1.1A) and as different anomers depending on the configuration on the anomeric carbon (Figure 1.1B). The complexity of glycans goes beyond just the number of available monosaccharide building blocks, with chain lengths ranging from single monosaccharide units through to polysaccharides containing 500,000 sugar units.

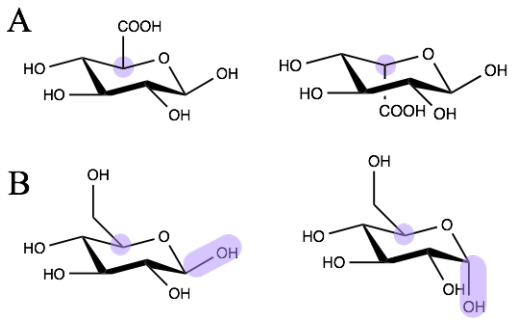


Figure 1.1 D- and L- optical isomers (left and right respectively) are determined by the arrangement around the chiral carbon (A). Monosaccharides can also exist as different anomers such as α and β glucose (left and right respectively). This is defined as the direction of the hydroxyl group on the anomeric carbon relative to the configuration around the chiral carbon (B).

Glycosidic linkages between monomers allow the formation of large glycan chains and these linkages can also show differences in both stereo- and regiochemistry adding another layer of complexity to glycan chains. Differences in stereochemistry occur whereby either an α or a β linkage can occur. An example of this difference is

shown by maltose and cellobiose, both of which are disaccharides of D-glucose (Figure 1.2).

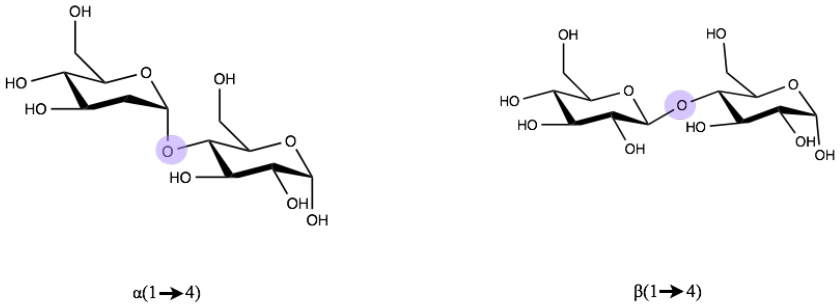


Figure 1.2 The glycosidic linkage between monosaccharides can show differences in stereochemistry with either an α linkage, such as that maltose, or a β linkage as in cellobiose (left and right respectively).

Differences in regiochemistry can also occur as the glycosidic bond can form between any of the hydroxyl groups as demonstrated by maltose and trehalose (Figure 1.3). These chains can also be highly branched utilising both α and β linkages. Often 3D structure of the carbohydrate, density, linker to the cell and topological arrangement all have to be maintained for recognition to occur.

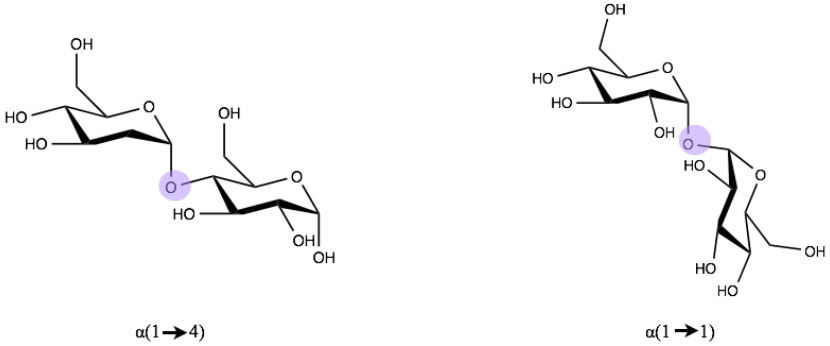


Figure 1.3 The glycosidic linkage can also show differences in regiochemistry as the linkage can form between any of the hydroxyl groups in the monosaccharide (examples shown are maltose and α, α-trehalose, left and right respectively).

After assembly by glycosyltransferases the glycan complexity is further increased by site-specific modifications of the monosaccharide units within a glycan chain. This modification can include sulfation, acylation or epimerisation of glucuronic acid to iduronic acid (Figure 1.4).

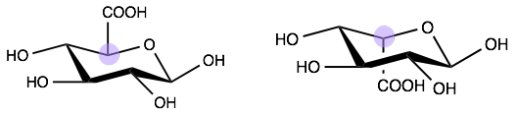


Figure 1.4 Epimerisation of glucuronic acid (left) to iduronic acid (right) is a common mammalian monosaccharide modification.

1.2.3 Mammalian Glycocalyx

Carbohydrates are covalently linked to membrane proteins and lipids to form a dense glycocalyx (Figure 1.5). Cell surface glycans are positioned optimally to mediate cell-cell interactions and protect the cell from the extracellular environment.[11, 12] They also act as recognition elements and binding sites for a multitude of other components and are crucial for preventing non-specific protein-protein interactions.[13, 14]

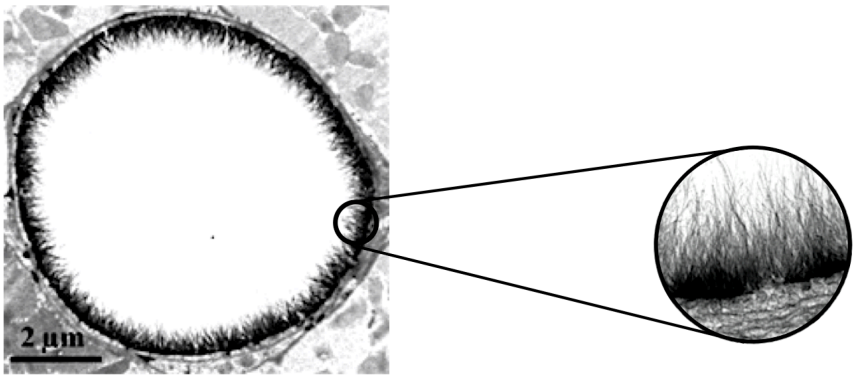


Figure 1.5 Electron microscopy image of a cross section of a rat myocardial capillary showing the glycocalyx of the endothelial cells (reprinted with permission from reference [15]).

The diversity of the glycans present in the glycocalyx is determined by the mammalian glycome. The glycome is comprised of ten key mammalian monosaccharides (Figure 1.6), which are added to other molecules in a non-template driven process mediated by the expression of several different glycosyltransferases (several of which have tissue specific isoforms).[16-18]

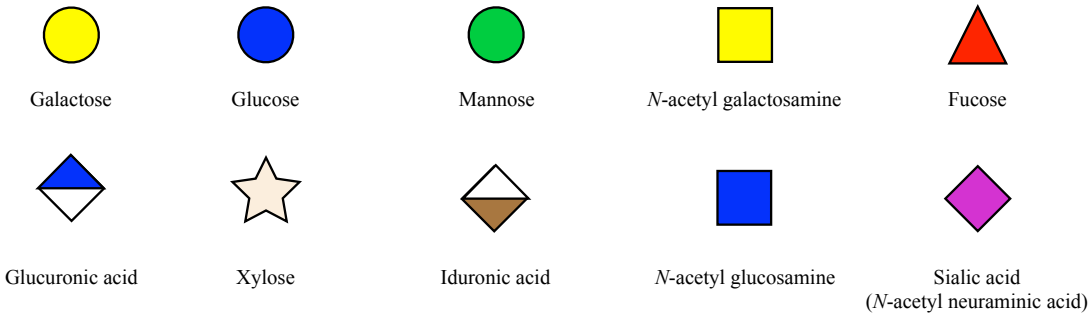


Figure 1.6 The ten mammalian monosaccharides displayed using standard consortium for functional glycomics (CFG) nomenclature.[19]

Whilst the theoretical glycospace (which is the number of all the possible combinations of monosaccharides) is vast only a small subspace is actually used in mammals[20] and it is thought that there are only around 2000 different mammalian glycans with many highly conserved core structures.[19] Prokaryotes utilise many more than ten monosaccharides and use a much larger subspace of their theoretical glycome as their outer membrane has to protect the cell from many harsh and changing environments.[20]

1.2.4 Glycosylation in disease susceptibility

As the glycocalyx is crucial in the interactions of a cell with its environment it plays a crucial role in disease susceptibility and pathogen interaction. For example, the ABO blood system is determined by differences in antigenic oligosaccharides on the surface of red blood cells (RBCs, Figure 1.7). Serological subgroups are further defined by the number of oligosaccharides per RBC highlighting the importance of both the specific antigen and the distribution of those antigens.[21, 22] Whilst the biological role of the system remains a mystery, anthropological studies on the population and geographic distributions of blood groups indicate that they may be related to group susceptibility to certain disease.[21] It is known that blood group determines susceptibility to or severity of diseases such as small pox, cholera and malaria.[23-25] The susceptibility of people with different blood groups to malaria is thought to be due to the invasion of the pathogen into uninfected RBCs.[25] It has been hypothesised that the individual variation of glycans on the surface of human tissues is a result of a selection pressure caused by past endemic or epidemic pathogens.[21]

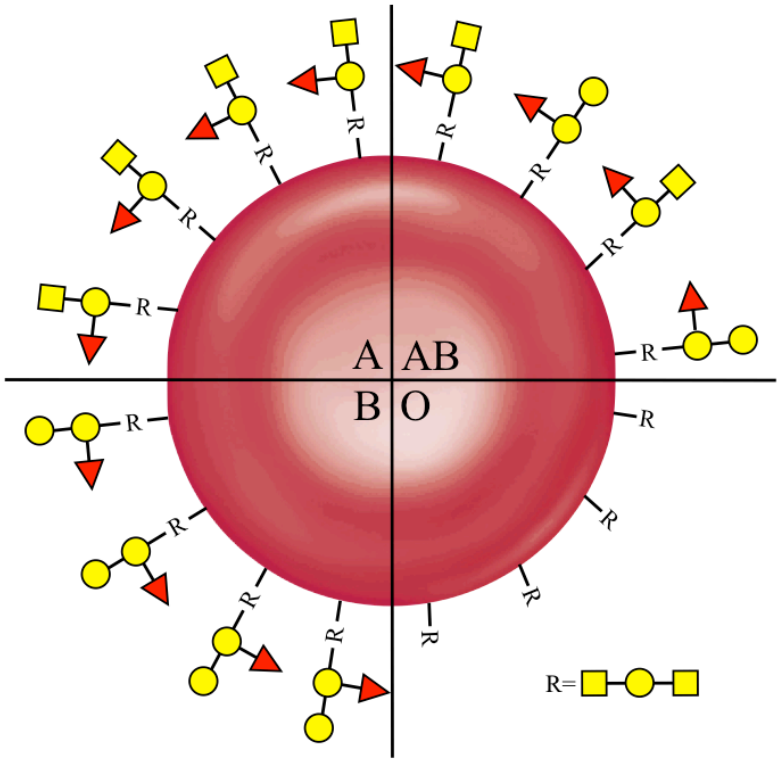


Figure 1.7 Serological blood groups are determined by oligosaccharides on the surface of red blood cells. Oligosaccharides shown are depicted using the standard CFG nomenclature.[19]

1.2.5 Altered glycosylation in disease states

Differences in glycosylation have been detected in diseased tissue and many different cancer cells. It is thought that changes in the expression of glycosyltransferases and glycosidases result in the changes in the glycans found on the surfaces of tumour cells, it has also been suggested that these differences affect interactions between lectins at the tumour surface which in turn determines a tumour's metastatic potential.[4, 26]

Cancer cells show a number of differential glycosylation patterns but these changes tend to be a specific set of permeations that result in malignant transformation and tumour progression. These changes have been selected for as a form of 'microevolution'. Most notable are the differences in glycosylation of mucins.[27] Mucins are large glycoproteins found on the surfaces of endothelial cells. They are also secreted by healthy cells into the lumens of organs in order to limit adhesion of pathogens (Figure 1.7B).[21] However, in tumours they have been found to be over-expressed[28] and a number of changes occur; firstly incomplete glycosylation of the mucins occurs resulting in truncated glycan patterns[29] or increased sialic acid moieties.[30] These truncated mucins can generate an immune response. The secreted mucins are also truncated and in some instances there is a total loss of glycosylation and the protein backbones are just secreted (Figure 1.8A).

The altered topology of the tumour itself means that the mucins are added to all surfaces of the cells whereas normally they are only expressed on the surface facing the lumen. It also results in mucins being secreted into places other than the lumen and can be secreted into the blood stream (Figure 1.8A). As in healthy cells, the mucins can still act in an anti-adhesion capacity but in cancerous tissue they can prevent the adhesion of cancer cells to neighbouring cancer cells resulting in cell migration and subsequently tumour metastases (Figure 1.8A).[30]

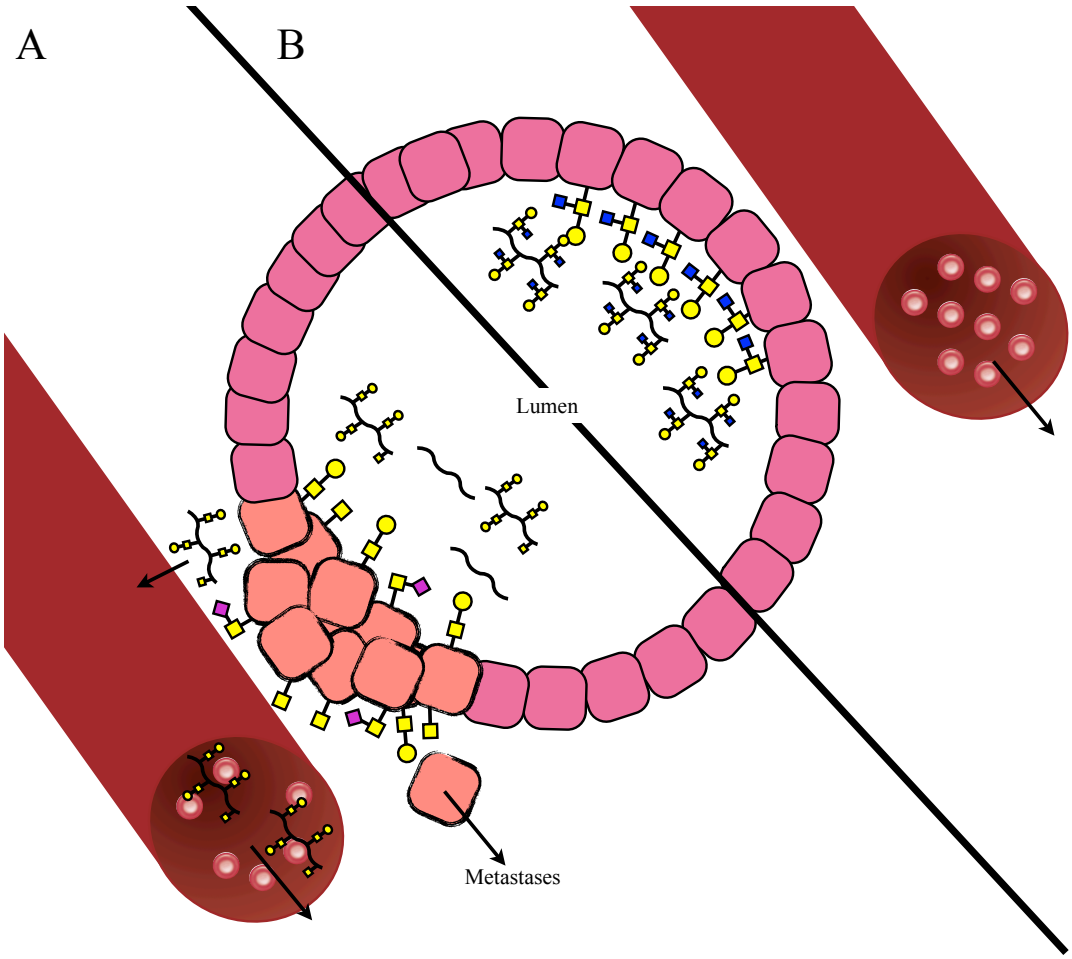


Figure 1.8 Glycosylation and secretion of mucins is altered in tumours (A) compared to normal tissue (B).

Other cancer-associated glycosylation changes include the increased β1-6 branching of N-linked glycans,[31] increased addition of sialic acid to surface glycans and polysialylation,[32] increased presentation of ligands specific to selectins[33] and increased presentation of sialylated blood-group associated glycans on cell surfaces (e.g. Lewis^x glycans).[34] Incomplete synthesis and altered expression of blood-group associated glycans coupled with the loss of normal AB blood group glycan expression is associated with poorer clinical outcome.[35]

Other diseases are also characterised by glycosylation differences. Mutations in certain glycosyltransferases will result in differences in glycosylation associated with many congenital diseases including galactosemia, muscle-eye brain disease and congenital muscular dystrophy.[18] Glycoforms (isoforms of a protein that differ only in glycosylation) of immunoglobulin-γ antibodies have been associated with rheumatoid arthritis.[36] Overexpression of mucins has also been associated with many chronic lung conditions including asthma, chronic-obstructive pulmonary disease, bronchitis and cystic fibrosis.[37]

1.2.6 Protein-Carbohydrate Interactions

Protein-carbohydrate interactions mediate many crucial biological processes including (but not limited to); cell-cell communication, fertilisation, pathogen recognition and response, protein folding and both passive and innate immunity.[6, 26] All known organisms are covered by either free or bound glycan structures and proteins known as lectins bind to these carbohydrates during signal transduction.[21] These interactions are typically weak with a dissociation constant (K_d) typically between 10^{-3} and 10^{-6} M. This is overcome in nature by presentation of multiple copies of the carbohydrate structures on the cells. The increase in affinity in binding is greater than the sum of the individual binding events that take place and this is known as the cluster glycoside effect.[38]

This multivalent interaction means that many lectins have multiple binding subunits, which can interact with the surface in one of two manners; either the bind-and-slide[39] method or face-to-face binding[40] (Figure 1.9). In the bind-and-slide model, a multivalent lectin will bind at one point and then once this complex dissociates the

lectin will slide along until it binds to another region and in the face-to-face model the subunits of the multivalent lectin face in opposite directions to each other and thus can engage in crosslinking.[41] Graded affinity binding can also increase the binding of a lectin to a surface where one subunit interacts with the native version of the ligand and other subunits interact more weakly with truncated versions of the ligand in order to increase the avidity of the binding event.[42]

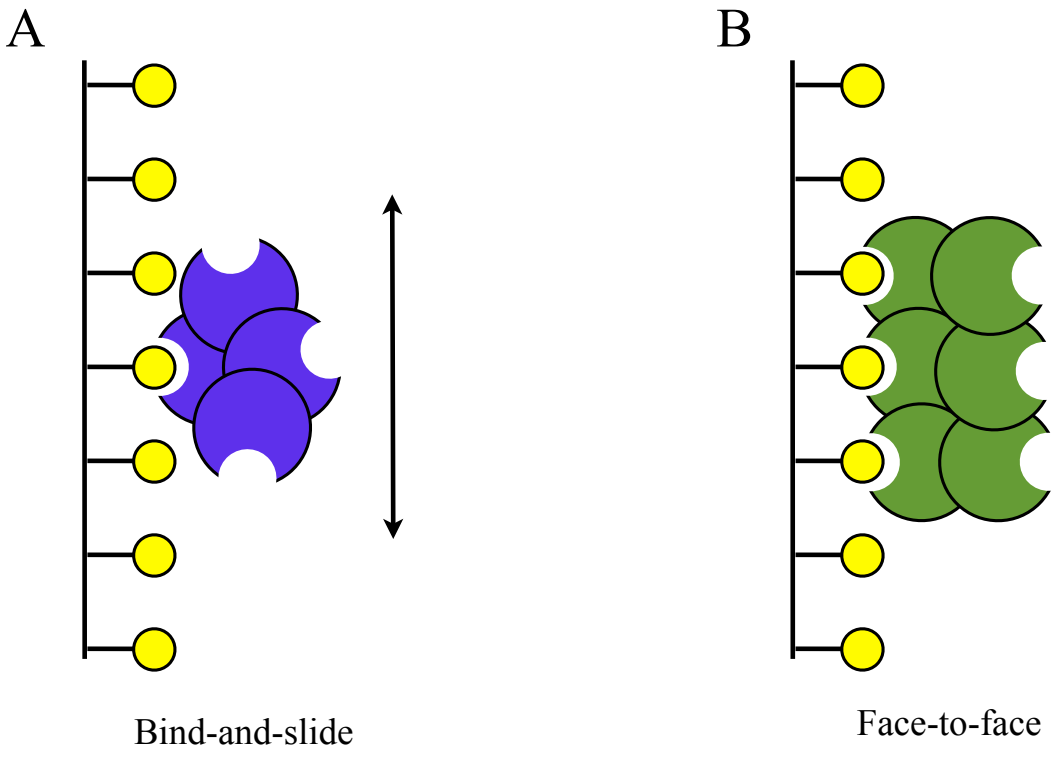


Figure 1.9 Mechanisms for multivalent binding can be achieved by either binding of a multivalent protein which then dissociates and binds to the next ligand effectively sliding along the chain of saccharides with each binding event essentially increasing the residence time of the protein within the area of the ligand (A) or through face to face binding of multiple dimers binding to multiple ligands on a surface (B).

1.2.7 Pathogen exploitation

The carbohydrates on a cell surface determine everything from blood type to disease susceptibility and they are often exploited by pathogens (Figure 1.10).[21] This exploitation falls in to three broad categories; mimicry of the host glycome to avoid detection and eradication by the host immune system, adhesion of the pathogen as a crucial step in infection or for internalisation, and production of lectins similar to those produced by the host's immune system in order to produce an immune-compromised niche for the pathogen.

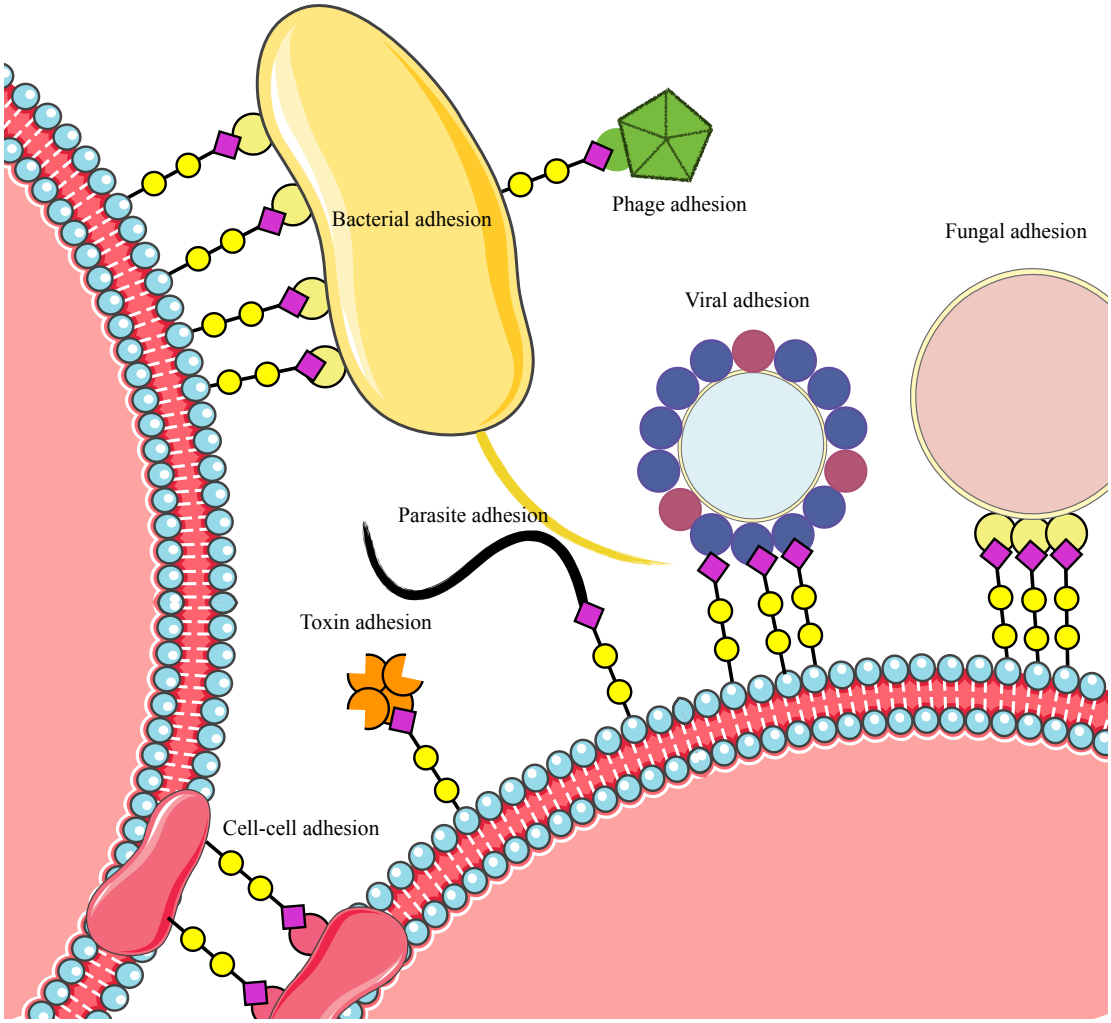


Figure 1.10 Protein-carbohydrate interactions are crucial for many biological processes including pathogen-associated adhesion.

1.2.7.1 Host glycome mimicry

Pathogenic bacteria often find themselves in a "dual glycan speedway" where they evolve to avoid phage recognition but also detection by the host.[43, 44] One mechanism by which they avoid host recognition is through mimicry of the host glycome and many species have evolved a variety of mechanisms for doing this. In general, many human pathogens have a glycome that is very similar to the mammalian glycome.[20] This avoids triggering the immune system by introducing antigenic glycans such as the non-mammalian monosaccharide D-galactofuranose, which is found in *Mycobacterium tuberculosis* in arabinogalactan chains and has been shown to be highly antigenic.[45]

Many mammalian glycans are either fucosylated or contain a sialic acid cap. This cap forms a crucial part of protecting host cells from the immune system. They mediate the binding of Factor H (which protects the host cells from attack by the complement immune system)[46] and also engage siglecs, which are sialic acid binding immunoglobulin-like lectins. Siglecs are known to have an inhibitory affect on immune cells thus protecting the cell they are bound to from the immune response.[47] As sialic acid mediates both of these processes it would be extremely beneficial for pathogenic organisms to adopt the sialic acid cap into their glycans to offer protection from the host immune system. Indeed, strains of *Haemophilus influenza* that have sialic acid incorporated in their glycome were shown to be more resistant to complement-mediated killing by human serum. Mutation of the gene controlling sialylation resulted in loss of sialic acid residues and a marked reduction in the tolerance to human serum.[48] A similar affect has been shown in pathogenic *Neisseria spp.* where sialylation of the lipopolysaccharide renders the bacteria resistant to

killing by human serum.[49] Whilst sialic acid is used in both of these examples not all bacteria have evolved to use it but many of those that have not will utilise bacterial mimics of sialic acid.[50-52]

L-fucose is another important terminal monosaccharide in mammalian glycans but this is only found in low abundance amongst bacterial species.[20] The exception to this is one class of bacterial species, which contain L-fucose in high abundance; and includes the pathogenic bacteria *Helicobacter pylori,* which contains large proportions of fucosylated Lewis A glycans in order to mimic the host glycome and avoid an immune response.[53]

Not all bacterial classes mimic the mammalian glycome, for example actinobacteria contain a large number of actinobacter specific monosaccharide residues.[20] This class contains a number of specialised soil bacteria crucial for nitrogen fixation or other elements of the soil cycle but also includes a number of human pathogens from the class of *Mycobacterium* including *M. tuberculosis*.[54] As already mentioned, this species utilises D-galactofuranose to form antigenic arabinogalactan. *M. tuberculosis* also contains very few sialic acid mimics and thus its glycome is a poor mimic of the host glycome but, in this instance, mimicry is less important for pathogen survival as *M. tuberculosis* is an intracellular pathogen.[20]

Many pathogenic bacterial species mimic the host glycome, at least in part, to avoid detection by the immune system or to aid in the binding of immunosuppressant lectins present in the host. Those that do not mimic the host or poorly mimic the host are

often intracellular based pathogens and thus do not experience the same selection pressures to evolve a glycome more similar to that of the host.

1.2.7.2 Toxin adhesion

As all cells are glycosylated, many pathogens exploit this glycosylation to mediate binding and internalisation of toxins. Cholera is caused by cell internalisation of a toxin produced by the bacteria *Vibrio cholerae*.[55, 56] The toxin is an AB_5 toxin, and is made up of one subunit that causes cell damage by constitutively activating a protein resulting in the excretion of ions from the cell into the lumen of the small intestine (Figure 1.11A) and five lectin subunits (Figure 1.11B) that bind the monosialyltetrahexosylganglioside (GM1) found on endothelial cells in the small intestine.[57]

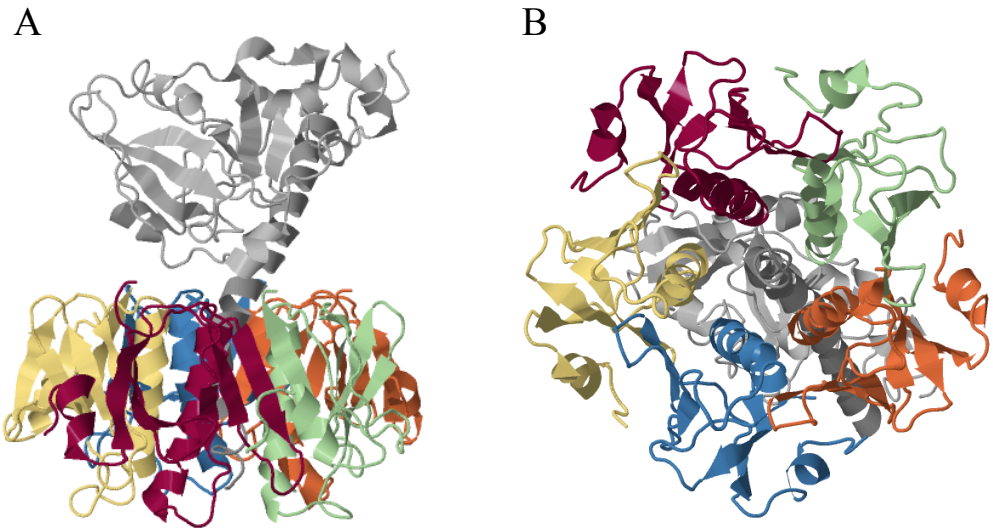


Figure 1.11 Crystal structures of cholera toxin which contains one enzymatic A subunit shown in grey (A) and five glycan binding B subunits (B).[58]

Ricin is a highly toxic lectin extracted from the seeds of the *Ricinus communis* plant. This toxin is an AB toxin and thus contains one glycan binding targeting domain (the B subunit) and an enzymatic A subunit. The targeting domain mediates the binding to galactose-rich glycans on the cell surface and upon binding the toxin is internalised by the cell. Once inside the cell the A subunit and the B subunit are cleaved at the disulphide bond (shown in yellow Figure 1.12) and the A subunit cleaves an adenine residue from the 28S ribosome effectively halting protein synthesis and triggering cell death.[59] Whilst somewhere between 10^6 and 10^8 molecules of ricin will bind to the cell surface, few will make it into the cytosol but entry of just a single A subunit of ricin is sufficient to deactivate 1500 ribosome molecules per second.[60]

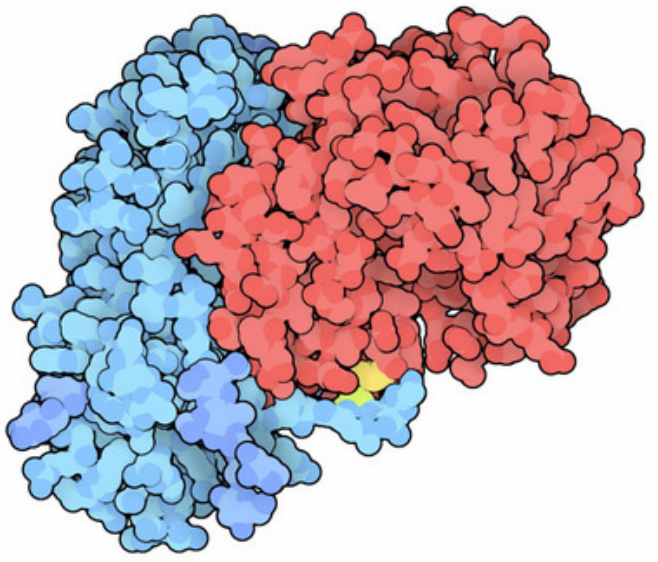


Figure 1.12 Crystal structure of Ricin, which is comprised of one A subunit (red) and a carbohydrate binding B subunit (blue). A disulphide bond connects the A subunit and the B subunit (yellow).[61]

Lectins are also found in many vegetables and consumption of improperly cooked vegetables can result in acute gastrointestinal (GI) tract distress.[62] Any plant lectins

that are not degraded during digestion can bind to the endothelial cells in the intestines and are often toxic.[63] Whilst in humans the affects tend to be nausea, vomiting and diarrhoea, long term consumption (in animal models) has been associated with endothelial cell necrosis and increased cell turn over.[64] Upon incubation of healthy cultured endothelial cells with plant based lectins, mitogenesis[65] and inhibition of exocytosis is observed.[66] Many of the endothelial cells in the GI tract experience mechanical stress resulting in damage to the plasma membrane.[67] Healthy cells can rapidly repair this damage.[68] However, lectin binding to these damaged cells inhibits this healing mechanism and results in cell necrosis and it has been proposed that this is responsible for plant-based food poisoning.[69]

1.2.7.3 Virus attachment

Many viruses exploit protein-carbohydrate interactions for the initial adhesion phase but by far the most studied example is that of hemagglutinin found in influenza A and B. This lectin is co-localised as trimmers onto the viral surface and mediates binding to the underlying surfaces through sialic acid terminated glycans. After binding, the viral envelope fuses to the cell membrane, the virus is internalised and replication occurs.

Viral specificity is determined by the linkage of the sialic acid to the terminal glycan and this differs between species; avian species have α(2-3) linked residues (Figure 1.13A), humans have predominantly α(2-6) linkages (Figure 1.13B) and porcine species have both α(2-3) and α(2-6) (Figure 1.13C). As such avian flu binds strongest to α(2-3) linked sialic acid residues and human strains of influenza A binds predominantly to α(2-6) linked sialic acid residues. In order for an avian influenza

virus to infect a human it must first infect a porcine species at the same time as either a porcine or human strain in order for the avian species to recombine with either the human or porcine viral DNA that allows binding to α(2-6) linked sialic acids.

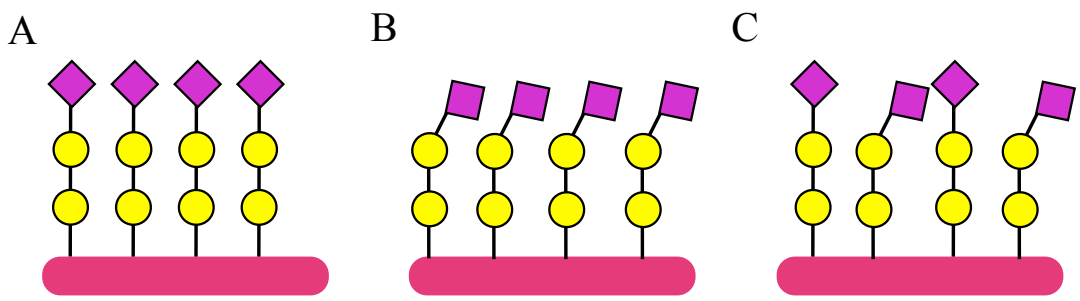


Figure 1.13 Avian tissue glycans are terminated by α(2-3) linked sialic acids (A), human tissue glycans are predominantly terminated by α(2-6) linked sialic acids (B) and porcine tissue glycans are terminated by a mixture of α(2-3) and α(2-6) linked sialic acids (C).

Interestingly, the carbohydrate binding properties of viruses are not limited to mammalian viruses but are also found amongst bacteriophage (which are viruses that target bacteria). The bacteriophage PL-1, which infects *Lactobacillus casei* was found to interact specifically with L-rhamonosyl residues in the lipopolysaccharide chains which cover the surface of this bacteria. Addition of a lectin capable of binding L-rhamanose resulted in a dose-dependent inhibition in phage absorption by this bacterial species.[70]

1.2.7.4 Bacterial adhesion and biofilm formation

Lectins play a crucial role in bacterial adhesion and colonisation both by pathogenic bacteria to host cells and subsequent biofilm formation, but also in symbiotic relationships where bacteria bind to plants and confer nitrogen-fixing properties.

During bacterial infection, a single bacterial cell will bind to the underlying host cell surface. The initial adhesion phase is often a protein-carbohydrate interaction either through bacterial adhesins binding to host glycans or *vice versa*.

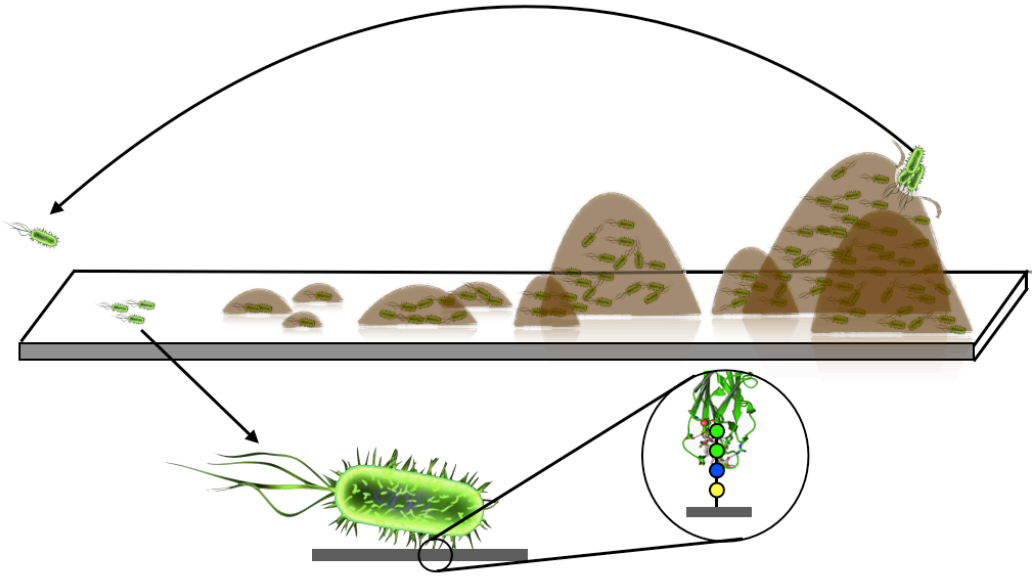


Figure 1.14 Schematic depicting bacterial biofilm formation. The example shown is the bacterial adhesin FimH interacting with a mannose rich tetrasaccharide.

For example the adhesin FimH is a crucial virulence factor found in uropathogenic *Escherichia coli* strains.[71, 72] FimH is found on the end of fimbrae and mediates the binding to mannose rich glycans on the cell surface (Figure 1.14). A combination of the adhesins present on the bacterial surface will define the range of susceptible tissues, which will differ between species and strains within the same species. For example, *E. coli* strains containing type 1 pili are responsible for initial colonisation and infection of the bladder whereas those that possess P-pili result in infection of the kidney through binding to glycans.[2] Loss or inhibition of these adhesins can prevent infection and strains that lack adhesins are non-pathogenic.

In the opposite case, binding of the symbiotic *Rhizobium spp.* to the root hairs of plants is mediated by polysaccharides on the bacterial cell (either the exopolysaccharide, the lipopolysaccharide or the capsular polysaccharide depending on the species). For example, on the roots of the Dutch clover plant (*Trifolium repens*) the protein trifoliin is expressed.[73] Trifoliin is a lectin that binds the exopolysaccharide of *Rhizobium trifolii*.[74] Colonisation of the plant's roots by the bacteria is the formation of a symbiotic relationship that gives the plant nitrogen-fixing properties.

1.2.7.5 Parasite exploitation

Many parasites are known to exploit protein-carbohydrate interactions during infection. The production of parasite specific lectins are thought to play a role in down-regulating the immune response upon parasite infection to create an immune-compromised niche for the parasite to live in. A lectin-like protein is also thought to play a key role in *Plasmodium falciparum* (one of the species that results in human malaria) infiltration of red blood cells.

Generation of an immune-compromised niche involve interactions with host c-type lectins. C-type lectins are calcium dependent lectins that play a crucial role in the mammalian immune system; encompassing molecules such as collectins (crucial for innate bacterial and viral recognition),[75] selectins (crucial for mediating adhesion of leukocytes at the site of inflammation)[33] and Natural Killer cell receptors (which bind to major histocompatability complex class 1 (MHC 1) molecules on cells to prevent

lysis and provide triggers for cell death upon detection of altered MHC 1 molecules)[76] to name but a few. Many parasites are known to produce c-type lectins upon infection or have had putative c-type lectin domains identified in their genome. *Necator americanus*[77] and *Nippostrongylus brasiliensis*[78] are both gastrointestinal worms with *N. americanus* a human pathogen and *N. brasiliensis* a rodent-specific parasite. Both of these parasites have been shown to produce c-type lectins, which show binding properties similar to mammalian selectins. It is thought that these play a role in suppressing the immune system during parasite infection.[78] The parasitic nematode *Toxocara canis* also secretes a glycoprotein, which has been shown to bind to mammalian cell surface glycans in a Ca^{2+} dependent manner and is thought to play a role in immune suppression.[79, 80]

Lectins have also been shown to have a more direct role in parasite infection. During infection with *Plasmodium falciparum* it has been shown that a lectin-like protein is involved in binding of the parasite to the surface of the red blood cells prior to internalisation of the parasite. This process was prevented upon the addition of high concentrations of *N*-Acetyl-D-Glucosamine to the media.[81] Many of the parasite-derived proteins that are expressed on the surface of an infected red blood cell (during a late stage in the infection process) have also been shown to contain putative carbohydrate binding domains.[82, 83]

1.2.7.6 Exploitation by Fungi

Lectins play an important role in the adhesion of pathogenic fungi and has been shown to occur in many species including *Chryosporium keratinophilum,*[84] *Macrophomina phaseolina,*[85] and *Anixiopsis stercoraria*.[86] Of particular interest is

Aspergillus fumigatus, which is the most prevalent airborne fungal pathogen of the developed world and causes life-threatening infections in immune-compromised people.[87] Infection involves the inhalation of conidia followed by adherence to the lung surface and then germination. This adhesion is known to be sialic acid dependent but many other lectins are also present on the surface of the fungi and play an important role in adhesion including a fucose binding lectin.[88]

1.2.8 Disease prevention

The prevalence of protein-carbohydrate interactions in pathogen mechanisms makes them an important target for disease prevention in the form of carbohydrate vaccine targets, new bacterial species-specific antibiotic targets and anti-adhesion therapy. These will now be addressed.

1.2.8.1 Carbohydrate vaccine targets

Despite many pathogens mimicking the host glycome, many still have pathogen specific oligosaccharide presentations, which could provide targets suitable for carbohydrate-based vaccines. For example, the antigenic response to arabinogalatan from *M. tuberculosis* is a potential vaccine candidate as the component galactofuranose is not present in mammalian cells; this minimising the chance of generating an autoimmune response.[20] Monoclonal antibodies have been generated that are directed at the cluster presented mannose-rich oligosaccharides present on human immunodeficiency virus (HIV) glycoprotein 120 and new broadly neutralizing carbohydrate based antibodies have also been extracted from patients.[89, 90] A carbohydrate specific immune response to the *Streptococcus pneumonia* bacteria has also been generated in mice in response to gold nanoparticles coated with a T-helper

peptide and the tetrasaccharide repeating unit of the *S. pneumoniae* capsular polysaccharide.[91]

Protein conjugated carbohydrate vaccines have also been developed for *H. influenza* type b, where meningitis caused by this bacteria has been eradicated in areas with a vaccine program.[92] A species specific tetrasaccharide has also been identified for *Bacillus anthracis*[93] and administration of this allowed the generation of antibodies specific to this bacteria although more work needs to be done to develop this into a usable vaccine against anthrax.[94] Proteins conjugated to the *P. falciparum* specific glycophosphatidylinositol (GPI) hexasaccharide have also been developed as a potential conjugate vaccine. This vaccine conferred reduced mortality to malaria in mice without any cross-reactivity with human GPI although more work to develop this is still on going.[95]

Carbohydrate vaccines are also being pursued as an anti-cancer treatment by utilising antibodies produced by the aberrant glycosylation of mucins on tumour cells. This involved the coupling of the sialylated Tn antigen (where the Tn Antigen is just *N*-Acetyl-D-Galactosamine coupled to a serine or a threonine residue) that is produced on aberrant mucins. This vaccine was able to generate a humoral and a cellular immune response to the cancer cells.[96]

1.2.8.2 Pathogen specific glycosyltransferase targets

As glycosylation is a process involving the expression of many glycosyltransferases and the existence of non-mammalian monosaccharides there will be pathogen specific glycosyltransferases. Drugs that target these enzymes may make the pathogen more

susceptible to the host's immune system by removing its ability to mimic host glycan presentation on its surface or by reducing its pathogenicity through other mechanisms.

Neuraminidases play a crucial role in influenza infection. Influenza particles adhere to the cell surface through adhesion to sialic acid residues. Upon creation of a new virus particle, the viral neuraminidase clips the terminal sialic acid residue from the cellular glycan thus allowing release of the viral particle from the cell surface where it can then go on and infect other cells.[97] This process is the target of neuraminidase inhibitors such as oseltamivir and zanamivir although resistance to such drugs is becoming more widespread.

V. cholerae can also produce a bacterial neuraminidase and this capability is found amongst all strains that cause the most severe cholera infection.[98] The role of this neuraminidase is to clip sialic acid residues from glycans on the cell surface in order to truncate the glycans to form GM1 (the natural binding ligand of the cholera toxin).[99, 100] This increases the cellular presentation of the cholera toxin binding ligand thereby increasing the amount of toxin binding and causing increased disease severity.[101] Inhibition of this neuraminidase could reduce disease severity in severe cholera infections.

As previously mentioned, the presence of sialic acid terminated glycans play a crucial role in the protection of *H. influenza* and *Neisseria spp.* from the complement immune system. As such, inhibition of the enzymes responsible for addition of the sialic acid cap to the glycans on the surface of these bacteria will make them more susceptible to killing by the immune system and offers a potential drug target (*vide infra*).

1.2.8.3 Anti-adhesion therapy

For those pathogens that have an essential carbohydrate mediated adhesion phase (Figure 1.15A), anti-adhesion therapy has potential for prevention of infection and treatment of disease. Many examples of carbohydrate based anti-adhesion compounds exist, but some examples include the use of glycopolymers for the inhibition of cholera toxin binding[102, 103] and for the specific inhibition of fimbriated (and thus pathogenic) strains of *E. coli*.[104 105] This mechanism is also used in nature to protect cells from bacterial infection, mucins which are found on the surface of epithelial cells are also secreted thus trapping pathogens before they can invade the underlying tissue (Figure 1.15B).[21, 106]

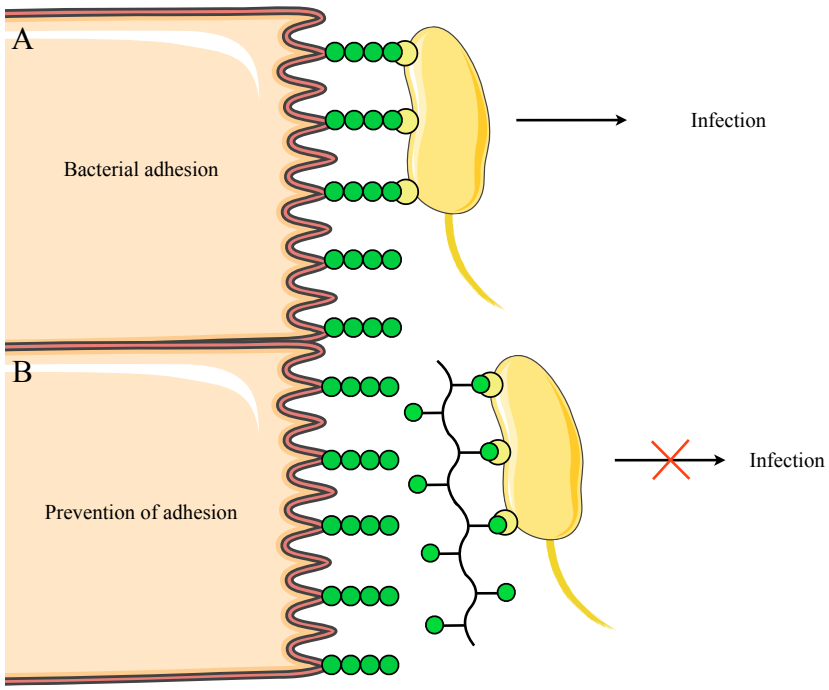


Figure 1.15 Schematic depicting anti-adhesion therapy. (A) Normally a bacteria will bind to the underlying cell glycans resulting in infection (B) whereas a drug or a mucin that mimcs these glycans can prevent binding and thus prevent infection.

A form of anti-adhesion therapy has also been proposed for the prevention of cancer metastases. Many cancer cells contain regions of aberrant glycosylation, one common modification is the presence of many selectin ligands. This would mask the cancer cells from the immune system as selectins are used to suppress the immune system and prevent it from attacking host cells. Several compounds have been designed to mimic the glycan-binding domain of selectins in order to prevent binding of selectins to tumour cells.[107-109]

1.2.8.4 Carbohydrate based drugs

Carbohydrates also form the basis of a number of other drug compounds including a number of antibiotics such as kanamycin, vancomycin and teicoplannin.[110] Heparin and related compounds such as dermatan sulphate form the basis of many anti-thrombotic agents.[111, 112] Glycoproteins have been used as enzyme replacement therapy in patients with Hurler and Hurler-Scheie forms of mucopolysaccharidosis I.[113]

1.2.9 Pathogen detection

As differential carbohydrate binding and glycosylation are often present in disease states or determine tissue specificity of pathogens they can be used as a detection mechanism too. Detection of disease specific glycosylation can be used in the detection of cancers or the presence of pathogens mimicking the host glycome and detection of pathogen specific lectins (either as secreted toxins or surface localised) can aid in the identification of pathogens.

1.2.9.1 Host detection of pathogen mimicry

Whilst many pathogens mimic the host glycome in the presentation of the glycans on their cell surface, the host immune system contains mechanisms for identifying mimicry. For example, sulfation is a site-specific modification that is common in mammals but very rare amongst bacterial species. Sulfation tends to occur on glycosaminoglycans and these sulfation events are recognised by immune lectins, which identify cells containing these compounds as belonging to the host. The sulfation of these glycans coupled with the epimerisation of glucuronic acid to iduronic acid allow host immune lectins to distinguish between host proteins and pathogen mimics.[20]

1.2.9.2 Detection of disease specific glycosylation

Disease specific changes in glycosylation can be detected through the use of labelled lectins or carbohydrate specific antibodies and many applications of this has been shown.

Lectins have been used for the staining of tissue in the detection of cancer for many years and hundreds of examples exist in the literature, to take but one example, the staining of tissue samples with the plant lectin leukoagglutinin (L-PHA) was able to detect cancerous cells through binding to the aberrant β(1-6) linkages present in oligosaccharides on their surfaces. Not only did L-PHA binding increase with cancerous tissues when compared to normal, but this increase was found to correlate with the pathological stage of the disease.[114]

Recently lectin microarrays have been used in the differentiation of gastric cancer tissue sections from gastric ulcers,[115] quantum dot conjugated lectins have been used for the evaluation of the glycophenotype of breast cancer[116] and a recombinant fungal lectin has been used in the detection of aberrant cancer associated glycosylation in tissue samples where the lectins were able to specifically bind cancer cells (reducing the risk of false positives).[117]

Not only has lectin binding been used to determine the cancerous state of tissues but this information has been used to inform treatment by identifying crucial cancer antigens. Lectin coated nanoparticles have also been shown to enhance drug delivery to the brain after intranasal administration of a drug with the aim of combatting Alzheimer's disease.[118]

Cancerous secretion of mucins into the bloodstream has been used as a marker for cancer through detection with monoclonal antibodies and many glycan-specific cancer-associated antibodies can be found in the blood of patients with cancer and these can be used for detection. Whilst many of these antibodies have become potential therapeutic targets their use may be limited by the high level of mucins coated in their glycan targets, which are secreted into the blood by cancerous tissues and thus would compete with tissue bound glycans for the drug.[31]

1.2.9.3 Detection of pathogen specific lectins

Many methods for detection of pathogen specific lectins have been explored in the literature. Glycan functionalised magnetic particles have been used for the detection of cancerous cells using magnetic resonance imaging[119], galabiose functionalised

magnetic particles for the detection of *Streptococcus suis* (a zoonotic pathogen),[2, 120] mannose conjugated gold nanoparticles for the detection of FimH positive (and thus pathogenic) strains of *E. coli*[104] and label-free real-time detection of lectins down to femtogram levels has also been achieved through the conjugation of unmodified carbohydrates to silicon nanowires.[121]

Whilst those examples were all particle based, carbohydrate microarrays were first used for mapping epitopes of HIV-related antibodies which has now been used to inform vaccine design.[122-125] They have also been used for the detection of food-based bacterial pathogens for which standard culturing methods are challenging. This methodology coupled a mono- and di-saccharide based microarray with surface plasmon resonance for a real-time label-free output.[126] Detection and typing of bacteria in the blood has also been achieved[127] along with; detection of oligosaccharide binding of avian and human influenza strains[128], identification of antibodies that protect against severe malaria[129] and epitope mapping of tumour-associated antigens.[130]

However, many of the techniques that are currently available for pathogen detection either have poor sensitivity as they utilise monosaccharides or those that are based on monosaccharides that have good sensitivity are often coupled to expensive detection methods such as surface plasmon resonance, mass spectrometry or quartz crystal microbalance. Long-chain oligosaccharides are often used as sensors in order to increase the sensitivity of detection systems but these are expensive or difficult to synthesise. Many other techniques require either expensive or synthetically challenging modified glycans for attachment to probes or complex and robust surface

chemistry. Whilst non-carbohydrate based-detection systems exist, many utilise proteomic or antibody based detection methods that require preparation, storage, expensive detection equipment or highly trained staff to implement and they are not always suitable for robust point-of-care diagnostics.

1.3 Summary

Whilst carbohydrates were previously thought of as lacking information density, they are now regarded as being unsurpassed in terms of structural diversity. Carbohydrates are covalently linked to membrane proteins and lipids to form a dense glycocalyx and whilst there are only ten mammalian monosaccharides it is thought that there are over 2000 different glycans present in the mammalian glycocalyx.

The glycocalyx is important for mediating many biological processes through protein carbohydrate interactions and also reflects disease states through aberrant glycosylation, such as the glycosylation patterns seen in mucins on cancerous cells. As they are crucial for many biological processes, protein carbohydrate interactions are also exploited by many pathogens during the initial stages of infection. They determine the binding of viral particles prior to cell internalisation, fungal cells, whole bacterial cells and their toxins and the initial binding of some parasites.

The mammalian glycocalyx also determines the binding of immune system lectins, which protect host cells from destruction by the immune system. As such, mimicry of the mammalian glycome to escape the host immune system is also prevalent amongst pathogens, with mimicry seen across a number of bacterial species to one degree or another. Those species that do not mimic the host glycome and thus express bacterial specific glycans on their surface are often intracellular pathogens or have developed other mechanisms of protection against the immune system.

Pathogen specific glycans and lectins form a mechanism for both the prevention of disease states and pathogen detection. Currently, carbohydrate based vaccines are

being developed (or in some cases already exist) and anti-adhesion therapy compounds are being developed for a number of different pathogenic organisms.

In terms of detection through glycan interactions, lectin staining of tissue samples in search of differential glycosylation patterns in disease states is a technique that has existed for many years and the detection of glycoproteins with aberrant glycosylation patterns or glycan specific antibodies is routinely used in testing for the presence of certain cancers. Conversely, the detection of pathogen specific lectins has also been used to develop many sensors for disease although currently many of the techniques require either complex oligosaccharide ligands which can be difficult to produce in large quantities and are expensive. These ligands can provide a very pathogen specific on/off response in terms of detection, where as monosaccharides which are readily available in large quantities can be used for much broader detection of multiple pathogens but with poor sensitivity and resolution unless coupled with expensive detection methods.

1.4 Aims and Thesis summary

Considering the above, the thesis aimed to couple unmodified readily available monosaccharides, facile surface functionalisation chemistry and a simple processing based methodology for improving the resolution in binding profiles without having to use expensive equipment or challenging techniques. This would then be used as a facile sensor for the detection of a multitude of pathogenic organisms.

In **Chapter 2** the information and binding profiles that are readily available through the Consortium for Functional Glycomics were utilised as a trial dataset to explore the information density that can be achieved when assessing carbohydrate-binding profiles. Furthermore, the power of two statistical techniques was assessed in order to improve the resolution of the information contained in monosaccharide binding profiles alone. **Chapter 3** then develops a facile surface coupling chemistry for the production of a cheap carbohydrate based assay. The key aim was to develop a simple methodology that showed specificity of binding but utilised unmodified saccharides for surface attachment in order for the technique to be cheap.

The combined approach of a monosaccharide surface coupled with a statistical analysis was then used in the detection of a variety of pathogens. In **Chapter 4**, the discrimination of pathogenic lectins of similar binding specificities was achieved through ratiometric profiling of binding to a variety of monosaccharides. **Chapter 5** then further develops this approach to allow the discrimination of bacterial species in order to improve point-of-care diagnostics. This chapter looked at the discrimination of a variety of bacteria including *Mycobacterium smegmatis*, *Mycobacterium marinum* and *Pseudomonas putida*, the first two being model organisms for

Mycobacterium tuberculosis (the causative agent of tuberculosis) and *P. putida* being a model organism for *Pseudomonas aeruginosa* which is a common cause of pneumonia in cystic fibrosis patients.

In **Chapter 6**, the potential of differential glycan binding by red blood cells infected with *Plasmodium falciparum* is described. *P. falciparum* is one of the parasites responsible for human malaria and is one of the main causes of global death from disease. With the spread of resistance to one of the main front line treatment drugs on the rise and microscopy remaining the gold standard for malaria diagnosis, the development of novel, facile and cheap diagnostics techniques that provide information about drug resistance is becoming an important area of research.

1.5 References

1. CDC, http://www.cdc.gov/TB/statistics/.
2. N. P. Pera and R. J. Pieters, *MedChemComm*, 2014.
3. G. M. Cook, *J Cell Sci*, 1986, **1986**, 45-70.
4. Y. J. Kim and A. Varki, *Glycoconj J*, 1997, **14**, 569-576.
5. D. Wang, *Proteomics*, 2003, **3**, 2167-2175.
6. T. Feizi, F. Fazio, W. Chai and C.-H. Wong, *Curr Opin Struct Biol*, 2003, **13**, 637-645.
7. K. Nicolaou and H. J. Mitchell, *Angew Chem Int Ed*, 2001, **40**, 1576-1624.
8. N. Douglas, S. Ley and S. Warriner, *J Chem Soc Perk T 1*, 1998, 51-66.
9. L. Yang, Q. Qin and X. S. Ye, *Asian J Org Chem*, 2013, **2**, 30-49.
10. N. V. Ganesh, K. Fujikawa, Y. H. Tan, K. J. Stine and A. V. Demchenko, *Org Lett*, 2012, **14**, 3036-3039.
11. T. Satomaa, A. Heiskanen, M. Mikkola, C. Olsson, M. Blomqvist, M. Tiittanen, T. Jaatinen, O. Aitio, A. Olonen, J. Helin, J. Hiltunen, J. Natunen, T. Tuuri, T. Otonkoski, J. Saarinen and J. Laine, *BMC Cell Biol*, 2009, **10**, 42.
12. A. Heiskanen, T. Hirvonen, H. Salo, U. Impola, A. Olonen, A. Laitinen, S. Tiitinen, S. Natunen, O. Aitio, H. Miller-Podraza, M. Wuhrer, A. M. Deelder, J. Natunen, J. Laine, P. Lehenkari, J. Saarinen, T. Satomaa and L. Valmu, *Glycoconj J*, 2009, **26**, 367-384.
13. S. C. Tao, Y. Li, J. Zhou, J. Qian, R. L. Schnaar, Y. Zhang, I. J. Goldstein, H. Zhu and J. P. Schneck, *Glycobiology*, 2008, **18**, 761-769.
14. V. Morad, M. Pevsner-Fischer, S. Barnees, A. Samokovlisky, L. Rousso-Noori, R. Rosenfeld and D. Zipori, *Stem Cells*, 2008, **26**, 2275-2286.
15. B. M. van den Berg, H. Vink and J. A. E. Spaan, *Circ Res*, 2003, **92**, 592-594.

16. K. F. Medzihradszky, K. Kaasik and R. J. Chalkley, *Mol Cell Proteomics*, 2015.

17. A. Lopez-Ferrer, C. Barranco and C. de Bolós, *Am J Clin Pathol*, 2002, **118**, 749-755.

18. J. B. Lowe and J. D. Marth, *Annu Rev Biochem*, 2003, **72**, 643-691.

19. CFG, http://www.functionalglycomics.org.

20. A. Adibekian, P. Stallforth, M.-L. Hecht, D. B. Werz, P. Gagneux and P. H. Seeberger, *Chem Sci*, 2011, **2**, 337-344.

21. A. Audfray, A. Varrot and A. Imberty, *Comptes Rendus Chimie*, 2013, **16**, 482-490.

22. E. Hosoi, *J Med Invest*, 2008, **55**, 174-182.

23. S. Berger, N. Young and S. Edberg, *Eur J Clin Microbiol*, 1989, **8**, 681-689.

24. D. J. Anstee, *Blood*, 2010, **115**, 4635-4643.

25. J. A. Rowe, D. H. Opi and T. N. Williams, *Curr Opin Hematol*, 2009, **16**, 480.

26. X. Zeng, C. A. Andrade, M. D. Oliveira and X.-L. Sun, *Anal Bioanal Chem*, 2012, **402**, 3161-3176.

27. M. A. Hollingsworth and B. J. Swanson, *Nat Rev Cancer*, 2004, **4**, 45-60.

28. M. Komatsu, S. Jepson, M. E. Arango, C. A. Carothers Carraway and K. L. Carraway, *Oncogene*, 2001, **20**, 461-470.

29. K. O. Lloyd, J. Burchell, V. Kudryashov, B. W. Yin and J. Taylor-Papadimitriou, *J Biol Chem*, 1996, **271**, 33325-33334.

30. S. Julien, M. A. Krzewinski-Recchi, A. Harduin-Lepers, V. Gouyer, G. Huet, X. Le Bourhis and P. Delannoy, *Glycoconj J*, 2001, **18**, 883-893.

31. M. M. Fuster and J. D. Esko, *Nat Rev Cancer*, 2005, **5**, 526-542.

32. T. Bogenrieder and M. Herlyn, *Oncogene*, 2003, **22**, 6524-6536.

33. A. Varki, *Proc Natl Acad Sci U S A*, 1994, **91**, 7390-7397.

34. A. Varki, *J Clin Invest*, 1997, **100**, S31-35.

35. S. Nakamori, M. Kameyama, S. Imaoka, H. Furukawa, O. Ishikawa, Y. Sasaki, Y. Izumi and T. Irimura, *Dis Colon Rectum*, 1997, **40**, 420-431.

36. R. Malhotra, M. R. Wormald, P. M. Rudd, P. B. Fischer, R. A. Dwek and R. B. Sim, *Nat Med*, 1995, **1**, 237-243.

37. M. C. Rose, *Am J Physiol*, 1992, **263**, L413-429.

38. J. J. Lundquist and E. J. Toone, *Chem Rev*, 2002, **102**, 555-578.

39. 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.

40. T. K. Dam, T. A. Gerken, B. S. Cavada, K. S. Nascimento, T. R. Moura and C. F. Brewer, *J Biol Chem*, 2007, **282**, 28256-28263.

41. K. Godula and C. R. Bertozzi, *J Am Chem Soc*, 2012, **134**, 15732-15742.

42. R. Raman, M. Venkataraman, S. Ramakrishnan, W. Lang, S. Raguram and R. Sasisekharan, *Glycobiology*, 2006, **16**, 82R-90R.

43. P. Gagneux and A. Varki, *Glycobiology*, 1999, **9**, 747-755.

44. I. W. Sutherland, K. A. Hughes, L. C. Skillman and K. Tait, *FEMS Microbiol Lett*, 2004, **232**, 1-6.

45. T. M. Daniel and G. R. Olds, *Clin Exp Immunol*, 1985, **60**, 249-258.

46. S. Rodriguez de Cordoba, J. Esparza-Gordillo, E. Goicoechea de Jorge, M. Lopez-Trascasa and P. Sanchez-Corral, *Mol Immunol*, 2004, **41**, 355-367.

47. P. R. Crocker, J. C. Paulson and A. Varki, *Nat Rev Immunol*, 2007, **7**, 255-266.

48. D. W. Hood, K. Makepeace, M. E. Deadman, R. F. Rest, P. Thibault, A. Martin, J. C. Richards and E. R. Moxon, *Mol Microbiol*, 1999, **33**, 679-692.

49. J. P. van Putten and B. D. Robertson, *Mol Microbiol*, 1995, **16**, 847-853.

50. Y. A. Knirel, E. T. Rietschel, R. Marre and U. Zahringer, *Eur J Biochem*, 1994, **221**, 239-245.

51. Y. A. Knirel, A. S. Shashkov, Y. E. Tsvetkov, P. E. Jansson and U. Zahringer, *Adv Carbohydr Chem Biochem*, 2003, **58**, 371-417.

52. P. Guerry and C. M. Szymanski, *Trends Microbiol*, 2008, **16**, 428-435.

53. C. Nilsson, A. Skoglund, A. P. Moran, H. Annuk, L. Engstrand and S. Normark, *Proc Natl Acad Sci U S A*, 2006, **103**, 2863-2868.

54. M. Goodfellow and S. T. Williams, *Annu Rev Microbiol*, 1983, **37**, 189-216.

55. C. J. O'Neal, M. G. Jobling, R. K. Holmes and W. G. Hol, *Science*, 2005, **309**, 1093-1096.

56. D. Cassel and T. Pfeuffer, *Proc Natl Acad Sci U S A*, 1978, **75**, 2669-2673.

57. S. V. Heyningen, *Science*, 1974, **183**, 656-657.

58. R. G. Zhang, D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook, *J Mol Biol*, 1995, **251**, 563-573.

59. J. M. Lord, L. M. Roberts and J. D. Robertus, *Faseb j*, 1994, **8**, 201-208.

60. F. Musshoff and B. Madea, *Drug Test Anal*, 2009, **1**, 184-191.

61. E. Rutenber, B. J. Katzin, S. Ernst, E. J. Collins, D. Mlsna, M. P. Ready and J. D. Robertus, *Proteins*, 1991, **10**, 240-250.

62. M. S. Nachbar and J. D. Oppenheim, *Am J Clin Nutr*, 1980, **33**, 2338-2345.

63. I. M. Vasconcelos and J. T. Oliveira, *Toxicon*, 2004, **44**, 385-403.

64. V. Lorenzsonn and W. A. Olsen, *Gastroenterology*, 1982, **82**, 838-848.

65. D. C. Kilpatrick, *Mol Biotechnol*, 1999, **11**, 55-65.

66. S. Boehm and S. Huck, *J Neurochem*, 1998, **71**, 2421-2430.

67. P. L. McNeil and S. Ito, *Gastroenterology*, 1989, **96**, 1238-1248.

68. P. L. McNeil and T. Kirchhausen, *Nat Rev Mol Cell Biol*, 2005, **6**, 499-505.

69. K. Miyake, T. Tanaka and P. L. McNeil, *PLoS ONE*, 2007, **2**, e687.

70. K. Ishibashi, S. Takesue, K. Watanabe and K. Oishi, *J Gen Microbiol*, 1982, **128**, 2251-2259.

71. M. Hartmann, H. Papavlassopoulos, V. Chandrasekaran, C. Grabosch, F. Beiroth, T. K. Lindhorst and C. Röhl, *FEBS letters*, 2012, **586**, 1459-1465.

72. L.-C. Xu and C. A. Siedlecki, *Acta biomaterialia*, 2012, **8**, 72-81.

73. J. E. Sherwood, G. L. Truchet and F. B. Dazzo, *Planta*, 1984, **162**, 540-547.

74. M. Janczarek, K. Rachwał, J. Cieśla, G. Ginalska and A. Bieganowski, *Plant and Soil*, 2015, **388**, 211-227.

75. J. Epstein, Q. Eichbaum, S. Sheriff and R. A. Ezekowitz, *Curr Opin Immunol*, 1996, **8**, 29-35.

76. J. C. Ryan and W. E. Seaman, *Immunol Rev*, 1997, **155**, 79-89.

77. J. Daub, A. Loukas, D. I. Pritchard and M. Blaxter, *Parasitology*, 2000, **120 (Pt 2)**, 171-184.

78. A. Loukas and R. M. Maizels, *Parasitology Today*, 2000, **16**, 333-339.

79. A. Loukas, A. Doedens, M. Hintz and R. M. Maizels, *Parasitology*, 2000, **121 Pt 5**, 545-554.

80. A. Loukas, N. P. Mullin, K. K. Tetteh, L. Moens and R. M. Maizels, *Curr Biol*, 1999, **9**, 825-828.

81. M. Jungery, G. Pasvol, C. I. Newbold and D. J. Weatherall, *Proc Natl Acad Sci U S A*, 1983, **80**, 1018-1022.

82. X.-z. Su, V. M. Heatwole, S. P. Wertheimer, F. Guinet, J. A. Herrfeldt, D. S. Peterson, J. A. Ravetch and T. E. Wellems, *Cell*, 1995, **82**, 89-100.

83. D. S. Peterson, L. H. Miller and T. E. Wellems, *Proc Natl Acad Sci U S A*, 1995, **92**, 7100-7104.

84. D. Chabasse and R. Robert, *Ann Inst Pasteur Microbiol*, 1986, **137b**, 187-193.

85. J. Bhowal, A. K. Guha and B. P. Chatterjee, *Carbohydr Res*, 2005, **340**, 1973-1982.

86. D. Chabasse, R. Robert, G. Tronchin and J. P. Bouchara, *Mycopathologia*, 1988, **103**, 81-85.

87. J.-P. Latgé, *Trends Microbiol*, 2001, **9**, 382-389.

88. J. Houser, J. Komarek, G. Cioci, A. Varrot, A. Imberty and M. Wimmerova, *Acta Crystallogr D Biol Crystallogr*, 2015, **71**, 442-453.

89. R. Pejchal, K. J. Doores, L. M. Walker, R. Khayat, P. S. Huang, S. K. Wang, R. L. Stanfield, J. P. Julien, A. Ramos, M. Crispin, R. Depetris, U. Katpally, A. Marozsan, A. Cupo, S. Maloveste, Y. Liu, R. McBride, Y. Ito, R. W. Sanders, C. Ogohara, J. C. Paulson, T. Feizi, C. N. Scanlan, C. H. Wong, J. P. Moore, W. C. Olson, A. B. Ward, P. Poignard, W. R. Schief, D. R. Burton and I. A. Wilson, *Science*, 2011, **334**, 1097-1103.

90. C. N. Scanlan, R. Pantophlet, M. R. Wormald, E. Ollmann Saphire, R. Stanfield, I. A. Wilson, H. Katinger, R. A. Dwek, P. M. Rudd and D. R. Burton, *J Virol*, 2002, **76**, 7306-7321.

91. D. Safari, M. Marradi, F. Chiodo, H. A. Th Dekker, Y. Shan, R. Adamo, S. Oscarson, G. T. Rijkers, M. Lahmann, J. P. Kamerling, S. Penades and H. Snippe, *Nanomedicine (Lond)*, 2012, **7**, 651-662.

92. G. Ada and D. Isaacs, *Clin Microbiol Infect*, 2003, **9**, 79-85.

93. J. M. Daubenspeck, H. Zeng, P. Chen, S. Dong, C. T. Steichen, N. R. Krishna, D. G. Pritchard and C. L. Turnbough, Jr., *J Biol Chem*, 2004, **279**, 30945-30953.

94. M. Tamborrini, D. B. Werz, J. Frey, G. Pluschke and P. H. Seeberger, *Angew Chem Int Ed Engl*, 2006, **45**, 6581-6582.

95. L. Schofield, M. C. Hewitt, K. Evans, M. A. Siomos and P. H. Seeberger, *Nature*, 2002, **418**, 785-789.

96. D. Miles, H. Roché, M. Martin, T. J. Perren, D. A. Cameron, J. Glaspy, D. Dodwell, J. Parker, J. Mayordomo, A. Tres, J. L. Murray, N. K. Ibrahim and G. the Theratope® Study, *The Oncologist*, 2011, **16**, 1092-1100.

97. N. A. Roberts, in *Prog Drug Res*, Springer, Editon edn., 2001, pp. 195-237.

98. I. Moustafa, H. Connaris, M. Taylor, V. Zaitsev, J. C. Wilson, M. J. Kiefel, M. von Itzstein and G. Taylor, *J Biol Chem*, 2004, **279**, 40819-40826.

99. G. M. Kuziemko, M. Stroh and R. C. Stevens, *Biochemistry*, 1996, **35**, 6375-6384.

100. D. J. S. Sharmila and K. Veluraja, *J Biomol Struct Dyn*, 2004, **21**, 591-613.

101. J. E. Galen, J. M. Ketley, A. Fasano, S. H. Richardson, S. S. Wasserman and J. B. Kaper, *Infect Immun*, 1992, **60**, 406-415.

102. M. Jones, L. Otten, S.-J. Richards, R. Lowery, D. Phillips, D. Haddleton and M. Gibson, *Chem Sc*, 2014, **5**, 1611-1616.

103. S. J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew Chem Int Ed*, 2012, **51**, 7812-7816.

104. S.-J. Richards, E. Fullam, G. S. Besra and M. I. Gibson, *J Mater Chem B*, 2014, **2**, 1490-1498.

105. Z. Han, J. S. Pinkner, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano and S. J. Hultgren, *J Med Chem*, 2010, **53**, 4779-4792.

106. M. A. McGuckin, S. K. Lindén, P. Sutton and T. H. Florin, *Nat Rev Micro*, 2011, **9**, 265-278.

107. R. K. Winn, D. Liggitt, N. B. Vedder, J. C. Paulson and J. M. Harlan, *J Clin Invest*, 1993, **92**, 2042-2047.

108. M. N. Fukuda, C. Ohyama, K. Lowitz, O. Matsuo, R. Pasqualini, E. Ruoslahti and M. Fukuda, *Cancer Res*, 2000, **60**, 450-456.

109. N. Kaila and B. E. t. Thomas, *Med Res Rev*, 2002, **22**, 566-601.

110. T. K. Ritter and C.-H. Wong, *Angew Chem Int Ed*, 2001, **40**, 3508-3533.

111. R. Merton and D. Thomas, *Thromb Haemost*, 1987, **58**, 839-842.

112. J. Hirsh and R. Raschke, *CHEST Journal*, 2004, **126**, 188S-203S.

113. H. M. I. Osborn, P. G. Evans, N. Gemmell and S. D. Osborne, *J Pharma Pharmacol*, 2004, **56**, 691-702.

114. B. Fernandes, U. Sagman, M. Auger, M. Demetrio and J. W. Dennis, *Cancer Res*, 1991, **51**, 718-723.

115. W.-L. Huang, Y.-G. Li, Y.-C. Lv, X.-H. Guan, H.-F. Ji and B.-R. Chi, *World J Gastroentero*, 2014, **20**, 5474-5482.

116. C. G. Andrade, P. E. Cabral Filho, D. P. L. Tenório, B. S. Santos, E. I. C. Beltrão, A. Fontes and L. B. Carvalho, *Int J Nanomedicine*, 2013, **8**, 4623-4629.

117. A. Audfray, M. Beldjoudi, A. Breiman, A. Hurbin, I. Boos, C. Unverzagt, M. Bouras, S. Lantuejoul, J. L. Coll, A. Varrot, J. Le Pendu, B. Busser and A. Imberty, *PLoS One*, 2015, **10**, e0128190.

118. C. Zhang, J. Chen, C. Feng, X. Shao, Q. Liu, Q. Zhang, Z. Pang and X. Jiang, *Int J Pharm*, 2014, **461**, 192-202.

119. K. El-Boubbou, D. C. Zhu, C. Vasileiou, B. Borhan, D. Prosperi, W. Li and X. Huang, *J Am Chem Soc*, 2010, **132**, 4490-4499.

120. N. P. Pera, A. Kouki, S. Haataja, H. M. Branderhorst, R. M. J. Liskamp, G. M. Visser, J. Finne and R. J. Pieters, *Org Biomol Chem*, 2010, **8**, 2425-2429.

121. G.-J. Zhang, M. J. Huang, J. A. J. Ang, Q. Yao and Y. Ning, *Anal Chem*, 2013, **85**, 4392-4397.

122. O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C. H. Wong and J. C. Paulson, *Proc Natl Acad Sci U S A*, 2004, **101**, 17033-17038.

123. E. W. Adams, D. M. Ratner, H. R. Bokesch, J. B. McMahon, B. R. O'Keefe and P. H. Seeberger, *Chem Biol*, 2004, **11**, 875-881.

124. C. N. Scanlan, R. Pantophlet, M. R. Wormald, E. O. Saphire, D. Calarese, R. Stanfield, I. A. Wilson, H. Katinger, R. A. Dwek, D. R. Burton and P. M. Rudd, *Adv Exp Med Biol*, 2003, **535**, 205-218.

125. P. M. Rudd, M. R. Wormald and R. A. Dwek, *Trends Biotechnol*, 2004, **22**, 524-530.

126. E. Bulard, A. Bouchet-Spinelli, P. Chaud, A. Roget, R. Calemczuk, S. Fort and T. Livache, *Biodevices*, 2015, DOI: 10.5220/0005254901210126.

127. M. D. Disney and P. H. Seeberger, *Chem Biol*, 2004, **11**, 1701-1707.

128. J. Stevens, O. Blixt, T. M. Tumpey, J. K. Taubenberger, J. C. Paulson and I. A. Wilson, *Science*, 2006, **312**, 404-410.

129. P. H. Seeberger and D. B. Werz, *Nature*, 2007, **446**, 1046-1051.

130. C. Y. Huang, D. A. Thayer, A. Y. Chang, M. D. Best, J. Hoffmann, S. Head and C. H. Wong, *Proc Natl Acad Sci U S A*, 2006, **103**, 15-20.

Chapter 2

The power and challenges of glycomics databases and the use of statistical tools in the extraction of crucial binding information; opportunities for chemists and biologists.

2.1 Abstract

Large amounts of lectin binding data are readily available in databases. However, considering the complex nature of carbohydrates it comes as no surprise that accessing and extracting information from these databases can be challenging. Despite the amount of data available, this information is also rarely used to inform experiments. As such this chapter highlights the complexity of the information available and also assesses the potential uses for this information in aiding in experimental design. Several statistical methodologies for the analysis of this data are also explored.

2.2 Introduction

Glycomics is faced with many unique challenges when compared to either proteomics or genomics mainly due to the complexity of the system being studied. Genomics involves the study of either DNA or RNA, which comprise four main building blocks. Proteomics analyses proteins, which are chains of amino acids of which there are 20 whereas there are 10 mammalian monosaccharides and even more found exclusively in bacterial species. Monosaccharides also have multiple attachment points, α or β conformations and are arranged in far more complex structures that those seen in either DNA or proteins. The building blocks in DNA, RNA and proteins can only be arranged in a linear manner whereas glycans often contain branched structures thus adding a further layer of complexity.

Further to the structural complexity of glycans themselves, there are additional challenges in the study of both glycosylation and protein-carbohydrate interactions. Glycosylation is not a template-driven process and is the product of the expression of multiple glycosyltransferases, which can have tissue specific isoforms leading to a wide variety in glycosylation and the degree of glycosylation between different tissues (Figure 2.1).[1-3] These different glycans can then interact with proteins but again these interactions are complex. Proteins that interact with carbohydrates make use of multivalency in order to overcome the naturally weak interactions.[4] Thus the properties of a protein-carbohydrate interaction will change dependent on the density of the target glycan.[5] This is then further complicated by the presence of graded affinity amongst these proteins, this means that the presence of the target glycan at low density can be overcome by additional weaker interactions of the protein with derivatives of the target carbohydrate chain.

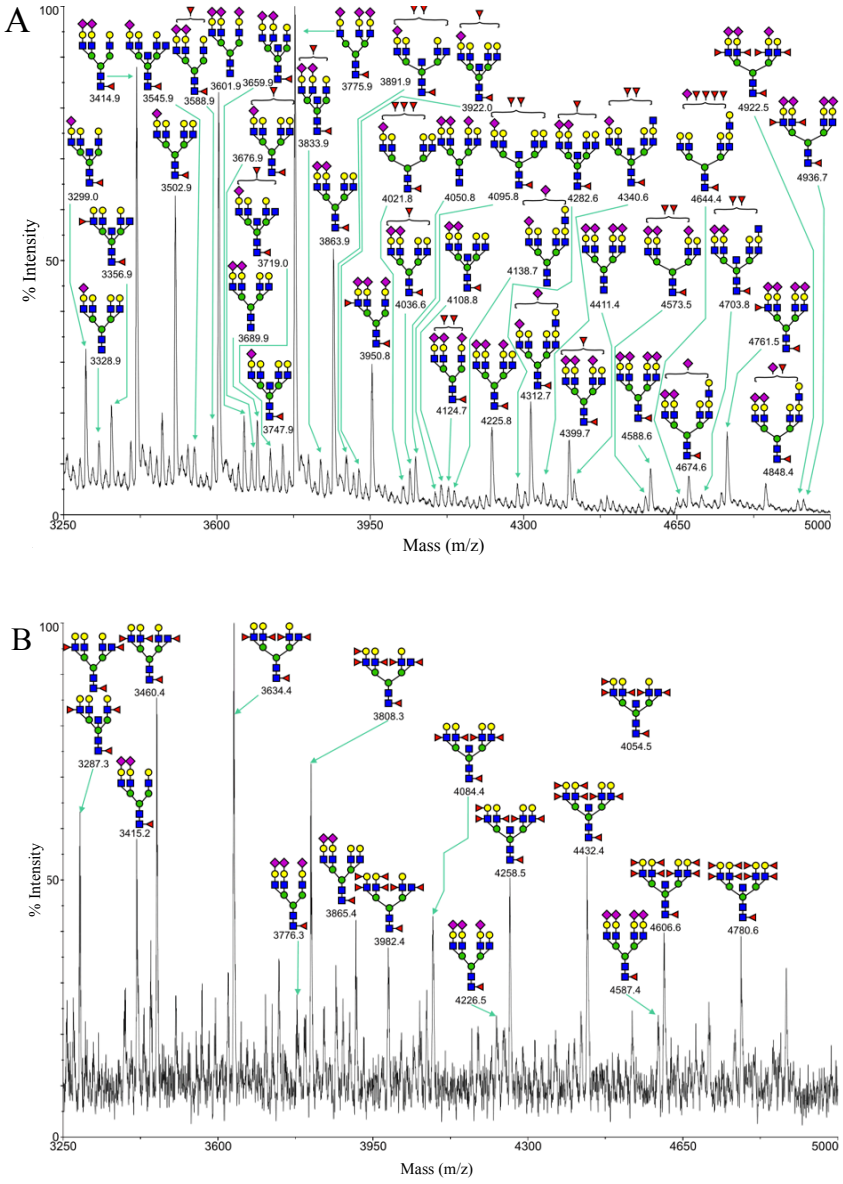


Figure 2.1 Mass spectrum showing tissue differences in both the amount of glycosylation and the nature of N-linked glycans, extracted from the consortium for functional glycomics (CFG) for A) human spleen tissue and B) human small bowel tissue. All glycan structures are depicted using standard CFG nomenclature.[6]

Given the complexity of glycan systems it is hardly surprising that the databases of information are themselves complicated and can often be challenging to use. There are many databases that consider only a small aspect of glycomics, for example

GlycomeDB,[7] which contains glycan structures including three-dimensional structural information and GlycoGene, which contains information about genes associated with glycan synthesis. Many other databases incorporate a much more integrative approach such as the consortium for functional glycomics (CFG), which contains information relating to the gene expression of glycosyltransferases and glycan binding proteins, phenotypic analysis of transgenic mice, mass spectrometry analysis of glycans isolated from various cell and tissue types and the screening of glycan affinity of proteins using a carbohydrate microarray approach.[8]

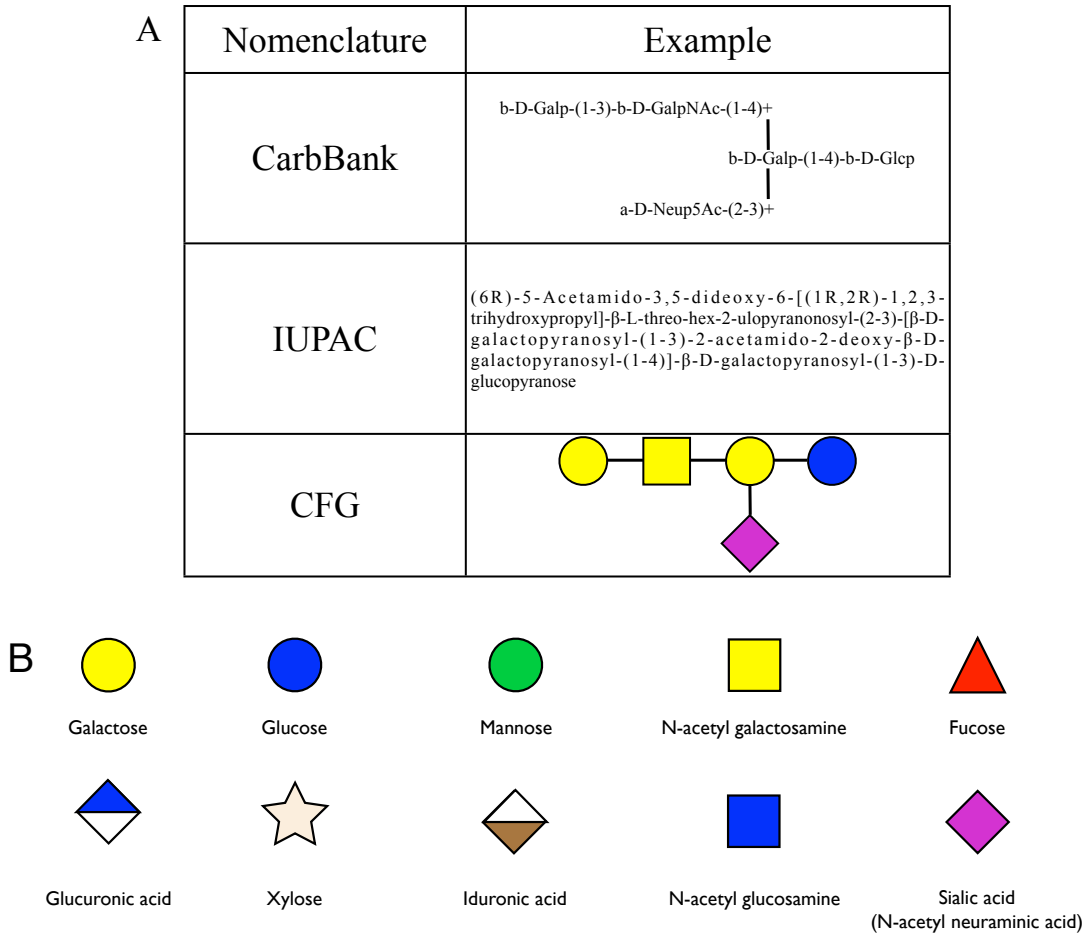


Figure 2.2 Many databases utilise their own nomenclature to represent glycans. A) Shows the CarbBank, IUPAC and CFG representation of the pentasaccharide component of monosialyltetrahexosylganglioside (GM1) and B) the standard CFG nomenclature for all 10 mammalian monosaccharides.

Using the output of one database as input information for another can also be challenging, as there is no uniform method for representing the structure of carbohydrates (Figure 2.2A) and many of the databases utilise their own nomenclature (Figure 2.2B). This means that after exporting information from one database the nomenclature will have to be changed in order to import into another database. This can be extremely time consuming. The differences in nomenclature can also make comparing information from two different databases challenging.

The CFG contains the most comprehensive information on the binding profile of glycan binding proteins with genomic, proteomic and glycomic information about every protein. All data is available in a processed and summarised format, raw data and experimental protocols are also available. Not only are there binding profiles for single proteins but there are also profiles for antibodies, viruses and whole bacterial cells. The binding profile to over 200 different glycans is stored for each, with the glycans themselves ranging from monosaccharides to branched oligosaccharides. These profiles were generated through the use of specialised carbohydrate microarrays comprising of glycans attached to an amino linker and covalently printed on *N*-hydroxysuccinimide activated glass slides.[8]

The wealth of information on protein and whole cell interactions with glycans can indicate the function of poorly understood proteins, and the mechanisms through which pathogenic lectins may have their function. This in turn can give the likely cell specificity of these lectins. Understanding mechanisms of action can be further utilised to guide inhibitor design and aid in the identification of lectins and whole

bacteria based on a ‘bar code’ approach. This approach requires only information about binding profiles for lectin identification.

Unfortunately much of the information in the CFG database is rarely used. As such we aim to show some of the information that can be extracted from this database to highlight the wealth of this resource for a variety of applications. Here, the power of this information in terms of aiding inhibitor design, sample classification and elucidating function based on binding profiles is highlighted. Some of the limitations of the information contained within these databases in terms of experimentally reproducing the data will be assessed.

2.3 Results and discussion

2.3.1 Toxin modes of action

Cholera toxin (CTx) is the toxic lectin produced by the bacteria *Vibrio cholerae* when it enters the small intestine. It is binding and internalisation of this lectin that results in constitutive activation of sodium channels in small intestine endothelial cells which then causes the symptoms associated with cholera infection.[9, 10] Ricin (RCA) is another toxic lectin, which is found in the seeds of the castor oil plant, binding and internalisation of this toxic lectin results in cell death.

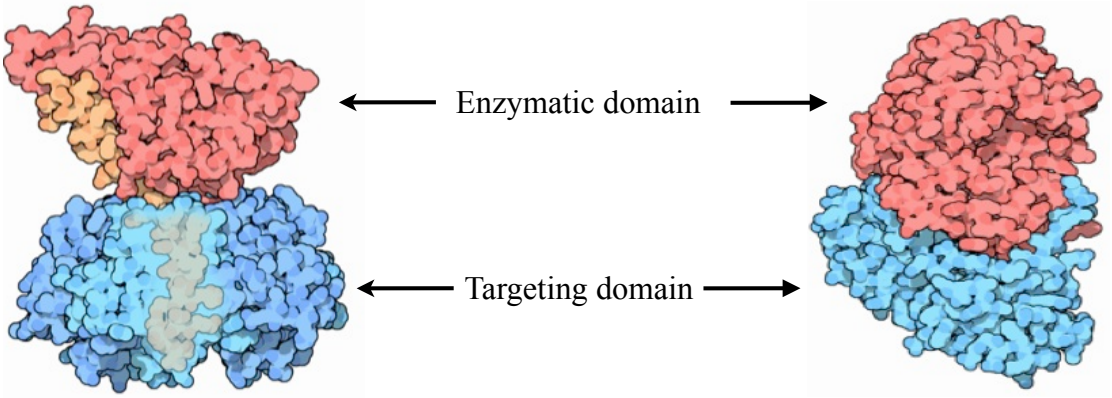


Figure 2.3 Crystal structures of the CTx (left) and RCA (right) highlighting the enzymatic domain shown in red and the carbohydrate-targeting domain shown in blue. Both structures are from the PDB with codes of 1XTC[11] and 2AAI[12] for cholera and ricin respectively.

Both of these lectins comprise an enzymatic domain and a targeting carbohydrate binding domain (Figure 2.3). Without binding of the targeting domain to carbohydrates on cell surfaces, the enzymatic subunit would not be internalised and the cell would not be affected; this forms the principle behind anti-adhesion therapy. Comparison of the carbohydrate binding regions of these toxins highlights the range of carbohydrates, which can result in toxin binding. This gives an indication of the cells that may be affected by binding. Binding of a subunit to a highly specific

oligosaccharide chain, as displayed by the b subunit of cholera toxin (CTxB, Figure 2.4), can be indicative of a very specific cellular target. The moiety that is strongly bound in this instance is the GM1 pentasaccharide (other moieties that show strong binding are truncated versions of this glycan), which is located on intestinal endothelial cells, the site of adhesion and internalisation of this toxin.[13]

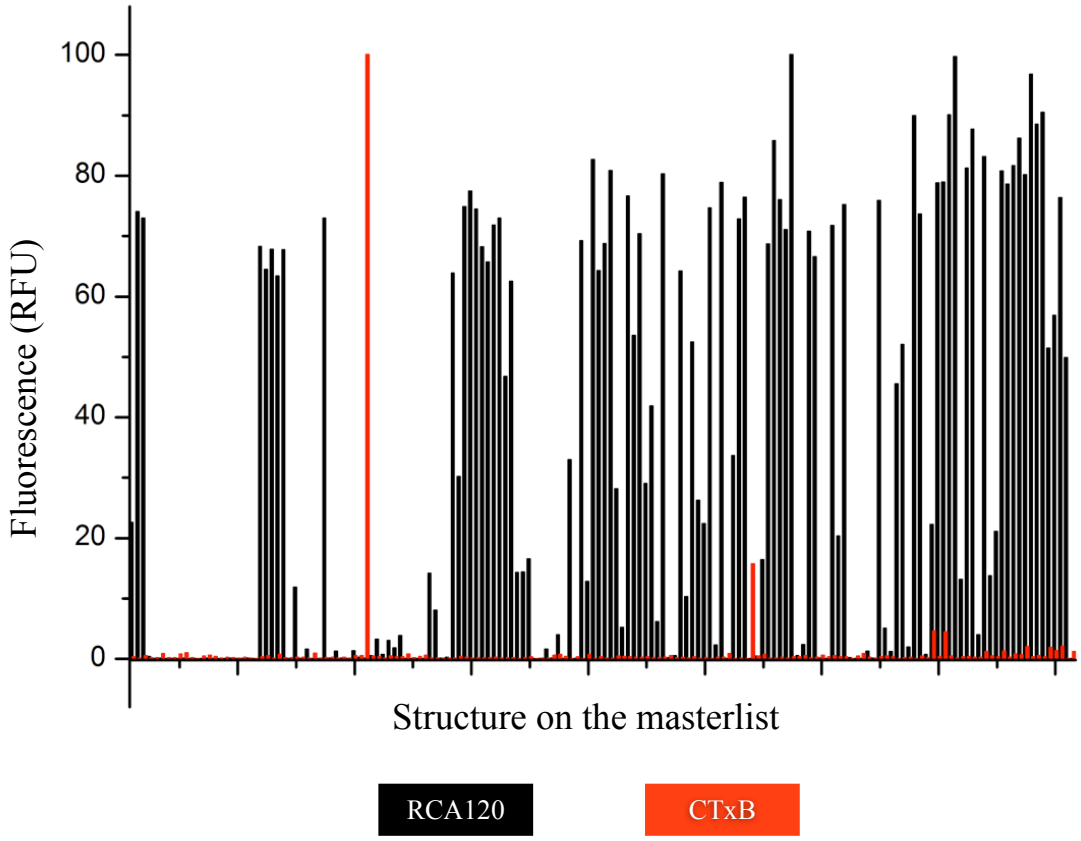


Figure 2.4 Binding profiles of cholera toxin (red) and ricin (black) to a variety of carbohydrate surfaces, all data extracted from the consortium for functional glycomics (the concentration for both is 10 μg.mL^{-1}) and binding is shown as a percentage relative to the maximum binding glycan. Fluorescence data is for all structures containing a terminal β-galactose residue.[6]

In comparison, the binding profile of the ricin targeting domain (RCA120, a non-toxic derivative of RCA) shows a much broader range of targets (Figure 2.4). All the glycans that are bound contain a terminal β-galactose moiety and this is the unit that is

required for ricin binding. As this is the only real similarity between the glycans that are strongly bound it indicates that ricin can target a wide range of carbohydrate chains as long as they contain a terminal β-galactose and as such could potentially target a wide variety of cells. Indeed ricin can affect any cell in the body. Upon binding the toxin enters the cell through endocytosis and the enzymatic A subunit is cleaved from the targeting B subunit. The A subunit then binds to the ribosome, halting protein synthesis and causing cell death.[14]

Whilst the modes of action of both these toxins have long been known, many of the lectins in the data base have been poorly studied or are difficult to obtain. These profiles could aid in the elucidation of the mode through which these lectins act. This is important for both lectins that are toxic or that play a role as part of a greater signalling pathway within an organism. Furthermore understanding modes of action can be used in the confirmation of proteomics studies in which putative glycan binding domains are identified.

2.3.2 Inhibitor design

Binding profiles can also be used to generate information that can be used in the design of inhibitors for targeted anti adhesion therapy. Upon identification of a toxin that binds a highly specific oligosaccharide, the binding profiles of truncated and modified versions of that oligosaccharide can be compared in order to identify which components are essential for binding. This can be used to aid in the production of inhibitors that are designed to mimic the target on the cell surface without using the cell surface glycan. This is vital as other native lectins in the body may use this glycan for signalling and as such would be affected by the inhibitor.

As has been shown, CTxB shows highly specific binding to the GM1 pentasaccharide (Figure 2.5). The binding profiles of a variety of truncated and modified versions of this glycan can guide the production of cholera toxin specific inhibitors.

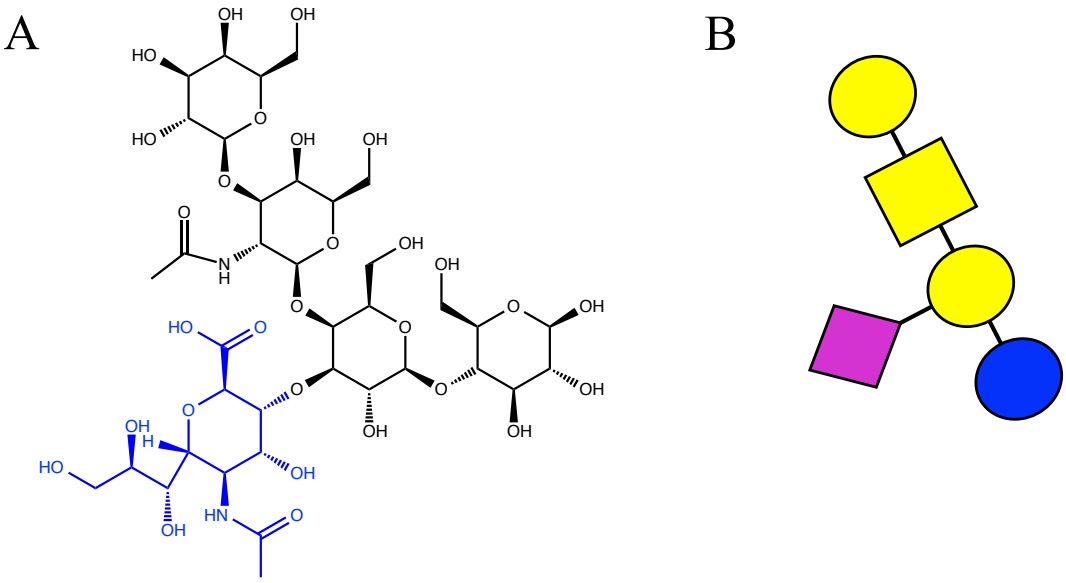


Figure 2.5 Structure of GM1 with the sugar backbone pictured in black and the side chain shown in blue (A). For comparison GM1 is also pictured using standard CFG nomenclature (B).

If we consider just the sugar backbone of GM1 (shown in black, Figure 2.5A), then we can base an inhibitor on anything from a monosaccharide up to a tetrasaccharide. The difference in binding to the terminal monosaccharide and the full tetrasaccharide (Table 2.1) is very small but the price difference in buying a monosaccharide and a tetrasaccharide is considerable. In order to form cheap inhibitors of cholera toxin, the terminal galactose residue is an attractive starting compound.

The terminal galactose has been combined with a variety of different scaffolds to generate a number of multivalent inhibitors of cholera toxin.[15-17] However, a number

of other lectins within the body, such as galectins, also work through binding to galactose moieties and a more specific inhibitor is needed. Adaption of the linker length between the galactose and the scaffold has been show to drastically improve the specificity of an inhibitor to cholera toxin when compared to another galactose binding lectin.[18] This was due to the deep binding pocket that cholera toxin has when compared to the surface exposed binding site for the other lectin. Thus galactose coupled with the appropriate length linker can form a good mimic of the GM1 sugar backbone.

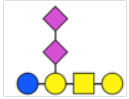
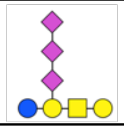
CFG structure	General name	Fluorescence	IC_{50}	Kd_{app}
	Galactose	146	N/A	N/A
	Sugar backbone	213	N/A	N/A
	GM1	46030	2.13 nM	1.5 nM
	GM1+one additional side chain unit	7215	12.3 nM	3.5 nM
	GM1+two additional side chain units	533	15.2 nM	12.9 nM
	Side chain (sialic acid)	114	N/A	N/A

Table 2.1 Table showing the relative fluorescence binding (RFU) of CTxB (concentration of 10 $\mu g.mL^{-1}$) to a variety of structures that represent either a modified or truncated version of GM1. Both a general name and the CFG structure are shown for both structures. IC_{50} values were determined by plotting the fluorescence data for a number of concentrations (all extracted from the consortium for functional glycomics database) and the apparent Kd (Kd_{app}) is calculated as described by Orsoz and Ovadi.[19]

However, if we consider the binding profile, the difference in binding between the GM1 sugar backbone and the full GM1 glycan is significant (Table 2.1). The only

difference in sequence between these glycans is the sialic acid side-chain, which sits in a hydrophobic binding pocket in the cholera toxin. Binding of the side chain on its own to cholera toxin is very poor. When coupled to the GM1 backbone the strength of binding is much greater than the sum of two individual component parts.

The importance of the side-chain is further highlighted by the large drop off observed in binding when the side chain is increased in length. The increase in the apparent dissociation constant (Kd_{app}) indicates that the sialic acid side chain is crucial for cholera toxin binding (Table 2.1). In fact GM1 with one additional side-chain sialic acid is found naturally on cell surfaces and severe cholera strains have evolved to secret a neuraminidase which clips this sialic acid, converting it into GM1 in order to increase the number of binding targets available to the cholera toxin.[20]

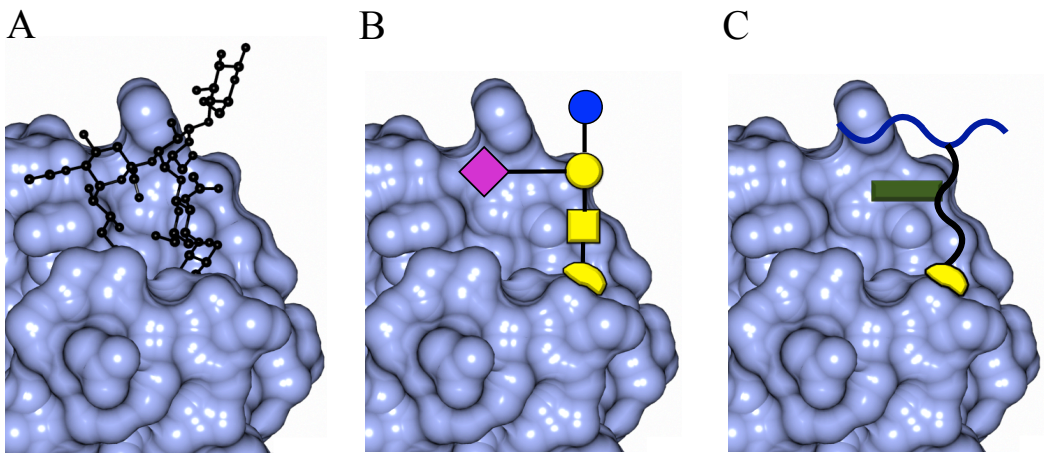


Figure 2.6 Crystal structure of GM1 bound to one CTxB as extracted from the PDB (A). A simplified version of A is shown using standardised CFG nomenclature (B). A schematic representation of the polymer inhibitor showing the proposed position of the hydrophobic secondary binding motif (C).

This observation can then be used to further guide the development of cholera toxin inhibitors. Polymers that contained a terminal galactose residue, coupled with the

idealised linker length for CTxB inhibition were further modified to test a variety of hydrophobic side chain compounds as demonstrated in Figure 2.6. Addition of a hydrophobic side group resulted in increased specificity for the inhibitor to CTxB (when compared to another galactose binding lectin).[21]

2.3.3 Identification of lectins from a carbohydrate binding 'bar code'

Many lectins are toxic and rapid identification of the toxin responsible for an infection can expedite treatment and save lives. Each lectin will have a unique binding fingerprint on the carbohydrate microarray (like a 'bar code') and this could be used as a tool for the rapid identification of toxins in samples.

A sample data set of 20 bacterial lectins with 6 replicates of every sample was extracted from the consortium for functional glycomics database in order to examine the possibility of lectin identification based on binding profiles. For each of the bacterial lectins, the binding intensity to a sample of 13 monosaccharides was extracted. Monosaccharides were selected as these are cheap and readily available and so, whilst they might be more challenging in discrimination of a bar code, they are the simplest platform for any identification sensor.

As can be seen from the raw binding profiles, identification of a single lectin based purely on its binding to these monosaccharides would be extremely challenging (Figure 2.7). However, there are a number of powerful statistics techniques available for the further analysis of this data, which may allow prediction of lectin samples. Two of those techniques, linear discriminant analysis and random forest, will now be explored and their identification power assessed.

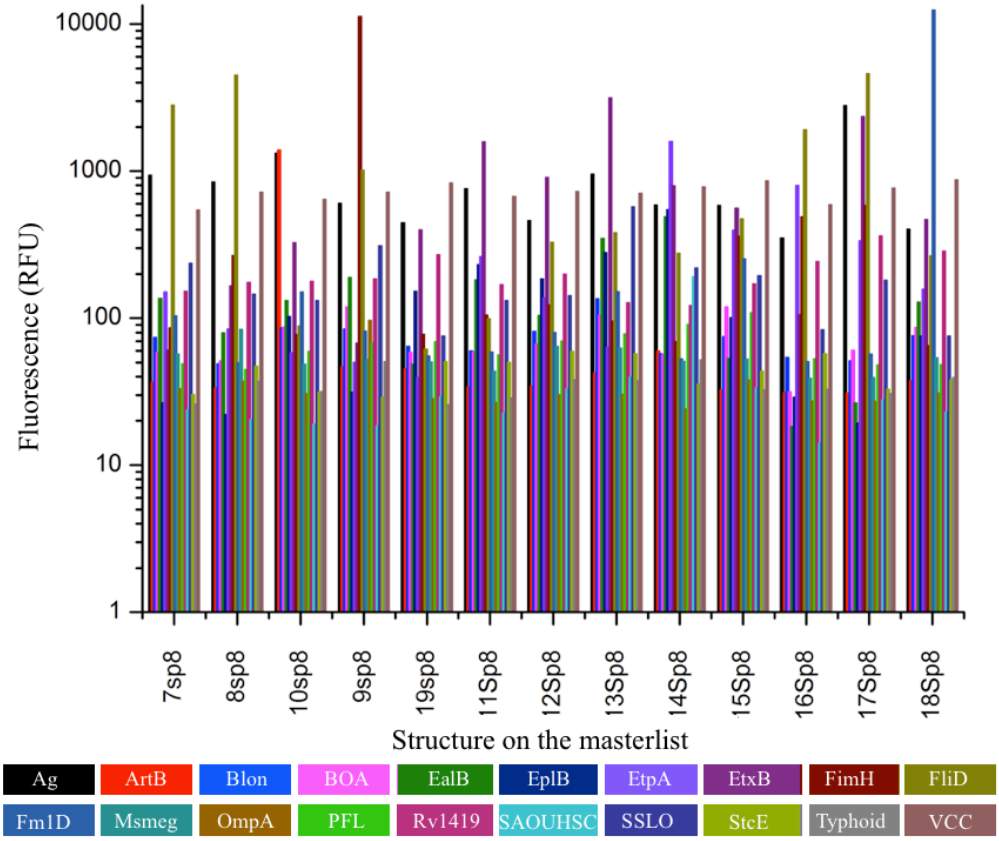


Figure 2.7 Relative binding profiles of 20 bacterial lectins to 13 monosaccharides as extracted from the consortium for functional glycomics. All monosaccharides are represented by their unique identifying code as used by the CFG.

2.3.3.1 Linear discriminant analysis

Linear discriminant analysis (LDA) is a clustering based machine-learning algorithm that can be used as a predictive tool for identifying group membership in unknown samples. Other clustering approaches such as principal component analysis focus on clustering data based on global variation rather than variation between groups. In comparison, linear discriminant analysis minimises variation between points within the same group whilst maximising variation between groups in order to allow better discrimination between groups (Figure 2.8). LDA has been previously used in the

identification of proteins based on binding to polymers and the identification of bacteria.[22, 23]

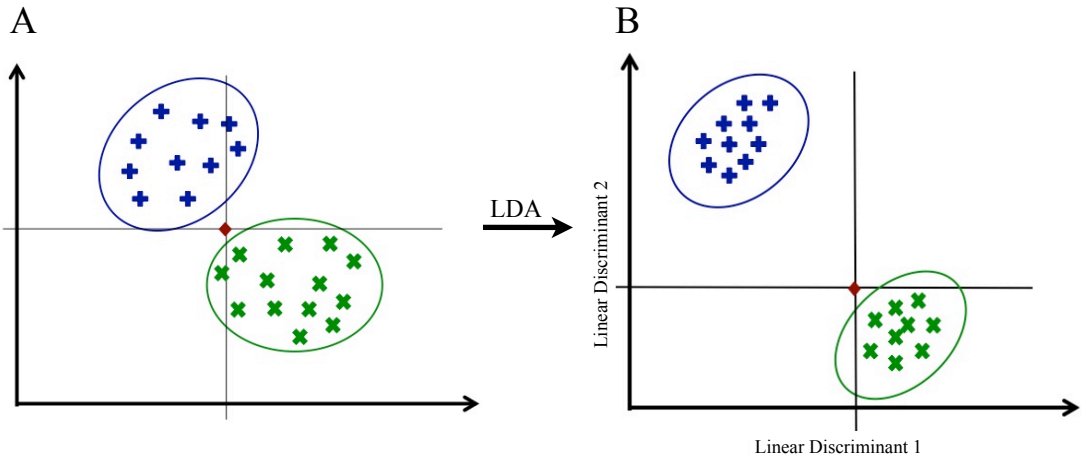


Figure 2.8 Example of linear discriminant analysis. A) Example of an initial data set segregated into two classes and showing an unknown sample for classification (the red diamond). By considering the x and y co-ordinates of the unknown sample it could be classified as belonging to either group. LDA takes this initial data set and transforms it to minimise the variation within each class whilst maximising the variation between classes to produce a model (B). When the x and y co-ordinates of the unknown sample are once more considered it can clearly be classified as belonging to the green class with much higher confidence.

One of the main challenges with this technique is the risk of overfitting (Figure 2.9, right), which results in a model that shows perfect discrimination for all groups submitted in the training matrix, but results in an excessively complex model that lacks the predictive capacity to classify new data. The opposite of this situation is also possible, in that too few components are retained (to avoid overfitting) and useful information is lost so the model lacks enough information to accurately describe the relationship (Figure 2.9, left). As a result the use of this technique requires many validation steps in order to find the perfect point, between being too complicated

(overfitting) and losing too much information (underfitting), known as the 'goldilocks point' (Figure 3.9, centre).

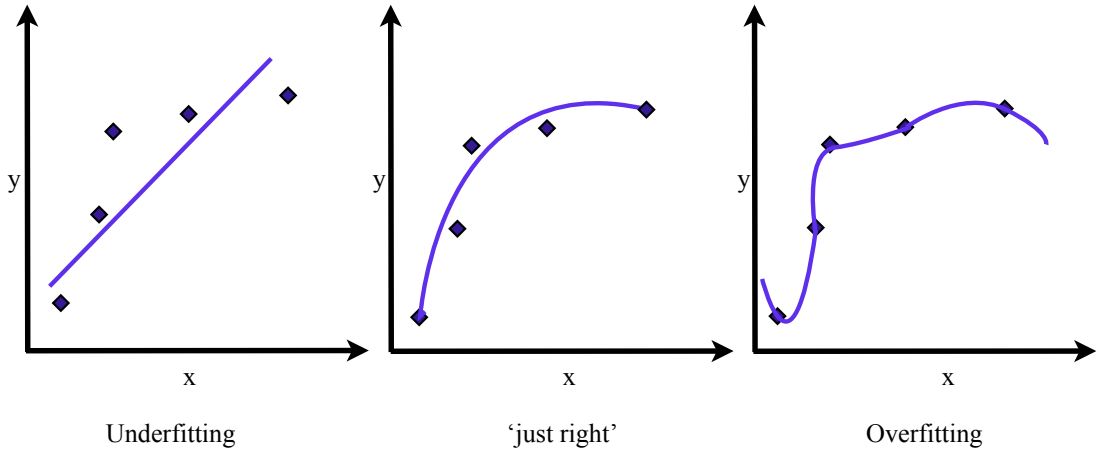


Figure 2.9 A model (represented by a purple line) can be fitted to data (navy diamonds) in order to describe the underlying relationship between x and y (centre). During underfitting, insufficient information is retained in the model and thus the model poorly describes the relationship (left). During overfitting, too much information is retained and whilst the model perfectly describes the data points it is excessively complex. Both underfitting and overfitting can result in the inability of the model to accurately predict a new point in the data.

Use of the previous data set (of 20 bacterial lectins, bound to 13 monosaccharide surfaces and 6 repeats for each lectin) was used as a training matrix for linear discriminant analysis. Cross-validation was performed in order to identify a suitable number of principal components (PCs) to retain in the final model (the 'goldilocks' number), this technique splits the data into two data sets; a training set (which contains 90 % of the original data) and a validation set (containing 10 % of the original data). The validation set is changed through random sampling and is classified by the model produced from the training set (with the number of retained principle components varied). This procedure is repeated 30 times for every number of retained principal components. For this data set 6 principal components were

selected as this explained most of the variation in the data (Figure 2.10A) and it achieved the highest successful prediction outcome score with the least number of PCs (Figure 2.10B).

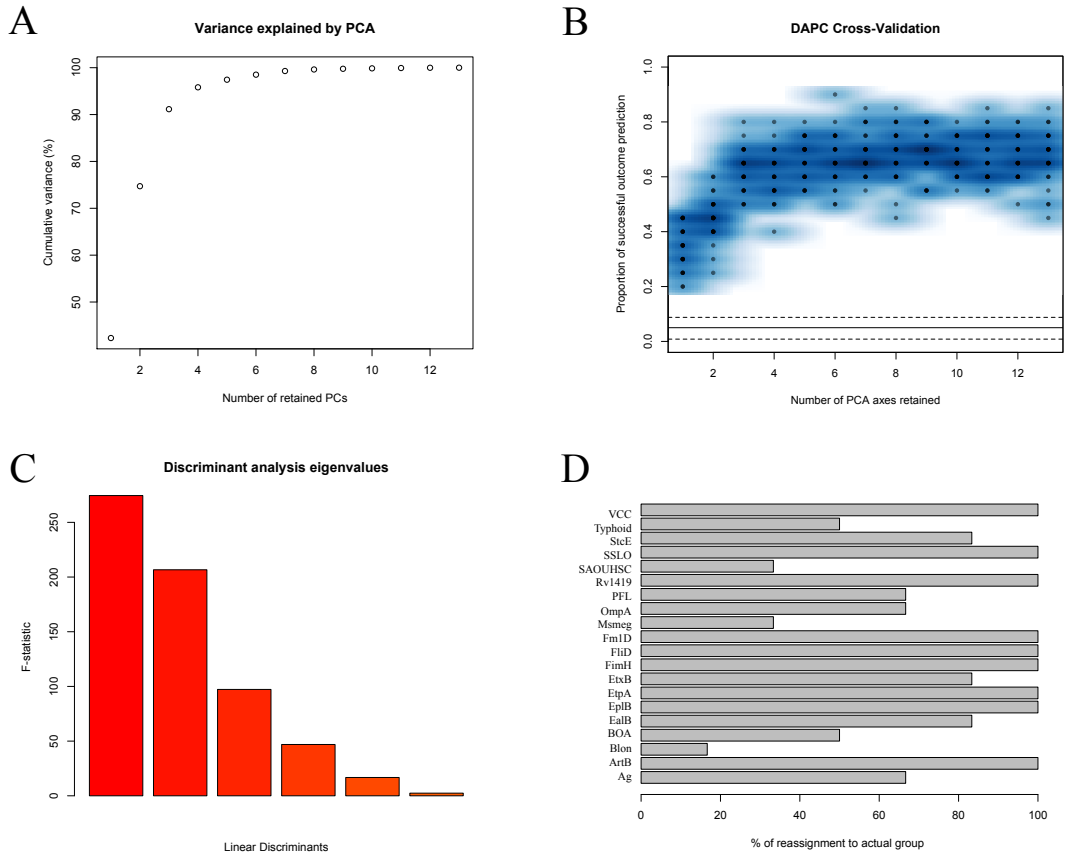


Figure 2.10 Variance explained by each principle component in the linear discriminant model (A). The successful prediction outcome when a validation data set is predicted from a model produced based on a training data set, the validation data set is randomly selected for 30 repeats and the number of principal components retained is varied (B). The variance explained by each of the linear discriminants for a model based on 6 principal components (C). The percentage of correct reassignment of a sample to its group by the LDA model produced in a 'leave-one-out' manner (D).

As only 6 principal components were retained, it was possible to retain all of the linear discriminants to produce the final model (for large numbers of principal components the model then becomes too complex to retain all of the linear

discriminants) and all linear discriminants contained a reasonable amount of information (Figure 2.10C). The validity of the model was then assessed using a 'leave-one-out' validation procedure in which every data point is left out of the model in an iterative process and the model is used to classify any points not used to generate the model. The model was able to correctly identify 76 % of all the samples analysed with several of the lectin groups scoring 100 % accuracy (Figure 2.10D). Whilst this is not a particularly high predictive power, it could be further improved by extending the analysis to include di- and oligosaccharides.

2.3.3.2 Random forest

Random forest is another classification technique that is generally assumed to be more powerful than linear discriminant analysis in terms of group identification. This analysis involves the production of multiple classification and regression trees based upon a random subsection of the data, each tree is also perturbed by a small amount in order to account for variance in any input data. In classification of a sample, each of the trees is used to classify the sample with the output determined as the majority class identified by the trees. This procedure has the benefit that it reduces the possibility of overfitting (where the model produced fits both the data and the underlying 'noise' in the dataset), which can be a problem when using LDA, by allowing variation within the model. It also copes with non-linearity in the data (where as LDA can only produce discriminants that are combined in a linear manner). This technique has previously been used for the prediction of disease risk and the identification of colon cancer biomarkers from microarray data.[24, 25]

Random forest was applied to the previous data set of 20 bacterial lectins on 13 monosaccharide surfaces (with 6 repeats for each lectin on every surface). This technique is generally thought to be more powerful a classification tool than LDA and as expected in a 'leave-one-out' analysis it correctly assigned samples to their lectin group in 77.5 % of all samples thus making it appear marginally more accurate than the 76 % achieved through LDA. However, the correct assignment when broken down to the individual lectins was worse than that achieved with LDA- whilst some of the lectins showed higher correct predictions with this technique, all samples for two of the lectin groups assessed were incorrectly classified (Figure 2.11).

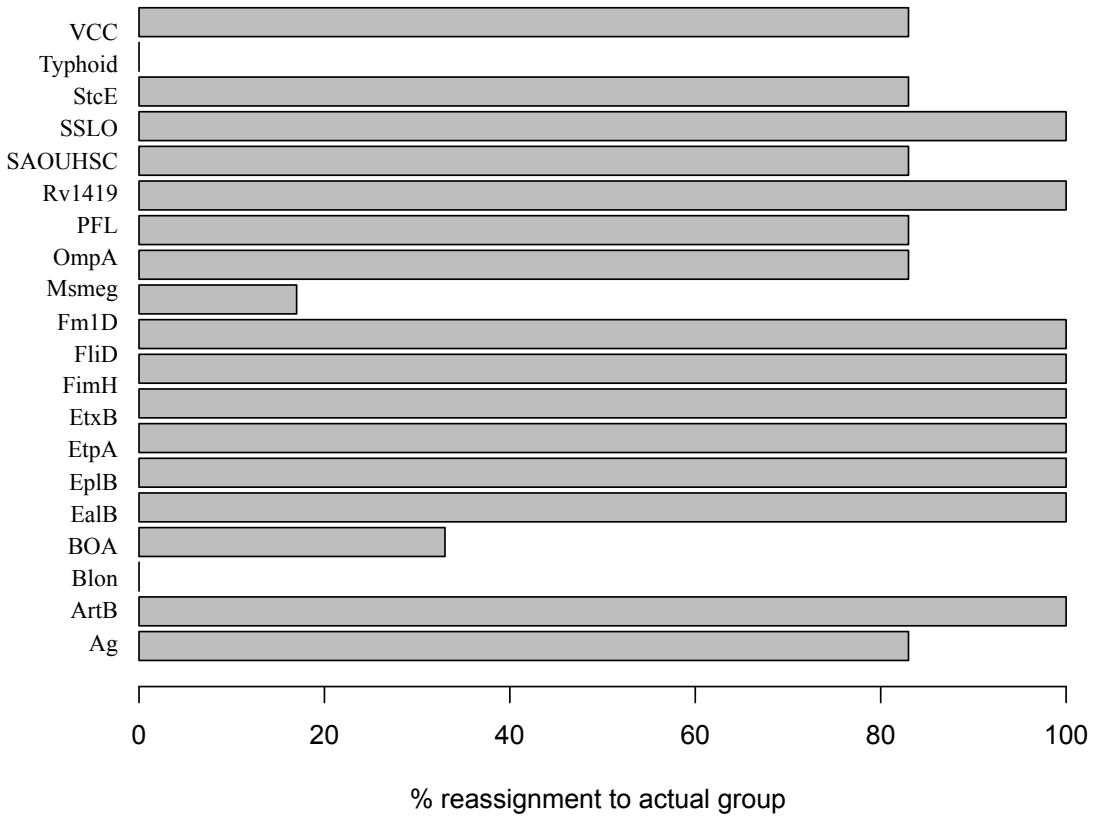


Figure 2.11 The percentage of correct reassignment of a sample to its group by the random forest model produced in a 'leave-one-out' manner.

2.3.4 Challenges when using database profiles

There are a number of challenges associated with using data from databases such as the consortium for functional glycomics. The CFG database includes attachment of a variety of glycans with a variety of spacers, which can impact on binding. The database entries are also produced by a variety of different laboratories, worldwide and, whilst all the detection values are fluorescence, the actual detection methods vary for some of the samples tested. This can have a significant impact on the binding profiles so only those with identical detection methods should be compared. Extraction of the data itself from the database can also be very challenging.

2.3.4.1 Spacer effect

The consortium for functional glycomics utilises a very large library of glycans (~ 200 glycans) on their surface. They also use a number of spacers in order to achieve surface attachment. Many of the glycans are attached using a number of different spacers but not all spacers are available for all gylcans. This presents a unique problem, in that the spacer itself can affect lectin binding.

For many of the spacers coupled to carbohydrates there is little difference between lectin binding to the same glycan utilising different spacers. But, for some lectins, there are large differences between spacers for the same glycan, and only for certain glycans (i.e. it is not a universal affect for that lectin). For example, changing the spacer from Sp0 ($-CH_2CH_2NH_2$) to Sp9 ($-CH_2CH_2CH_2CH_2CH_2NH_2$) resulted in a two-fold increase in IC_{50} for cholera toxin binding to the demonstrated hexasaccharide (Figure 2.12). This effect is unsurprising as it is also used in nature to modulate binding to glycophospholipids where the binding strength to the

carbohydrate moiety can be affected by the fatty acid chain.[26] Whilst the effect of the spacer could provide information about the depth of the binding pocket or other characteristics it does create a problem when comparing glycan binding between lectins as the affect of the spacer on binding can be significant.

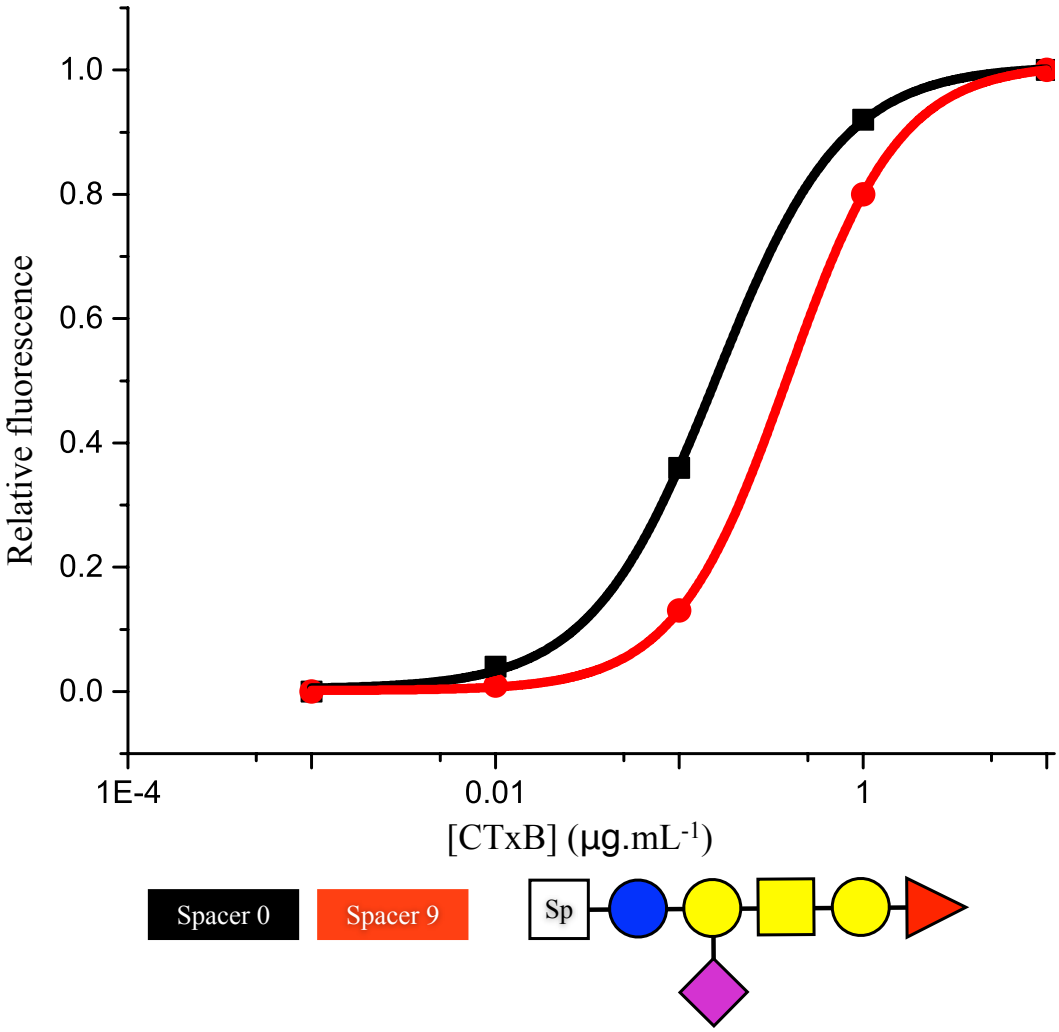


Figure 2.12 The relative fluorescence for various concentrations of cholera toxin (CTxB) binding to the displayed glycan for spacer 0 and spacer 9 was extracted from the CFG database. Binding to the glycan attached to spacer 0 (Sp0) showed stronger binding than binding to the same glycan to spacer 9 (Sp9) with IC_{50} values of 1.8 nM and 4.5 nM respectively.

The presence of different spacers means that the data in the CFG can only be compared with lectin binding using a CFG microarray or binding onto a surface

functionalised in an identical way. Even the use of identical spacers on a different surface may produce different results as density of surface functionalisation can impact on lectin binding (as they often rely on multivalent interactions).

2.3.4.2 Data comparison

Other than the affect of the spacer, there are many other aspects that limit the use of the CFG database for lectin comparisons. Many of the entries in the database utilise different labelling or detection protocols, which may affect the detection limit of the assay. Only those with identical labelling protocols should really be compared and even then the labelling efficiency itself needs to be considered as this can vary drastically between proteins.[27]

Fluorescent labelling of proteins has also been shown to impact upon glycan binding profiles of certain proteins and thus data within this database should not be compared to label free techniques and may not be an accurate representation of natural binding of the protein.[28]

Extraction of entries themselves can be challenging as several of the entries are in a non-standard format and thus extraction of the data can be laborious. Once the extraction is complete then further formatting of the data may be required depending on future analysis protocols. A standardised array for all glycans and a standard visualisation technique could ensure that all data are comparable between experiments.

2.4 Conclusion

In this chapter, we have demonstrated that large amounts of lectin binding information is readily available from databases such as the consortium for functional glycomics. This data can be used to highlight important elements involved in binding of a lectin to a known glycan, which can then be used to inform inhibitor design for anti-adhesion therapy.

Powerful statistics tools such as random forest and linear discriminant analysis coupled with the information within these databases highlights the potential for using large data sets to assist in classification of unknown lectin samples based on their binding profiles. This could then be used to identify toxic environmental lectins, such as cholera toxin and ricin based on binding alone, which is currently challenging, as samples may be contaminated with environmental lectins that show similar binding to either ricin or cholera toxin.

However, the use of these databases, in informing real-world experiments is limited by the poor accessibility of the data, non-uniform formatting of data sets, variations between data sets in experimental protocol and challenges caused by the effect of the spacers on lectin binding meaning. This means the data in the CFG can only be compared to lectin binding to exactly the same microarray platform. This is unfavourable as these microarrays are expensive.

In conclusion, whilst there is a wealth of information available in terms of carbohydrate binding profiles, its use has many limitations not least of which is the expense of using their microarray platforms. It is favourable to combine many of the

analysis techniques examined within this chapter (such as linear discriminant analysis) with a much cheaper more customisable surface modification technique. Coupled with standardised techniques for assessing binding (be that fluorescence or other 'label free' methodologies) this could be used to generate a database of profiles that can be accessed and reproduced around the world with the aim of using the database for sample classification based on binding profiles.

2.5 Methods

Consortium for functional glycomics data extraction: The following data files were downloaded from the consortium for functional glycomics and all data plotted are the average of the 6 repeats and error bars represent the standard error (Table 2.2). All graphs are plotted in OriginPro and binding isotherms are fitted using the nonlinear curve fitting function.

Abbreviation in text	**Sample name in CFG database**
RCA120	RCA120 (10 μg.mL^{-1})
CTxB	CTxB (0.001 μg.mL^{-1}, 0.01 μg.mL^{-1}, 0.1 μg.mL^{-1}, 1 μg.mL^{-1}, 10 μg.mL^{-1})
Ag	Ag I/II Vhelical-GFP: Ag I/II Vhelical-GFP
ArtB	ArtB toxin: ArtB bacterial toxin- 200 μg.mL^{-1}
Blon	Blon 2468 mut
BOA	*Burkholdeira oklahomensis* agglutinin (BOA): BOA lectin
EalB	EalB-h6:EalB-h6 ADP- ribosylating toxin
EplB	EplB-h6: EplB-h6 ADP ribosylating enzyme
EtpA	recombinant EtpA glycoptorein: rEtpA
EtxB	EtxB:etxAB-100 μg.mL^{-1}
FimH	FimH:FimH type 1 fimbriae 10 μg.mL^{-1}
Fm1D	FimD:FimD (200 μg.mL^{-1})
FliD	FliD:FliD
Msmeg	MSMEG_3662:MSMEG3662-2B
OmpA	OmpA+*E. coli*:OmpA+*E. coli*
PFL	PFL:PFL
Rv1419	Iron regulated heparin binding hemmaglutinin from *Mycobacterium tuberculosis*
SAOUHSC	SAOUHSC_00176
SSLO	SSLO:SSLO-488 (200 μg.mL^{-1})
StcE	StcE:StcE
Typhoid	Typhoid toxin: WT_Typhoid toxin- 20 μg.mL^{-1}

Table 2.2 Table linking the lectin abbreviation in the text to the file name within the CFG database. Where more than one concentration is stated then the concentration used for the method is stated.

Linear discriminant analysis: Six repeats of the relative fluorescence for each of the bacterial lectins binding to each of the chosen monosaccharide surfaces was extracted from the CFG. This data formed a training matrix which was subjected to classical linear discriminant analysis using the 'dapc' function in the 'adegenet' package (version 1.4-2)[29] in the open source statistical package R (version 3.1.3).[30]

Cross-validation was performed using the 'xvalDapc' function whereby 10 % of the data set is 'left out' of the model and then the model produced is used to classify the 'left out' data. This is repeated multiple times with the 'left out' data being randomly selected.

Random forest analysis: The dataset of six repeats of the binding of 20 bacterial lectins to 13 monosaccharide surfaces as extracted from the CFG database was used to produce a random forest model using the 'randomForest' function (version 4.6.10)[31] in the open source statistical package R (version 3.1.3).[30] The model produced was the average of 500 trees. Data were cross-validated using a 'leave-one-out' approach and the percentage of correct reassignment for each bacterial lectin was calculated.

Apparent Kd calculations: Kd_{app} calculations were performed as described by Orosz and Ovadi[19] but, briefly; relative fluorescence data produced upon binding of different concentrations of cholera toxin to the various glycans assessed was extracted from the consortium for functional glycomics.

Fluorescence levels were then modified to put them on a relative scale with 1 indicating the maximum fluorescence with 0 indicating the lowest fluorescence. Points were then plotted in OriginPro and a binding isotherm fitted to the data using the nonlinear curve fit tool (as shown in Figure 2.13A). Concentration of cholera toxin at a variety of fluorescence levels (i) was extracted from the fitted curve. All values were then modified and plotted as shown in Figure 2.13B and the gradient of the fitted line (m) was extracted.

$$Kd_{app} = 1/m \qquad \text{(Eq. 1)}$$

The Kd_{app} can then be determined by using Eq. 1. For any data set where the binding isotherm failed to converge, the Kd_{app} was unable to be calculated and this is reported as NA.

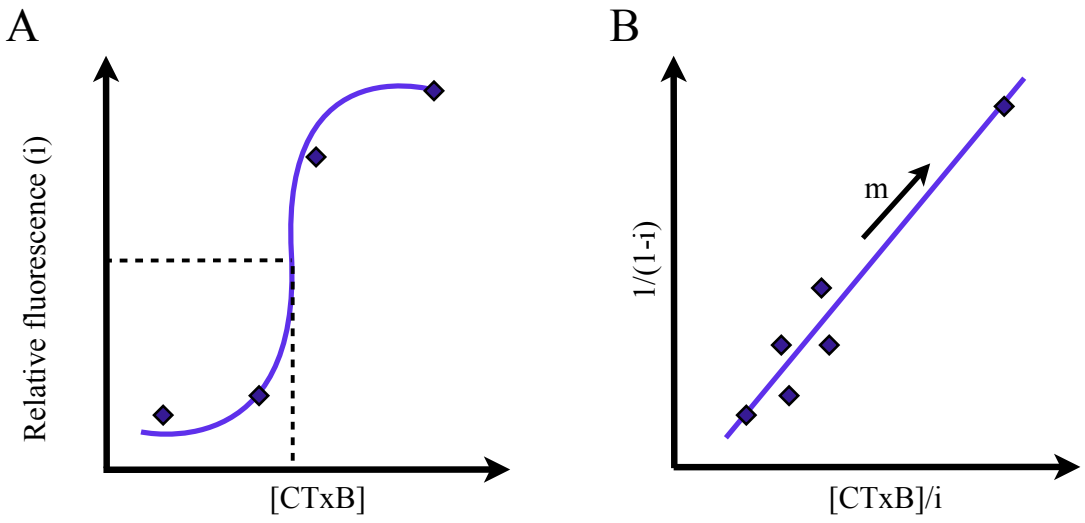


Figure 2.13 Example data set showing transformation needed to calculate Kd_{app} as described by Orosz and Ovadi.[19] A) Relative fluorescence (i) is plotted for various concentrations of cholera toxin and a binding isotherm fitted to the data. The modelled isotherm is then used to estimate cholera toxin concentration at various fluorescence (i) values. These extracted values are then further transformed and plotted as indicated in B. The gradient of the fitted line (m) is extracted.

2.6 References

1. K. F. Medzihradszky, K. Kaasik and R. J. Chalkley, *Mol Cell Proteomics*, 2015.
2. A. Lopez-Ferrer, C. Barranco and C. de Bolós, *American journal of clinical pathology*, 2002, **118**, 749-755.
3. J. B. Lowe and J. D. Marth, *Annual review of biochemistry*, 2003, **72**, 643-691.
4. J. J. Lundquist and E. J. Toone, *Chemical reviews*, 2002, **102**, 555-578.
5. T. K. Dam and C. F. Brewer, *Glycobiology*, 2010, **20**, 270-279.
6. CFG, http://www.functionalglycomics.org.
7. R. Ranzinger, S. Herget, C. W. von der Lieth and M. Frank, *Nucleic Acids Res*, 2011, **39**, D373-376.
8. R. Raman, M. Venkataraman, S. Ramakrishnan, W. Lang, S. Raguram and R. Sasisekharan, *Glycobiology*, 2006, **16**, 82R-90R.
9. C. J. O'Neal, M. G. Jobling, R. K. Holmes and W. G. Hol, *Science*, 2005, **309**, 1093-1096.
10. D. Cassel and T. Pfeuffer, *Proc Natl Acad Sci U S A*, 1978, **75**, 2669-2673.
11. R. G. Zhang, D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook, *J Mol Biol*, 1995, **251**, 563-573.
12. E. Rutenber, B. J. Katzin, S. Ernst, E. J. Collins, D. Mlsna, M. P. Ready and J. D. Robertus, *Proteins*, 1991, **10**, 240-250.
13. S. V. Heyningen, *Science*, 1974, **183**, 656-657.
14. J. M. Lord, L. M. Roberts and J. D. Robertus, *FASEB J*, 1994, **8**, 201-208.
15. M. Mattarella, J. Garcia-Hartjes, T. Wennekes, H. Zuilhof and J. S. Siegel, *Org Biomol Chem*, 2013, **11**, 4333-4339.

16. 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.

17. T. R. Branson, T. E. McAllister, J. Garcia-Hartjes, M. A. Fascione, J. F. Ross, S. L. Warriner, T. Wennekes, H. Zuilhof and W. B. Turnbull, *Angew Chem Int Ed Engl*, 2014, **53**, 8323-8327.

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

19. F. Orosz and J. Ovadi, *J Immunol Methods*, 2002, **270**, 155-162.

20. I. Moustafa, H. Connaris, M. Taylor, V. Zaitsev, J. C. Wilson, M. J. Kiefel, M. von Itzstein and G. Taylor, *J Biol Chem*, 2004, **279**, 40819-40826.

21. M. Jones, L. Otten, S.-J. Richards, R. Lowery, D. Phillips, D. Haddleton and M. Gibson, *Chemical Science*, 2014, **5**, 1611-1616.

22. R. L. Phillips, O. R. Miranda, C. C. You, V. M. Rotello and U. H. Bunz, *Angew Chem Int Ed Engl*, 2008, **47**, 2590-2594.

23. C.-C. You, O. R. Miranda, B. Gider, P. S. Ghosh, I.-B. Kim, B. Erdogan, S. A. Krovi, U. H. F. Bunz and V. M. Rotello, *Nat Nano*, 2007, **2**, 318-323.

24. Z. Yan, J. Li, Y. Xiong, W. Xu and G. Zheng, *Oncol Rep*, 2012, **28**, 1036-1042.

25. M. Khalilia, S. Chakraborty and M. Popescu, *BMC Med Inform Decis Mak*, 2011, **11**, 51.

26. E. Arigi, O. Blixt, K. Buschard, H. Clausen and S. B. Levery, *Glycoconj J*, 2012, **29**, 1-12.

27. T. Kodadek, *Chem Biol*, 2001, **8**, 105-115.

28. Y. Fei, Y. S. Sun, Y. Li, K. Lau, H. Yu, H. A. Chokhawala, S. Huang, J. P. Landry, X. Chen and X. Zhu, *Mol Biosyst*, 2011, **7**, 3343-3352.

29. T. Jombart, C. Collins, P. Solymos, I. Ahmed, F. Calboli and A. Cori, adegenet: an R package for the exploratory analysis of genetic and genomic data, http://adegenet.r-forge.r-project.org/).

30. R. D. C. Team, R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing, http://www.R-project.org/).

31. A. Liaw and M. Wiener, *R news*, 2002, **2**, 18-22.

Chapter 3

Optimisation of glycan-surface conjugation

3.1 Abstract

Carbohydrate microarrays are an important technique for the study of pathogens such as bacteria and their toxins and in the generation of inhibitory compounds for anti-adhesion therapy. Whilst carbohydrate microarrays are commercially available they are expensive and limited in the number of surfaces available. In this chapter, we examine two methods of surface functionalisation and adapt them in order to generate microwell based carbohydrate microarrays in a cheap and facile manner. Functionalisation of 96 well plates for lectin assays by using covalent attachment of unmodified glycans through the addition of cyanuric chloride to the surface was unsuccessful due to poor binding resolution. Attachment of reducing glycans to commercially available hydrazide funtionalised microwell plates was found to be a much easier technique and was shown to be successful through the use of a modified drop shape analysis technique and showed far better resolution for lectin binding.

3.2 Introduction

3.2.1 Carbohydrate Microarrays

Many microarrays are commercially available, for example the consortium for functional glycomics (CFG) have used glycans with an amino linker printed onto *N*-hydroxysuccinimide (NHS) coated glass to generate a vast database of binding profiles for hundreds of lectins and whole cell lines to an array of 200 glycans.[1] These carbohydrates include everything from simple mono- and di-saccharides to large and branched oligosaccharides.

A variety of different microarray platforms have been developed to tackle a number of different challenges including; the detection of food-based bacterial pathogens[2], the detection and typing of bacteria in blood samples[3], the identification of human and avian influenza strains[4] and identification of antibodies that protect against severe malaria.[5] Grabosch *et al.* used carbohydrate functionalised 96 well plates in order to test inhibitors of type 1 fimbriae-mediated bacterial using *Escherichia coli*.[6] Microarrays have also been used in the development of toxin resistant cell lines by selecting cells that lack the surface-bound carbohydrates necessary for lectin binding (or were more resistant to lectin binding).[7]

However, many of the microarrays currently available have limitations. Those that utilise monsaccharides either lack sensitivity due to the promiscuous nature of lectin binding at the monosaccharide level or are combined with expensive techniques such as surface plasmon resonance in order to overcome this limitation. The use of larger oligosaccharides in microarrays results in much better lectin specificity but these

compounds can be expensive or synthetically challenging, especially if the glycans require modification for surface attachment. As carbohydrate microarrays can play an important role in the detection of pathogens, the development of cheap and facile functionalisation techniques is crucial.

3.2.2 Surface functionalisation

In order to develop carbohydrate microarrays, carbohydrates have to be attached to a surface. There are many different methods for this and broadly speaking they can be split into four categories (Table 3.1).[8] There are many published examples of all of these methods for attachment and each category has advantages and limitations when compared to the others.

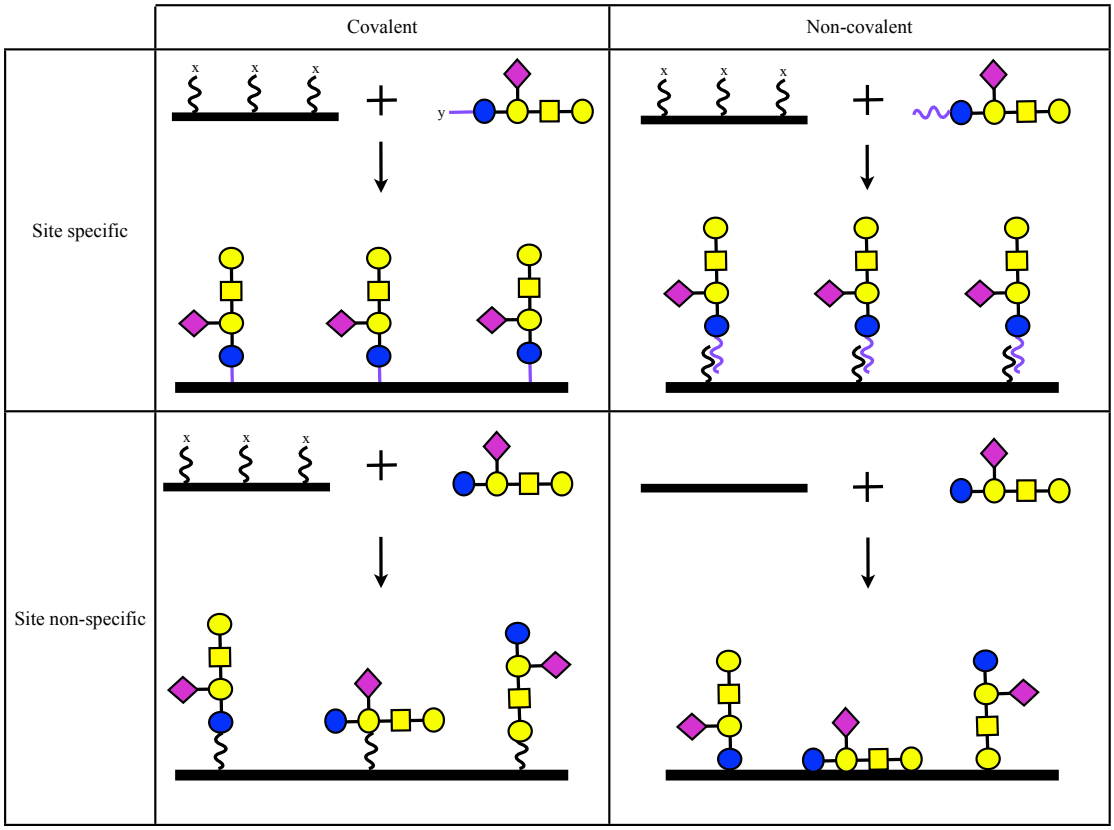


Table 3.1 Generalised methods of surface functionalisation in the development of carbohydrate microarrays. The pentasaccharide displayed is drawn using standard CFG nomenclature.

3.2.2.1 Site non-specific and non-covalent attachment

At first glance this is a very attractive technique as it allows the carbohydrates to be in their natural form and does not require any expensive or laborious preparation of the surface. Many examples of this type of method exist in the literature including the non-covalent absorption of polysaccharides to black polystyrene and the immobilisation of polysaccharides on nitrocellulose.[9, 10]

However, it has been observed that the immobilisation process is dependent on the molecular weight of the compound being immobilised, where immobilisation is most efficient for larger molecules.[9] This is a major drawback of this method as it means that the technique is only applicable to large polysaccharides and so smaller mono- or disaccharides cannot be probed. It also means that differential protein binding between polysaccharides may be caused by differences in polysaccharide immobilisation.

Non-covalent (or passive) absorption of polysaccharides onto a surface also has another drawback; deposition of the polysaccharide will be random and us such the molecule will be presented on the surface in a variety of orientations the majority of which will not be biologically relevant thus a lower signal may just be as a result of a low number of molecules presented in the biologically relevant conformation.[11]

3.2.2.2 Site non-specific and covalent attachment

This approach is appealing as it allows the use of unmodified glycans, which are bound to a modified surface. An example of this mechanism is the attachment of free

glycans to a surface modified with 4-azido-2,3,5,6-tetrafluorophenyl under UV light or the photochemical ligation of perfluorophenylazide derivatised carbohydrates to a poly(ethylene oxide) surface.[12, 13] Whilst several other examples of this technique exist in the literature, one of the main limitations of this technique is the site non-specific nature of the attachment, which means that the glycans can attach in any orientation to the surface.

In 2012, Liang and Chen described the attachment of unmodified mono- di- and polysaccharides to modified glass and gold slides.[14] The process they described initially involved the hydroxylation of either a glass or gold surface followed by addition of cyanuric chloride, this compound contains three chlorine atoms all of which can undergo hydrolysis and bind to hydroxyl containing molecules. The cyanuric chloride binds to the surface and then carbohydrates will then bind to the remaining binding sites of the compound in any orientation.[14] Whilst this means that any orientation of molecule can be presented on the surface, lectin specificity was retained implying that the natural presentation was present on the surface and this means that this technique may overcome the challenges of the other methods that utilise the site non-specific, covalent attachment of glycans.

3.2.2.3 Site specific and non-covalent attachment

Examples of this approach include the use of biotin-conjugated glycans coupled with a streptavidin surface,[15] the attachment of fluorous tagged carbohydrates to fluorous derivatised glass slides[16] and the attachment of carbohydrates conjugated to either bovine serum albumin or human serum albumin to epoxide functionalised microscope slides.[17]

Whilst this technique ensures that orientation of the glycan is controlled, it has the key disadvantage that it requires both a chemically modified surface and modified glycans, which can be expensive, and synthetically challenging, thus limiting the use of techniques based on this approach. Also the storage of protein functionalised surfaces is crucial to avoid protein denaturation and loss of surface functionalisation.

3.2.2.4 Site specific and covalent attachment

This method is the most desirable as it ensures that all glycans are presented in the same orientation and whilst many methods require both surface and glycan modifications there are some that can utilise unmodified glycans. Many methods that involve the modification of the surface and the glycans have been described, for example, the attachment of modified carbohydrates in the form of glycosylamines or malemide conjugated carbohydrates onto thiol modified glass slides.[8] This approach is also utilised by the CFG in the production of a carbohydrate microarray, which uses glycans with an amino linker, which are covalently attached to NHS coated glass slides.[1, 15, 18]

However these methods require the modification of carbohydrates which is challenging, time consuming and often requires multiple purification steps.[19] Site-specific and covalent immobilisation of chemically unmodified carbohydrates on a modified surface has also been described and this eliminates the need for modification of the glycan thus making it the most desirable technique. However, this technique is also the most challenging and few examples exist in the literature. Reductive animation is the most common choice for attaching unmodified carbohydrates to a

surface, this involves the use of a hydrazide functionalised surface which is commercially available in the form of 96 well plates.[20] The reducing end of the sugar reacts with the hydrazide on the surface, after which the Schiff base is reduced. One example of this method is the attachment of unmodified oligosaccharides to hydrazide modified self-assembled monolayers on a gold surface.[11] This method has also been used in the attachment of unmodified glycans to hydrazide functionalised polymers.[21] Whilst this technique allows the site-specific attachment of unmodified glycans (and thus allows biological orientation to be conserved) its key disadvantage is that only reducing sugars can be attached to the surface, which limits the number of glycans that can be analysed.

Recently a new technique has come to light in which unmodified carbohydrates were attached to a silicon nanowire (SiNW). This has the advantage that interactions can be detected in a label free manner and real-time binding can be observed. The technique was also significantly more sensitive than any other techniques currently available (the authors were able to detect lectin binding down to 100 $fg.mL^{-1}$). Currently the synthesis of these SiNW biosensors is challenging, making the technique inaccessible, but there is the potential to commercially produce arrays of these sensors, which would greatly improve the sensitivity of protein-carbohydrate interaction detection that is currently observed.[22]

3.2.3 Surface functionalisation considerations

As there are many techniques for modifying the surface it should be noted that initial modification of the surface before the addition of the carbohydrates is important and should be kept constant between different carbohydrates as the linker attaching the

carbohydrate to the surface can modulate lectin binding. This is also occasionally seen in nature so is not always a negative thing; glycophospholipids are carbohydrates attached to long fatty acid chains, their binding is often examined using only the carbohydrate sequences but it has been shown that in some instances the fatty acid chain plays a role in determining binding strength in biological conditions.[23]

Linker length is also crucial as it has been shown that the length of the chain on the surface can have a massive impact on binding; the initial chain length should be between 10 and 14 carbons long to allow binding of the lectin (Figure 3.1).[24] Not only does the spacer length and nature play a crucial role in lectin binding but surface density is also crucial, if the glycans are too close together then a reduction in lectin binding is observed but if they are too far apart then the lectins cannot utilise multivalent binding which would result in the protein-carbohydrate interactions being very weak. Multivalent binding is utilised in nature to overcome the weak protein-carbohydrate interaction and most lectins have multiple binding domains, it has been shown that the simultaneous occurrence of two binding events can increase the strength of an interaction by between 100 and 10,000-fold.[20]

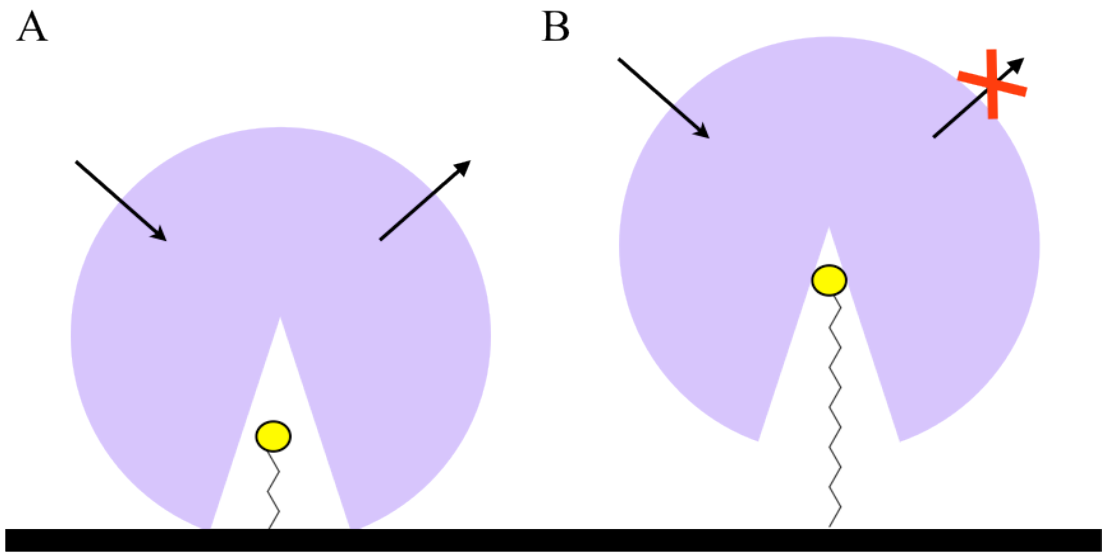


Figure 3.1 Surface functionalisation where the glycan is presented at less than 10 carbon atoms' distance from the surface is insufficient for lectin binding (A) where as at 10 carbons or more, from the surface lectin binding can occur (B).

Here we describe the development of a facile surface functionalisation technique that allows the use of unmodified carbohydrates to be added to a modified surface for the production of a microwell based carbohydrate microarray. Modifications of the method for the attachment of native glycans to a cyanuric chloride functionalised surface or the addition of unmodified glycans to a commercially available hydrazide functionalised 96-well plate were both explored in detail.

3.3 Results and Discussion

The aim of this work was to develop a surface functionalisation technique that allows the modification of surfaces with unmodified glycans for the production of a 96-well plate based carbohydrate microarray.

3.3.1 Solvent Testing

As the method described by Liang and Chen allowed the addition of any unmodified glycan this technique was very attractive and so work was done to attempt to modify this method for 96-well plates. The initial step of the Liang and Chen method involves the formation of a hydroxylated surface.[14] The aim was to achieve this through non-specific absorption of the carbohydrate GM1 to a high-binding plate. However the next step in the process involves adding cyanuric chloride (Figure 3.2) dissolved in acetone, which would dissolve the polystyrene high-binding plate. As such diethyl ether and lutidine were tested as potential solvents for the cyanuric chloride. Diethyl ether was able to dissolve the cyanuric chloride but was found to rapidly dissolve the high-binding plate.

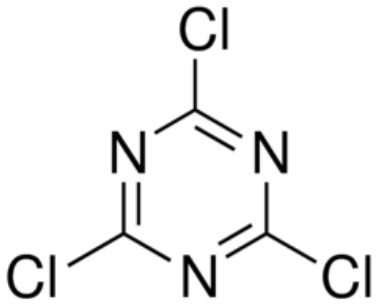


Figure 3.2 Structure of cyanuric chloride.

Lutidine was found to both dissolve the cyanuric chloride and be relatively stable in the high-binding plates. The cyanuric chloride solution has to be incubated in the

wells for approximately 6 hours and this resulted in the formation of a tar like substance that proved impossible to remove from the wells and eventually corroded the bottom. The attempt to find another solvent to dissolve the cyanuric chloride was abandoned and the plate type was changed.

3.3.2 Hydroxylation of Polypropylene Surface with GM1

Polypropylene 96-well plates were then used as the surface for modification. This material has the benefit in that it does not react with the acetone that the cyanuric chloride is dissolved in. The polypropylene plates were functionalised first with GM1 to produce a hydroxylated surface and then with cyanuric chloride followed by a variety of glycans.

In order to confirm surface functionalisation, polypropylene slides were also functionalised in the same manner (initially with GM1 then cyanuric chloride followed by a glycan). These slides were then analysed using drop shape analysis (DSA) for changes in the nature of the surface angle. DSA involves adding a droplet of water to the surface and then using a microscope to generate an image and measure the contact angle of the water droplet. Changes in the nature of the surface would indicate that surface functionalisation had occurred. As can be seen by Figure 3.3 there is a significant difference in the hydrophobicity of the surfaces at each step of the functionalisation indicating a change in the chemical composition of the surface.

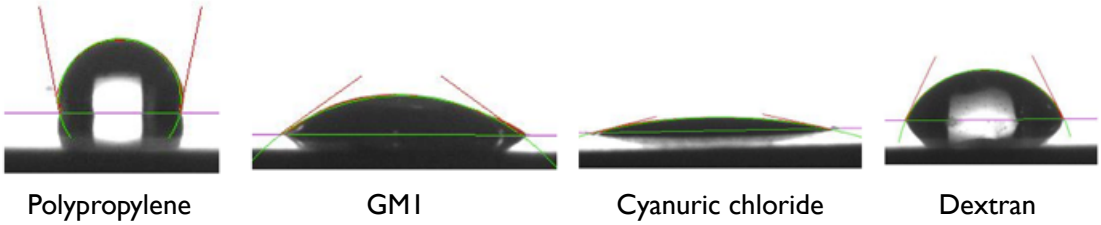


Figure 3.3 Polypropylene slides were prepared for DSA to determine the characteristics of the surface. There was a difference in the hydrophobicity of the surface between each stage of the functionalisation indicating a change in the nature of the surface.

After surface functionalisation had been confirmed on the slides the surface specificity was tested using fluorescently labelled Concanavalin A (Con A) and Peanut Agglutinin (PNA) in the 96-well plates. These lectins have been extensively studied and their binding specificity has been well characterised; PNA is a β-D-galactose binding lectin that also binds lactose whereas ConA is a α-D-mannose binding lectin that also has some affinity towards α-D-glucose. As such the following four sugars were functionalised onto the surface of the modified plates for testing; D-glucose, D-galactose, D-lactose and D-mannose.

Comparison of differential PNA binding between these four sugars revealed there to be a significant difference between binding to mannose compared to binding to galactose but the difference was very small and there was little difference between maximum binding seen. This indicates the presence of sugars on the surface but the poor resolution limits the usefulness of this technique.

3.3.3 Surface Blocking

Several blocking agents were trialled on a variety of the surfaces in order to try and improve the specificity of the surface. Initially ethanolamine was used as described by Liang and Chen,[14] and did appear to produce some improvement in surface specificity (Figure 3.4) but, the improvement was not reproducible. Indeed in some instances the addition of ethanolamine prevented any protein binding.

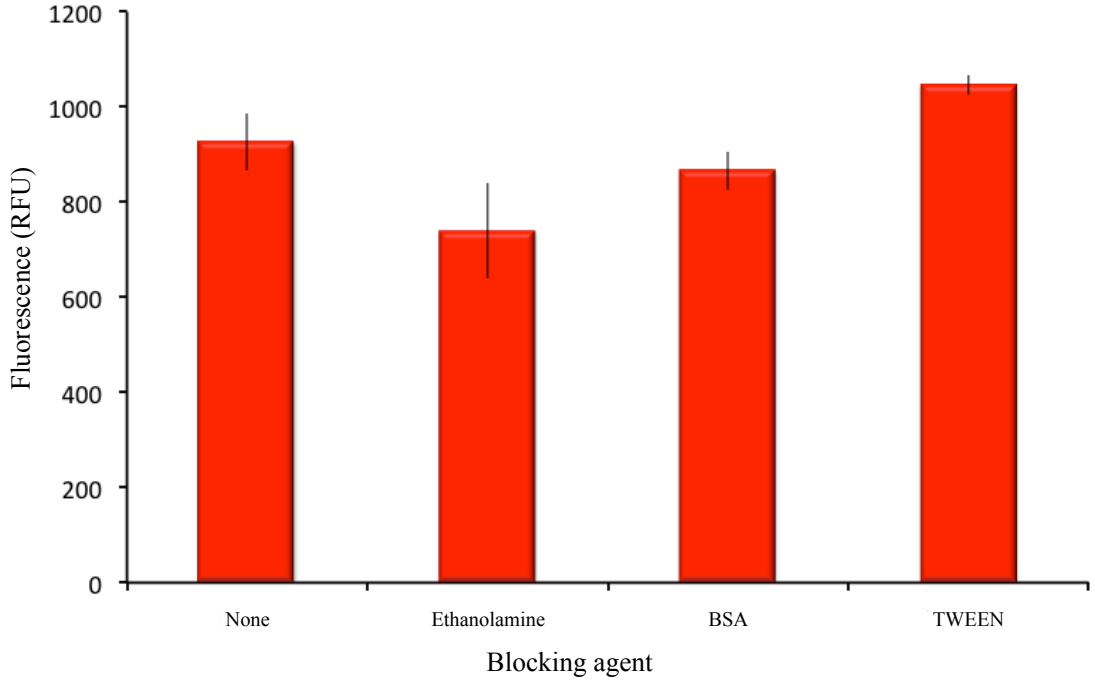


Figure 3.4 Blocking of a mannose surface with a variety of blocking agents prior to incubation with fluorescently labelled peanut agglutinin (to represent non-specific binding). Bars represent the average of three measurements and the error bars are the standard error.

Bovine serum albumin (BSA) was then utilised as it is a cheap protein source available in large quantities and often used in blocking non-specific binding in biological assays. There did not appear to be much improvement in the binding profiles (Figure 3.4) produced and this material would be unfavourable as a blocking

reagent as interaction between BSA and carbohydrates has been poorly studied and there is some evidence to show that BSA will bind certain carbohydrates thus competing with any lectins being studied.

Tween was also tested as a potential blocking agent as it is often used in carbohydrate microarrays and has the benefit that it is not a protein and does not interact with the carbohydrates on the surface at all but it resulted in no improvement of the specificity of the surface (Figure 3.4).

3.3.4 Hydroxylation of Polypropylene Surface with Octanol

As no reproducible improvement was seen after use of several blocking agents, a new method of surface hydroxylation was trialled. The aim of this was to increase the density of hydroxyl compounds on the surface, which would then increase the cyanuric chloride functionalisation and subsequently the number of bound sugar residues. Greater sugar functionalisation should reduce non-specific binding whilst increasing specific binding thus resulting in better resolution.

A paper by Grabosch *et al.* describes the use of compounds with long carbon chains for functionalisation of polypropylene.[6] It was therefore reasoned that long chain alcohols could be used to introduce hydroxyl functionality to the polypropylene plates.

Alcohols have several advantages; they will be more readily bound to the polypropylene than the GM1 resulting in greater initial functionalisation and consequently greater functionalisation at later steps. They also do not interact with the

lectins for testing (unlike GM1 which is the target carbohydrate for cholera toxin lectin and has a terminal galactose that has been shown to interact with galactose binding lectins), they are also cheap and readily available in large quantities.

Initially, octanol was added to the surface for the hydroxylation stage before the cyanuric chloride step and the addition of the sugars. However, there was found to be no lectin binding upon testing for the presence of the sugar. This does not necessarily mean that the sugar is not present as it has previously been shown that a chain length of less than ten is not optimal for lectin binding to surface carbohydrates as the carbohydrate would not have the flexibility to enter the binding pocket.

3.3.5 Hydroxylation of Polypropylene Surface with Dodecan-1-ol

As octanol hydroxylation was insufficient for lectin binding the chain length of the alcohol was increased and dodecan-1-ol was used for the initial functionalisation step before addition of the cyanuric chloride solution and the sugars as shown in Figure 3.5.

To test surface characteristics, polypropylene slides were prepared by initially coating them in dodecan-1-ol before further functionalisation with the cyanuric chloride and the following sugars; sucrose, galactose, lactose and cellobiose. As well as this, small 8 x 8 mm polypropylene chips were functionalised in the same way for analysis by variable angle XPS.

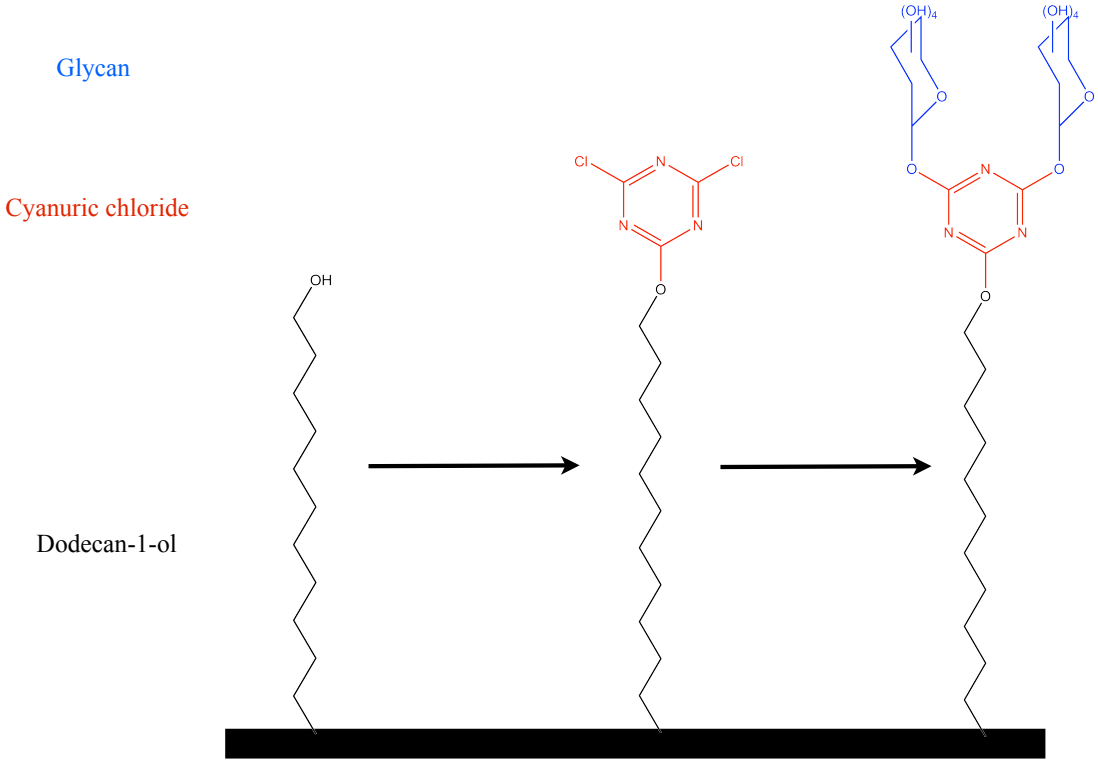


Figure 3.5 Schematic showing the three steps of surface functionalisation based on the method described by Liang and Chen.[14] Initially a hydroxylated surface was produced through non-specific, non-covalent addition of dodecan-1-ol to the surface. Cyanuric chloride was then bound to the hydroxyl groups on the surface allowing addition of carbohydrates in the final step.

Successful surface modification was determined using variable angle XPS and DSA. DSA showed there to be a significant change in contact angle between all of the steps of the modification process indicating a change in the hydrophilicity of the surface (Figure 3.6).

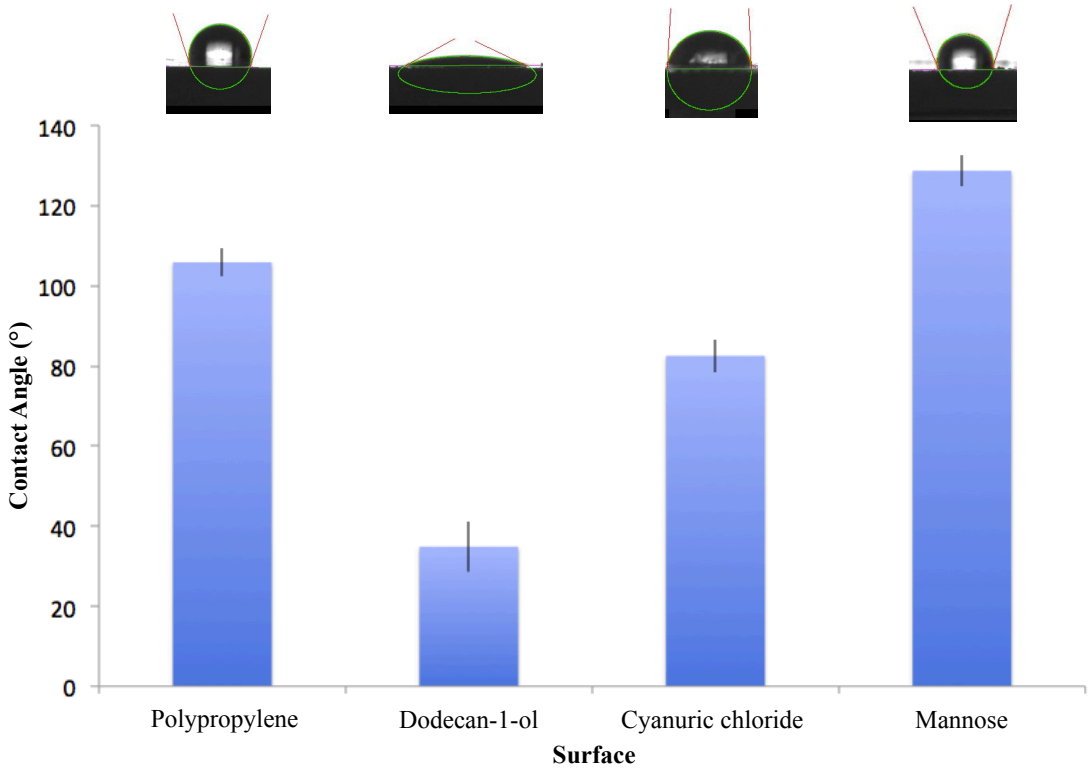


Figure 3.6 A slide was prepared for each step of the functionalisation and DSA was performed. Each measurement was repeated three times and the average is shown, the error bars represent the standard deviation. As can be clearly seen there is a change in surface hydrophobicity at each step indicating that functionalisation has been successful.

The variable angle XPS results also confirm the functionalisation of the surface (Figure 3.7). The polypropylene and dodecan-1-ol surfaces only contained carbon and oxygen (the large change in contact angle in the DSA confirms the presence of the dodecan-1-ol). After addition of the cyanuric chloride both nitrogen and chlorine can be found on the surface (indicating the presence of the cyanuric chloride) the higher ratios of the nitrogen and chlorine fractions at 60° also indicates that they are more prevalent on the top layer of the surface and the carbon and oxygen fraction underneath. After the addition of the sugar, the loss of the chlorine whilst no loss of

nitrogen indicates the loss of the chlorine groups after reaction with the carbohydrates. This is further confirmed by the fact that the nitrogen fraction is under the carbon and oxygen fractions indicating the presence of the sugar on the surface.

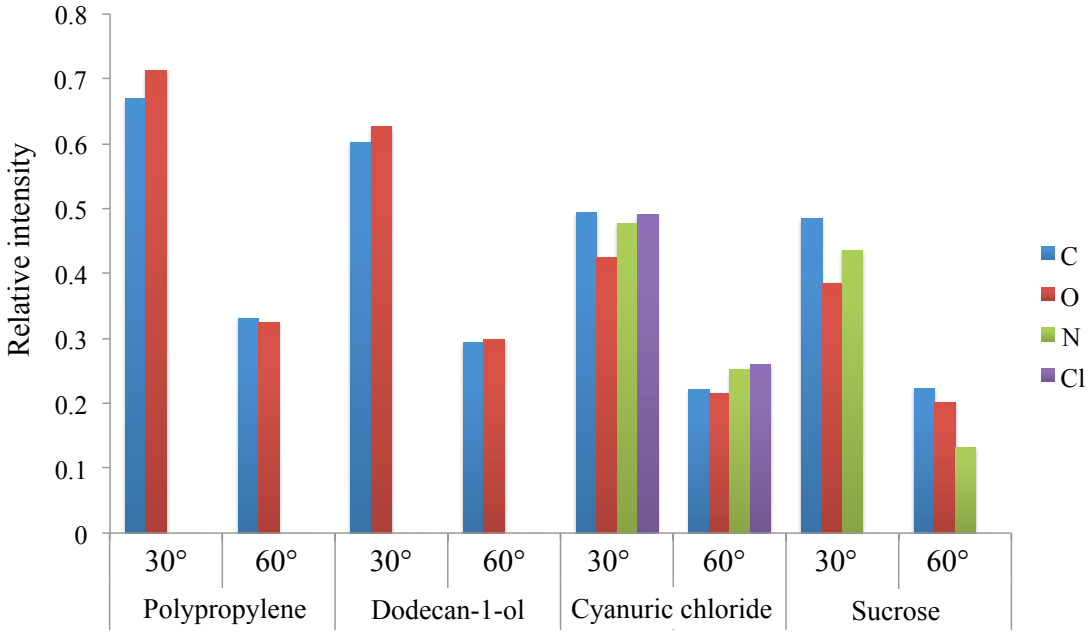


Figure 3.7 Variable angle XPS was performed on chips after each step of functionalisation. The bars represent one reading.

After the surface had been characterised, the surface specificity was tested using the same well-characterised lectins as previously described. It was found that, while again there was differential binding, there was little difference between maximum binding and no binding (Figure 3.8) and despite attempts at blocking the rest of the surface with either BSA (10 $mg.mL^{-1}$) or TWEEN (5 %) the difference between the maximum and no binding did not improve.

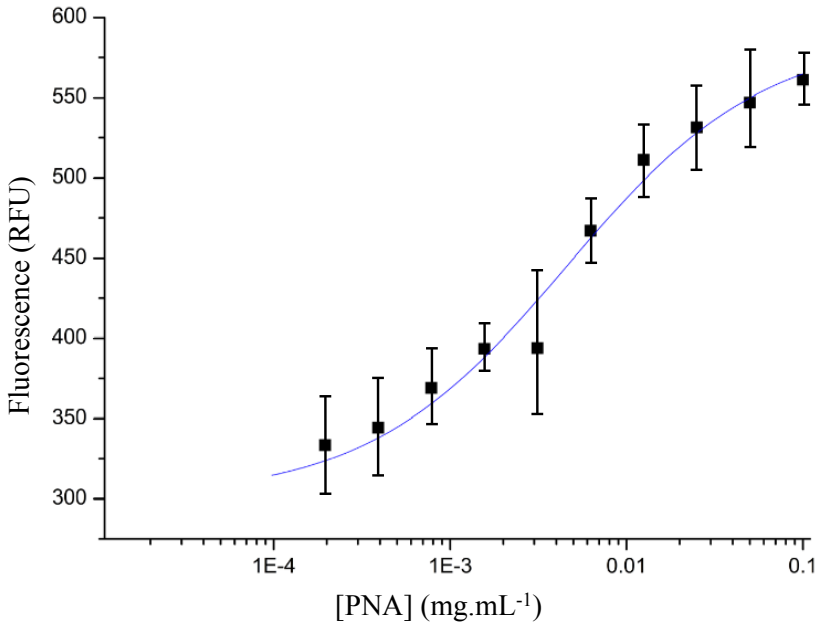


Figure 3.8 Addition of a serial dilution of PNA to a galactose functionalised surface revealed there to be little difference between maximum binding (550 RFU) and no binding (350 RFU).

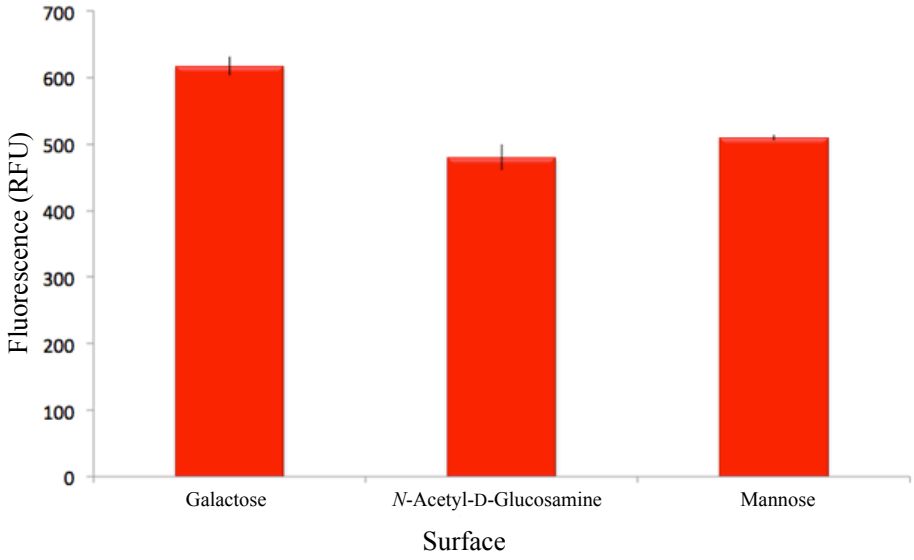


Figure 3.9 Functionalisation of the surface with 3 different carbohydrates (galactose, *N*-acetyl-D-glucosamine and mannose) resulted in differences in PNA binding with the highest binding shown on the galactose surface (as would be expected for a galactose specific lectin).

Binding was shown to be specific as binding of PNA on different sugar surfaces was shown to be distinct, thus indicating the presence of the carbohydrate on the surface (Figure 3.9). This showed that whilst there was surface functionalisation it was not highly functionalised. It may be that the surface is functionalised but not densely enough to allow multivalent binding so only weak interactions are seen but, it may be that this technique is applicable to bacteria as they are larger and so can undergo multivalent binding as they can bridge the gap between sugar units (Figure 3.10).

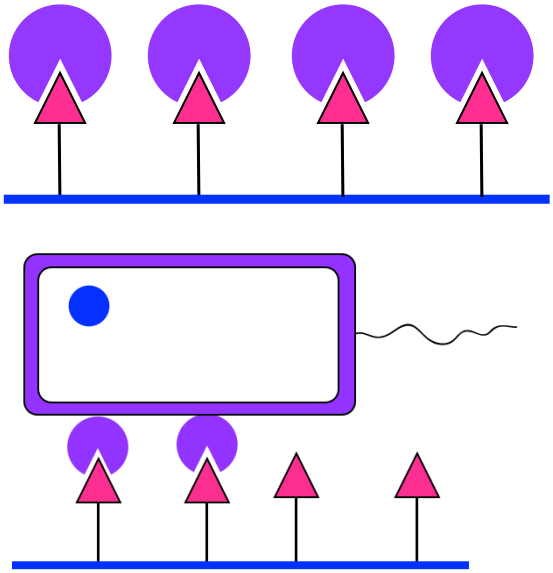


Figure 3.10 It is hypothesised that the functionalisation of the surface with carbohydrates may be insufficient for lectins to show multivalent binding due to too great a distance between sugar units thus resulting in only weak binding. Bacteria may be able to bridge this gap as they are larger and thus can still undergo multivalent binding. This means that the technique may be applicable as a bacterial binding assay.

Despite discontinuing this technique in favour of another surface functionalisation method its versatility has been shown through the analysis of the surface; it could be applied either to 96 well plates or a sheet of polypropylene can be coated in the same way and can be cut to almost any size allowing the production of samples that can be

analysed in many different ways including microscope slide sized samples for DSA and small chips for variable angle XPS.

3.3.6 Hydrazide Functionalised 96-Well Plates

Commercially available hydrazide functionalised 96-well plates can be further functionalised with sugars (Figure 3.11A).[21] Owing to the high numbers of hydrazide groups on the well surface this technique might be expected to result in improved carbohydrate functionalisation and thus greater lectin specificity. This technique has previously been well-characterised and results in attachment of glycans predominantly in the ring closed (pyranose) β-D-anomeric form.[21] Modification of the surface in this manner has a disadvantage over the previously mentioned technique in that only reducing glycans can be attached to the surface.

Functionalisation was confirmed by using a modified drop shape analysis technique, which characterised the spreading of a water droplet (containing resorufin dye) on the surface before and after functionalisation with either galactose or glyceraldehyde (as a positive control) to gain an estimate of the surface hydrophilicity. As can be seen by Figure 3.11B, the surface hydrophilicity was significantly increased after surface functionalisation with galactose (when compared to the unfunctionalised hydrazide surface) and this was even further increased for the surface functionalised with glyceraldehyde.

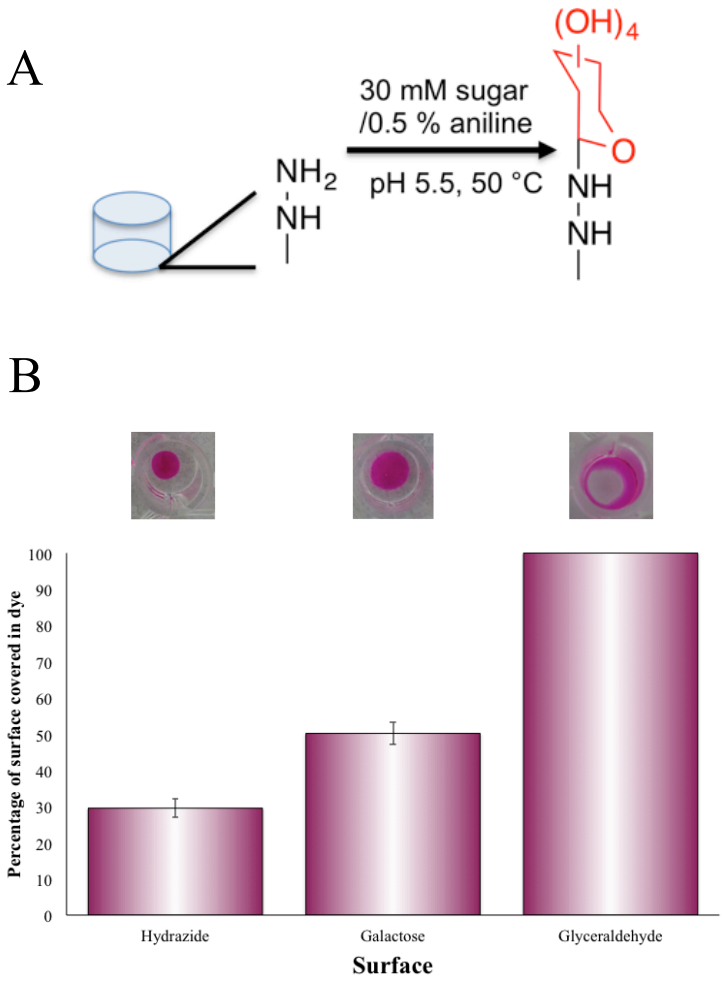


Figure 3.11 Glycosylated 96-well plates were generated through the reaction of commercially available 96-well plates with glycans (A). Functionalisation was then confirmed using a modified DSA technique where surface spreading of a droplet of water was determined at every step (B). The bars represent the average of 3 independent replicates and the error bars represent the standard deviation. Representative images of droplet spreading shown above the bar.

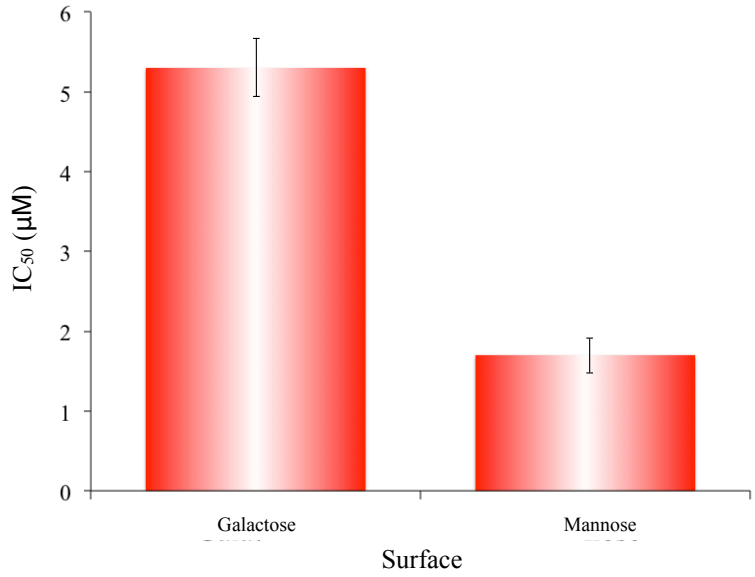


Figure 3.12 A serial dilution of fluorescently labelled Con A was added to both a galactose and a mannose functionalised surface and fluorescence was measured. Con A was found to bind the mannose surface (1.7 μM) more strongly than the galactose surface (5.3 μM). This IC_{50} value is determined from a curve fitting to an average binding curve based on three triplicate measurements and error bars represent the standard error.

Whilst the change in hydrophilicity indicates a change in the nature of the surface that can be explained by the presence of galactose on the surface, further probing with lectins was then performed in order to further characterise the surface. Fluorescently labelled Con A (α-D-mannose specific) and PNA (β-D-galactose specific) were incubated with a mannose or a galactose functionalised surface and Con A was found to bind the mannose surface more strongly than the galactose surface (Figure 3.12) and the reverse was also found to be true for PNA (Figure 3.13). This selective lectin binding indicated the success of the surface functionalisation step.

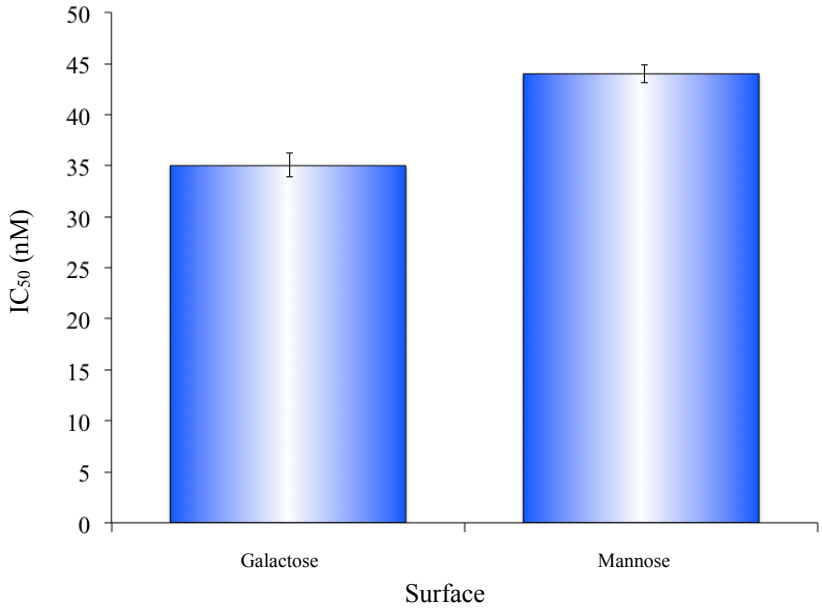


Figure 3.13 A serial dilution of fluorescently labelled PNA was added to both a galactose and a mannose functionalised surface and fluorescence measured. PNA was found to bind the galactose surface (35 nM) more strongly than the mannose surface (44 nM). This IC_{50} value is determined from a curve fitting to an average binding curve based on three triplicate measurements and error bars represent the standard error.

3.3.7 Hydrazide surface blocking

The blocking of microarray surfaces in order to prevent non-specific adhesion is routinely performed, as this would increase the specificity of the surface. Functionalised surfaces are routinely blocked with BSA and various other agents such as Tween to prevent non-specific binding of proteins to the underlying surface. However, little work has been done to identify possible interactions between the BSA and the carbohydrates on the surface or the lectins being tested. Blocking should only really be performed if it will prevent a significant amount of non-specific carbohydrate adhesion.

Addition of a serial dilution of fluorescently labelled BSA to both a carbohydrate functionalised surface and an unfunctionalised surface made of the same material highlighted that at the concentrations of lectin being analysed there was no significant non-specific adhesion on the functionalised surfaces (highlighted by the lack of FITC-BSA binding on these surfaces compared to a non-functionalised surface, Figure 3.14). Owing to the lack of knowledge about interactions the BSA may have with the lectins being tested and the limited amount of non-specific binding that could occur at the concentrations of protein being used, the hydrazide surfaces were not blocked during further experiments.

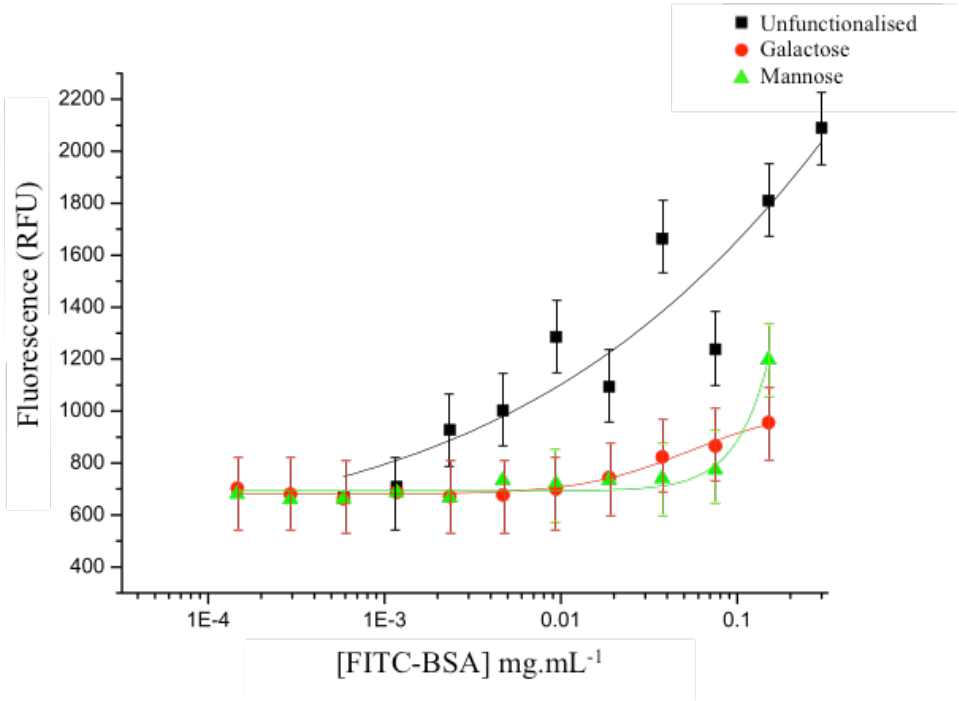


Figure 3.14 Addition of a serial dilution of FITC labelled BSA to a hydrazide surface and surfaces functionlised with two different carbohydrates highlights that up until about 0.01 $mg.mL^{-1}$ of protein there is no significant non-specific binding of the BSA to functionalised surfaces when compared to unfunctionalised surfaces. Points represent individual measurements and error bars represent machine error.

3.4 Conclusion

Carbohydrate microarrays play a crucial role in the identification and classification of toxins and whole bacteria and they play a pivotal role in the testing of inhibitory compounds. Whilst commercial carbohydrate microarrays are available, they are expensive and the surfaces are limited to those commercially available and thus it is advantageous to use a surface functionalisation technique in order to be able to generate the surfaces of interest.

Many surface functionalisation techniques exist and have been used in the generation of carbohydrate microarrays. The best techniques use unmodified carbohydrates which are attached covalently in a site-specific manner to a modified surface, this is because unmodified carbohydrates are significantly cheaper than modified ones and site-specific attachment results in one orientation of the carbohydrate being present on the surface. Spacers and linkers between the surface and the carbohydrate can impact significantly on binding so, ideally, these should be the same for all the carbohydrates being attached.

Attempts to modify the Liang and Chen method for attachment of any unmodified glycan to a surface functionalised with cyanuric chloride residues for use in 96- well plates proved unsuccessful. The technique was shown to be adaptable to a number of surface analysis techniques through the use of polypropylene slides and chips and surface functionalisation was confirmed through the use of DSA and variable angle XPS. The resolution of lectin binding proved to be insufficient despite changing both the hydroxylation steps used and testing a number of surface blocking techniques.

Commercially available hydrazide functionalised microwell plates were employed to generate 96 well plate based carbohydrate microarrays. This method of attachment has been very well characterised previously and it is known that reducing glycans immobilised through this method are predominantly in the β-D-anomeric pyranoside form and all are attached in identical orientation through the reducing terminus.[21] Surface functionalisation was confirmed using a modified DSA technique and the carbohydrate specific adhesion of two well-characterised lectins. This simple and cheap method for generating carbohydrate microarrays represents a method through which any (reducing) glycan of interest can be immobilised for testing of either lectin binding, the binding of whole cells or the testing of inhibitory compounds.

3.5 Materials and methods

Materials: All chemicals were used as supplied unless otherwise stated. High binding clear flat-bottomed polystyrene microplates were purchased from Greiner Bio-one. Polypropylene flat-bottomed microplates, polypropylene sheet, cyanuric chloride, acetone, aniline (99.5 % ACS reagent), monosialylganglioside GM1 from bovine brain, D- (+)cellobiose, D- (+)galactose, D- (+)mannose, DL- glyceraldehyde, bovine serum albumin, HEPES, lactose, α-D-glucose, *N*-Acetyl-D-Glucosamine, resorufin, octanol, dodecan-1-ol, ethanol, phosphate buffered saline (PBS) tablets, sodium acetate anhydrous, acetic acid glacial, fluorescein isothiocyanate isomer I (≥ 90 % HPLC grade), PD10 desalting columns, Tween20 and sucrose were all purchased from Sigma-Aldrich. FITC labelled Concanavalin A and Peanut Agglutinin from *Arachis hypogaea* were purchased from Vector Labs. 100 mM acetate buffer (pH 5.5) with 1 mM aniline was prepared in 200 mL of milliQ water (with a resistance >19 mΩ.cm at 25 °C). 10 mM HEPES buffer was prepared in 200 mL of milliQ water (with a resistivity >18 $M\Omega.cm^{-1}$ at 25 °C) with added 0.15 M NaCl, 0.1 mM $CaCl_2$ and 0.01 mM Mn^{2+} (pH 7.5) and is now referred to as HEPES in the text.

GM1 hydroxylation: 100 μL of GM1 solution (0.1 $mg.mL^{-1}$ in PBS) was added to every well of the 96 well plate before incubation for 16 hours at 4 °C. Unbound GM1 solution was then removed and every well washed three times with MilliQ water (resistivity >18 $M\Omega.cm^{-1}$ at 25 °C) before drying under N_2 stream. Functionalised 96 well plates were either used directly or stored at -20 °C prior to use.

Cyanuric chloride and sugar functionalisation: Hydroxylated surfaces were further functionalised with cyanuric chloride by addition of 100 μL of cyanuric chloride

solution (100 mM in acetone) to every well and incubation at 4 °C for 6 hours. Unbound solution was then removed and wells were extensively washed with acetone and then a further three times with HEPES buffer. 100 μL of glycan solutions (30 mM for small glycans or 3 $mg.ml^{-1}$ for polysaccharides, in water adjusted to pH 9 with 1 M NaOH) was added to every well before incubation for 10 hours at room temperature. Unbound solution was then removed and wells washed three times with MilliQ water (resistivity >18 $M\Omega.cm^{-1}$) before drying under N_2 stream. Functionalised 96-well plates were then either used directly or stored at -20 °C prior to use.

Cyanuric chloride blocking: After preparation of a mannose functionalised surface the surfaces were further incubated with either bovine serum albumin (at 10 $mg.mL^{-1}$ in PBS), 5% Tween (in PBS) at room temperature for 12 hours. Blocking with ethanolamine was performed by incubating with a 1 M aqueous solution of ethanolamine (adjusted to pH 9) for 3 hours. After surface blocking, the wells were incubated with fluorescently labelled PNA (0.1 $mg.mL^{-1}$in HEPES) at 37 °C for 30 minutes before removal of unbound protein solution and extensive washing with HEPES. Fluorescence was then measured using a BioTek Synergy HT multi-detection microplate reader and Gen5 software with excitation and emission wavelengths of 485 and 528 nm respectively.

Drop shape analysis: Water contact analysis experiments were performed at room temperature using a Krüss drop shape analysis system (DSA100) equipped with a moveable sample stage and a microliter syringe. Samples were placed on the sample stage using tweezers and aligned within the field of view of the camera. Then the microliter syringe was advanced until a drop of 10 μL of deionized water was formed

and suspended at the end of the needle. The sample stage was then elevated, until the sample touched the bottom of the drop, causing it to detach from the needle and form on the surface. The sample stage was then moved back to the initial position and an image of the droplet was recorded. The baseline contact angles were then computed from the image.

Octanol and Dodecan-1-ol hydroxylation: 100 μL of either octanol or dodecan-1-ol solution (10 mM in ethanol) was added to every well of a polypropylene plate before incubation at room temperature for 24 hours. Any remaining unbound solution was then removed before washing of every well with ethanol and twice with PBS.

Variable angle XPS: 8 x 8 mm polypropylene chips were prepared at various stages of functionalisation including unfunctionalised, hydoxylated with dodecan-1-ol, after addition of cyanuric chloride and then after addition of a glycan. All variable angle XPS was performed by Stephen Edmondson, School of Materials Science at the University of Manchester.

Lectin binding assay: Surfaces were incubated with either fluorescently labelled Concanavalin A or Peanut Agglutinin in a serial dilution from a starting concentration of 0.1 $mg.mL^{-1}$ (in HEPES) were incubated for 37 °C for 30 minutes before removal of unbound protein solution and extensive washing with HEPES. Fluorescence was then measured using a BioTek Synergy HT multi-detection microplate reader and Gen5 software with excitation and emission wavelengths of 485 and 528 nm respectively.

Hydrazide functionalisation: 100 μL of 30 mM glycan solutions (in 100 mM acetate buffer with 1 mM aniline, pH 5.5) was added to each well of a Carbo-BIND^{TM} 96-well plate. Plates were then covered in foil and incubated at 50 °C for 24 hours. After incubation, any unbound solution was removed and the well washed thoroughly with PBS three times. Plates were then either used immediately or stored at -20 °C prior to their use.

Modified drop shape analysis: 10 μL drops of resorufin (0.01 mg.mL^{-1} in milliQ water) were added to hydrazide surfaces prior to functionalisaton and after functionalisation with either galactose or glyceraldehyde. A digital camera was then used to take a photograph (Figure 3.15A) and the image was then converted into a Hue, Saturation, Brightness (HSB) stack in ImageJ. The area of the well in pixels was determined initially by drawing a region of interest around the well (Figure 3.15B). A region of interest was then drawn around the droplet and the area that the droplet occupied (Figure 3.15C) was compared to the area of the well to determine the percentage of the surface covered by the droplet. The greater the hydrophilicity of the surface, the more the droplet will spread and the greater the percentage of the well area covered by the droplet.

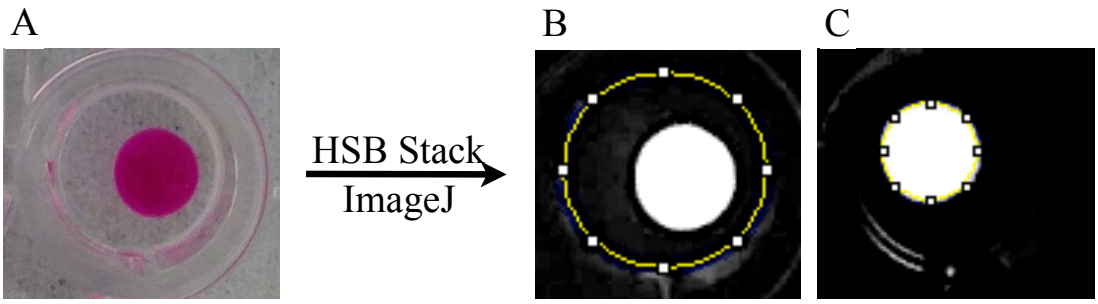


Figure 3.15 A digital photograph is taken of the 10 μL dye droplets on the well surface (A), this is then uploaded into ImageJ and converted into a HSB stack. A region of interest is drawn around the entire well (B) and the area in pixels is measured. A second region of interest is then drawn around the droplet (C) and the area occupied by the droplet in pixels is measured. The two values are then compared to determine the percentage of the surface covered by the dye droplet.

BSA labelling: A stock solution of 1 $mg.mL^{-1}$ BSA in PBS was prepared and 30 μL of fluorescein isothiocyanate isomer I (at 10 $mg.mL^{-1}$ in DMSO) was added to every 1 mL of stock solution. This was then incubated at room temperature in the dark for 1 hour, with continuous stirring. Before use, the solution was washed through a PD10 desalting column in order to remove unlabelled protein and unbound FITC.

Non-fouling surface analysis: 100 μL of FITC-BSA was added as a serial dilution (from 1 $mg.mL^{-1}$ in PBS) to the wells of either an unfunctionalised hydrazide surface, or those further functionalised with either D-galactose or D-glucose. Fluorescence was then measured using a BioTek Synergy HT multi-detection microplate reader and Gen5 software with excitation and emission wavelengths of 485 and 528 nm respectively.

Data analysis: All protein-binding curves were plotted in OriginPro (version 8.6) and the built in nonlinear curve-fitting tool was used to fit a logistic curve to the data and to calculate IC_{50} values. All other graphs were also plotted in OriginPro and averages and standard deviations were calculated using the open source statistical package R (version 3.1.3).[25]

3.6 References

1. CFG, http://www.functionalglycomics.org.
2. E. Bulard, A. Bouchet-Spinelli, P. Chaud, A. Roget, R. Calemczuk, S. Fort and T. Livache, *Biodevices*, 2015, DOI: 10.5220/0005254901210126.
3. M. D. Disney and P. H. Seeberger, *Chem Biol*, 2004, **11**, 1701-1707.
4. J. Stevens, O. Blixt, T. M. Tumpey, J. K. Taubenberger, J. C. Paulson and I. A. Wilson, *Science*, 2006, **312**, 404-410.
5. P. H. Seeberger and D. B. Werz, *Nature*, 2007, **446**, 1046-1051.
6. C. Grabosch, K. Kolbe and T. K. Lindhorst, *ChemBioChem*, 2012, **13**, 1874-1879.
7. D. Wang, *Proteomics*, 2003, **3**, 2167-2175.
8. S. Park and I. Shin, *Angew Chem Int Ed*, 2002, **41**, 3180-3182.
9. D. Wang, S. Liu, B. J. Trummer, C. Deng and A. Wang, *Nat Biotech*, 2002, **20**, 275-281.
10. W. G. Willats, S. E. Rasmussen, T. Kristensen, J. D. Mikkelsen and J. P. Knox, *Proteomics*, 2002, **2**, 1666-1671.
11. Z.-l. Zhi, A. K. Powell and J. E. Turnbull, *Anal Chem*, 2006, **78**, 4786-4793.
12. A. Tyagi, X. Wang, L. Deng, O. Ramström and M. Yan, *Biosens Bioelectron*, 2010, **26**, 344-350.
13. Z. Pei, H. Yu, M. Theurer, A. Waldén, P. Nilsson, M. Yan and O. Ramström, *ChemBioChem*, 2007, **8**, 166-168.
14. K. Liang and Y. Chen, *Bioconjugate Chem*, 2012, **23**, 1300-1308.
15. R. Raman, M. Venkataraman, S. Ramakrishnan, W. Lang, S. Raguram and R. Sasisekharan, *Glycobiology*, 2006, **16**, 82R-90R.

16. S. K. Mamidyala, K.-S. Ko, F. A. Jaipuri, G. Park and N. L. Pohl, *J Fluorine Chem*, 2006, **127**, 571-579.

17. J. C. Manimala, T. A. Roach, Z. Li and J. C. Gildersleeve, *Angew Chem Int Ed*, 2006, **118**, 3689-3692.

18. H. O. Yosief, A. A. Weiss and S. S. Iyer, *ChemBioChem*, 2013, **14**, 251-259.

19. I. Shin, S. Park and M. r. Lee, *Chem Eur J*, 2005, **11**, 2894-2901.

20. X. Zeng, C. A. Andrade, M. D. Oliveira and X.-L. Sun, *Anal Bioanal Chem*, 2012, **402**, 3161-3176.

21. K. Godula and C. R. Bertozzi, *J Am Chem Soc*, 2010, **132**, 9963-9965.

22. G.-J. Zhang, M. J. Huang, J. A. J. Ang, Q. Yao and Y. Ning, *Anal Chem*, 2013, **85**, 4392-4397.

23. E. Arigi, O. Blixt, K. Buschard, H. Clausen and S. B. Levery, *Glycoconj J*, 2012, **29**, 1-12.

24. T. Abel, J. I. Cohen, R. Engel, M. Filshtinskaya, A. Melkonian and K. Melkonian, *Carbohydrate Res*, 2002, **337**, 2495-2499.

25. R Core Team, *R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing*, http://www.R-project.org/.

Chapter 4

Discrimination between lectins with similar specificities by ratiometric profiling of binding to glycans

4.1 Abstract

Carbohydrate-lectin interactions mediate a number of crucial biological processes. Their prevalence in nature means they are often exploited by pathogens and as such have recently been utilised for diagnostic applications. Whilst lectin binding to long chain glycans can be highly specific these compounds are challenging to access and often expensive. The use of readily available and cheap monosaccharides is complicated by the promiscuous nature of lectin binding to these compounds. Herein the use of monosaccharide-coated surfaces is described, coupled with a machine-learning algorithm for the discrimination of lectins of similar binding specificities. This approach is then further expanded to a solution based 'label-free' approach and the use of heterogeneous glyco-environments to move towards a more 'real-world' system that can not only discriminate between lectins but can do so in a concentration independent manner.

4.2 Introduction

Protein carbohydrate interactions are crucial for many biological processes including cell-cell communication, innate immunity, and fertilisation.[1-4] These processes are mediated by carbohydrate binding proteins known as lectins. Carbohydrate-lectin interactions are moderated by a combination of the carbohydrate itself, the linker between the carbohydrate and the cell surface and the precise 3D presentation of carbohydrates on the cell surface.[5, 6]

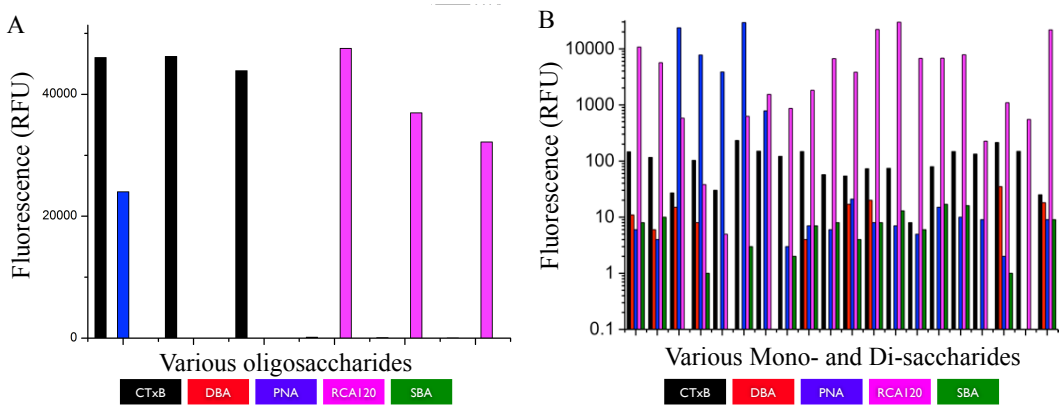


Figure 4.1 Lectin binding to six oligosaccharides extracted from the consortium for functional glycomics (A). Lectin binding to all mono- and di-saccharides extracted from the consortium for functional glycomics (B). The lectins are as follows; cholera toxin B subunit (CTxB), *Dolichos biflorus* agglutinin (DBA), peanut agglutinin from *Arachis hypogaea* (PNA), *Ricinus communis* agglutinin 120 (RCA120) and Soybean agglutinin (SBA).[7]

Many lectins show highly specific binding at the oligosaccharide level (Figure 4.1A) but highly promiscuous binding on the mono- and di-saccharide level (Figure 4.1B). For example, Peanut Agglutinin from *Arachis hypogaea* (PNA) is described as being galactose (Gal) specific but microarray analysis reveals it will readily bind to all monosaccharides to one degree or another.[7] Similarly, the B subunit of cholera toxin (CTxB) is described as binding specifically to monosialyltetrahexosylganglioside

(GM1) found on intestinal endothelial cells but it will indiscriminately bind all monosaccharides (although in a completely different order of magnitude compared to GM1 binding).[7]

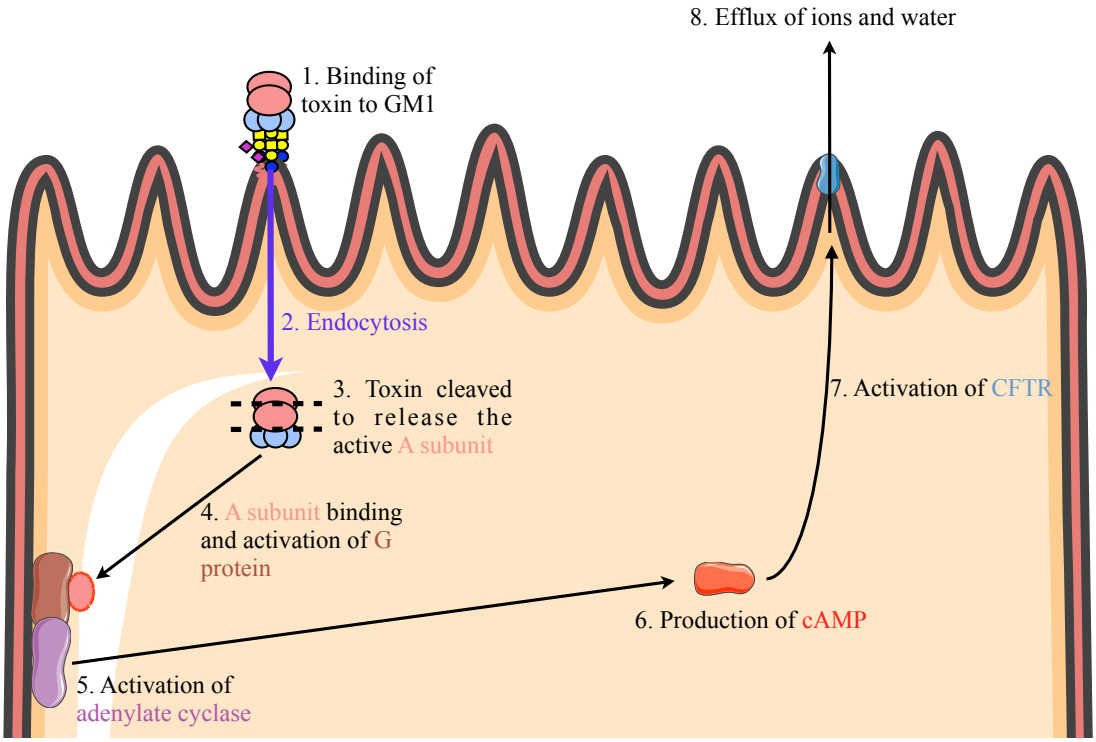


Figure 4.2 Schematic of cholera toxin action.

The wide variety of roles played by glycans in the body's innate processes and the prevalence of these interactions in nature means that they are readily exploited by pathogens. For example, FimH is a lectin that mediates the binding of uropathogenic *Escherichia coli* to mannose rich residues thus initiating infection and is a crucial virulence factor.[8, 9] The intraerythrocytic phase of *Plasmodium falciparum* (the parasite responsible for malaria) is mediated by a lectin-like protein, which allows it to bind to glycans on the surface of the red blood cell.[10] Cholera is caused by cell internalisation of an AB_5 toxin secreted by the bacteria *Vibrio cholera*.[11, 12] This toxin comprises of one enzymatic A subunit and 5 lectin B subunits (CTxB) which bind to the GM1 ganglioside present on intestinal endothelial cells.[13] After binding and

internalisation of the toxin, the B subunits are cleaved from the A subunit and the A subunit is further cleaved to form an active enzyme. This enzyme then enters the cytosol of the cell where it binds to a G protein locking it in its active state so it continually stimulates adenylate cyclase to generate cyclic adenosine monophosphate (cAMP). cAMP binds to cystic fibrosis transmembrane conductance regulator (CFTR) causing an efflux of ions and water (Figure 4.2). Ricin is a highly toxic lectin extracted from *Ricinus communis* seeds. It comprises an A subunit and a D-galactose binding B subunit. After binding of the toxin to the cell surface and internalisation the two subunits are cleaved apart and the A subunit goes on to remove an adenine residue from the 28S ribosomal RNA thus rendering the cell incapable of protein synthesis and ultimately resulting in cell death (Figure 4.3).[14] Differences in glycosylation have also been detected in tumour cells and can determine the metatastic potential of cancers whilst changes of glycosylation patterns on red blood cells have been linked to susceptibility towards, and severity of, diseases such as smallpox, cholera and malaria.[2, 4, 15-17]

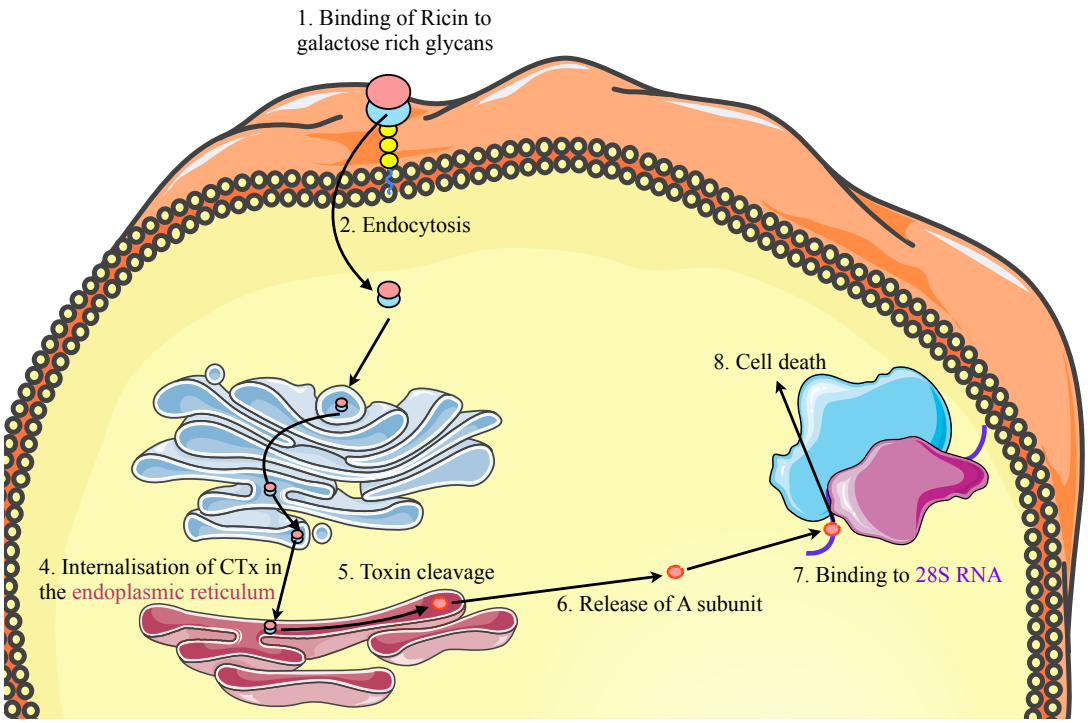


Figure 4.3 Schematic of ricin action.

As they are so crucial for a multitude of pathogenic processes, the rapid detection of lectins can aid in the early identification and prevention of diseases. Whilst proteomic and antibody based techniques can be used in the identification of lectins these techniques are not always suitable for robust point-of-care applications. These methods often require infrastructure for preparation and storage, expensive equipment and/or highly staff. To address this, multiplexed nanoparticles for biosensing have attracted much interest, especially for diagnostics. For example, Rotello *et al.* have developed a system using differentially functionalised gold nanoparticles, which allowed detection of 52 different mixtures of 7 proteins using just 6 different nanoparticles.[18] Glycosylated magnetic particles have also been used for MRI based detection of cancerous cells and gold particles functionalised with three different thiols allowed identification of cancerous cells from normal cells.[19, 20] Jayawardena *et al.* described the use of gold nanoparticles functionalised with a number of glycans.

These gold nanoparticles were used to discriminate between a number of different lectins based on the characteristic shift in SPR readings upon lectin binding to the gold nanoparticles. In this instance the lectins assessed had very different binding specificities and they could be discriminated based on binding to a single glycan so a multiplexed approach was not needed.[21]

Many lectins will have similar binding specificities especially at the monosaccharide level where binding can be fairly promiscuous. The aim of this research was to assess the use of monosaccharides as multiplexed sensors to enable the discrimination of lectins with similar binding specificities. Rapid diagnostic techniques based on discriminant lectin binding would allow for rapid detection of environmental toxins and by utilising monosaccharides the need for a single high affinity binder for each target is removed. This would allow the discrimination of a broader range of targets.

4.3 Results and Discussion

The key aim of this work was to evaluate the use of monosaccharides (as these are cheap and readily available) as multiplexed sensors to enable discrimination between different lectins.

4.3.1 Lectin binding profiles

Hydrazide functionalised 96-well plates were further reacted with a range of monosaccharides to produce a number of monosaccharide surfaces (as described in Chapter 2). To highlight the challenges faced in identification of lectins, a panel of five fluorescently labelled D-galactose (Gal) or *N*-Acetyl-D-Galactosamine (GalNAc) binding lectins were selected including two toxins; *Ricinus communis* Agglutinin-120 (RCA120, a non-toxic surrogate for the ricin B subunit) and cholera toxin (CTxB) and three environmental plant lectins; Soybean Agglutinin, *Dolichos biflorus* Agglutinin and Peanut Agglutinin from *Arachis hypogaea* (SBA, DBA and PNA respectively).

These lectins were then incubated with a D-galactose functionalised surface and after removal of unbound lectin and washing the fluorescence spectra were read. The binding profile for most of the lectins were within error of each other and as such identifying any one lectin based on binding to D-galactose alone would be challenging (Figure 4.4).

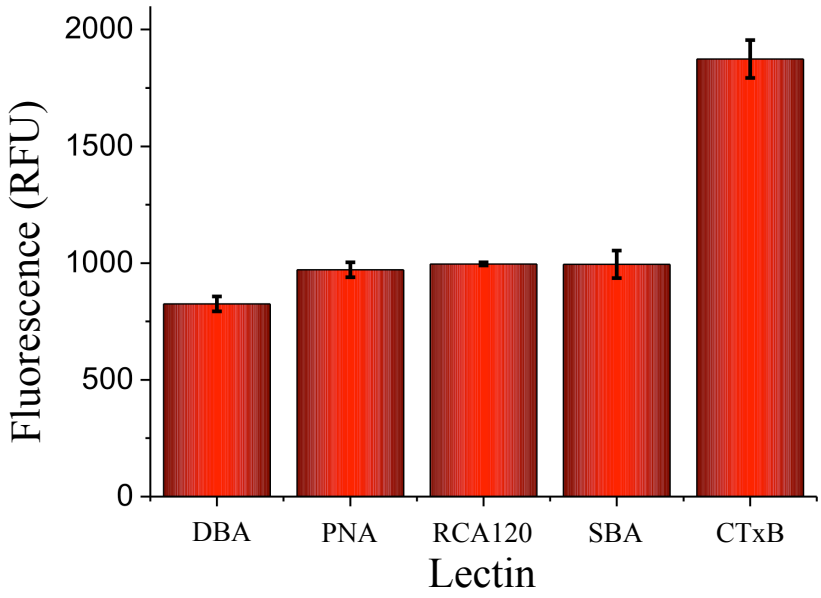


Figure 4.4 Addition of five fluorescently labelled lectins (at a concentration of 0.01 $mg.mL^{-1}$) to a D-galactose functionalised surface resulted in very similar binding profiles for most of the lectins. Bars represent the average of three measurements and the error bars are the standard error of the measurements.

Increasing the number of monosaccharide surfaces to include D-glucose (Glc), D-mannose (Man) and a 1:1 mixture of Gal:Man (Gal/Man) resulted in a unique binding profile for each lectin (Figure 4.5). For example RCA120 displayed significantly higher binding to Gal and the Gal/Man mixture with respect to Glc binding. Conversely, SBA showed depressed binding to the Gal/Man surface.

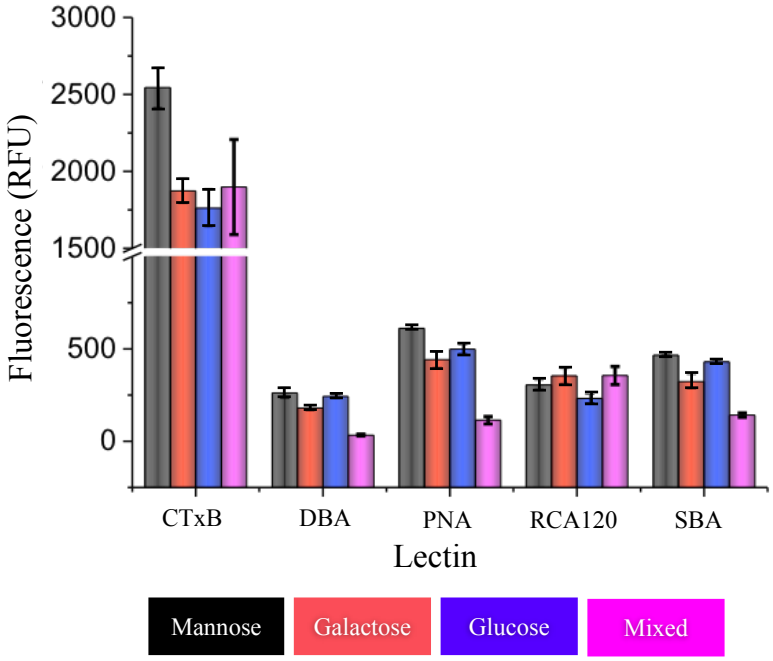


Figure 4.5 Addition of five fluorescently labelled lectins (at a concentration of 0.01 $mg.mL^{-1}$) to a surface functionalised with either Gal, Glc, mannose or a mixed surface of 1:1 Gal:Man. Bars represent the average of 4 independent replicates with error bars representing the standard error.

Despite the differences in the binding profiles it would still be challenging to identify a single lectin based only on its binding profile as many of the differences between the carbohydrates are within error of each other and many of the differences between the lectins are also very small or within error. A better analysis method is needed in order to improve the resolution between the profiles to aid in the classification of unknown samples.

4.3.2 Linear Discriminant analysis

Linear discriminant analysis (LDA) is a statistical technique that can aid in the classification of samples as it minimises the variation within the class (in this instance

the lectin) and maximises the variation between classes thus increasing the resolution of the profiles generated. Initial binding profiles are used as a 'training matrix' to produce a model in which the resolution between lectin classes is maximised.

The LDA model produced showed excellent resolution between groups (Figure 4.6A). Whilst it appears that all other lectin classes are well resolved from CTxB but poorly separated from each other this is not actually the case. Excluding CTxB, the model shows good separation between the other lectin classes has been achieved (Figure 4.6B). Validation of the model in a leave-one-out manner (where a sample is left out and the model is generated before classification of the 'left out' samples) gave a reclassification accuracy of 100 %.

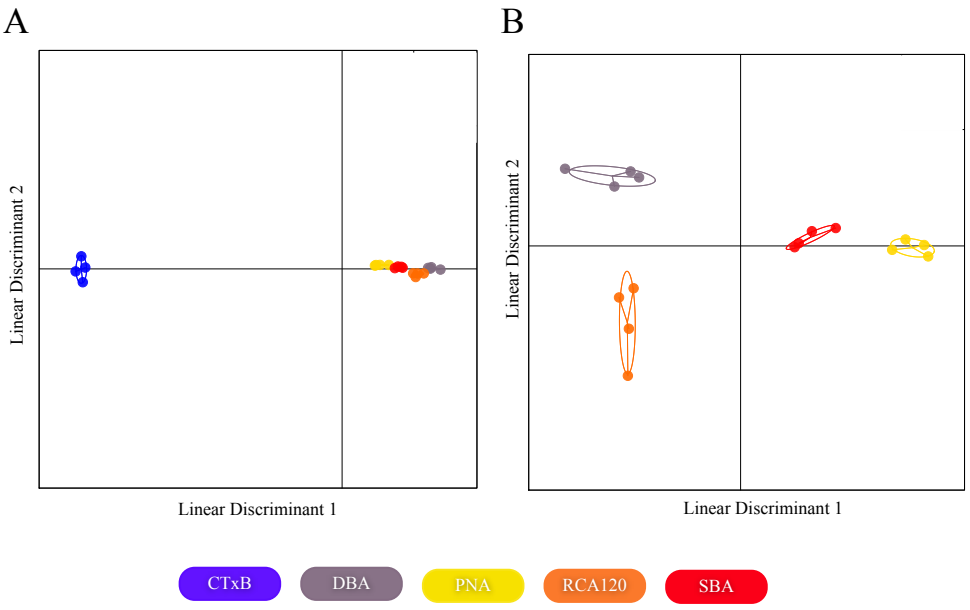


Figure 4.6 The LDA model produced from the surface binding profiles showed excellent resolution amongst the lectin groups (A). Whilst it appears that DBA, PNA, RCA120 and SBA are tightly clustered, they are also well separated from each other (B). Each point represents the binding profile of one lectin to all 4 surfaces and the ellipse represents the standard deviation from the average response.

The LDA model was then used to classify a blind sample of each of the five lectins binding to each of the four surfaces. All blind samples were correctly classified by the model based on their binding profiles with 100 % accuracy. This highlights the potential of a carbohydrate based identification sensor.

4.3.3 Identification of mixed samples

The model was able to correctly identify blind samples of pure lectins but in environmental samples it is likely that there will be contamination of other lectins and as such the differentiation between a mixed sample of the two pathogenic lectins (RCA120 and CTxB) was investigated.

Mixed samples of RCA120 and CTxB were added to the four sugar surfaces and binding profiles were determined (Figure 4.7). The LDA model was then used to classify the binding profiles in order to determine the lectin responsible. When the lectin was present at greater than 50 % (by mass) of the sample it was correctly classified, so at > 50 % CTxB the model classified the profile as CTxB with the reverse case also being true.

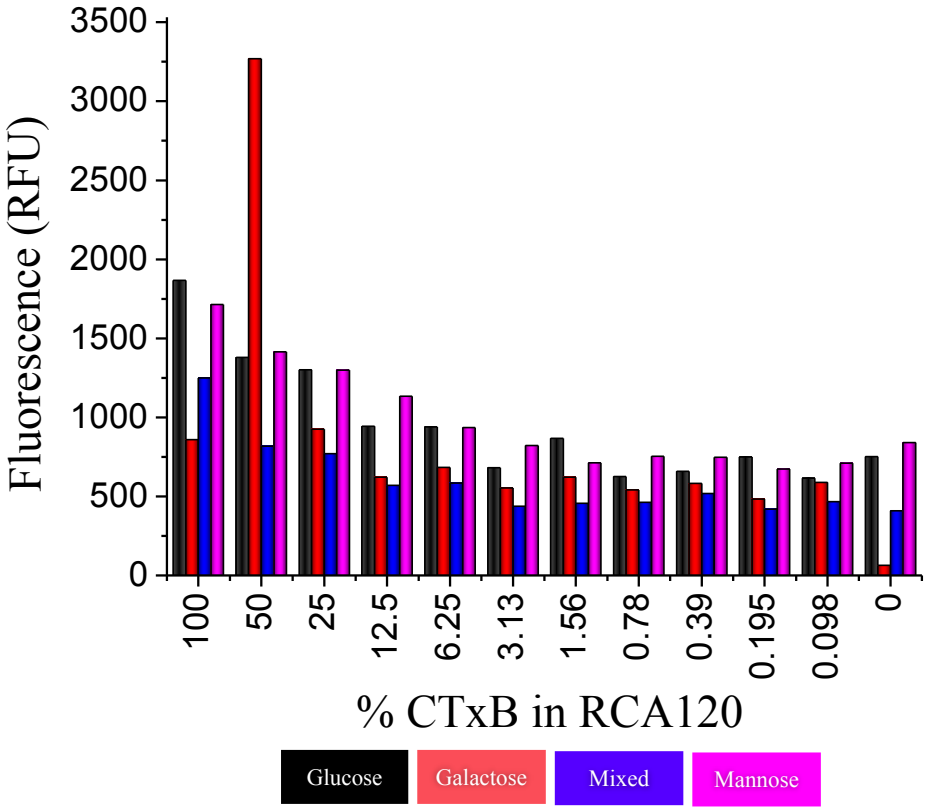


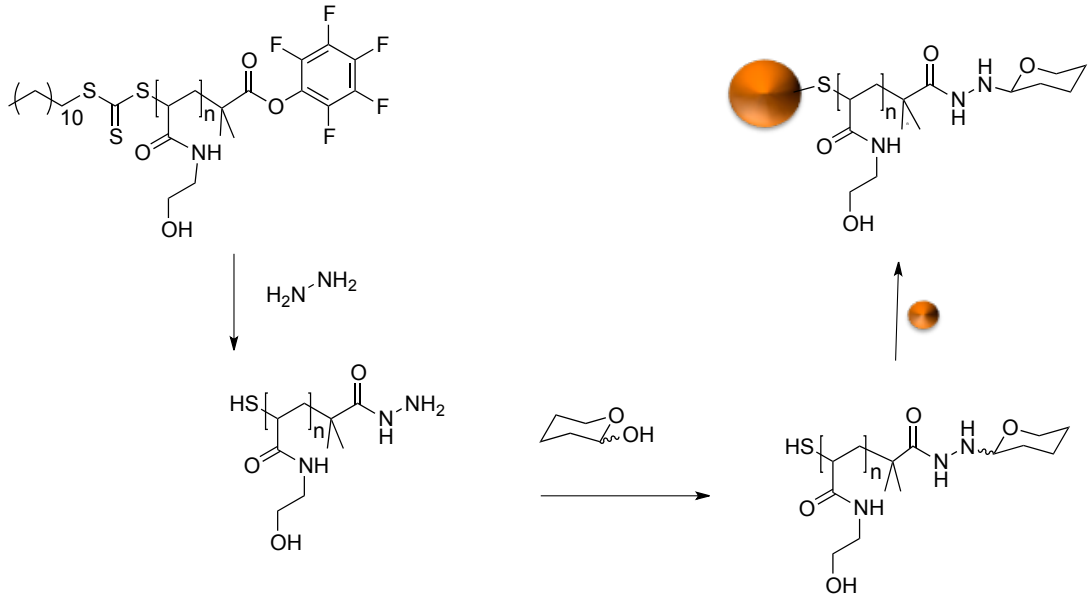
Figure 4.7 Binding of a variety of samples containing different ratios of CTxB to RCA120 to the 4 sugar surfaces. The bars show the binding of a single sample.

4.3.4 Solution based assay

Classification of non-pure lectin samples was possible with the mixed surfaces technique and, whilst that is a step-towards a 'real-world' sensing application, it still utilises fluorescently labelled lectins, which would not be available in all settings. As such LDA was applied to the binding profiles generated from a label-free solution based assay.

This assay uses sugar functionalised gold nanoparticles (GlycoAuNPs, Scheme 4.1). In solution, these appear pink to the naked eye but upon addition of a lectin and aggregation of the GlycoAuNPs the solution undergoes a colour change from pink to

blue (Figure 4.8). This colour change can also be seen in the absorbance spectrum as a decrease in absorbance at 530 nm and an increase at 700 nm. Using these changes glycan binding affinity can be estimated.



Scheme 4.1 Synthetic route to carbohydrate functionalised gold nanoparticles (60 nm). Particles were synthesised by Sarah-Jane Richards.

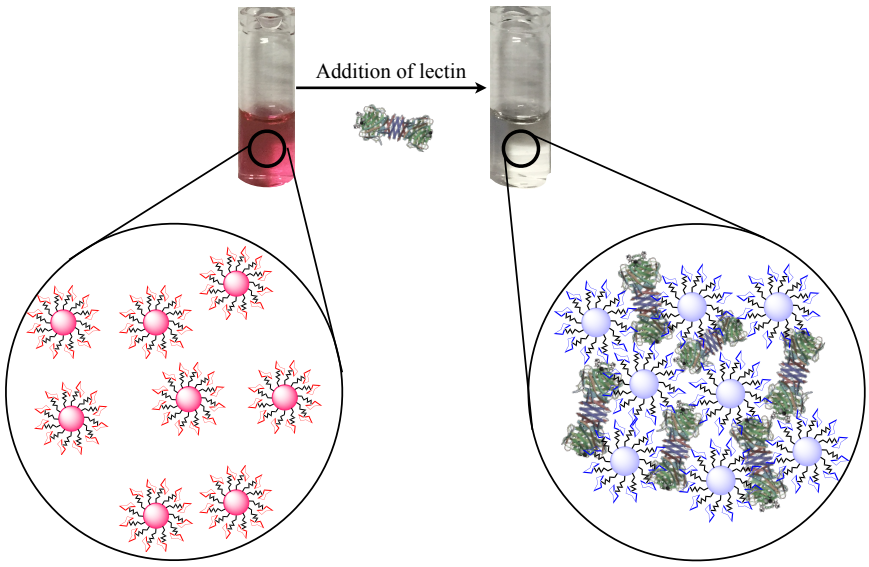


Figure 4.8 Upon addition of a lectin to GlycoAuNPs aggregation occurs which results in a colour change from pink to blue.

Addition of six lectins to 6 different GlycoAuNPs and measurement of the absorbance at 700 nm resulted in the production of binding profiles for all of the lectins (Figure 4.9A). All of the lectins assessed had unique binding profiles. However, many of the binding profiles for one lectin fell within error of the profile of another lectin, consequently, binding profiles alone could not be used to predict unknown samples.

The binding profiles were used as a 'training matrix' to produce a LDA model (Figure 4.9B). The LDA model produced showed excellent resolution between all lectin classes and validation of the model using a leave-one-out approach indicated the reclassification accuracy of the model was 100 %.

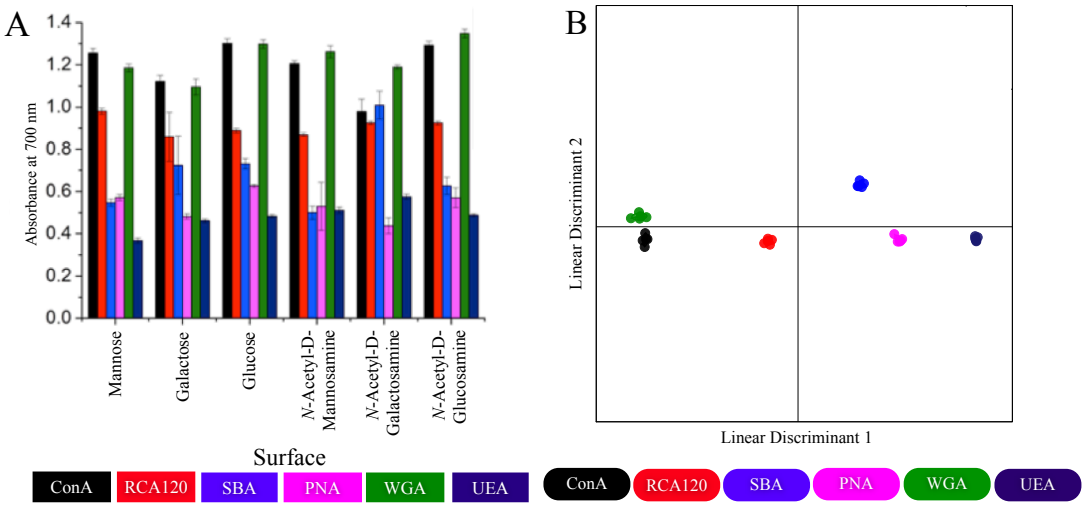


Figure 4.9 After incubation of the lectin with each of the GlycoAuNPS absorbance was measured at 700 nm resulting in the production of distinct binding profiles for all lectins assessed where the bars represent the average of 4 independent replicates and the error bars are the standard error (A). The binding profiles were then used as a 'training matrix' to generate a LDA model which showed excellent resolution between all lectins assessed (B).

The differential binding of lectins to monosaccharides coupled with LDA was able to produce a model with excellent resolution using this label free solution based technique. This highlights the potential of the method for application in a more realistic setting where lectins for detection would not be fluorescently labelled. Whilst we have utilised absorbance data as the 'training matrix' we have previously shown that the colourimetric change from pink to blue can also be graphically interpreted using a digital camera and analysis in ImageJ to determine pixel intensity, which correlates well with absorbance.[22] This approach is label free, cheap and highly accessible.

4.3.5 Mixed surfaces

All of the approaches mentioned so far have used a known concentration of lectin in the generation of binding profiles and subsequent detection of blind samples have been at the same concentration. In a 'real-world' setting the concentration of the lectin in a sample would be unknown and efforts to create a concentration independent sensor should also be made.

Increasing the number of subunits in the glycan chain could be utilised to achieve this sensing capability. However, the synthetic complexity and thus price of glycans generally increases as the chain length increases. Increasing the information density without increasing the number of saccharides used would be advantageous. It is known that the use of variable density glycan mixtures can result in non-linear responses in terms of glycan binding.[23-25] It can therefore be rationalised that the use of mixed surfaces can increase the information density of a sensor without resorting to increasingly complex glycans.[26]

To this end GlycoAuNPs were prepared with varying ratios of *N*-Acetyl-D-Mannosamine (ManNAc) and GalNAc from 100 % GalNAc to 100 % ManNAc in steps of 10 (Figure 4.10). Once again the shift in absorbance at 700 nm was used as the response and each of the different GlycoAuNPs was incubated with seven different concentrations of three different lectins. The binding profiles were used as a 'training matrix' and a LDA model was produced (Figure 4.11). Within the LDA model the lectins are at several different concentrations but only the lectin class was considered.

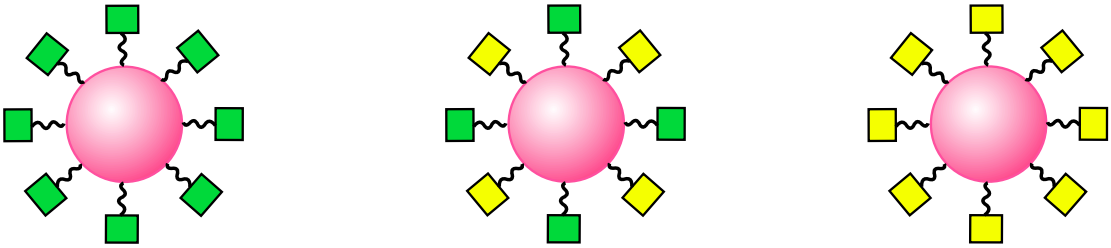


Figure 4.10 GlycoAuNPs were functionalised with varying ratios of ManNAc and GalNAc from 100 % ManNAc (left) to 100 % GalNAc (right).

The LDA model showed good resolution between the groups but with some overlap between the lectin classes. Despite this, analysis in a leave-one-out manner revealed the reclassification accuracy to be 86 %. The overlap between the classes was due to very similar binding profiles for the lectins when the concentration was low.

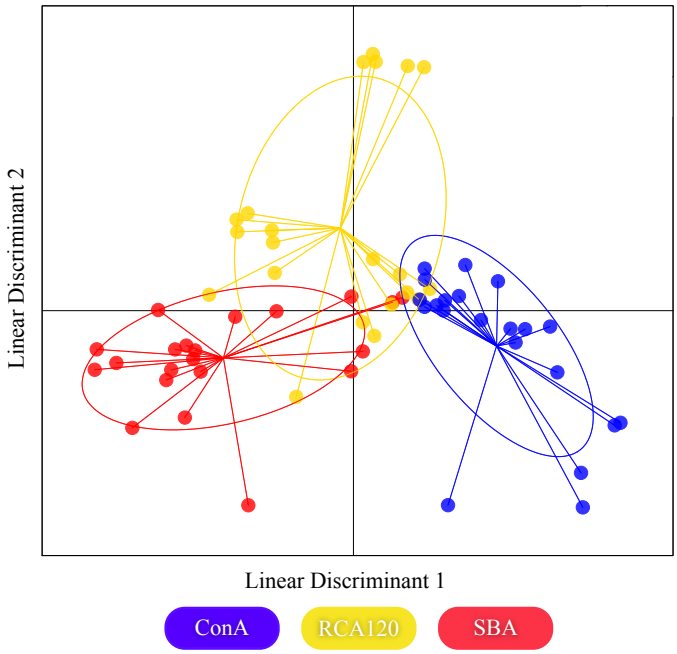


Figure 4.11 Shift in absorbance at 700 nm after incubation of the lectin at various concentrations was used as a 'training matrix' to produce a LDA model where each point represents the binding of a single lectin (at any of the seven concentrations) to all GlycoAuNPs and the ellipses represent the standard deviation from the average response.

Whilst the model only had an accuracy of 86 % it should be noted that this is based on the binding profiles considering only two monosaccharides (GalNAc and ManNAc). If we consider a LDA model containing information from only the 100 % GalNAc and 100 % ManNAc particles then the reclassification accuracy of the model decreases to 68 % and, by including the varying ratios, the information density of the assay has been increased.

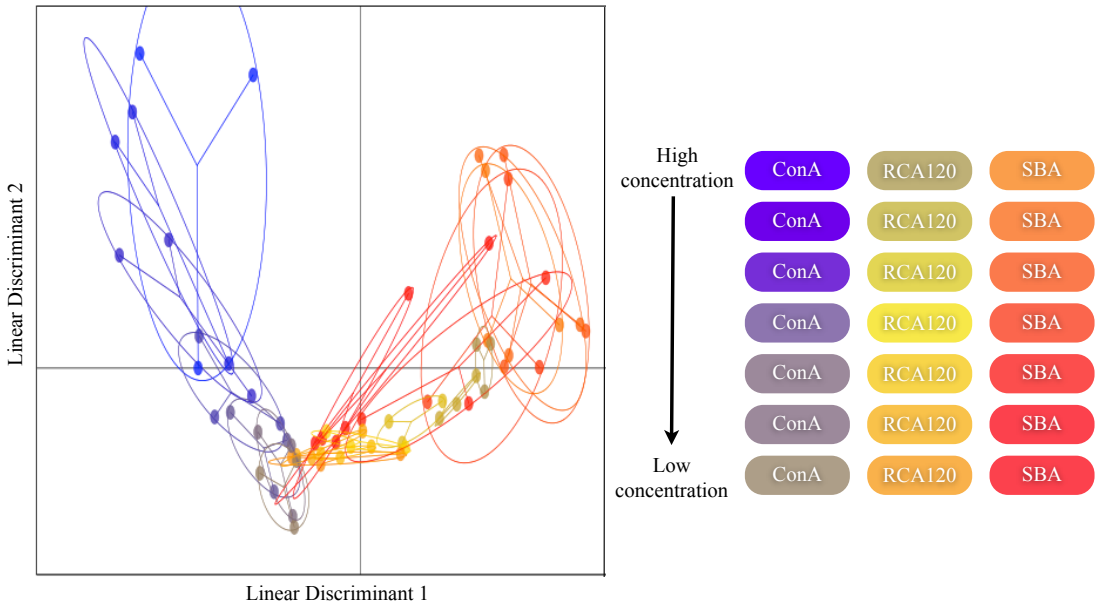


Figure 4.12 The concentration dependent LDA model produced showed significant overlap both between different concentrations of the same lectin and between different lectin classes.

If we separate the lectin classes into lectin and concentration then the model has the potential to classify a sample based on both lectin class and concentration of the lectin thus providing more information. A LDA model produced that is separated in this manner showed considerable overlap both between different concentrations of the same lectin and between lectin classes (Figure 4.12). The classification power of this model is significantly reduced when compared to the concentration independent model, giving an accuracy of only 27 %.

Generally the predictive power of the model is higher for the higher concentrations and generally declines as the concentration decreases. If we consider only the profiles for lectin concentration of greater than 100 nM then the profiles show excellent resolution between the classes and reasonable resolution between the different

concentrations within each lectin class (Figure 4.13A). However, if we consider the profiles for which the lectin concentration was less than 100 nM then there is significant overlap between concentrations and between lectin classes resulting in the lower classification accuracy of this model (Figure 4.13B).

The accuracy of both the concentration independent and concentration dependent model could be increased by increasing the number of sugars, increasing glycan chain length or with the addition of varying surfaces thus increasing the accuracy of the model. This approach has the potential for lectin sample classification in both a concentration independent manner and also classification of not just lectin but also concentration of lectin within a sample.

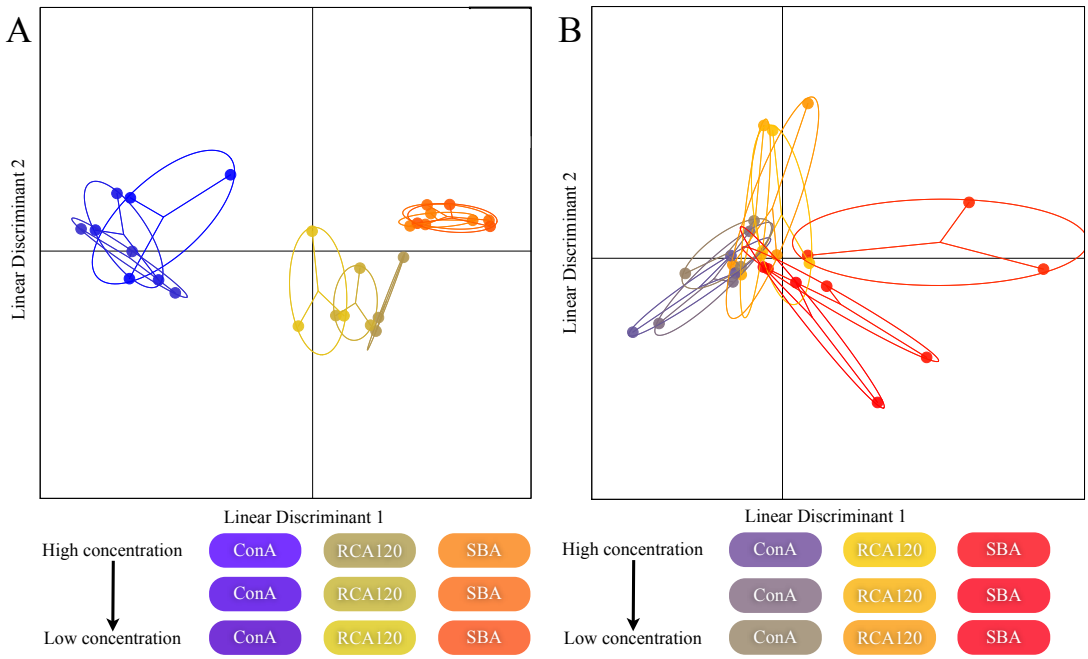


Figure 4.13 The concentration dependent LDA model considering just those profiles with a concentration above 100 nM showed excellent differentiation of lectin classes but poorly resolved for different concentrations (A). At lower concentrations (< 100 nM) the profiles for both different concentrations and different lectins overlapped (B).

4.4 Conclusion

Here we have described the use of four monosaccharide-coated surfaces combined with linear discriminant analysis for the discrimination of five lectins of similar binding specificity. This technique was able to correctly identify blind samples with 100 % accuracy and could even determine one lectin in the presence of another as long as the lectin was present at greater than 50 % (by mass). The coupling of monosaccharide binding profiles with LDA was further applied to a 'label free' solution based assay utilising glycan functionalised gold nanoparticles. The use of mixed glycan surfaces coupled with LDA highlighted a manner through which increasing information density can be achieved without using expensive (bio)synthetically challenging oligosaccharides. This highlighted the potential for a monosaccharide based, 'label-free' detection system that can discriminate between lectins of similar binding specificities in a concentration independent manner, which would enable rapid lectin detection in a real world situation.

4.5 Materials and methods

Materials: All chemicals were used as supplied unless otherwise stated. Corning® 96 well clear flat bottom polystyrene Carbo-BIND^{TM} microplates, aniline (99.5 % ACS reagent), D-(+)mannose, α-D-glucose, D-(+)galactose, DMSO, phosphate buffered saline tablets and FITC conjugated Cholera Toxin B subunit were purchased from Sigma-Aldrich. 100 mM acetate buffer with 1 mM aniline (pH 5.5) was prepared in 200 mL of milliQ water (with a resistivity >18 MΩ.cm^{-1}). 10 mmol HEPES buffer containing 0.05 M NaCl, 0.1 mM $CaCl_2$ and 0.01 mM $MnCl_2$ (pH 7.5, HEPES) was prepared in 200 mL of milliQ water (with a resistivity >18 MΩ.cm^{-1}). FITC labelled *Dolichos biflorus* agglutinin, Soybean agglutinin, *Ricinus communis* Agglutinin I and Peanut Agglutinin from *Arachis hypogaea* were purchased from Vector Labs.

Preparation of glycosylated surfaces: 30 mM monosaccharide solutions were prepared in 100 mM acetate buffer with 1 mM aniline. 100 µL of monosaccharide solution was then added to every well of a hydrazide functionalised 96 well microtitre plate before incubation at 50 °C for 24 hours. After incubation wells were extensively washed three times with MilliQ water before being allowed to dry. Plates were functionalised with galactose, glucose, mannose or a 1:1 solution of galactose and mannose and were either used immediately or stored at -20 °C.

Lectin binding assays: Various fluorescently labelled lectins (DBA, CTx, PNA, RCA120 and SBA) were dissolved to a concentration of 0.01 mg.mL^{-1} in 10 mM HEPES with 0.15 M NaCl, 0.1 mM $CaCl_2$ and 0.01 mM $MnCl_2$ (pH 7.5). 100 µL of the lectins were then added to each well of a microtitre plate before incubation in the dark at 37 °C for 30 minutes. After incubation, each well was rigorously washed 3

times with water before fluorescence readings were taken. Fluorescence readings were taken using a BioTek Synergy HT multi-detection microplate reader and Gen5 software with excitation and emission wavelengths of 485 and 528 nm respectively.

Blind sample assays: Blind samples were prepared in the same method as the lectin binding assays. Briefly, 100 μL of each blind sample (with a lectin concentration of 0.01 mg/mL) was added to each well and after incubation in the dark at 37 °C for 30 minutes wells were washed with water before fluorescence readings were taken. Fluorescence readings were taken using a BioTek Synergy HT multi-detection microplate reader and Gen5 software with excitation and emission wavelengths of 485 and 528 nm respectively.

Linear discriminant analysis: Every lectin was added to each surface as described in the lectin binding assay section. This was repeated 4 times to generate a training data matrix of 4 surfaces x 5 lectin x 4 replicates, which was then subjected to a classical linear discriminant analysis (LDA)[27] in the open source statistical package R (version 2.14.1).[28] The model produced in this analysis was used to predict the nature of the random samples using the predict function in R.[27]

Linear discriminant analysis plots: LDA plots were drawn in R (version 2.14.1)[28] using the plot function. Each data point represents the fluorescence readings for a single lectin transformed into linear discriminants using the LDA model. Ellipses were added using the ellipse function and represent the mean linear discriminant values for each lectin category +/- the standard deviation of the linear discriminants.

4.6 References

1. C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357-2364.
2. X. Zeng, C. A. Andrade, M. D. Oliveira and X.-L. Sun, *Anal Bioanal Chem*, 2012, **402**, 3161-3176.
3. T. Feizi, F. Fazio, W. Chai and C.-H. Wong, *Curr Opin Struc Biol*, 2003, **13**, 637-645.
4. Y. J. Kim and A. Varki, *Glycoconj J*, 1997, **14**, 569-576.
5. A. Bernardi, J. Jimenez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K. E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penades, F. Peri, R. J. Pieters, O. Renaudet, J. L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schaffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof and A. Imberty, *Chem Soc Rev*, 2013, **42**, 4709-4727.
6. H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis, *Science*, 2001, **294**, 2163-2166.
7. CFG, http://www.functionalglycomics.org.
8. N. P. Pera and R. J. Pieters, *MedChemComm*, 2014.
9. M. Hartmann, H. Papavlassopoulos, V. Chandrasekaran, C. Grabosch, F. Beiroth, T. K. Lindhorst and C. Röhl, *FEBS letters*, 2012, **586**, 1459-1465.
10. M. Jungery, G. Pasvol, C. I. Newbold and D. J. Weatherall, *Proc Natl Acad Sci U S A*, 1983, **80**, 1018-1022.
11. C. J. O'Neal, M. G. Jobling, R. K. Holmes and W. G. Hol, *Science*, 2005, **309**, 1093-1096.
12. D. Cassel and T. Pfeuffer, *Proc Natl Acad Sci U S A*, 1978, **75**, 2669-2673.

13. S. V. Heyningen, *Science*, 1974, **183**, 656-657.

14. J. M. Lord, L. M. Roberts and J. D. Robertus, *Faseb j*, 1994, **8**, 201-208.

15. E. Hosoi, *J Med Invest*, 2008, **55**, 174-182.

16. D. J. Anstee, *Blood*, 2010, **115**, 4635-4643.

17. J. A. Rowe, D. H. Opi and T. N. Williams, *Curr Opin Hematol*, 2009, **16**, 480.

18. C.-C. You, O. R. Miranda, B. Gider, P. S. Ghosh, I.-B. Kim, B. Erdogan, S. A. Krovi, U. H. F. Bunz and V. M. Rotello, *Nat Nano*, 2007, **2**, 318-323.

19. K. El-Boubbou, D. C. Zhu, C. Vasileiou, B. Borhan, D. Prosperi, W. Li and X. Huang, *J Am Chem Soc*, 2010, **132**, 4490-4499.

20. A. Bajaj, O. R. Miranda, I. B. Kim, R. L. Phillips, D. J. Jerry, U. H. Bunz and V. M. Rotello, *Proc Natl Acad Sci U S A*, 2009, **106**, 10912-10916.

21. H. S. Jayawardena, X. Wang and M. Yan, *Anal Chem*, 2013, **85**, 10277-10281.

22. L. Otten, S.-J. Richards, E. Fullam, G. S. Besra and M. I. Gibson, *J Mater Chem B*, 2013, **1**, 2665-2672.

23. M. Gómez-García, J. M. Benito, A. P. Butera, C. O. Mellet, J. M. G. Fernández and J. L. J. Blanco, *J Org Chem*, 2012, **77**, 1273-1288.

24. J. L. Jimenez Blanco, C. Ortiz Mellet and J. M. Garcia Fernandez, *Chem Soc Rev*, 2013, **42**, 4518-4531.

25. D. Ponader, P. Maffre, J. Aretz, D. Pussak, N. M. Ninnemann, S. Schmidt, P. H. Seeberger, C. Rademacher, G. U. Nienhaus and L. Hartmann, *J Am Chem Soc*, 2014, **136**, 2008-2016.

26. S. J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew Chem Int Ed*, 2012, **51**, 7812-7816.

27. Jonathan Chang (2011). *R package version 1.3.1,* http://CRAN.R-project.org/package=lda

28. R Core Team, R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing, http://www.R-project.org/.

Chapter 5

Carbohydrate functionalised surfaces for the rapid identification of bacterial species to improve point-of-care diagnostics

5.1 Abstract

Antibiotic resistance remains a global health concern with deaths as a result of antimicrobial resistance due to outnumber those caused by cancer by 2050. There is a growing concern that without an increase in new treatment methods, or a decline in the spread of resistance amongst bacterial species, there will be a return to a 'dark age' of medicine during which procedures that are currently routine will once again become impossible. Current bacterial infection identification methods are often challenging, time-consuming or expensive and most treatment is carried out in the absence of microbial information. Here we describe the use of carbohydrate functionalised surfaces coupled with a powerful statistical technique for the rapid identification of a variety of bacterial species in order to improve point-of-care diagnostics of bacteria responsible for infection.

5.2 Introduction

5.2.1 Bacterial infections

Bacterial infections are responsible for millions of deaths a year. Although exact statistics are hard to come by, the number of childhood deaths is often used as a 'surrogate marker' for the bacterial infection-derived mortality worldwide. It is estimated that bacterial infections contribute to one third of all childhood deaths, approximately 1.6 million deaths a year, with the majority occurring in South Asia and Sub-Saharan Africa.[1]

Bacteria cause diseases such as meningitis, pneumonia, diarrhoea and are responsible for many hospital acquired infections such as *Staphylococcus aureus*[2] and *Clostridium difficile*.[3] Moreover, 'old' illnesses such as tuberculosis, caused by the bacteria *Mycobacterium tuberculosis,* for which antibiotics have proven particularly effective, are once again on the rise.[4, 5]

5.2.2 Antibiotic resistance

Penicillin was the first antibiotic discovered and was first used clinically in 1940 where it revolutionised the treatment of gram-positive infections.[6] Several other antibiotic agents came to light soon after this but, despite this early progress, bacterial infections are still rife today due to the adaptable nature of the bacteria. Bacteria often develop resistance to the antibiotics used to treat them. The spread of resistance can be a very rapid process within a population of bacteria, for example Methicillin was first clinically used in 1960 and Methicillin resistant *S. aureus* cases were first reported a short while after (Figure 5.1).[7]

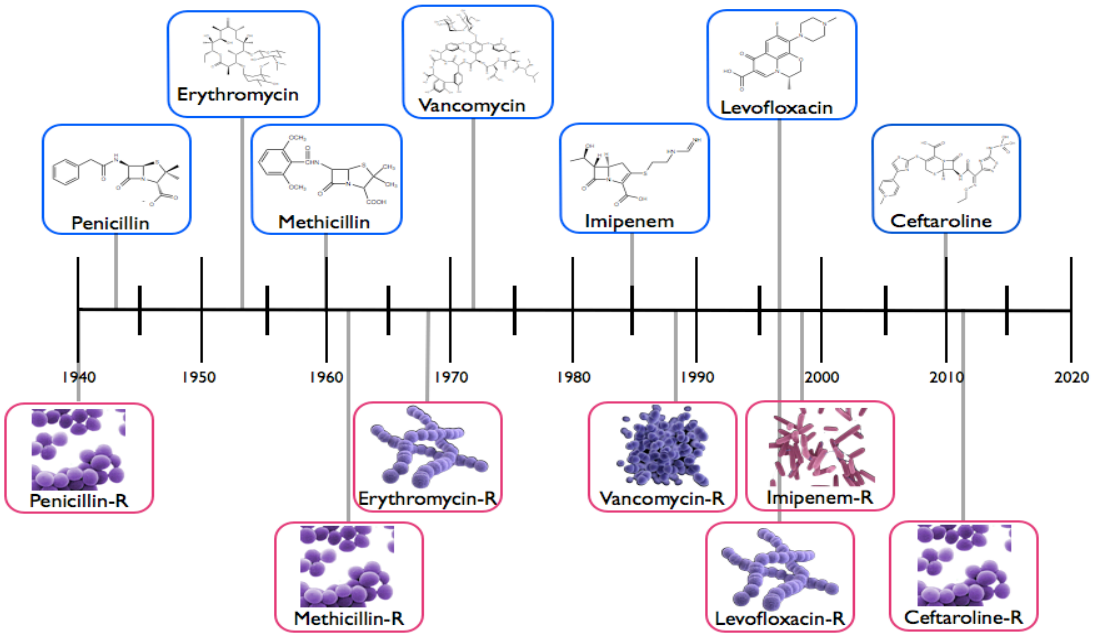


Figure 5.1 Time line of antibiotic compound discoveries (shown in blue boxes) and development of resistant bacterial species (shown in red boxes). The resistant bacterial species depicted are as follows; *Staphylococcus aureus* (Penicillin-R, Methicillin-R and Ceftaroline-R), *Streptococcus pneumonia* (Erythromycin-R and Levofloaxin-R), *Enterococcus faecalis* (Vancomycin-R) and *Pseudomonas aeruginosa* (Imipenem-R).

Resistance occurs as a result of the development of one of a number of protective mechanisms including efflux pumps to remove the antibiotic (such as MexXY which is responsible for multidrug resistance in *Pseudomonas aeurginosa*),[8] antibiotic degrading enzymes (such as the β-lactamases that confer resistance to the β-lactam class of antibiotics),[9] and antibiotic altering enzymes (resistance to aminoglycosides is caused by modifying enzymes),[10] all of which render the antibiotics ineffective. The genes responsible for these mechanisms confer an evolutionary benefit towards bacteria living in an environment containing antibiotics and as such are readily passed

throughout the population through the production of daughter cells, this is known as vertical gene transfer. Furthermore, bacteria can undergo horizontal gene transfer, where plasmids containing genes responsible for antibiotic resistance can be transferred between both neighbouring bacterial cells and even between closely related species.[11] This increases the spread of resistance within a population and is one of the reasons for the rapid spread in resistance genes with a bacterial population.

Antibiotic resistance now affects almost all species of pathogenic bacteria and is increasingly becoming a concern in terms of treatment of infections with increasing numbers of resistance cases (Figure 5.2). It has been predicted that by 2050, the number of deaths associated with antibiotic resistance will exceed those caused by cancer and many medical procedures that are currently routine will be impossible due to the increased risk of death from infection.[12, 13]

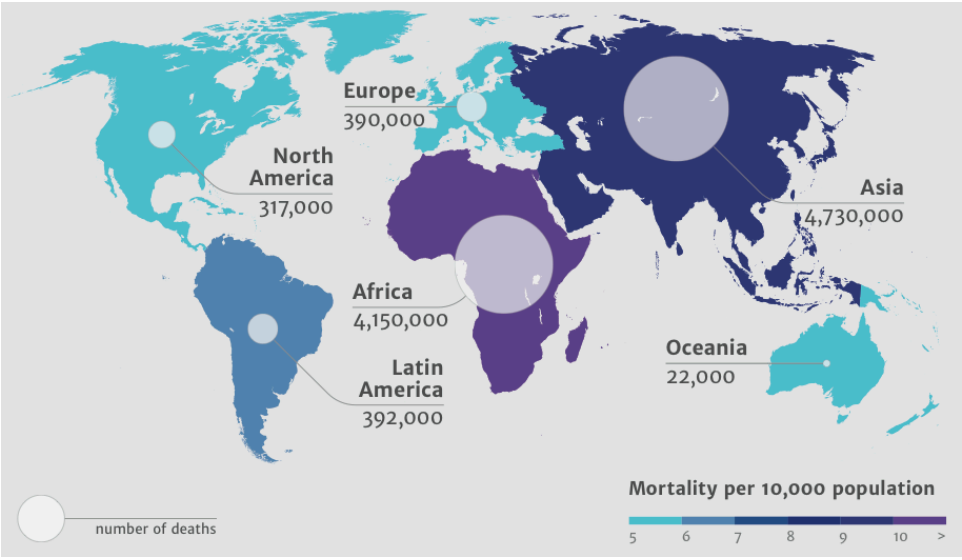


Figure 5.2 Predicted number of deaths associated with antibiotic resistance every year by 2050 where colour indicates mortality per 10,000 population. Figure adapted from reference [12].

5.2.3 Point-of-care diagnostics

The problem of antibiotic resistant bacteria is being propagated by the lack of rapid point-of-care diagnostic tools for detection of the bacterial species responsible for an infection. Whilst current tools such as culturing are accurate they are often time consuming and detection of certain species can be complicated by the fact that they enter a viable but not-culturable (VBNC) phase; meaning that they are un-culturable and thus un-detectable by conventional methods.[14] As a result, treatment decisions are often made in the absence of microbiological information.

Whilst several current methods exist for identification of bacterial species, most of the Food and Drug Administration (FDA) approved detection methods are based on culturing techniques with some exceptions for specific bacteria, which are known to be challenging to culture. These bacteria are detected through the use of Polymerase Chain Reaction (PCR) based methods.

Culturing for the purpose of identifying bacteria was first described over 140 years ago by Robert Koch, for the identification of planktonic bacteria responsible for acute epidemic infections and the technique has hardly changed since.[14, 15] These types of bacterial infections are now very uncommon due to the production of vaccines and their susceptibility to broad-spectrum antibiotics.[16] Biofilm based infections (e.g. tuberculosis, *vide infra*) are now a widespread problem with these bacteria proving incredibly challenging to culture. It is estimated that only 1 % of bacteria are culturable in standard culture media and several of the common species that cause infection in a hospital setting can enter a VBNC phase where they are not detectable

through culturing techniques. This has been shown to occur in enterohemorrhagic *Escherichia coli*, *Salmonella enterica* and *S. aureus*.[14, 17]

PCR based processes include: (i) analysis of the full genome of microbial species and identification through comparison with a database[18, 19] and (ii) IBIS a novel method which involves selective PCR of a species-specific element of DNA which is flanked by highly conserved regions of DNA. This section of DNA is then analysed with mass spectrometry and compared to a database in order to identify the species. PCR based techniques can be rapid, with IBIS reportedly taking only 6 hours, but they can also be incredibly expensive and often large numbers of bacterial cells are required for good PCR analysis. PCR techniques also rely on extraction of nucleic acids from within bacteria, which can be problematic in Gram-positive bacteria due to their thick peptidoglycan wall, which is difficult to lyse.[16]

With most approved detection methods unable to detect a multitude of the bacteria responsible for biofilm based infections there is increasing interest in the discovery of new, facile and rapid techniques for point-of-care bacterial detection. Developing countries are also likely to be hardest hit in terms of deaths associated with antibiotic resistance. For example, it is predicted that by 2050 one in every four deaths in Nigeria[12] will be due to antibiotic resistant microorganisms and, new techniques should be cheap in order to make them accessible to all countries.

5.2.4 Bacterial Biofilm formation

Biofilm based bacterial infections are becoming increasingly common. They are responsible for chronic bacteria associated conditions such as pneumonia[20] and

tuberculosis[21] and result in millions of deaths every year (tuberculosis caused 1.3 million deaths in 2012).[22] As previously discussed, detection of these bacteria using current detection methods is challenging and, even upon detection, treatment can be difficult as the biofilm protects the infection from both the host's immune system and penetration by antibiotics. The process of biofilm formation is an important area of research in the hope that greater understanding will result in increased treatment options.

The initial step in bacterial infection is the process of adhesion, mediated by bacterial proteins (adhesins), which bind to carbohydrates on cells. After adhesion, a proliferation step occurs resulting in biofilm formation once the colony reaches a certain size. Once the population in the biofilm reaches this critical level, some of the bacteria leave the biofilm and travel further around the body to start new colonies in other places (Figure 5.3). Biofilm formation is a biomedically important step in infection as it allows the bacterial population to expand whilst offering protection from both the hosts immune system and antibiotics.[23]

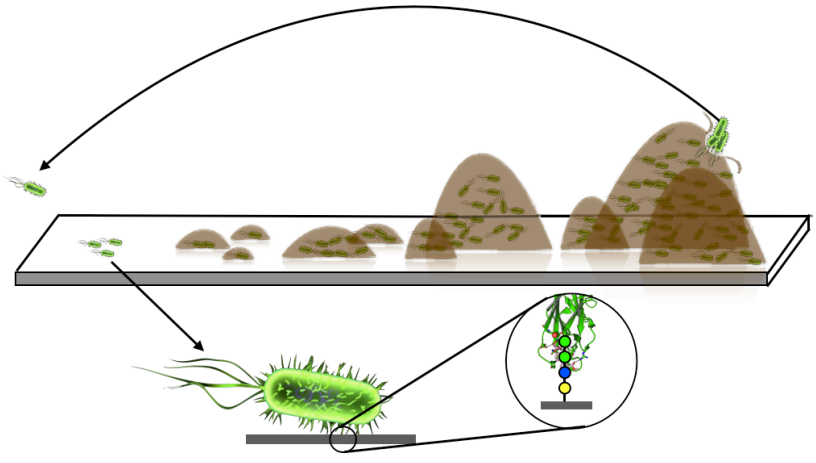


Figure 5.3 Schematic depicting bacterial biofilm formation. The example shown is the bacterial adhesin FimH interacting with a mannose rich tetrasaccharide.[24]

5.2.5 Role of protein-carbohydrate interactions in bacterial infection

Protein-carbohydrate interactions are a target for the prevention of bacterial infection. If the initial adhesion phase (mediated by these interactions) does not happen then a biofilm cannot be formed, so any bacteria in the body will be exposed to antimicrobial agents. This is of interest owing to the increase in bacterial antibiotic resistance; if critical bacterial adhesins could be identified and targeted then the initial adhesion phase could be prevented (anti-adhesion therapy).[25-28] This mechanism is also used in nature to protect cells from bacterial infection.[29] Carbohydrates known as mucins found on epithelial cells have been shown to be targets for pathogens. However, these cells also secrete the same carbohydrates on their surface thus trapping pathogens before they can invade the underlying tissue.[30]

Adhesins are thus essential for bacterial pathogenicity and detection through the initial adhesion phase could be key to selectively identifying pathogenic bacteria, because binding specificity is intrinsically linked to bacterial virulence.[20] The adhesin FimH is a crticial virulence factor found in uropathogenic *E. coli* and among enterobacteriaceae (a large family of bacteria including many pathogenic bacteria such as *E. coli*).[31] FimH is a mannose binding adhesin and so targets cells coated in mannose-rich carbohydrates. Upon binding, colonisation and biofilm production can occur.[32] Inhibitors of FimH have been identified, containing a mannose-based structure.[33] These inhibitors were then further assessed for cytotoxicity where their potency in inhibiting bacterial cells binding to highly glycosylated mammalian cells was found to be very different from that on a surface coated in mannan (polysaccharide of repeating D-mannose units). For all compounds tested, a narrow range between the IC_{50} of the drug (the concentration at which the drug prevented

adhesion of half of the bacterial cells) and the concentration at which 50 % cell death occurred (in the mammalian cells used) was observed thus potentially limiting their therapeutic potential.[31]

Identification of bacterial adhesins that are crucial for pathogenicity could result in faster detection of potentially pathogenic bacteria. They may also reveal potential drug targets to prevent bacteria from binding to mammalian cells, thereby preventing biofilm formation. However, whilst for some bacteria (such as *E. coli*) these adhesins have been well characterised for others (such as those found in *M. tuberculosis*) very little research has been done. Despite the existence of cheap and readily available drugs to cure tuberculosis, there were 8.6 million new cases and 1.3 million deaths reported in 2012 (figure 5.4A).[34] Increasingly, drug resistant strains have been identified and recently there have been several outbreaks of multi-, extensively- and totally-drug resistant (and thus incurable) tuberculosis and these strains (along with co-infection with HIV) are largely responsible for the resurgence of tuberculosis as a global health epidemic (figure 5.4B).[35, 36]

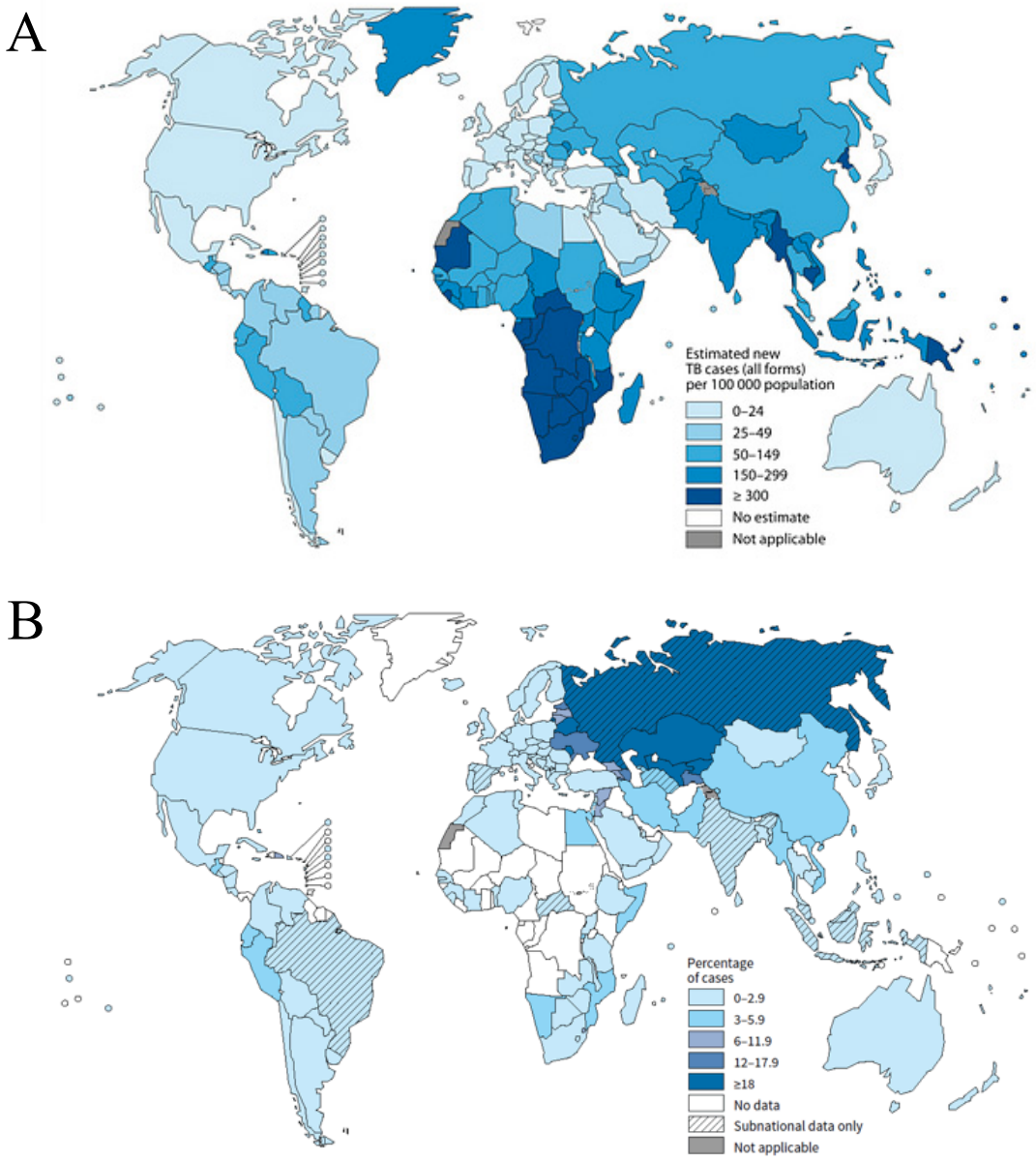


Figure 5.4 (A) Estimated number of new tuberculosis cases diagnosed in 2011. Colours depicted show the number of cases per 100 000 people (B) The number of new tuberculosis cases that are multi-drug resistant (the year the percentage was reported varies from country to country and the colours represent the percentage of new tuberculosis cases reported as being multi-drug resistant).[37]

Genotyping of the bacteria in an attempt to elucidate those antibiotics which are effective before treatment has demonstrated limited success. It has been shown that

strains of *M. tuberculosis* that are supposed to be susceptible to an antibiotic (after genotyping) can grow on media containing that antibiotic, and a better mechanism for identifying drug resistance in tuberculosis infection is needed.[38]

Carbohydrate adhesion patterns could provide a better detection system and any changes could help identify differences amongst the drug resistant strains not necessarily present in their genotype. Glycosylation patterns may be different amongst the different strains and presentation of adhesins on the bacterial surface will also be important. The bacteria may all express the same adhesins but as the interactions are reliant on multivalency to achieve strong adhesion it may be that those that present certain adhesins in patches will be reliant on one type of adhesin for infection where as others may rely on other adhesin proteins.

5.2.6 Carbohydrate interactions for bacterial detection

Detection of bacteria through their carbohydrate binding specificity is a promising area of research. As discussed earlier, the interactions between surface carbohydrates and bacterial lectins is a crucial initial step in infection and adhesins are crucial for pathogenicity. Detection of bacteria through these interactions could allow specific detection of pathogenic bacteria.

Initial studies in the use of these interactions as a method for detection have shown some promise. It has been shown that different species show distinct adhesion specificities. For example, *E. coli* strains contain different adhesins, which allow them to target different organs. *E. coli* containing type 1 pili are responsible for initial

colonisation and infection in the bladder whereas those species with P-pili are responsible for infection in the kidneys.[20]

Modification of systems with known carbohydrate binding targets of specific bacteria can allow the production of cheap detection systems. For example, functionalisation of streptavidin coated magnetic beads with biotin-linked galabiose (α-D-Galactose (1-4) β-D-Galactose) resulted in the detection of *Streptococcus suis* (a zoonotic pathogen commonly associated with meningitis and pneumonia in pigs, human infection with this bacteria is life-threatening and an emerging public health concern) down to a detection limit of 10^4 bacterial cells per mL.[20, 39]

Mannose conjugated gold nanoparticles have been shown to selectively detect FimH positive strains of *E. coli* down to a detection limit of 1.5×10^7 colony forming units per mL.[40] This detection was selective for FimH positive strains and did not detect other strains of *E. coli* without this adhesin. This system showed selective detection of uropathogenic *E. coli* through appropriate selection of the carbohydrate unit used in the detection system.

Whilst these systems are all particle-based, it highlights the potential for carbohydrates in detection of both pathogenic bacteria and bacterial toxins. Several examples of bacterial detection through the use of carbohydrate microarrays exist, though many utilise expensive commercially available microarrays and as the importance of appropriate carbohydrate units for detection has been highlighted, this is not always possible if using commercially available surfaces. Thus utilising a

surface functionalisation technique would increase the choice of carbohydrate units allowing for production of a targeted detection system.

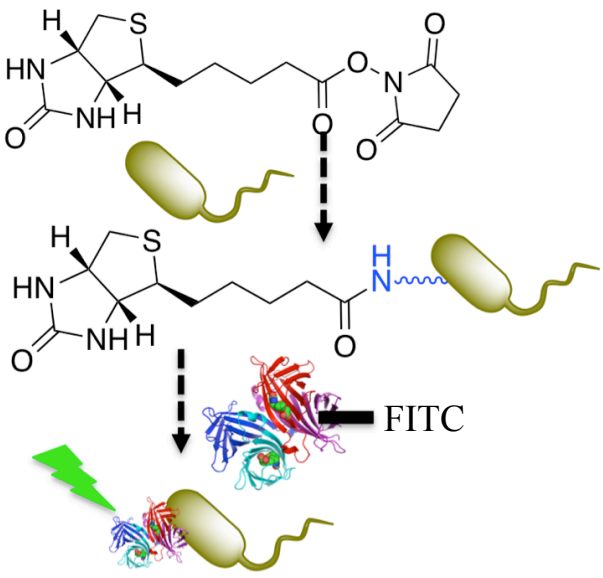
Here we describe the combination of hydrazide coated 96-well plates, which have been further functionalised with a variety of glycans, and a powerful machine learning algorithm for the rapid identification of five bacterial species (both pathogenic and non-pathogenic), Gram-positive and Gram-negative and surrogates for *Mycobacterium tuberculosis*, for which there is an urgent need for low cost, point of care diagnostics to address the emerging resistance time bomb.

5.3. Results and discussion

The key aim of this work was to identify whether differential binding to carbohydrate functionalised Carbo-BIND^{TM} surfaces can be used as a rapid tool for identifying bacterial species.

5.3.1 Bacterial cell membrane labelling

In order to develop a detection system, bacteria first needed to be labelled, which was achieved using a two-step process. Initially bacterial cells were biotinylated through reaction of primary amines found in proteins in the cell membrane with (+)- Biotin *N*-hydroxysuccinimide ester (Biotin-NHS). Biotinylated cells were then incubated with fluorescein conjugated to streptavidin resulting in fluorescent labelling of the cell membrane (Scheme 5.1).



Scheme 5.1 Reaction of Biotin-NHS with primary amines (found in proteins on the cell membrane of bacterial cells) results in the formation biotinylated proteins and *N*-hydroxysuccinimide.

Labelling of all the bacterial species was confirmed using a number of techniques. Biotinylation of the cells was confirmed through detection of a side product of the reaction between Biotin-NHS and the primary amine groups. During this process, *N*-hydroxysuccinimide is produced as a side product. This compound shows absorbance in the region of 260-280 nm and as such samples were removed and absorbance readings were recorded. Absorbance was found to increase for after incubation (Figure 5.5).

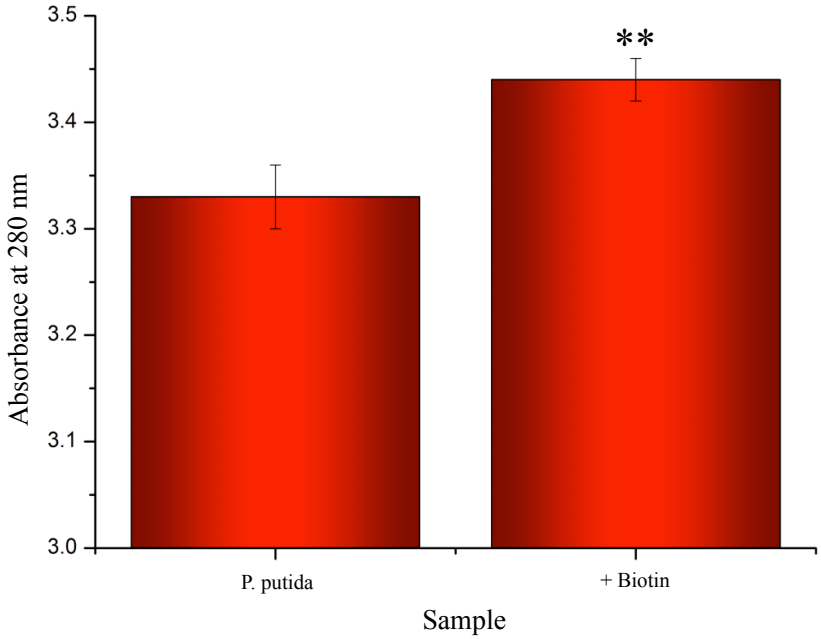


Figure 5.5 Absorbance at 280 nm for *P. putida* after a 2.5 hour incubation with Biotin-NHS at room temperature was significantly higher than an unlabelled sample. Bars represent the average absorbance of three repeats and the error bars represent the standard error. Significance is indicated by an asterix above the bar where ** indicates $p<0.01$ as determined by a Student's t-test.

After incubation of the biotinylated cells with the fluorescein-streptavidin conjugate the labelled cells were viewed under ultraviolet light as fluorescein is known to

fluoresce at this wavelength. Prior to washing, the solution containing the biotinylated bacteria and the fluorescein-streptavidin conjugate appeared yellow when compared to biotinylated bacterial cells (Figure 5.6A, left and right respectively), when viewed under UV light the sample containing the fluorescein-streptavidin fluoresced (Figure 5.6B). After washing three times with PBS to remove unconjugated fluorescein-streptavidin neither the fluorescein labelled cells or the biotinylated sample were found to fluoresce under UV light (Figure 5.6C). Pleasingly the fluorescein labelled cells did appear a more yellow colour when compared to biotinylated cells (Figure 5.6D) indicating the success of the labelling procedure.

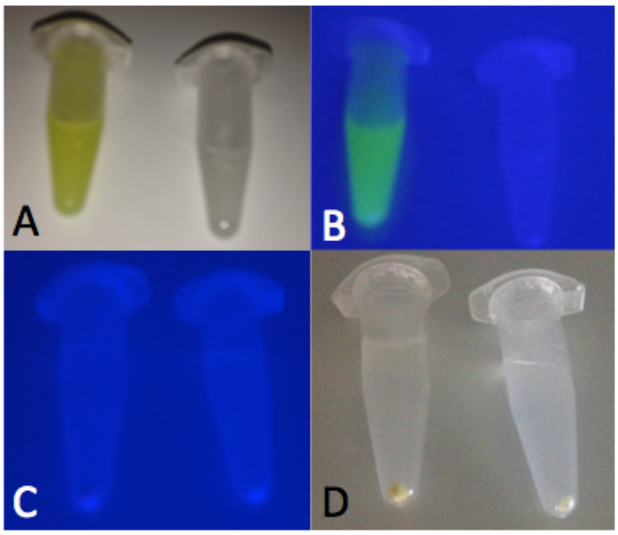


Figure 5.6 Samples under normal lighting (A and D) and UV lighting (B and C). A and B represent labelled (left) and unlabelled (right) samples under normal and UV lighting (in solution). The labelled sample glows under UV light showing the presence of fluorescein. Samples pictured in C and D are the same samples after 3 wash steps to remove unbound fluorescein. As can be seen they do not glow under UV lighting but the bacteria in the labelled sample under normal lighting (D, left) is visibly more yellow than unlabelled (D, right). All images shown are for Top10 (OD=1).

This was then further confirmed by comparison of the fluorescence intensity of a blank well, biotinylated bacteria and bacteria after the addition of the avidin-fluorescein conjugate. The fluorescence was found to be significantly higher for those bacterial cells that had been incubated with the avidin-fluorescein conjugate when compared to either the biotin labelled cells or the blank well indicating the success of the fluorescence step (Figure 5.7).

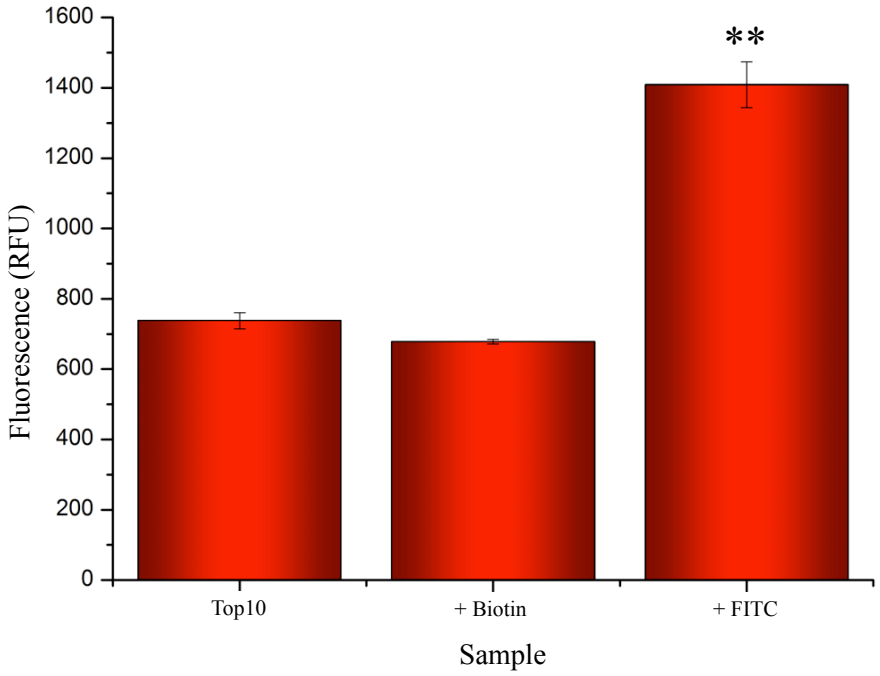


Figure 5.7 Comparison of the fluorescence of a blank unfunctionalised well to a well after incubation with biotinylated bacterial cells (and subsequent washing and removal of unbound cells) or fluorescein conjugated bacterial cells shows a significantly higher fluorescence reading for those that had been incubated with the fluorescein-avidin conjugate. Bars are the average of four replicates, the error bars represent the standard error. A Student's t-test was performed to determine differences, ** represents $p<0.01$.

5.3.2 Bacterial differential binding

Binding of 5 different bacterial species to 9 functionalised surfaces was analysed. The bacterial species chosen were; two strains of *E. coli* Top10 and K12, *Pseudomonas putida*, *Mycobacterium smegmatis* and *Mycobacterium marinum*. Of the two *E. coli* strains, Top10 is FimH negative and K12 is FimH positive and thus pathogenic. *P. putida* is similar in nature to *Pseudomonas aeruginosa*, which is the pathogenic organism responsible for pneumonia (particularly in cystic fibrosis patients).[41, 42] *M. marinum* and *M. smegmatis* are both model organisms for *Mycobacterium tuberculosis,* the causative agent of tuberculosis, with *M. marinum* sharing around 3000 orthologs with over 85 % amino acid identity with *M. tuberculosis*.[43] The species assessed include Gram-positive and Gram-negative species, pathogenic and non-pathogenic species and different strains for one species of bacteria.

The binding profiles produced after incubation of fluorescently labelled bacteria with the surfaces revealed unique binding profiles for each of the bacterial species (Figure 5.8). However most of the differences observed are within error or other species for each surface and so it would be challenging to identify an unknown sample based on its binding to these surfaces. As such linear discriminant analysis was utilised to improve the discrimination.

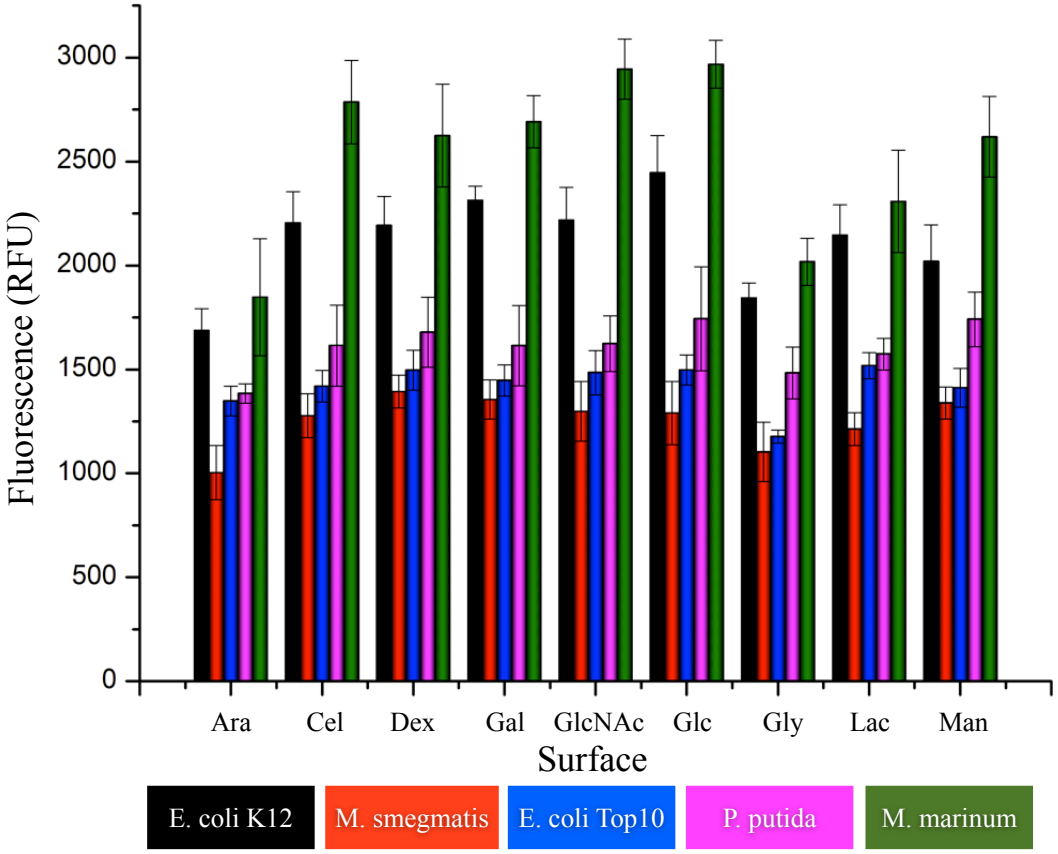


Figure 5.8 Binding profiles for a variety of carbohydrate functionalised surfaces with one of five fluorescently labelled bacterial species: *E. coli* K12 (black), *Mycobacterium smegmatis* (red), *E. coli* Top10 (blue), *Pseudomonas putida* (pink) or *Mycobacterium marinum* (green). The carbohydrate surfaces analysed are as follows; L-arabinose (Ara), cellobiose (Cel), dextran (Dex), D-galactose (Gal), *N*-Acetyl-D-glucosamine (GlcNAc), D-glucose (Glc), DL-glyceraldehyde (Gly), lactose (Lac) and D-mannose (Man). Bars represent an average of 8 replicates and the error bars represent the standard error.

5.3.3 Specificity of bacterial binding

Specificity of bacterial binding to the surfaces was determined by addition of a serial dilution of bacteria to a sugar coated surface and then measurement of fluorescence.

The bacterial binding at various concentrations confirmed that bacterial adhesion was specific and not random deposition (Figure 5.9).

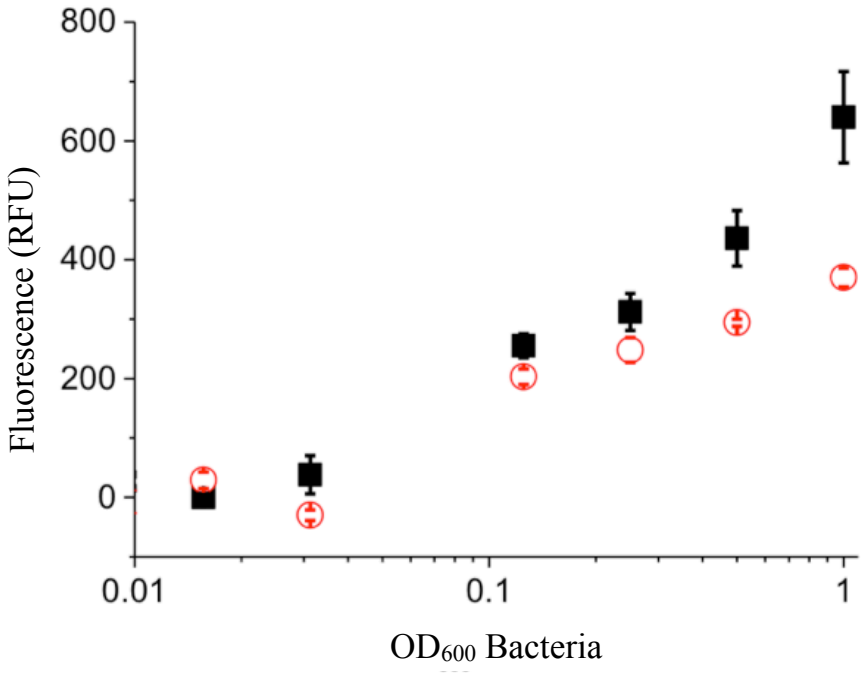


Figure 5.9 Binding of a serial dilution of *E. coli* Top10 (red circles) and K12 (black squares) to a glucose functionalised surface showed that the greatest binding of bacteria corresponded to the greatest concentration of added bacteria. Every point represents the average of 3 measurements and the error bars represent the standard deviation.

To further determine if the affect was specific and not due to charge interactions with the surface, bacteria were incubated with a glucose surface in the presence of glucose in the buffer (from 100 mM) as an inhibitory compound. High glucose showed the lowest bacterial binding and *vice versa* thus indicating a specific interaction (Figure 5.10).

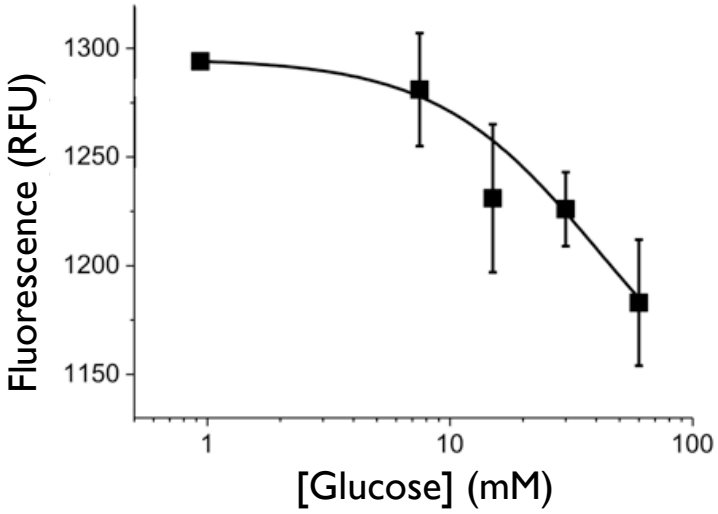


Figure 5.10 Binding of *M. smegmatis* to a glucose functionalised surface in the presence of inhibitory glucose in the buffer (a serial dilution of glucose from 100 mM in PBS). Each point is the average of 3 repeats and the error bars represent the standard error.

5.3.4 Bacterial species discrimination

The binding profiles were utilised as a training matrix for linear discriminant analysis (LDA). This analysis separates the data based on class in order to achieve the best possible separation between all classes of data in the training matrix (see Chapter 2). All five bacterial species analysed were shown to have excellent resolution between species in the model produced (Figure 5.11).

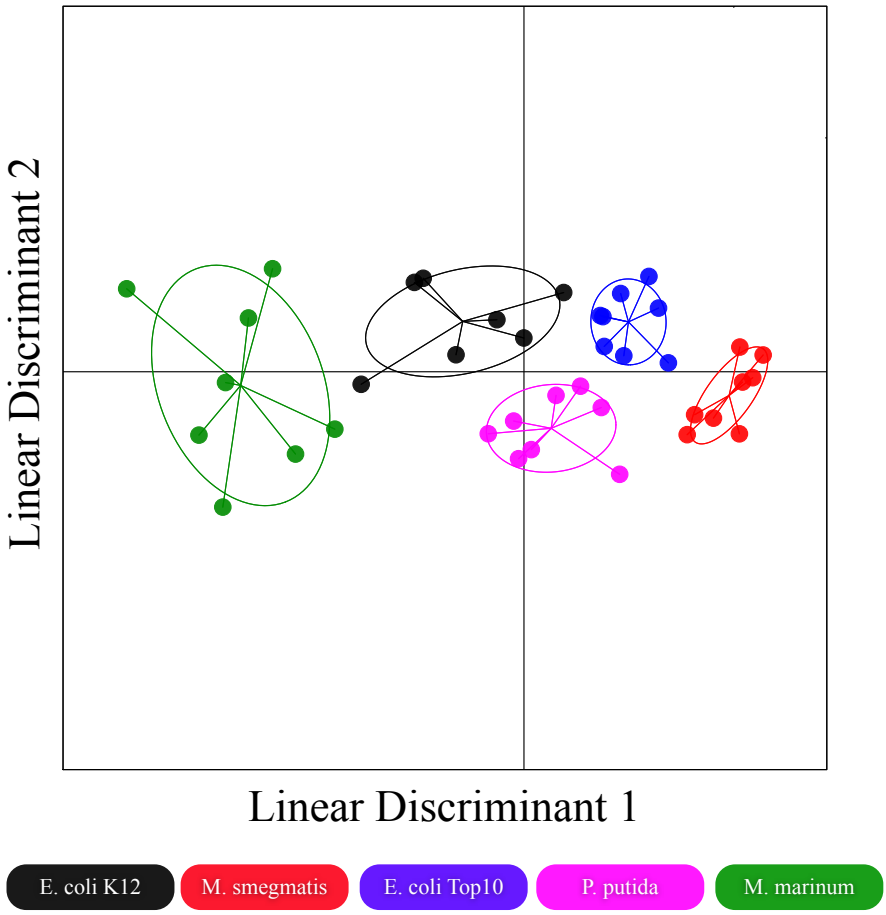


Figure 5.11 The binding profiles were used to generate a linear discriminant model, which showed good discrimination between bacterial species and was able to predict unknown samples with an accuracy of 92 %. Each point represents one full binding profile, the colour represents the species responsible and the ellipses represent the standard deviation from the average for each species.

The validity of the model was further assessed by using a 'leave one out' protocol, in which each binding profile is removed from the training matrix in an iterative manner and the LDA model produced was then used to predict the bacterial species responsible for that profile. This was performed for all repeats, all bacterial species and the model was found to correctly identify the bacteria responsible 92 % of the time. The only false classifications were classifications of K12 as Top10 and as both of these are strains of the same species that is not surprising.

Random Forest was also used as a comparative classification technique, but this produced worse classification than the LDA model. The overall correct classification after a 'leave one out' analysis was performed was 71 % and the only bacterial species correctly predicted for all samples was *M. smegmatis* (Figure 5.12).

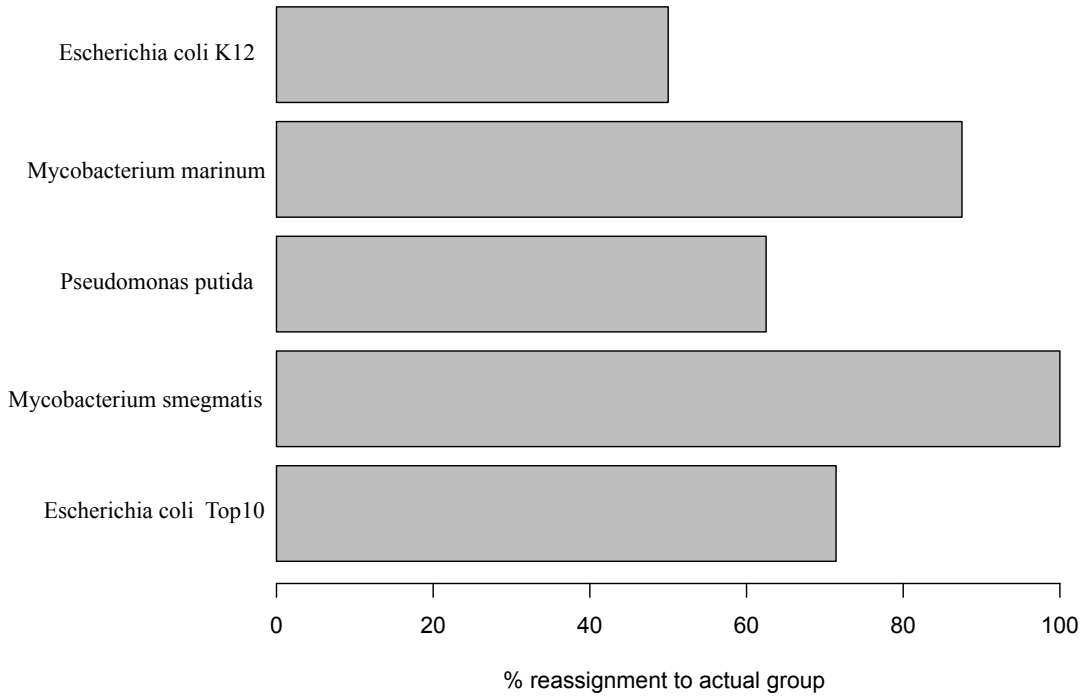


Figure 5.12 The binding profiles were used to generate a Random Forest model, which showed poor discrimination between bacterial species and was only able to predict unknown samples with an accuracy of 71 %. Only *M. smegmatis* was correctly predicted 100 % of the time for all samples.

5.3.4 Blind sample identification

As the LDA model showed greater predictive power than the model produced with Random Forest, this model was used to determine the bacterial species responsible for

a blind sample of one of the five bacterial species in the model. Addition of the sample to each surface in triplicate was used to determine an average reading for the binding of the sample to the surfaces. The response was correctly classified by the LDA model as *E. coli* Top10 with 96 % certainty. The turquoise dot in Figure 5.13 represents the blind sample and whilst it appears to fall between several categories it should be noted that the model shown is only a 2 dimensional representation of a multidimensional model and the certainty in the classification was still very high.

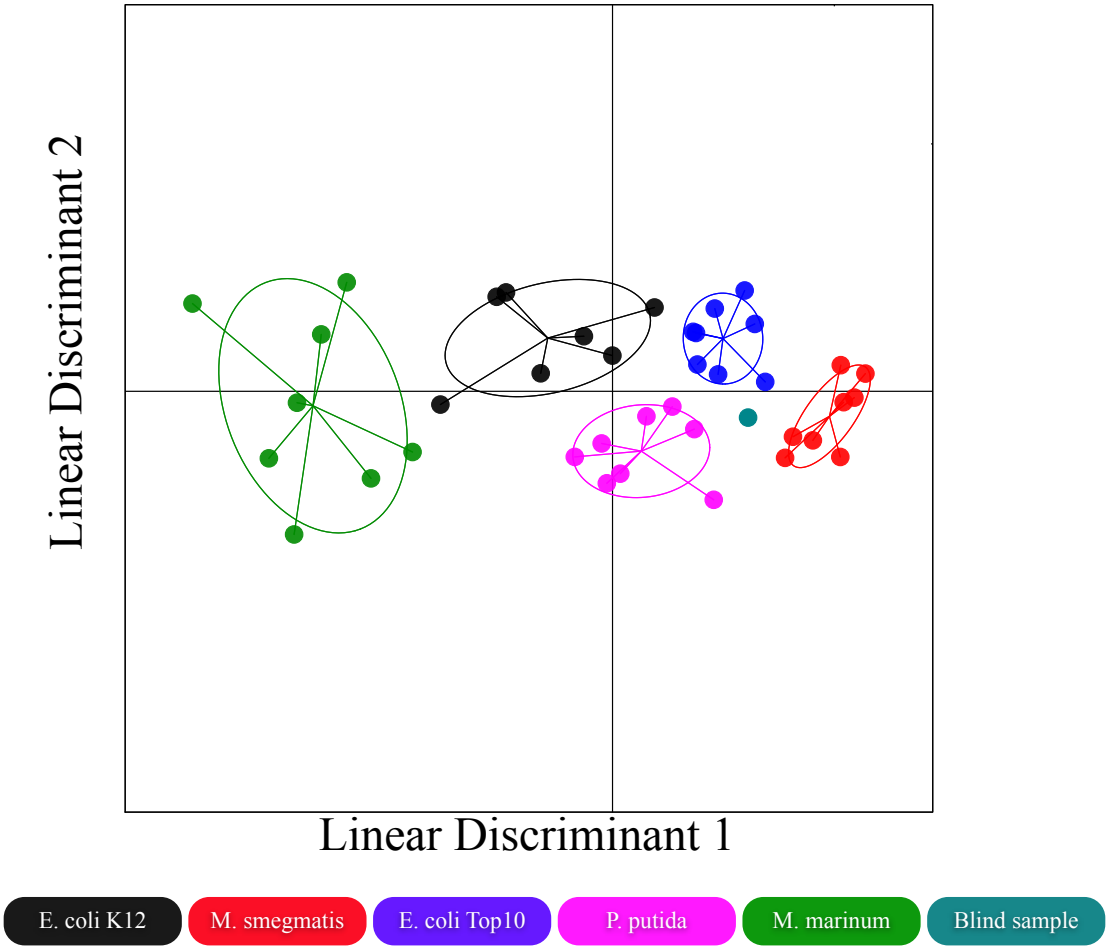


Figure 5.13 The LDA model was used to classify a blind sample of one of the 5 bacterial species present in the model. The turquoise circle represents the average profile of the blind sample added in triplicate to every surface. The blind sample was classified as being *E. coli* Top10 with 96 % confidence.

5.4 Conclusions

Antibiotic resistance is a growing global health concern and the numbers of deaths associated with it are on the rise, coinciding with a reduction in the number of new antibiotics being identified each year. There is a concern that without a change in current medical practice there will be a return to the 'dark age' of medicine during which procedures that are standard with only limited risk currently will become far too risky in the future due to the high chance of infection.

Current bacterial identification methods either involve culturing, which is time consuming and thought to be only applicable to around 1 % of all bacterial species, or expensive PCR based techniques. Owing to these limitations, antibiotics are often prescribed in the absence of microbial information. There is a high likelihood that antibiotics will be prescribed when the bacteria causing the infection are resistant to that particular antibiotic, this would also further propagate the spread of resistance to other bacterial species which may be present in the body (although not at a clinically pathogenic level). The development of rapid point-of-care diagnostics tools is an important area of research in combating this.

We have shown that by combining carbohydrate functionalised 96-well plates with a powerful machine-learning algorithm we can exploit the mechanism through which bacteria achieve their initial adhesion (prior to establishing infection and biofilm formation) in order to detect bacterial species present in samples. The binding of five bacterial species to nine different functionalised surfaces was used to develop a linear discriminant analysis model that could identify bacterial species responsible for carbohydrate binding profiles with 92 % accuracy. The model was then utilised to

correctly classify the bacterial species present in a blind sample as *Escherichia coli* Top10 with 96 % certainty. Whilst this is only a limited number of bacterial species, it highlights the potential for this method in rapid point-of-care species identification as the entire procedure from labelling to identification can be achieved in under five hours. Furthermore, it requires no prior knowledge of the bacterial species to achieve in that all steps are identical for all bacterial species.

5.5 Materials and methods

Materials: All chemicals were used as supplied unless otherwise stated. Corning® 96 well clear flat bottomed polystyrene Carbo-BIND™ microplates, ExtrAvidin®-FITC, (+)-Biotin *N*-hydroxysuccinimide ester, DMSO, L-arabinose, D-(+)-cellobiose, Dextran from *Leuconostoc mesenteroides*, D-(+)galactose, *N*-Acetyl-D-glucosamine, α-D-glucose, DL-glyceraldehyde, lactose, D-(+)mannose, phosphate buffered saline tablets, sodium acetate anhydrous, acetic acid glacial and aniline (99.5 % ACS reagent) were all purchased from Sigma-Aldrich. 100 mM acetate buffer (pH 5.5) with 1 mM aniline was prepared in 200 mL of milliQ water (with a resistivity >18 $M\Omega.cm^{-1}$). K12 JM109 (referred to as K12 in the text) and Top10 were both grown in LB media from frozen stocks prior to use.

Functionalisation of hydrazide surfaces: 100 μL of 30 mM sugar solutions (in 100 mM acetate buffer with 1 mM aniline, pH 5.5) was added to each well of a Carbo-BIND 96-well plate. Plates were then covered in foil and incubated at 50 °C for 24 hours. After incubation, any unbound solution was removed and the well washed thoroughly with PBS three times. Plates were then either used immediately or stored at -20 °C prior to their use.

Biotinylation of bacterial species: Bacterial species in cell culture media were centrifuged at 6800 rpm for 10 minutes and then resuspended in PBS with a final OD of 1. Then 2.5 μL of (+)-Biotin *N*-hydroxysuccinimide ester (10 $mg.mL^{-1}$ in DMSO) was added to every 1 mL of bacterial solution before incubation at room temperature for 2.5 hours. A 200 μL sample was then removed and absorbance at 280 nm measured using a BioTek Synergy HT multi-detection microplate reader to determine

the success of the biotinylation procedure. The remaining bacterial solutions were then centrifuged at 6800 rpm for 10 minutes and resuspended in PBS three times to remove unbound (+)-Biotin *N*-hydroxysuccinimide ester.

Carbohydrate binding assay: 50 μL of biotin labelled bacterial cells was added to every sugar functionalised surface in triplicate before incubation at 37 °C for 30 minutes. Unbound solution is then removed from each well prior to washing three times with PBS. 100 μL ExtrAvidin®-FITC (in a 1 in 200 dilution in PBS) was then added to every well prior to incubation at 37 °C for 1 hour. Unbound solution was then removed and each well washed thoroughly three times with PBS before fluorescence readings were taken using a BioTek Synergy HT multi-detection microplate reader with excitation and emission wavelengths of 485 and 528 nm respectively.

Linear discriminant analysis: Every bacterial species/strain was added to every surface as described in the carbohydrate binding assay section. This was repeated 8 times for K12, Top10, *M. smegmaits*, *P. putida* and *M. marinum* to generate a training matrix, which was then subjected to classical linear discriminant analysis using the 'dapc' function in the 'adegenet' package (version 1.4-2)[44] in the open source statistical package R (version 3.1.3). [45]

Random Forest model: The carbohydrate binding profiles were used to produce a random forest model using the 'RandomForest' function (version 4.6.10) in the open source statistical package R (version 3.1.3). The model produced was the average of

500 trees and the percentage of correct reassignment for each bacterial species was calculated.[45]

Blind culture identification: A blind culture of one of the bacterial species was prepared and biotinylated as described. Biotinylated samples were then added to all the functionalised surfaces in triplicate and incubated at 37 °C for 30 minutes. Unbound solution is then removed from each well prior to washing three times with PBS. 100 μL ExtrAvidin®-FITC (in a 1 in 200 dilution in PBS) was then added to every well prior to incubation at 37 °C for 1 hour. After fluorescence measurements were taken as described the average binding profile of the blind sample was classified by using the 'predict.dapc' function in the adegenet package.

5.6 References

1. A. Zaidi, W. Huskins, D. Thaver, Z. Bhutta, Z. Abbas and D. Goldmann, *Lancet*, 2005, **365**, 1175-1188.
2. M. T. Holden, L. Y. Hsu, K. Kurt, L. A. Weinert, A. E. Mather, S. R. Harris, B. Strommenger, F. Layer, W. Witte, H. de Lencastre, R. Skov, H. Westh, H. Zemlickova, G. Coombs, A. M. Kearns, R. L. Hill, J. Edgeworth, I. Gould, V. Gant, J. Cooke, G. F. Edwards, P. R. McAdam, K. E. Templeton, A. McCann, Z. Zhou, S. Castillo-Ramirez, E. J. Feil, L. O. Hudson, M. C. Enright, F. Balloux, D. M. Aanensen, B. G. Spratt, J. R. Fitzgerald, J. Parkhill, M. Achtman, S. D. Bentley and U. Nubel, *Genome Res*, 2013, **23**, 653-664.
3. C. M. Surawicz, L. J. Brandt, D. G. Binion, A. N. Ananthakrishnan, S. R. Curry, P. H. Gilligan, L. V. McFarland, M. Mellow and B. S. Zuckerbraun, *Am J Gastroenterol*, 2013, **108**, 478-498.
4. J. M. Q. Hiram, F. Ricardo, C. Miguel, A. Jose, C. B. Maria, C. M. S. Juan, R. Pedro Torellas and C. Veronica, in *B48. Case Reports on Tuberculosis*, American Thoracic Society, Editon edn., 2015, pp. A3275-A3275.
5. N. Rezaei, A. Aghamohammadi, D. Mansouri, N. Parvaneh and J. L. Casanova, *Expert Rev Clin Immunol*, 2011, **7**, 129-131.
6. C. M. Grossman, *Ann Intern Med*, 2008, **149**, 135-136.
7. D. Bethune, R. Blowers, M. Parker, E. Pask, O. Jessen, K. Rosendal, P. Bulow, V. Faber, K. Eriksen and K. Hiramatsu, *Brit Med J*, 1961, **124**, 124-125.
8. Y. Morita, J. Tomida and Y. Kawamura, *Front Microbiol*, 2012, **3**, 408.
9. T. Chuma, D. Miyasako, H. Dahshan, T. Takayama, Y. Nakamoto, F. Shahada, M. Akiba and K. Okamoto, *Front Microbiol*, 2013, **4**, 113.

10. K. Shi, S. J. Caldwell, D. H. Fong and A. M. Berghuis, *Front Cell Infect Microbiol*, 2013, **3**, 22.

11. H. Ochman, J. G. Lawrence and E. A. Groisman, *Nature*, 2000, **405**, 299-304.

12. J. O'Neil, *Antimicrobial resistance: Tackling a crisis for the health and wealth of nations.*, http://amr-review.org/home.

13. B. Spellberg, R. Guidos, D. Gilbert, J. Bradley, H. W. Boucher, W. M. Scheld, J. G. Bartlett, J. Edwards and t. I. D. S. o. America, *Clin Infect Dis*, 2008, **46**, 155-164.

14. S. Tan, S. Chew, S. Tan, M. Givskov and L. Yang, *Curr Opin Biotech*, 2014, **26**, 1-6.

15. R. Koch and T. Brock, *Milestones in microbiology: 1556 to 1940*, 1884.

16. J. W. Costerton, J. C. Post, G. D. Ehrlich, F. Z. Hu, R. Kreft, L. Nistico, S. Kathju, P. Stoodley, L. Hall-Stoodley, G. Maale, G. James, N. Sotereanos and P. DeMeo, *FEMS Immunol Med Mic*, 2011, **61**, 133-140.

17. S. Pasquaroli, G. Zandri, C. Vignaroli, C. Vuotto, G. Donelli and F. Biavasco, *J Antimicrob Chemoth*, 2013, **68**, 1812-1817.

18. E. L. Doughty, M. J. Sergeant, I. Adetifa, M. Antonio and M. J. Pallen, *PeerJ*, 2014, **2**, e585.

19. M. J. Pallen, *Parasitology*, 2014, **141**, 1856-1862.

20. N. P. Pera and R. J. Pieters, *MedChemComm*, 2014.

21. A. K. Ojha, A. D. Baughn, D. Sambandan, T. Hsu, X. Trivelli, Y. Guerardel, A. Alahari, L. Kremer, W. R. Jacobs and G. F. Hatfull, *Mol Microbiol*, 2008, **69**, 164-174.

22. CDC, http://www.cdc.gov/TB/statistics/.

23. L.-C. Xu and C. A. Siedlecki, *Acta biomater*, 2012, **8**, 72-81.

24. D. Monroe, *PLoS biology*, 2007, **5**, e307.

25. C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357-2364.

26. S. J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew Chem Int Ed*, 2012, **51**, 7812-7816.

27. S. G. Spain and N. R. Cameron, *Polym Chem*, 2011, **2**, 60-68.

28. M. Jones, L. Otten, S.-J. Richards, R. Lowery, D. Phillips, D. Haddleton and M. Gibson, *Chem Sci*, 2014, **5**, 1611-1616.

29. M. A. McGuckin, S. K. Lindén, P. Sutton and T. H. Florin, *Nat Rev Micro*, 2011, **9**, 265-278.

30. A. Audfray, A. Varrot and A. Imberty, *Comptes Rendus Chimie*, 2013, **16**, 482-490.

31. M. Hartmann, H. Papavlassopoulos, V. Chandrasekaran, C. Grabosch, F. Beiroth, T. K. Lindhorst and C. Röhl, *FEBS letters*, 2012, **586**, 1459-1465.

32. S. Sakarya and S. Oncu, *Med. Sci. Monit*, 2003, **9**.

33. Z. Han, J. S. Pinkner, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano and S. J. Hultgren, *J Med Chem*, 2010, **53**, 4779-4792.

34. A. Zumla, R. Atun, M. Maeurer, P. Mwaba, Z. Ma, J. O'Grady, M. Bates, K. Dheda, M. Hoelscher and J. Grange, *Trop Med Int Health*, 2011, **16**, 79-83.

35. A. A. Velayati, M. R. Masjedi, P. Farnia, P. Tabarsi, J. Ghanavi, A. H. ZiaZarifi and S. E. Hoffner, *Chest J*, 2009, **136**, 420-425.

36. V. S. Govender, S. Ramsugit and M. Pillay, *Microbiology*, 2014, **160**, 1821-1831.

37. O. World Health, *Global tuberculosis control: WHO report 2010*, World Health Organization, 2010.

38. S. M. Fortune, *Journal of Infect Dis*, 2012, jis603.

39. N. P. Pera, A. Kouki, S. Haataja, H. M. Branderhorst, R. M. J. Liskamp, G. M. Visser, J. Finne and R. J. Pieters, *Org Biomol Chem*, 2010, **8**, 2425-2429.

40. S.-J. Richards, E. Fullam, G. S. Besra and M. I. Gibson, *J Mater Chem B*, 2014, **2**, 1490-1498.

41. A. Wittgens, T. Tiso, T. T. Arndt, P. Wenk, J. Hemmerich, C. Muller, R. Wichmann, B. Kupper, M. Zwick, S. Wilhelm, R. Hausmann, C. Syldatk, F. Rosenau and L. M. Blank, *Microb Cell Fact*, 2011, **10**, 80.

42. N. Hoiby, E. W. Flensborg, B. Beck, B. Friis, S. V. Jacobsen and L. Jacobsen, *Scand J Respir Dis*, 1977, **58**, 65-79.

43. T. P. Stinear, T. Seemann, P. F. Harrison, G. A. Jenkin, J. K. Davies, P. D. Johnson, Z. Abdellah, C. Arrowsmith, T. Chillingworth, C. Churcher, K. Clarke, A. Cronin, P. Davis, I. Goodhead, N. Holroyd, K. Jagels, A. Lord, S. Moule, K. Mungall, H. Norbertczak, M. A. Quail, E. Rabbinowitsch, D. Walker, B. White, S. Whitehead, P. L. Small, R. Brosch, L. Ramakrishnan, M. A. Fischbach, J. Parkhill and S. T. Cole, *Genome Res*, 2008, **18**, 729-741.

44. T. Jombart, C. Collins, P. Solymos, I. Ahmed, F. Calboli and A. Cori, *adegenet: an R package for the exploratory analysis of genetic and genomic data*, http://adegenet.r-forge.r-project.org/.

45. R. D. C. Team, *R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing*, http://www.R-project.org/.

Chapter 6

Differential glycan binding for the detection of *Plasmodium falciparum* infected erythrocytes and identification of drug resistance.

6.1 Abstract

Malaria is one of the main causes of global death from disease and with the spread of resistance to Artemisinin (the basis for most modern malaria treatment) this looks set to remain unchanged. Microscopy currently remains the gold standard for diagnosis of malaria, but this offers no information about potential drug resistance of the parasite thus allowing the spread of drug resistance through administration of ineffective treatment. There are four species of parasite that are responsible for malaria in humans, but most of the deaths are associated with *Plasmodium falciparum* infection due to the complications caused by cytoadhesion in the later stages of the parasite development cycle. This cytoadhesion mechanism is highly complex and it is poorly understood. Here we highlight the use of facile carbohydrate functionalised surfaces for detection of *P. falciparum* infected red blood cells. These differential binding patterns were used in the life-stage specific detection of parasites in a sample and allowed identification of drug resistance strains.

6.2 Introduction

Malaria is one of the main causes of global death from disease with approximately 198 million cases occurring in 2013 resulting in 584,000 deaths predominantly among African children. It is estimated that one African child dies of malaria every minute.[1] Malaria is caused by parasites of the genus *Plasmodium* of which there are five species that can infect humans; *P. falciparum*, *P. malariae*, *P. vivax*, *P. ovale* and *P. knowlsei*. *Plasmodium falciparum* and *Plasmodium vivax* are the main causes of disease and deaths from malaria with *P. falciparum* responsible for almost all deaths associated with malaria.[2]

P. falciparum is the most lethal of the *Plasmodium spp.* that can infect humans as the parasites can infect red blood cells (RBCs) at all stages whereas the other species can often only infect red blood cells at specific erythrocyte developmental stages. For example, *P. vivax* can only infect reticulocytes.[3] *P. falciparum* also has a very high replication rate. Both of these factors result in a higher parasite burden within the blood, which is associated with high mortality levels.[4] In addition to this, whilst both *P. falciparum* and *P. vivax* can cause severe anaemia, only *P. falciparum* can cause cerebral malaria, which has many complications including; hypoglycaemia, respiratory distress and metabolic acidosis.[3]

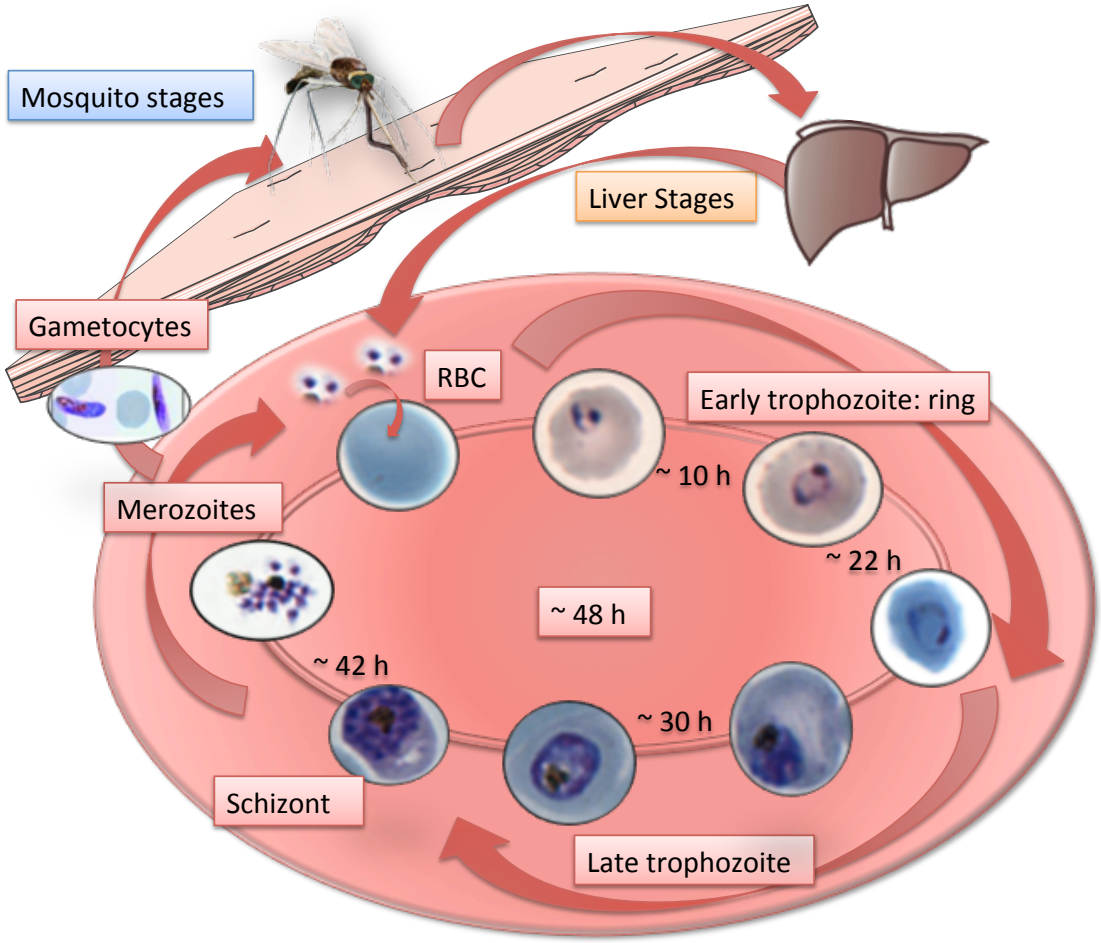


Figure 6.1 *P. falciparum* life cycle. During the ~48 hour intraerythrocytic life stage, a single cell (merozoite) invades a red blood cell (RBC) and takes on a ring-shaped form (early trophozoite: ring) that rapidly expands and drastically changes in morphology resulting a vastly enlarged nucleus, which divides into up to 36 daughter cells. A small fraction of these differentiate into gametocytes that continue the malaria life cycle in the mosquito. Images shown for the intraerythrocytic stages are microscopy images.

Parasites (sporozoites) are mainly injected into the body through the bite of the female *Anopheles* mosquito. Once in the body, sporozoites travel directly to the liver where they enter hepatocytes[5] until the parasites develop into merozoites. Merozoites are then released from the liver where they go on to infect RBCs (Figure 6.1). Within the

RBC, the parasite enters the asexual cycle where initially they are known as the ring stage as the parasite engulfs some cytoplasm upon entering the RBC giving it a ring like appearance (Figure 6.1). Ring stage parasites then develop into trophozoites and finally merozoites (Figure 6.1). These merozoites are released back into the blood stream to infect uninfected RBCs. Some of the asexual parasites go on to form gametocytes, which are essential for transmission of the disease by mosquitoes (Figure 6.1). From initial infection to lysis of the RBC and release back into the blood stream as merozoites takes approximately 48 hours, the final phase coinciding with the characteristic paroxysms associated with malaria (sudden onset of chills, followed by fever and sweating).[6]

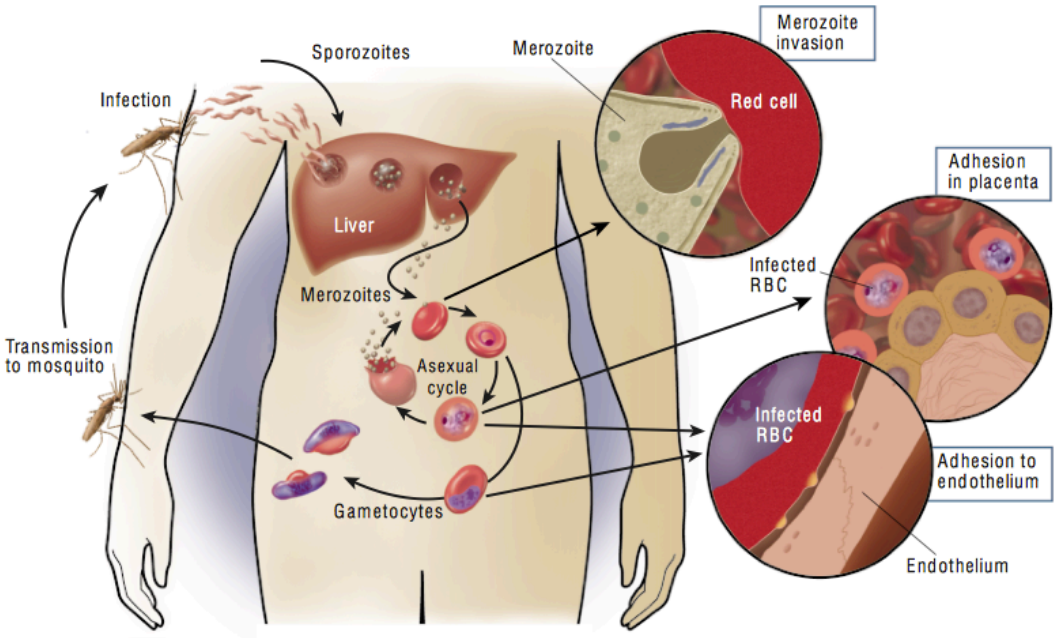


Figure 6.2 Schematic showing the cytoadhesion of *P. falciparum*. Initial infection occurs in the liver where sporozoites develop into merozoites, which go on to infected red blood cells thus beginning the intraerythrocytic life stage. During this stage, infected cells undergo adhesion to the endothelium and the placenta. Reprinted with permission from reference 3.

Once *P. falciparum* parasites reach the mature trophozoite stage the infected RBCs can be cleared by the spleen. To prevent this from occurring *P. falciparum* infected erythrocytes undergo cytoadherance (Figure 6.2) during this phase as is evinced by the fact that none of the cells in this stage of infection will be detected in peripheral circulation in an infected person but ring-stage infected RBCs will be present.[7] RBCs in this advanced stage of the parasite growth cycle undergo membrane alterations resulting in the formation of knob like protrusions which act as points of attachment for these cells to tissue vascular beds and endothelial cells.[8]

The knob like protrusions are the site of localisation of many parasite derived proteins including those from the *var* gene family. The *var* gene family encodes for 200-350 kDa proteins responsible for antigenic variation and cytoadherance of *P. falciparum* and accounts for an estimated 2-6 % of *P. falciparum* genomic DNA with between 50 and 150 copies per parasite. A number of different proteins are encoded by the *var* gene family and differences in gene transcription have been shown to correlate with changes in cytoadherance and antigenic phenotypes (resulting in waves of parasitemia in infected people).

The expressed proteins are collectively known as *P. falciparum* erythrocyte membrane protein 1 (PfEMP1). Many of these proteins have not been fully characterised, but have been shown to have domains homologous to the binding domains of a glycophorin binding molecule and ligands involved in the invasion of Duffy blood group positive erythrocytes.[8-10] Other proteins (which interact with host receptors) not encoded by the *var* gene family have also been indicated to be involved

in cytoadherance but these proteins are relatively consistent amongst parasite lines and don't show antigenic variation.[7]

Cytoadherance to the microvasculature results in reduced blood flow to organs and causes the symptoms associated with severe malaria.[3] Whilst the role of PfEMP1 in cytoadherance of the parasitised RBCs has long been known, it is still not understood why some strains of *P. falciparum* result in sequestering and adhesion in the microvasculature of the brain causing cerebral malaria whereas others are localised to different tissue types.[3] One mechanism through which this could occur is through protein-carbohydrate interactions. Many of the proteins in PfEMP1 have lectin like properties and it is known that a lectin like protein plays a key role in the initial infection of erythrocytes by merozoites.[9-11] It may also highlight a mechanism through which infected RBCs sequester to different tissue types. Whilst endothelial cells all have fairly similar carbohydrates on their surface there are known differences between tissue types both in density of the carbohydrate residues and the linkages involved.[12] For example, whilst many of the endothelial surface carbohydrates contain terminal sialic acid residues, it has been shown that there are tissue differences in whether the sialic acid residues are attached through an α(2-3) linkage or an α(2-6) linkage.[7]

Very little work has been done looking at parasitised red blood cell binding to carbohydrates previously, but they form a potential mechanism for tissue specificity of the parasitised erythrocytes. Some research has been done looking at glucose transport in infected RBCs. As the parasite lives within a host erythrocyte, it is dependent on the hosts ability to provide nutrients, most specifically glucose

(although there is some evidence that fucose can replace glucose in the parasites metabolic pathways). During infection, *P. falciparum* increases glucose transport into the red blood cell by up to 50 fold, this increase is maximal at the trophozoite stage.[13, 14] From as little as 6 hours post infection the very nature of the erythrocyte membrane is modified in order to support this increase.[15] The membrane dramatically increases in permeability to small molecules such as sugars, sugar-alcohols, amino acids and nucleotides all to support the growing parasite.[16] It is possible that these changes in permeability could be detected through the use of carbohydrate microarrays, which may elucidate some of the important carbohydrate residues for RBC sequestering and may also form an important tool for malaria diagnosis.

Malaria must be promptly recognised in order to rapidly begin treatment thus maximising the possibility of a positive patient outcome and preventing further spread. Currently, treatment is based on Artemisinin-based combination therapies, but resistance to Artemisinin is on the rise and threatens global efforts to control and eliminate malaria.[17, 18] As such, rapid diagnosis of malaria and identification of drug resistance is crucial in order to stop the spread of resistance and ensure positive patient outcome.

However, diagnosis of malaria can be challenging and is usually based upon clinical symptoms such as fever, muscle pain and exhaustion, which are not specific to malaria and are common among many other diseases. Currently microscopy remains the gold standard for identification of malaria This method is both cheap and sensitive with a detection limit as low as 4 parasites per μL of blood.[19] However, transport of microscopes to remote areas can be difficult and it takes considerable skill to achieve

such a good level of sensitivity and the diagnostic power of this technique depends on operator performance and so prone to errors.[20] This technique also provides no information about drug resistance and so may increase the likelihood of the spread of resistant malaria due to ineffective treatment.

In addition to this it can be challenging to identify the species involved in the infection through microscopy. The Centre for Disease Control (CDC) provides a detailed chart, which involves looking at characteristics of infection throughout the parasites life-cycle in order to definitively identify species involved.[21] Species identification is crucial as some strains, such as *P. falciparum*, are more dangerous than others. Differential carbohydrate binding may be a mechanism by which species identification can be improved as *P. falciparum* exploits very different binding pathways from the other malaria strains and is also the only species that has a cytoadhesion phase.[22]

Microscopy is not the only diagnostic tool available for the detection of malaria. Immunochromatography based 'dipstick' tests are also available. These techniques detect the presence of malarial antigens in the blood rather than infected RBCs and are rapid and easy to use, but their sensitivity is significantly below that achieved through microscopy (at 100 parasites per μL of blood), the technique is expensive and sensitivity can be affected by adverse field conditions.[20] The output of the test does not provide a quantitative indication of the level of infection (as the strength of the response to the test varies from person to person).[23] The level of antigens in the blood also declines as the infection progresses and this method cannot be used to determine whether a treatment is effective, but can be used to detect recurring waves of

infection. Whilst some immunochromatography methods exist which detect host antibodies, as they remain in the blood stream even after the infection has passed, this method cannot be used to determine an infection in someone who may have suffered a previous infection.[24]

There is also a growing investment in molecular diagnosis tools. These techniques have comparable sensitivity to microscopy and can give quantitative signals, which correlate with level of infection. Additional information such as species and drug resistance can also be determined using this technique thus making it significantly more powerful than microscopy. This technique is currently being used for the tracking of Artemisinin resistant malaria cases.[25] However, this technique requires careful sample preparation, is expensive and the results are often not rapidly available. Technique sensitivity may also suffer due to the field conditions.[26]

All of the current methods that exist have limitations and, with the rise of population movement from areas of high malaria transmission to low transmission due to tourism and socioeconomic factors, has resulted in a rise in the number of cases of malaria with a parallel rise in mortality due to late or misdiagnosis there is a need for a rapid, facile and cheap diagnostics method.[27, 28] It would be advantageous for new methods to be able to detect drug resistance in order to combat its spread.

In this chapter, we describe a number of carbohydrate moieties for which differential binding is observed upon incubation with parasitised RBCs when compared to native uninfected erythrocytes. This is the first time such a technique has ever been used in

the detection of the presence of *P. falciparum* in RBCs. This technique was further expanded to the detection of drug resistant strains of *P. falciparum*.

6.3 Results and discussion

The key aim of this work was to determine differential carbohydrate binding of RBCs infected with *P. falciparum* and native uninfected RBCs as a potential diagnostic tool, to highlight a mechanism through which tissue specific sequestering of *P. falciparum* infected RBCs may occur and to identify whether differing levels of drug resistance can be detected through glycan binding. As such we employed facile coupling chemistry to functionalise commercially available Carbo-BIND™ 96-well plates to generate a number of different glycan functionalised surfaces as previously described.[29]

6.3.1 Level of infection in infected red blood cell samples

Prior to use on surfaces all samples, were checked for contamination and an estimate of the percentage of infected cells was determined (Figure 6.3). The developmental stage of the parasite was also confirmed. Four stages of infected cells were analysed and the level of infection was found to be 10 % for all stages.

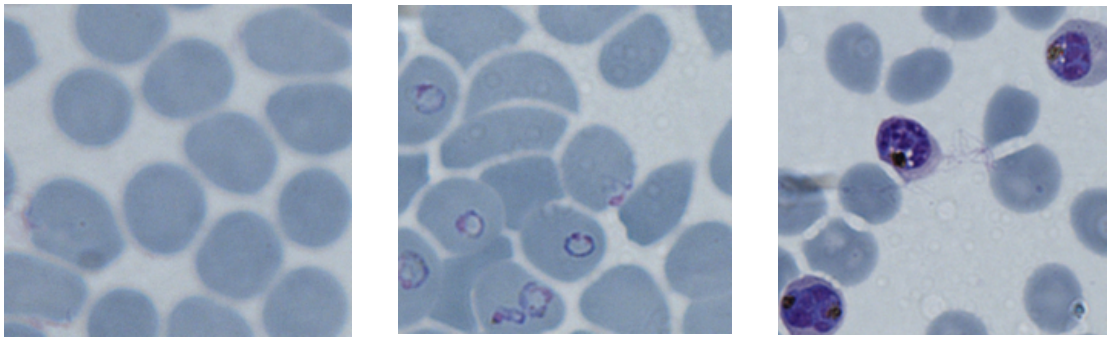


Figure 6.3 Giemsa stained RBCs (viewed under 100 X oil immersion objective) showing uninfected (left), ring stage (centre) and mature trophozoite (right) stages of infection.

6.3.2 Labelling techniques

A number of techniques for the detection of RBCs were analysed. Initially, two label-free methodologies were assessed (Figure 6.4A and 6.4B) before comparing two fluorescent labelling methodologies (Figure 6.4C and D).

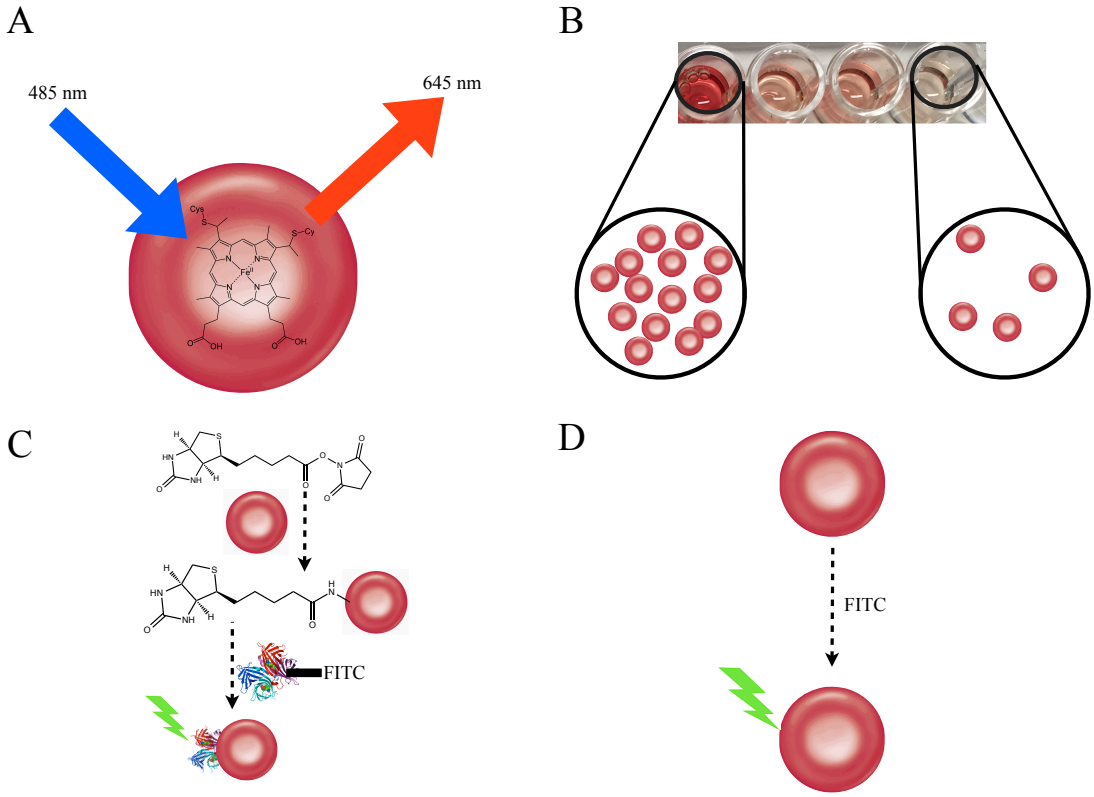


Figure 6.4 Schematic highlighting the four detection strategies assessed. Autofluorescence of haem under certain wavelengths of light (A). At various dilutions of RBCs the red colour changes and this can be used to estimate the RBC concentration (B). Biotinylation of the RBC was followed by addition of fluorescein-labelled streptavidin, which results in fluorescent labelling of the RBC membrane (C). Direct labelling of the RBC membrane through reaction with fluorescein isothiocyanate (FITC) (D).

6.3.2.1 Autofluorescence

Firstly, RBCs undergo a phenomenon known as autofluorescence due to the haem component (Figure 6.5A). However, upon incubation with the surfaces and subsequent wash steps the autofluorescence methodology was found to be insufficient for detecting differences in binding upon addition of red blood cells in a serial dilution to a glucose functionalised surface (Figure 6.5B). This method also had the additional complication that autofluorescence of red blood cells increases with cell age, so all samples added to the surface would have to be the same age in order to ensure that differences observed are due to changes in sugar binding during parasite phases and not to cell age. Additionally, during the trophozoite phase of infection (also known as mature infected cells) the haem component of the red blood cells becomes surrounded by the parasite and its daughter cells. This may result in changes in autofluorescence and as such this technique was not used.

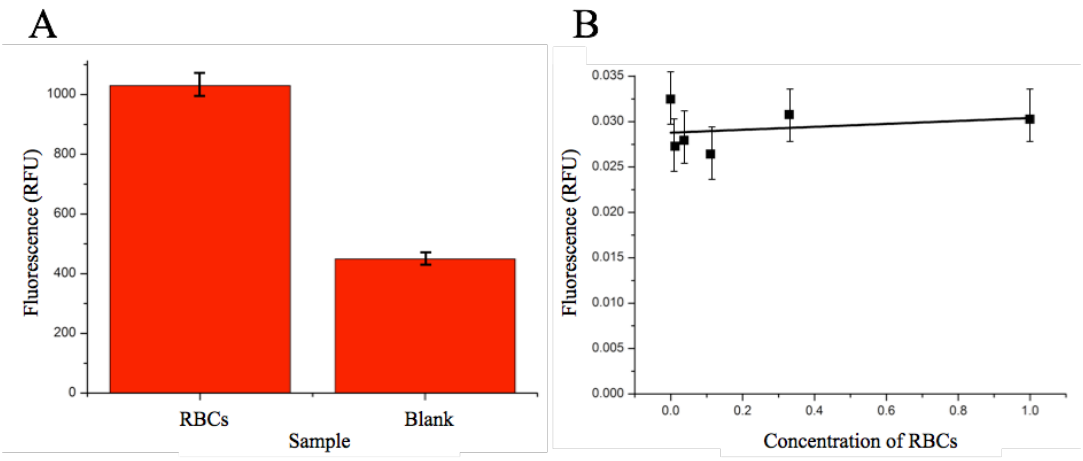


Figure 6.5 Autofluorescence of RBCs where bars represent an average of three measurements and error bars are the standard error (A). Autofluorescence analysis of a serial dilution of red blood cells on a glucose surface (B).

6.3.2.2 Colorimetric analysis

The second label free methodology analysed was the use of ImageJ to compare changes in colour upon binding of red blood cells to the cell surface. An addition of a serial dilution of RBCs to a 96-well plate results in different shades of red, which can be analysed in ImageJ to produce a rough dilution curve (Figure 6.6). Disappointingly, there was no observed colour change after wash steps in sugar functionalised surfaces following addition of red blood cells, thus ruling this out as a potential detection technique.

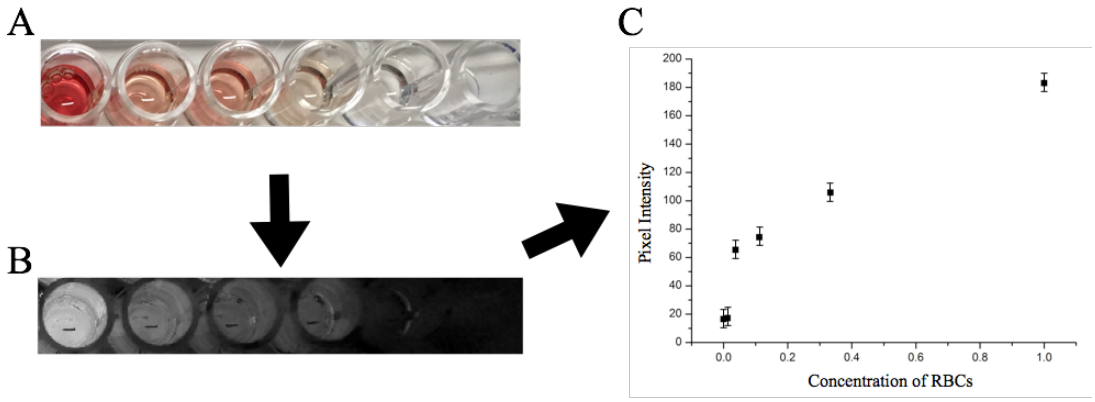


Figure 6.6 A photograph of the colourimetric dilution (A) can be converted into a Hue, Saturation, Brightness (HSB) stack in ImageJ (B). The saturation image in this stack can then be used and the pixel intensity in each well measured to give an approximate measurement of the colour in the well, error bars represent the standard error (C).

6.3.2.3 NHS-Biotin:Avidin-FITC labelling

As neither of the label free methodologies were successful, two labelling methods were utilised in order to aid in the detection of differential binding. Initially the NHS-Biotin:Avidin-FITC system for labelling of the cell surface was used for cell

detection. This method has been successfully utilised in Chapter 5 for the detection of bacterial cells. This method failed in the labelling of RBCs. This was due to the fact that the reaction involves a 2.5 hour incubation step with the NHS-Biotin. For this duration the parasites would have died if they had been in PBS and thus had to be kept in cell culture media. However, cell culture media also contains primary amines, which would have competed with those in proteins on the cell surface for the NHS-Biotin label thus reducing the overall efficiency of the labelling of the cells. Increasing the concentration of NHS-Biotin added to the cells may have improved the labelling efficiency but RBCs are very sensitive to DMSO, in which the NHS-Biotin is dissolved, so addition of more NHS-Biotin would have resulted in increasing levels of cell lysis.

6.3.2.4 FITC labelling

Finally, direct labelling of RBCs with fluorescein isothiocyanate (FITC) was used. This technique is commonly used for labelling RBCs in blood flow experiments in animals and it was theorised that it would be successful in cell culture media.[30] A serial dilution of FITC labelled RBCs added to glucose and galactose surfaces showed differential binding that was detectable (Figure 6.7).

Further to this, successful FITC labelling of each stage was also confirmed through the use of fluorescence microscopy (Figure 6.8). There appeared to be some differences in the labelling efficiency seen in the fluorescence microscopy images with parasitized cells labelling much more than uninfected cells, which was further confirmed by the measurement of fluorescence of samples. All samples were

corrected to account for labelling efficiency and this labelling method was utilised for all further experiments.

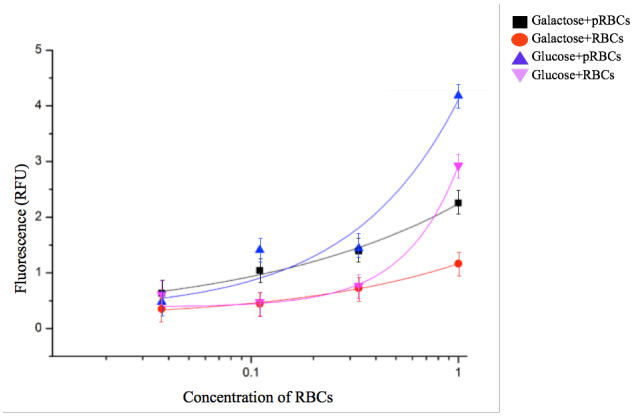


Figure 6.7 Addition of a serial dilution of FITC labelled uninfected (RBC) and trophozoite stage infected (pRBC) cells to a glucose or galactose surface showed increases in binding as concentration increased. The concentration shown is from 1, which represents the concentration of cells in a pellet of RBCs (~1.9x10^6 cells per µL) and the serial dilution was performed in cell culture media. Error bars represent machine error.

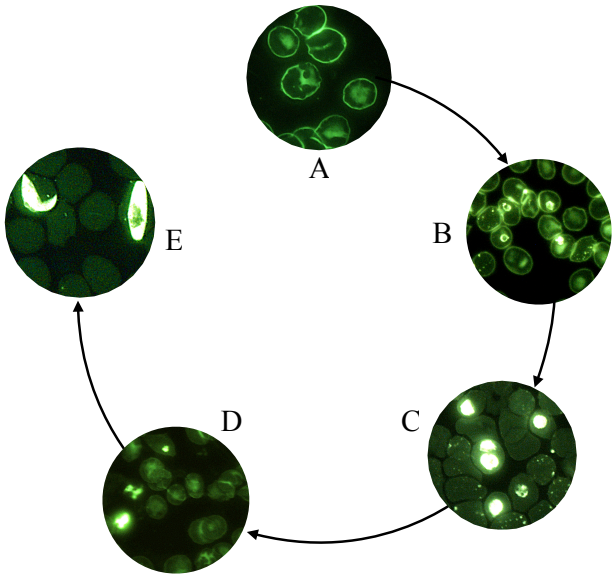


Figure 6.8 Fluorescence microscopy images of FITC labelled uninfected RBCs (A), ring stage infected RBCs (B), trophozoite stage infected RBCs (C), early stage gametocytes (D) and late stage gametocytes (E).

6.3.3 Differential binding on carbohydrate surfaces

With the RBC labelling method in hand, surfaces were functionalised with six different carbohydrates; D-galactose, D-glucose, *N*-Acetyl-D-Galactosamine (GalNAc), *N*-Acetyl-D-Glucosamine (GlcNAc), α(2-3) Sialyllactose and α(2-6) Sialyllactose. Galactose was selected as permeability of the RBC membrane to small molecules is known to drastically increase upon infection. Glucose would also be affected by this, and it is known that the intraerythrocytotic stage of parasite development is characterised by a 50-100 fold increase in glucose absorption to meet the metabolic demands of the parasite.[13, 14] This increase is maximal during the trophozoite stage. Rosetting (the process whereby infected RBCs bind to uninfected RBCs) is known to occur during the later stages of infection, GlcNAc, GalNAc and galactose were selected as these saccharides are found on the surface of RBCs. Both α(2-3) sialic acid and α(2-6) sialic acid residues are found on the surface of endothelial cells in varying ratios. As late stage infected RBCs undergo cytoadhesion to the endothelial surface of the vasculature α(2-3) sialyllactose and α(2-6) sialyllactose were also selected to represent α(2-3) sialic acid and α(2-6) sialic acid residues.

RBCs infected with *Plasmodium falciparum* 3D7, at four different stages in the intraerythrocytotic development cycle of the parasite and uninfected RBCs were fluorescently labelled and incubated with a carbohydrate functionalised Carbo-BIND™ plate. Differential binding was observed for many different carbohydrates at different stages in the development cycle despite the level of parasitemia being only around 10 % for all stages (Figure 6.9).

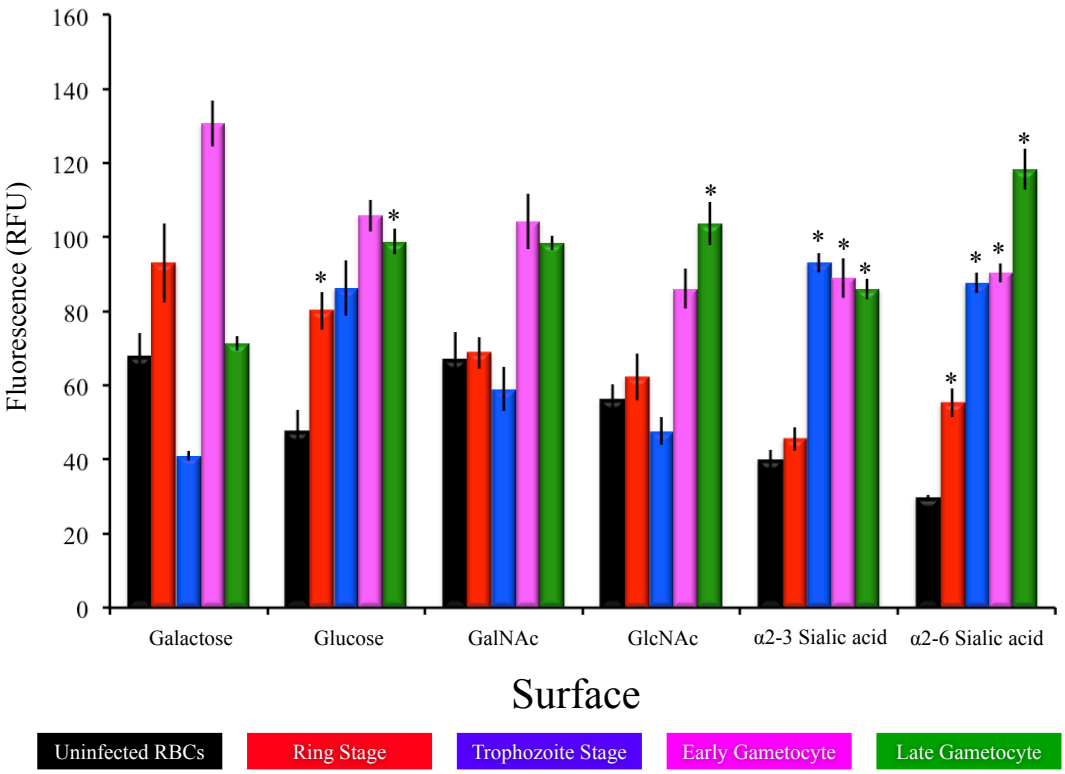


Figure 6.9 Binding of FITC labelled ring stage infected RBCs (green), uninfected (red) RBCs and trophozoite stage infected RBCs (blue) to several carbohydrate surfaces showed differential binding. Significance is indicated by an asterix above the bar where * indicates $p<0.05$ relative to uninfected RBCs. Error bars represent the standard error.

Ring stage RBCs are the initial stage immediately after parasite infection and these showed increased binding to both the glucose surface and the α(2-6) linked sialic acid surface. The metabolic demand of hosting the parasite would increase the absorption of glucose by the infected RBCs and thus an increase in glucose binding during infection would be expected (*vide infra*).

It has been shown that whilst endothelial cells in macrovascular structures, such as the pulmonary artery contain terminal α(2-3) and α(2-6) linked sialic acid residues,

pulmonary microvascular structures only contain α(2-3) linked sialic acid residues.[31] As such differential binding to these surfaces may indicate a mechanism for targeted sequestering of infected RBCs in the microvasculature.

Ring stage infected RBCs are generally found in peripheral circulation and are not thought to undergo cytoadhesion and as such the increase in binding to α(2-6) linked sialic acid residues is surprising. The lack of increase in binding to α(2-3) linked sialic acid residues would prevent binding of these cells in the microvasculature and, whilst adhesion in the microvasculature may be possible, it is possible that binding to α(2-6) alone is insufficient to support binding to the microvasculature and so RBCs in the ring stage of infection are found in circulating blood.

Trophozoite stage RBCs also showed increased binding to α(2-6) linked sialic acid residues as well as an increase in α(2-3) linked residues. This may indicate a mechanism for binding of RBCs in the trophozoite stage to the vasculature. Specifically, the increase in α(2-3) binding may allow adhesion of these cells to the microvasculature.

Gametocytes are split into five distinct phases, but extraction of gametocytes at very specific stages is challenging and just two samples were assessed; early gametocytes (stages I to III of gametocyte development) and late gametocytes (stages IV and V). The early gametocytes showed increased binding to both α(2-3) and α(2-6) linked sialic acid residues. It is known that, similarly to trophozoites, early stage gametocytes undergo cytoadhesion and sequestering in order to escape eradication by the spleen.[32] Increased binding to α(2-3) and α(2-6) residues may indicate a

mechanism for cytoadhesion and sequestering of gametocytes in this early developmental stage.

Late stage gametocytes (stages IV and V) showed significantly higher binding to glucose, GlcNAc, α(2-6) and α(2-3) linked sialic acid. As only mature stage V gametocytes are found in peripheral blood it is not surprising that these samples still showed adhesion to α(2-3) and α(2-6) residues, as this would allow sequestering of the stage IV parasites present within the sample. It would be expected that this binding would be lost in stage V RBCs, as this would allow release of the parasite into the circulation where it can be taken up by mosquitoes (*vide infra*).

6.3.4 Differential binding for diagnostics

Whilst several of the stages showed statistically significant differences for some glycans, classification of a sample based on binding profile can still be assisted by use of linear discriminant analysis (LDA). LDA is a statistical tool, which can be used for the classification of samples. Since it maximises the difference between groups in the original data set and minimises differences within the groups, it is sensitive to picking up small differences in binding profiles that are not necessarily statistically significant.

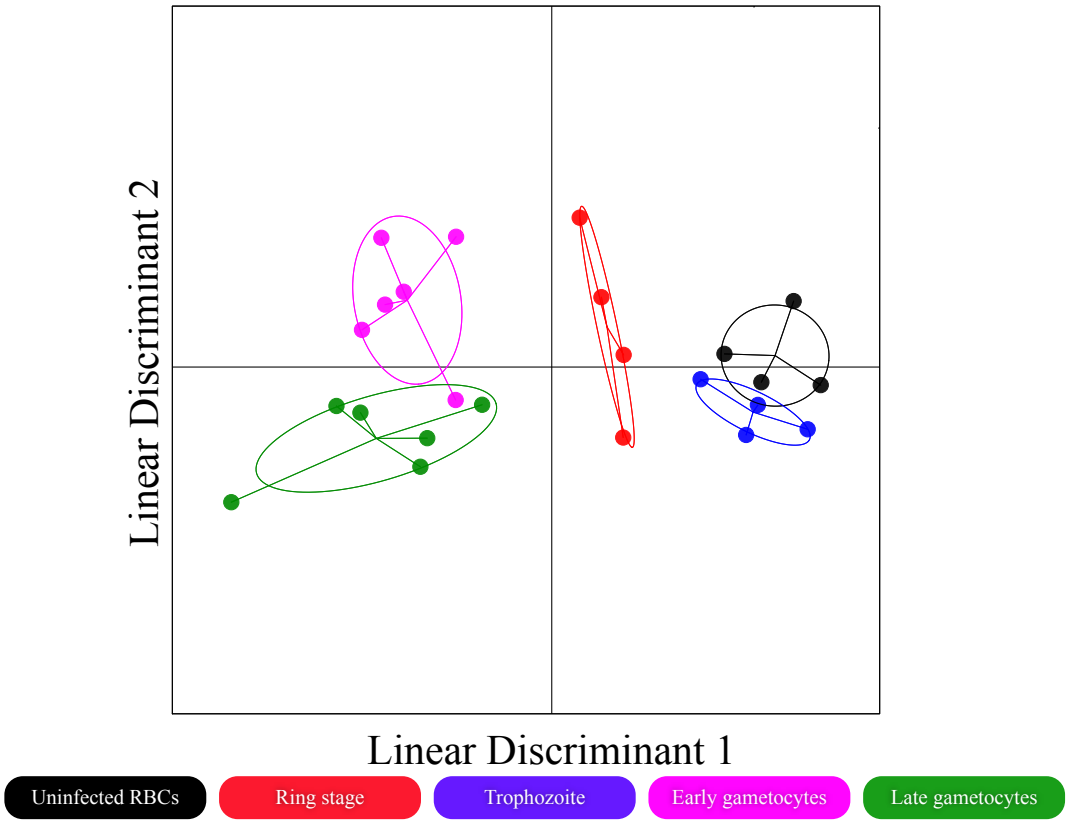


Figure 6.10 Linear discriminant analysis model produced shows good discrimination between most cell types.

As can be seen, the LDA model produced shows excellent discrimination between RBCs at different stages in the parasite life cycle (Figure 6.10). During cross-validation the model was able to correctly classify samples with an accuracy of 96 % and the model was able to correctly identify the nature of two blind samples (Figure 6.11). Whilst the samples do not seem to fall into any of the categories in the model it should be noted that the model is a two-dimensional representation of a multidimensional model and a high classification accuracy was still achieved.

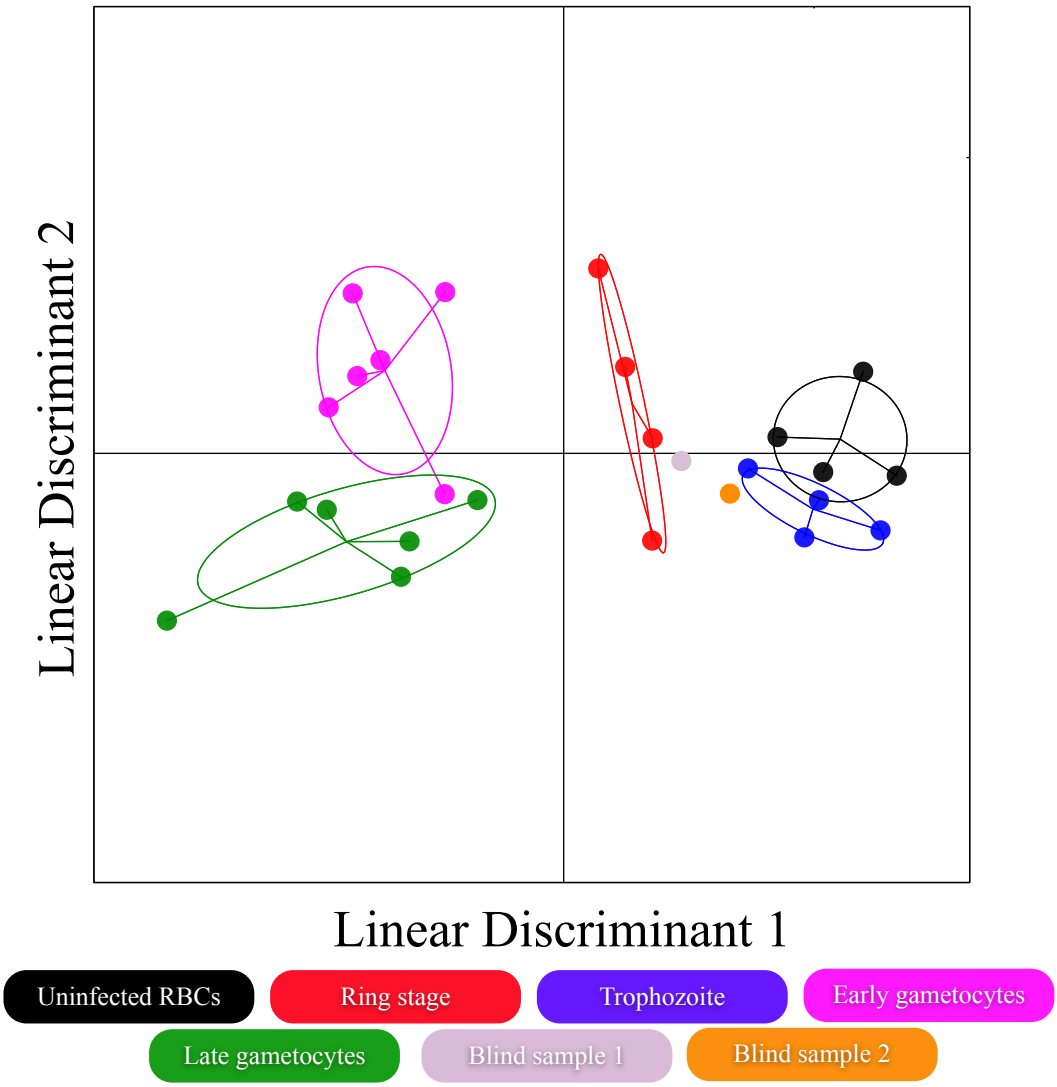


Figure 6.11 Linear discriminant analysis model correctly classified two different blind samples as ring stage and trophozoite respectively. Blind samples represent the average glycan binding profile of four repeats.

Clinically relevant levels of parasitemia vary wildly, a patient with 1 % parasitemia could die but a patient with 40 % parasitemia could be fine. This large disparity is often caused by the complications associated with the trophozoite adhesion phase.[4] 10 % parasitemia is generally described as 'hyperparasitemia' and as such differential binding was determined for two lower levels of parasitemia (5 % and 1%). All parasitemia levels were for the ring stage only, as this is generally the only stage

present in peripheral circulation in patients infected with *P. falciparum*.[22] Mature stages are sometimes found in peripheral circulation, but it is thought that this is linked to disease severity.[33] In this system neither of these samples were correctly classified thus limiting the potential as a diagnostic tool. In order to improve the sensitivity of the model, increasing the number of glycans or the chain length of the glycans could increase the differential response of the different cell types, which would improve the sensitivity of the system.

Whilst this lack of sensitivity would limit the use of this system in terms of a diagnostic tool, it still represents the first time that differential glycan binding in parasitized RBCs has ever been investigated and the ability of this technique to classify blind samples in a stage specific manner highlights the potential of this system. This methodology is only really a proof of principle and there is significant potential for the expansion of the system and increasing sensitivity through coupling to other current glycan based techniques such as a label-free gold nanoparticle-based system.

6.3.5 Detection of drug resistance

Whilst sensitivity of the system was too low to be of real use diagnostically, drug resistance is becoming a real problem amongst *P. falciparum* and detection of drug resistance clinically is currently achieved using molecular markers.[25] Artemisinin based treatments are currently the main treatment for *P. falciparum* infections (due to high levels of drug resistance to other drugs). Artemisinin treatment results in increased clearance of ring stage parasites by the spleen thus reducing the number of parasitized RBCs in circulation.[34] Resistance is increasingly becoming a concern and

is being found in increasing numbers of countries thus endangering the eradication of malaria.[17, 35-38] Resistance is known to be associated with very specific single nucleotide polymorphisms occurring in the *K13* gene. This results in mutations in the primary amino acid sequence of the propeller region of the kelch domain-carrying protein K13.[18] It is currently not known how this affects Artemisinin sensitivity but resistance is characterised by reduced clearance of ring stage parasites by the spleen.[39]

As trophozoites also escape eradication by the spleen and seem to show increased glycan binding which may mediate tissue cytoadhesion and sequestering then it can be reasoned that perhaps glycan binding plays a role in Artemisinin resistance. Tissue sequestration of the ring stage parasites that display resistance to Artemisinin may explain why ring stage parasites are not eradicated by the spleen.

P. falciparum 3D7 is a purely laboratory strain of *P. falciparum* and as such there is no data concerning drug resistance so a different strain was used for the drug resistance experiments. Four strains of *P. falciparum* TH004 with differing levels of resistance to Artemisinin were selected with strain 4 representing clearance time (and thus Artemisinin susceptibility) closest to the wild type *P. falciparum*. Both strains 23 and 36 show roughly the same clearance times, which are approximately three fold higher than 4. Strain 64 showed the highest levels of resistance with a clearance time over four fold higher than the wild type (personal communication with Matthew Phanchana).

The different strains showed differential binding at both the ring stage (Figure 6.12) and the trophozoite stage of parasite development (Figure 6.13). At the ring stage

those with the lowest level of resistance (4) showed increased binding to glucose, galactose, α(2-3) and α(2-6) sialic acid residues when compared to uninfected RBCs. This may indicate that these RBCs can adhere to both the macro and microvasculature, which would allow them to escape eradication by the spleen. Increased glucose would occur due to the increased metabolic demands of the host and increased galactose may occur as a result of the increase in permeability of the RBCs membrane to small molecules, which is associated with infection.

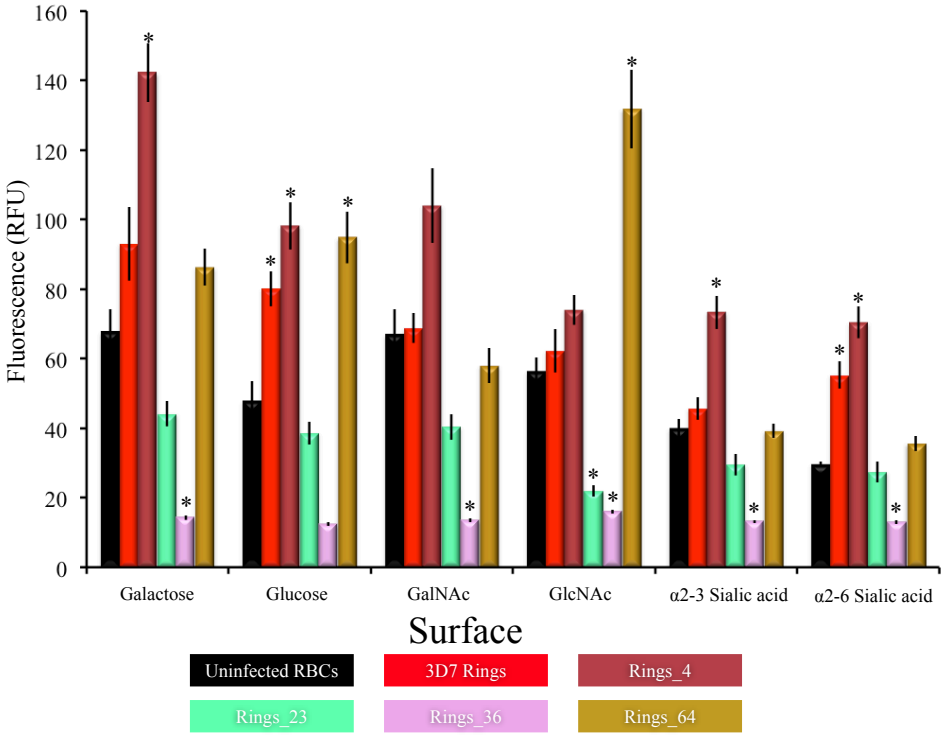


Figure 6.12 Differential binding seen between strains with different drug resistance levels at the ring stage. Significance is indicated by an asterix above the bar where * indicates $p<0.05$ relative to uninfected RBCs. Error bars represent the standard error.

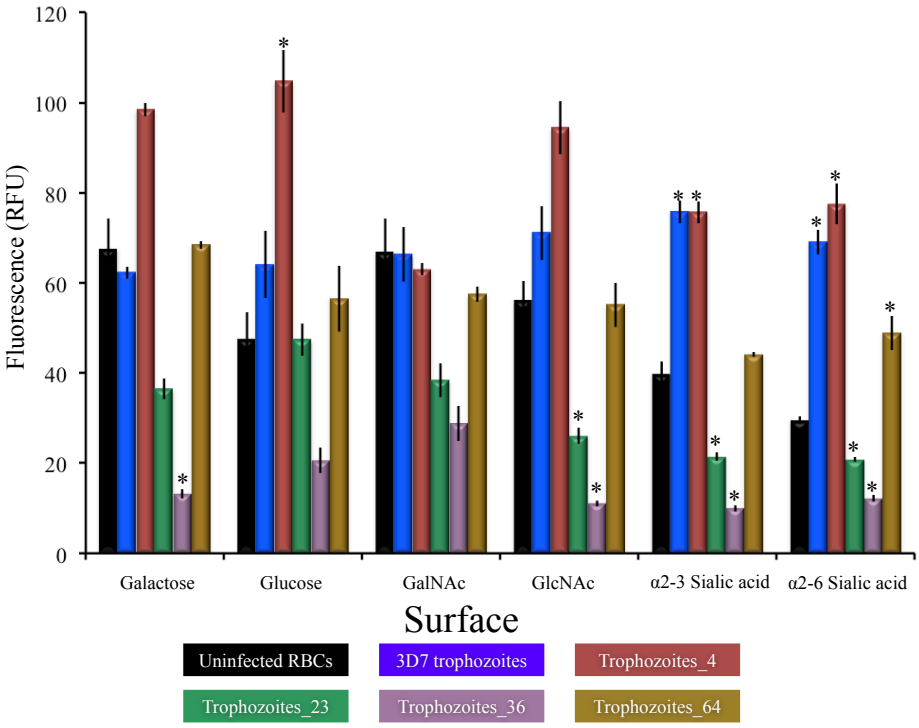


Figure 6.13 Differential binding seen between strains with different drug resistance levels at the trophozoite stage. Significance is indicated by an asterix above the bar where * indicates $p<0.05$ relative to uninfected RBCs. Error bars represent the standard error.

Ring stage RBCs infected with parasites 23 and 36 showed significantly reduced binding to almost all glycans assessed. As these parasites have similar clearance times (and thus similar levels of drug resistance) it is possible that they escape spleen clearance through another mechanism that does not involve glycan binding. Whilst genetically the main mutational difference between these strains and 4 is within the *K13* gene, it is possible for glycan binding to be disrupted purely by re-patterning of glycan binding proteins within a membrane (as this would disrupt multivalent binding) and this could explain the decrease in glycan binding observed for these strains.

The most drug resistant strain (64) showed increased binding to glucose and GlcNAc. Increased glucose could be explained by the metabolic demand of hosting the parasite whereas GlcNAc binding may indicate binding to uninfected RBCs. This would protect the parasitized cell from the immune system and also prevent eradication by the spleen.

For *P. falciparum* 3D7, the development of the parasite from the ring stage to the trophozoite stage was characterised by a reduction in binding to glucose and an increase in binding to sialic acid residues (when compared to uninfected RBCs). Once again, strains 23 and 36 showed reduced binding to many of the glycan residues highlighting that other mechanisms may be involved in sequestering of RBCs infected with these strains.

Strain 4 trophozoites retained the ring-stage characteristic of increased α(2-3) and α(2-6) sialic acid residue binding which would allow them to sequester in the microvasculature thus escaping the spleen. Whilst glucose binding remained high for the trophozoite stage, there was a reduction in galactose binding that may indicate a loss of ring stage membrane permeability.

Similarly to 3D7, strain 64 showed a reduction in glucose binding when in the trophozoite stage. GlcNAc binding was also significantly reduced and there was an increase in binding to α(2-6) sialic acid residues, which may indicate targeting of these cells towards the vasculature system. Although this strain lacks the increased α(2-3) sialic acid binding shown by 3D7 it may be that this strain does not bind in the

microvasculature during the trophozoite stage as not all strains of *P. falciparum* display this characteristic.

As many of the development stages showed differential binding for different strains (summarised in Table 6.1) the power of LDA as a diagnostic tool was assessed in order to determine whether drug resistant strains could be identified based on their glycan binding profiles. The LDA model produced when considering all strains and the two separate parasite development stages showed good separation between groups (Figure 6.14A). The model was able to classify samples with an accuracy of 78 %, with overlaps generally limited to the different development stages of the same strain.

Development stage	Strain	Glycan on surface					
		Galactose	Glucose	GalNAc	GlcNAc	α2-3 Sialic acid	α2-6 Sialic acid
Ring	3D7		↑				↑
	4	↑	↑			↑	↑
	23				↓		
	36	↓		↓	↓	↓	↓
	64		↑		↑		
Trophozoite	3D7					↑	↑
	4		↑				↑
	23				↓	↓	↓
	36	↓			↓	↓	↓
	64						↑

Table 6.1 Summary of differential glycan binding for all resistance strains of *P. falciparum* in which an upward arrow indicates a significant increase and a downward arrow a significant decrease in binding relative to uninfected RBCs.

This is further highlighted when we split the model into development stages. Considering just the ring stage, the classification power of the model can be increased to 85 % (Figure 6.14B). 100 % reclassification can be achieved if just the binding profiles of trophozoites are considered (Figure 6.14C) but, these late stages in the parasites development generally sequester themselves within tissues and are only rarely found in circulation. The ring stage model is most useful for classification of samples.

Whilst ring stage parasites are most likely to be found in circulation, other developmental stages may be present in samples especially in cases of severe malaria. Identification of the developmental stage of the parasites within a sample would not be as important as identifying drug resistance and so a model was developed using the binding profiles of resistant strains at either the ring or trophozoite stage. The LDA model only considered the strain and contained no information about the developmental stage of the parasite that produced that binding profile. This was able to increase the accuracy of the initial model (which considered both developmental stage and strain) to 83 % classification of strain based on binding profile alone (Figure 6.15).

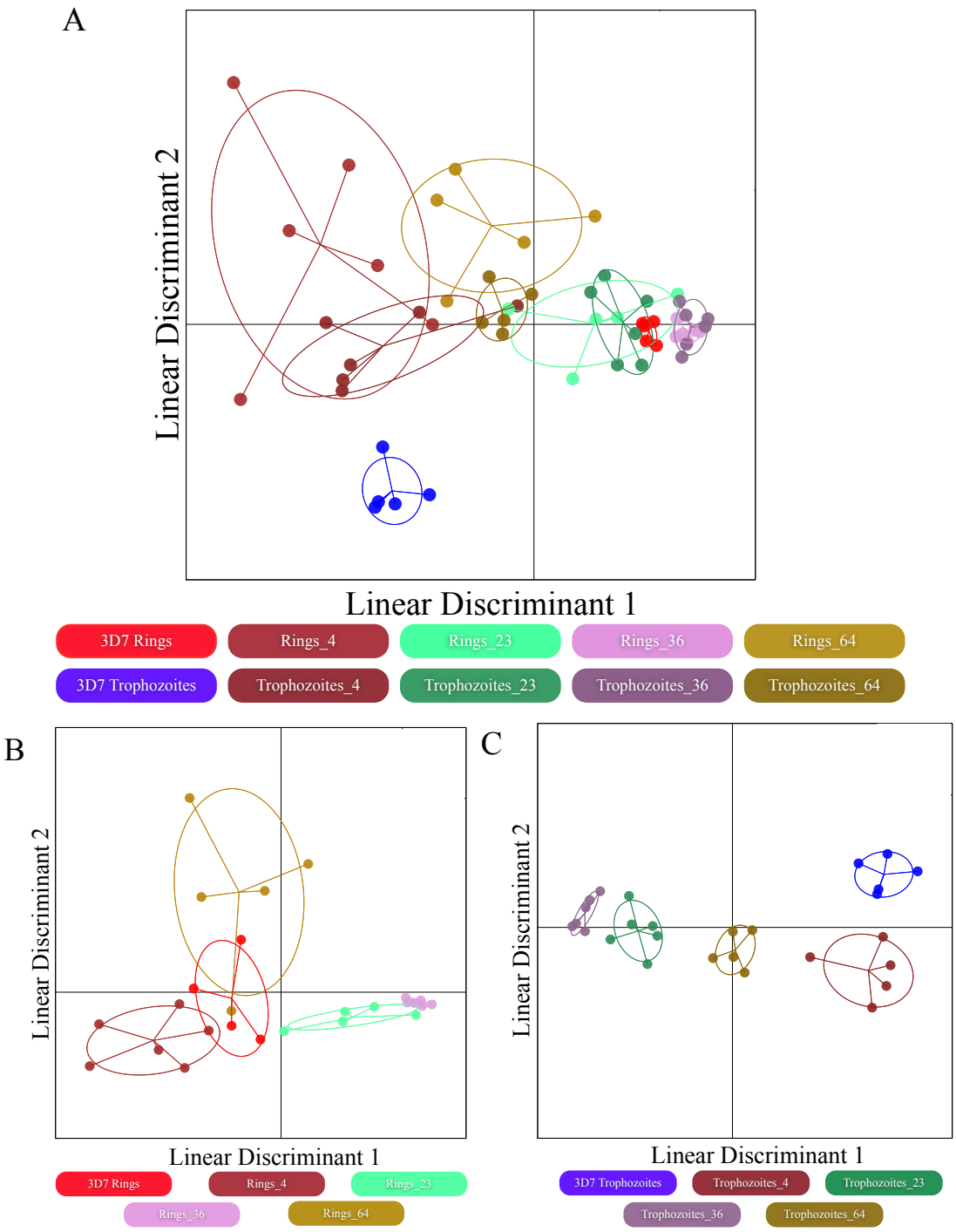


Figure 6.14 Linear discriminant analysis model produced shows good discrimination between all resistance strains and was able to correctly identify both strain and stage with an accuracy of 78 % (A). Further splitting into ring stage (B) and trophozoite stage (C) allowed prediction of resistance strain with an accuracy of 85 and 100 % respectively.

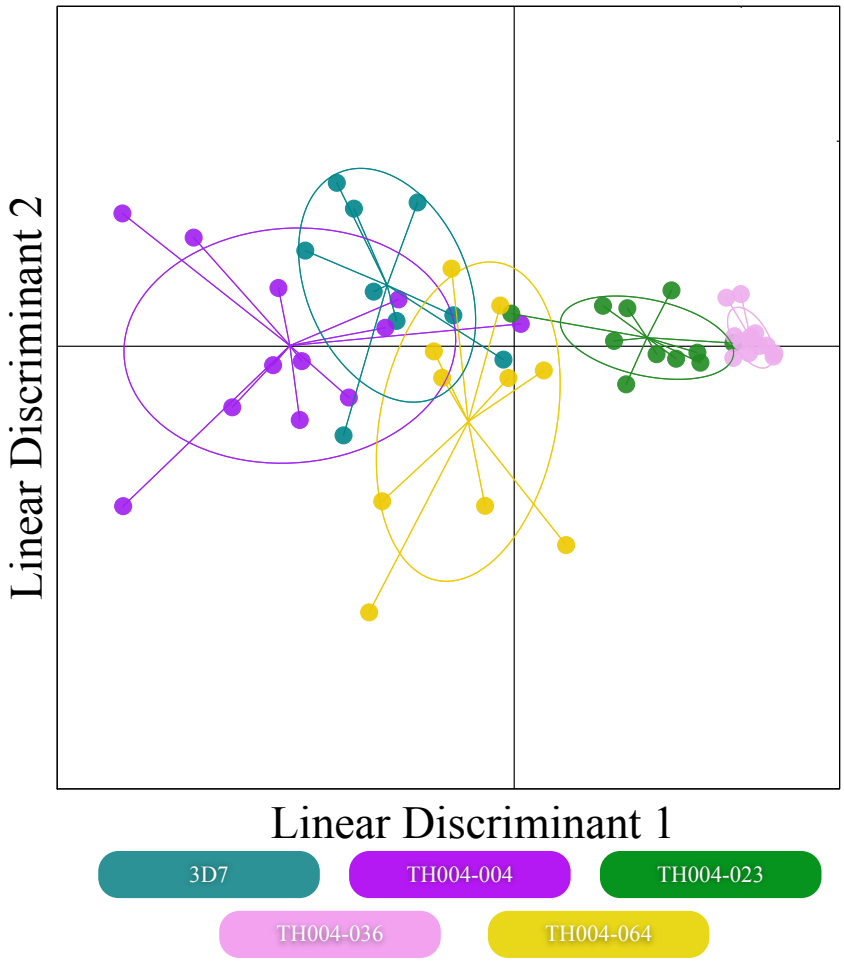


Figure 6.15 Linear discriminant analysis model in which only the strains and not the development stage were considered showed reasonable separation between resistant strains and was able to re-classify data in a 'leave-one-out' method with an accuracy of 83 %.

The ability of this facile system to detect differences between resistance strains highlights its power as a diagnostic tool. Whilst the sensitivity of the system is still low this is a proof of principle study and sensitivity could be improved through a number of different methods such as increasing the number of glycans studies, increasing glycan chain length or coupling with another detection system such as gold nanoparticles.

6.4 Conclusion

We have developed a rapid and simple diagnostic tool for the detection of RBCs infected with the malaria parasite through binding to carbohydrate functionalised surfaces, the first time such a method has ever been used for the detection of malaria. Uninfected and parasitised RBCs were successfully labelled using FITC to allow detection of binding.

Ring stage RBCs were found to bind glucose and α(2-6) sialic acid residues significantly more than uninfected RBCs. Trophozoites and early gametocytes showed increased binding to α(2-3) and α(2-6) sialic acid residues. Late stage gametocytes showed increased binding to glucose, GlcNAc, α(2-3) and α(2-6) sialic acid residues.

Moreover we have shown that these differences fit within the supporting literature evidence and may indicate a mechanism through which trophozoite infected RBCs can sequester specifically to the microvasculature through adhesion to α(2-3) linked sialic acid residues found predominantly within the microvasculature.

These differential binding profiles were used as a training matrix to produce a LDA model with a classification accuracy of 96 % and was able to correctly classify two blind samples in a developmental stage specific manner. Furthermore, ring stage samples exhibiting reduced parasitemia (5 % and 1 %) were also analysed to determine whether more clinically relevant levels of parasite infection could be detected. Disappointingly, neither of these samples were correctly classified. The detection limit of this system was found to be ~2.4 $x10^4$ parasitsed cells per μL of

blood. Whilst other currently available techniques are more sensitive, the detection limit could be improve through the incorporation of more complex glycan structures onto the surfaces, the use of a wider range of sugars or the use of mixed sugar surfaces.

Further to this we analysed the ability of this system to detect differences in glycan binding between Artemisinin resistant strains of *P. falciparum*. Differential binding was observed both between different strains and between life stages of the same strain. A LDA model produced using these binding profiles was analysed in a 'leave-one-out' manner and was correctly able to classify a sample based on both life stage and strain with an accuracy of 78 %. Accuracy at classifying strain increased markedly to 85 % and 100 % when samples were separated into either ring or trophozoite stage respectively. A further model was produced in which only the strain of the sample was considered (and thus all samples included were either ring or trophozoite stage) and this was able to classify strain (and thus resistance level) with an accuracy of 83 %.

As Artemisinin is currently the basis of almost all standard malaria treatment regimes, the increase in Artemisinin resistance amongst *P. falciparum* is of great concern but, current malaria diagnosis uses microscopy, which provides no information about drug resistance. The spread of resistance is currently monitored using molecular markers, but this technique requires careful sample preparation, is often expensive and results can take time to process. As such this glycan based detection system offers a much more accessible technique that can identify differences between infected and uninfected RBCs as well as identifying drug resistant strains in under two hours.

Whilst sensitivity is currently a limitation of this assay, there are many routes through which this could be improved and this work represents a proof of principle. This represents the first time that differential glycan binding has been used to detect the presence of a parasite. Although the technique is in its early stages, we have presented a cheap, facile and stage specific technique for the identification of *P. falciparum* infected red blood cells and detection of drug resistant strains.

6.5 Materials and methods

Materials: All chemicals were used as supplied unless otherwise stated. Corning® 96 well clear flat bottomed polystyrene Carbo-BIND™ microplates, ExtrAvidin®-FITC, (+)-Biotin *N*-hydroxysuccinimide ester, Fluorescein isothiocyanate isomer I (≥ 90 % HPLC), aniline (99.5 % ACS reagent), sodium acetate anhydrous, acetic acid glacial, phosphate buffered saline tablets, D- (+)galactose, *N*-Acetyl-D-galactosamine, *N*-Acetyl-D-glucosamine, α-D-glucose and DMSO were all purchased from Sigma-Aldrich. α(2-3) D-sialyllactose and α(2-6) D-sialyllactose were both purchased from Carbosynth. 100 mM acetate buffer with 1 mM aniline was prepared in 200 mL of milliQ water (with a resistivity of >18 $M\Omega.cm^{-1}$). *Plasmodium falciparum* 3D7 were cultured in RPMI 1640 R8758 (Sigma-Aldrich) supplemented with 25 mM HEPES (pH 7.4, Amresco), 0.25 % Albumax I (Sigma-Aldrich), 0.4 mM hypoxanthine (Sigma-Aldrich) and 0.020 mM gentamycin sulphate (Sigma-Aldrich). *Plasmodium falciparum* TH004 were cultured in RPMI 1640 supplemented with 25 mM HEPES (pH 7.4, Amresco), pooled human AB+ serum (50 mL per 500 mL of RPMI 1640), 0.4 mM hypoxanthine (Sigma-Aldrich) and 0.020 mM gentamycin sulphate (Sigma-Aldrich). Parasite work was carried out with Eva Caamaño-Gutiérrez in the Department of Parasitology at the Liverpool School of Tropical Medicine with the assistance of Ahmed Saif and Matthew Phanchana.

NHS-Biotin:Avidin-FITC labelling of red blood cells: 50 mL of infected and uninfected RBCs in cell culture media were spun down by centrifugation at 2000 rpm for 5 minutes before being resuspended in 10 mL of cell culture media. 2.5 μL of NHS-Biotin (10 $mg.mL^{-1}$ in DMSO) was then added to every 10 mL of cell suspension. RBCs were then added to cell culture flasks and incubated at 37 °C for

2.5 hours before washing and resuspension in cell culture media. 50 μL of biotin labelled red blood cells were then added in a serial dilution (in cell culture media) to every well of various sugar functionalised 96-well plates from a maximum concentration of that achieved in the pellet formed post-centrifugation.

96-well plates with RBCs were then incubated at 37 °C for 30 minutes before removal of unbound cells and washing of every well with PBS. 100 uL of ExtrAvidin®-FITC (0.5 % in PBS) was then added to every well before incubation at 37 °C for 1 hour. Unbound ExtrAvidin®-FITC was then removed from every well before washing with PBS. All wells were then incubated with cell lysis buffer for 1 hour prior to fluorescence detection. Fluoresence measurements were conducted on a Thermo Electron Varioskan spectral scanning multimode reader using excitation and emission wavelengths of 475 and 525 nm respectively.

Fluorescein Isothiocyanate labelling of red blood cells: 50 mL of infected and uninfected RBCs in cell culture media was spun down by centrifugation at 500 g before being resuspended in 12 mL of cell culture media. Subsequently FITC was added to produce a final concentration of 0.1 $mg.mL^{-1}$ (and 0.35 % DMSO). The cells were then incubated for 30 minutes at 37 °C as previously described by Fujii *et al.*[30] before centrifugation and resuspension of cells in fresh cell culture media at 11 % haematocrit.

Autofluorescence: 50 uL of RBCs were added in a serial dilution (in cell culture media) from the maximum pellet concentration to every well of a sugar functionalised 96-well plate. This was then incubated at 37 °C for 30 minutes prior to removal of

unbound red blood cells and washing with PBS. Fluorescence was then measured with excitation and emission wavelengths of 485 nm and 645 nm respectively.

Colorimetric assay and ImageJ analysis: Dilution assays of RBCs in culture media were photographed using a camera and images were uploaded into ImageJ. Each well had a region of interest drawn around it before conversion of the image into a HSB (Hue, Saturation, Brightness) stack. The region of interest was added to the saturation image from the HSB stack and pixel intensity in each well was measured before exporting the data into Excel and plotting in OriginPro.

Fluorescence microscopy: Smears of blood drops from either infected cells (early or late trophozoite stage) or uninfected cells after incubation with FITC were observed under a fluorescence microscope (Olympus b x-60) with a blue filter (495-570 nm) and digitalised with the aid of a Nikon camera.

Carbohydrate surface preparation: 100 μL of 30 mM sugar solutions (in 100 mM acetate buffer with 1 mM aniline, pH 5.5) was added to each well of a Carbo-BIND 96-well plate. Plates were then covered in foil and incubated in the dark at 50 °C for 24 hours. After incubation, any unbound solution was removed and the well washed with PBS. Plates were either used immediately or stored at -20 °C prior to their use.

Carbohydrate binding assay: FITC labelled RBCs were diluted to one third of the pellet concentration (~6.3×10^5 cells per μL) and 50 μL of this cell suspension was added to every well of a sugar functionalised Carbo-BIND™ plate and incubated for 30 minutes at 37 °C, followed by removal of unbound cell suspension and washing

with PBS. Finally 1X dilution of lysis buffer consisting of 5 mM EDTA, 0.08 % Triton x100, 20 mM Tris and 0.08 % saponin was added to each well and the fluorescence was measured using SkanIt in a Thermo Electron Varioskan spectral scanning multimode reader (excitation and emission wavelengths of 475 and 525 nm respectively).

Cell culturing method: *Plasmodium falciparum* 3D7 were cultured in RPMI 1640 R8758 supplemented with 25 mM HEPES (pH 7.4), 0.25 % Albumax I, 0.4 mM hypoxanthine and 0.020 mM gentamycin sulphate. *Plasmodium falciparum* TH004 were cultured in RPMI 1640 supplemented with 25 mM HEPES (pH 7.4), pooled human AB+ serum (50 mL per 500 mL of RPMI 1640), 0.4 mM hypoxanthine and 0.020 mM gentamycin sulphate. Cultures were grown under a gaseous headspace of 3 % O_2 and 4 % CO_2 in N_2 at 37 °C.[40] Parasite growth was synchronised by treatment with sorbitol. [41]

Gametocyte preparation: *Plasmodium falciparum* 3D7 were cultured in RPMI 1640 culture media containing L-glutamine and sodium bicarbonate. Media was supplemented with 25 mM HEPES (pH 7.4), 0.4 mM hypoxanthine, pooled human AB+ serum (50 mL per 500 mL of RPMI 1640) and 0.020 mM gentamycin sulphate. For the first 4 days the media was supplemented with 50 mM *N*-Acetyl-D-glucosamine. For late stage gametocytes, complete culture media was used for the remaining 3 days and for early stage gametocytes complete media supplemented with 50 mM *N*-Acetyl-D-glucosamine was used. Cultures were grown under a gaseous headspace of 3 % O_2 and 4 % CO_2 in N_2 at 37 °C.[40]

Slide preparation for red blood cells: Samples of RBCs in culture media were prepared on slides for light microscopy. Each slide was fixed with methanol, dried and then stained with Giemsa before washing and drying again. Slides were viewed at 100X magnification using an oil immersion objective.

Statistical analysis: Differences in carbohydrate binding were analysed using a two-sample one-sided Student's t-test performed in the open source statistical package R (version 3.1.3) [42] after equal variance was confirmed by performing a Fisher test for variance and normality of data was confirmed using the Shapiro-Wilk normality test. For those samples were equal variance was not observed a Welch t-test was performed and for those where data was not normally distributed, the Wilcoxon Rank Sum test was employed. All graphs were produced in OriginPro. LDA was performed using the 'dapc' function in the 'adegenet' package (version 1.4-2)[43] in R.

6.6 References

1. World Health Organisation, *Malaria fact sheet*, http://www.who.int/mediacentre/factsheets/fs094/en/.
2. Centre for Disease Control, *Comparison of Malaria species*, http://www.cdc.gov/dpdx/malaria/dx.html - benchaids.
3. L. H. Miller, D. I. Baruch, K. Marsh and O. K. Doumbo, *Nature*, 2002, **415**, 673-679.
4. K. Chotivanich, R. Udomsangpetch, J. A. Simpson, P. Newton, S. Pukrittayakamee, S. Looareesuwan and N. J. White, *J Infect Dis*, 2000, **181**, 1206-1209.
5. K. J. Robson, U. Frevert, I. Reckmann, G. Cowan, J. Beier, I. G. Scragg, K. Takehara, D. H. Bishop, G. Pradel and R. Sinden, *The EMBO Journal*, 1995, **14**, 3883-3894.
6. National Institute of Health, *Understanding malaria*, http://www.niaid.nih.gov/topics/malaria/understandingmalaria/pages/symptoms.aspx.
7. D. I. Baruch, *Best Pract Res Cl Ha*, 1999, **12**, 747-761.
8. Q. Chen, M. Schlichtherle and M. Wahlgren, *Clin Microbiol Rev*, 2000, **13**, 439-450.
9. X.-z. Su, V. M. Heatwole, S. P. Wertheimer, F. Guinet, J. A. Herrfeldt, D. S. Peterson, J. A. Ravetch and T. E. Wellems, *Cell*, 1995, **82**, 89-100.
10. D. S. Peterson, L. H. Miller and T. E. Wellems, *Proc Natl Acad Sci U S A*, 1995, **92**, 7100-7104.
11. M. Jungery, G. Pasvol, C. I. Newbold and D. J. Weatherall, *Proc Natl Acad Sci U S A*, 1983, **80**, 1018-1022.

12. S. Shen, K. D. Bryant, J. Sun, S. M. Brown, A. Troupes, N. Pulicherla and A. Asokan, *J Virol*, 2012, **86**, 10408-10417.

13. M. Mehta, H. M. Sonawat and S. Sharma, *FEBS Letters*, 2005, **579**, 6151-6158.

14. K. Tanabe, *Parasitology today*, 1990, **6**, 225-229.

15. K. W. Deitsch and T. E. Wellems, *Mol Biochem Parasit*, 1996, **76**, 1-10.

16. K. Kirk, H. A. Horner, B. C. Elford, J. C. Ellory and C. I. Newbold, *J Biol Chem*, 1994, **269**, 3339-3347.

17. K. M. Tun, M. Imwong, K. M. Lwin, A. A. Win, T. M. Hlaing, T. Hlaing, K. Lin, M. P. Kyaw, K. Plewes, M. A. Faiz, M. Dhorda, P. Y. Cheah, S. Pukrittayakamee, E. A. Ashley, T. J. C. Anderson, S. Nair, M. McDew-White, J. A. Flegg, E. P. M. Grist, P. Guerin, R. J. Maude, F. Smithuis, A. M. Dondorp, N. P. J. Day, F. Nosten, N. J. White and C. J. Woodrow, *Lancet Infect Dis*, 2015, **15**, 415-421.

18. S. Mok, E. A. Ashley, P. E. Ferreira, L. Zhu, Z. Lin, T. Yeo, K. Chotivanich, M. Imwong, S. Pukrittayakamee, M. Dhorda, C. Nguon, P. Lim, C. Amaratunga, S. Suon, T. T. Hien, Y. Htut, M. A. Faiz, M. A. Onyamboko, M. Mayxay, P. N. Newton, R. Tripura, C. J. Woodrow, O. Miotto, D. P. Kwiatkowski, F. Nosten, N. P. Day, P. R. Preiser, N. J. White, A. M. Dondorp, R. M. Fairhurst and Z. Bozdech, *Science*, 2015, **347**, 431-435.

19. P. Gascoyne, M. Satayavivad J Fau - Ruchirawat and M. Ruchirawat, *Acta Trop*, 2004, **89**, 357-369.

20. A. Moody, *Clin Microbiol Rev*, 2002, **15**, 66-78.

21. Centre for Disease Control, *Comparison of Malaria species*, http://www.cdc.gov/dpdx/malaria/dx.html - benchaids.

22. B. Pouvelle, P. A. Buffet, C. Lepolard, A. Scherf and J. Gysin, *Nat Med*, 2000, **6**, 1264-1268.

23. D. C. Richardson, M. Ciach, K. J. Y. Zhong, I. Crandall and K. C. Kain, *J Clin Microbiol*, 2002, **40**, 4528-4530.

24. J. Iqbal, N. Khalid and P. R. Hira, *Adv Exp Med Biol*, 2003, **531**, 135-148.

25. C. Roper, M. Alifrangis, F. Ariey, A. Talisuna, D. Menard, O. Mercereau-Puijalon and P. Ringwald, *Lancet Infect Dis*, 2014, **14**, 668-670.

26. T. Hanscheid and M. P. Grobusch, *Trends Parasitol*, 2002, **18**, 395-398.

27. J. M. Rubio, A. Benito, P. J. Berzosa, J. Roche, S. Puente, M. Subirats, R. López-Vélez, L. García and J. Alvar, *J Clin Microbiol*, 1999, **37**, 3260-3264.

28. L. N. Wangai, M. G. Karau, P. N. Njiruh, O. Sabah, F. T. Kimani, G. Magoma and N. Kiambo, *Afr J Infect Dis*, 2011, **5**, 1-6.

29. L. Otten and M. I. Gibson, *RSC Advances*, 2015, **5**, 53911-53914.

30. A. Fujii, K. Yoshikawa, T. Iwata, M. Akashi, H. Noguchi and Y. Takumi, *Masui*, 1993, **42**, 545-551.

31. A. C. Eugene, F. A. Diego, P. Subha and L. C. Donna, in *C26. Regulators of the Lung Endothelial Barrier*, American Thoracic Society, Editon edn., 2011, pp. A4177-A4177.

32. F. Silvestrini, M. Tibúrcio, L. Bertuccini and P. Alano, *PLoS ONE*, 2012, **7**, e31567.

33. T. Sun and C. Chakrabarti, *Ann Clin Lab Sci*, 1985, **15**, 465-469.

34. N. J. White, *Malar J*, 2011, **10**, 278.

35. A. P. Phyo, S. Nkhoma, K. Stepniewska, E. A. Ashley, S. Nair, R. McGready, C. ler Moo, S. Al-Saai, A. M. Dondorp and K. M. Lwin, *Lancet*, 2012, **379**, 1960-1966.

36. K. Mugittu, B. Genton, H. Mshinda and H. P. Beck, *Malar J*, 2006, **5**, 3-5.

37. E. A. Ashley, M. Dhorda, R. M. Fairhurst, C. Amaratunga, P. Lim, S. Suon, S. Sreng, J. M. Anderson, S. Mao and B. Sam, *New England Journal of Medicine*, 2014, **371**, 411-423.

38. A. M. Dondorp, F. Nosten, P. Yi, D. Das, A. P. Phyo, J. Tarning, K. M. Lwin, F. Ariey, W. Hanpithakpong and S. J. Lee, *New Engl J Med*, 2009, **361**, 455-467.

39. B. Witkowski, N. Khim, P. Chim, S. Kim, S. Ke, N. Kloeung, S. Chy, S. Duong, R. Leang and P. Ringwald, *Antimicrob Agents Ch*, 2013, **57**, 914-923.

40. W. Trager and J. B. Jensen, *Science*, 1976, **193**, 673-675.

41. C. Lambros and J. P. Vanderberg, *J Parasitol*, 1979, 418-420.

42. R. D. C. Team, *R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing*, http://www.R-project.org/.

43. T. Jombart, C. Collins, P. Solymos, I. Ahmed, F. Calboli and A. Cori, *adegenet: an R package for the exploratory analysis of genetic and genomic data*, http://adegenet.r-forge.r-project.org/.

Chapter 7

Conclusions

Carbohydrate arrays represent a powerful tool for the detection and identification of pathogenic organisms in point-of-care diagnostics. Currently the poor sensitivity of binding at the monosaccharide level has limited the potential of these devices as increased sensitivity is achieved by using expensive and synthetically challenging oligosaccharides or by coupling to more sensitive detection methods. These methods include; surface plasmon resonance, quartz crystal microbalance and mass spectrometry to name but a few. Such techniques are often challenging, require high sample purity, highly skilled technicians, expensive machinery and processing of results can be both challenging and time-consuming.

In this work, we have described the use of simple and synthetically accessible glycans as multiplexed sensors after analysis with linear discriminant analysis (LDA). LDA not only improved the resolution between groups based on binding but allowed the identification of unknown samples based on binding.

The huge wealth of binding data that is already available from the consortium for functional glycomics has been achieved using an expensive microarray. Such information density can be matched by the coupling of a multiplexed monosaccharide

sensor with LDA. Cheap and facile surface chemistry allowed the production of a 96-well based carbohydrate array through coupling of reducing glycans to a commercially available hydrazide functionalised 96-well plate. Functionalisation was confirmed using a modified drop shape analysis technique where the spreading of a droplet of dye within a well was compared at each stage of the functionalisation in order to determine surface hydrophobicity. Functionalisation was further confirmed by the specific binding of two well-characterised lectins to carbohydrate functionalised surfaces. Concanavalin A (α-mannose specific) was found to bind more strongly to the mannose surface than the galactose surface and the reverse was found to be true for peanut agglutinin (a β-galactose specific lectin).

This functionalisation methodology was then used to generate four monosaccharide surfaces (galactose, mannose, glucose and a 50 % galactose 50 % mannose mixed surface) and the binding of five galactose specific lectins was assessed. Whilst binding profiles alone were not sufficient for lectin discrimination (due to the similarity in their binding), combination of these profiles with LDA allowed excellent resolution between all lectin groups and 100 % correct classification of blind samples for all the lectins.

This methodology was also able to correctly identify the presence of a toxin when mixed with another lectin as long as it was present at greater than 50 % of the sample thus paving the way towards identification of toxins in impure samples. Fluorescently labelled lectins were used for this method but as these are not going to be present in real biological samples, the use of LDA for the analysis of lectin binding couple with a label-free detection method was also examined. Gold nanoparticles were coupled to

monosaccharide functionalised polymers (GlycoAuNPs) for label free detection of lectin binding. The characteristic absorbance shift in response to lectin binding for six different monosaccharide functionalised particles was measured. Identification of any of the six lectins assessed based on binding to the surfaces would be challenging based purely on their binding profile but the LDA model showed excellent resolution between lectins thus allowing clear identification of any of the lectins assessed.

Combining GlycoAuNPs with varying rations of *N*-Acetyl-D-Mannosamine and *N*-Acetyl-D-Galactosamine functionalisation, highlighted the potential for ratiometric surfaces to improve the information density of the sensor without increasing glycan number or complexity. This approach allowed concentration independent lectin discrimination (with an accuracy of 86 %) that was significantly more accurate than considering just 100 % mannose and 100 % galactose functionalised particles (which had an accuracy of 68 %). This would allow classification of biological samples with unknown levels of lectin thus highlighting the potential for this methodology in a 'real-world' situation.

The combination of facile monosaccharide functionalised surfaces and LDA was also utilised in the discrimination of bacterial species. Identification of bacterial species to determine appropriate point-of-care treatment is crucial to avoid the spread of antibiotic resistance strains. Five bacterial species, encompassing pathogenic and non-pathogenic strains, gram positive and gram-negative species and strains of the same species were successfully fluorescently labelled. After labelling and incubation with nine different functionalised surfaces and production of a LDA model it was shown that each bacterial species showed good resolution from the other species and the

model could be utilised in the identification of a blind bacterial sample of one of the species assessed.

This approach also allowed the life-stage specific identification of red blood cells (RBCs) infected with *Plasmodium falciparum* (the most lethal of the parasite species responsible for human malaria). RBCs infected with parasites at different stages of development showed distinct binding patterns when compared to uninfected RBCs. Production of an LDA model highlighted distinct profiles for all life-stages and allowed not just the identification of infected *versus* uninfected but the identification of the life stage of the parasite within the RBC. Furthermore, the application of this methodology to *P. falciparum* strains of differing levels of Artemisinin (a drug that underpins many malarial treatment therapies) resistance allowed the identification of differing resistance levels based on their carbohydrate binding profiles. This allowed not just the identification of resistance but was able to detect differences between strains with differing levels of resistance with a high accuracy. This represents the first time *P. falciparum* infection has been detected based on carbohydrate adhesion and a promising avenue of research for a simple method for identification of drug resistant strains.

One of the key limitations of this approach is the lack of techniques available for surface characterisation. Due to this, the exact conformation of the glycans on the surface is unknown, the topology of the surface and their exact surface density. This results in the possibility for false negatives in binding whereby a lectin or cell fails to bind a surface that it binds to natively due to non-optimum surface presentation of glycans. Despite this, the technique represents a facile methodology for surface

coupling of glycans and further optimisation of the reaction conditions and development of surface characterisation techniques may aid in improving the level of characterisation of the surface.

Point-of-care detection would currently be limited by the sensitivity of the technique, especially when considering contaminated biological samples. Whilst a significant amount of work would have to be done in order to improve the technique to the point where it could be used as a diagnostic tool, it represents a powerful tool for research in other fields. For example, this tool could be used to look at differential glycan binding characteristics of bacteria at different growth phases or in different nutritional conditions. This may elucidate some potential binding targets that occur only in specific growth phases or in certain nutrient availability conditions that may be more similar to that found within the host.

Whilst much current work in this field is focused towards improving specific binding interactions through improved surface functionalisation (specifically with longer chain glycans), this combined technique highlights a mechanism through which better analysis of low resolution data (such as monosaccharide binding) can still provide information in the discrimination of samples. As such, future work within this field should focus on the use of more powerful statistical tools such as linear discriminant analysis and Random Forest for the analysis of binding data to ensure maximum information density of glycan binding is utilised. Whilst sensitivity is currently a limitation for the real-world application of this approach we have shown methods for improving this through the use of mixed glycan surfaces and coupling to a ‘label-free’

approach such as gold nanoparticles further improves the real world application of this methodology.

Appendices

Appendix 1: Gold nanoparticle-linked analysis of carbohydrate-protein interactions and polymeric inhibitors, using unlabelled proteins; easy measurements using a 'simple' digital camera

Journal of
Materials Chemistry B

RSC Publishing

PAPER

View Article Online
View Journal | View Issue

Gold nanoparticle-linked analysis of carbohydrate–protein interactions, and polymeric inhibitors, using unlabelled proteins; easy measurements using a 'simple' digital camera†

Cite this: *J. Mater. Chem. B*, 2013, **1**, 2665

Lucienne Otten,[a] Sarah-Jane Richards,[a] Elizabeth Fullam,[bc] Gurdyal S. Besra[b] and Matthew I. Gibson*[a]

Received 21st February 2013
Accepted 10th April 2013

DOI: 10.1039/c3tb20259c

www.rsc.org/MaterialsB

Traditional methods of measuring the affinity of lectins (or other carbohydrate-binding proteins) to their target carbohydrate ligand rely on the use of chemically/recombinantly modified proteins in sorbent assays, microarrays or the use of expensive label-free methods such as surface plasmon resonance spectrometry. In this work we exploit the extremely high extinction coefficient (*i.e.* colour) of gold nanoparticles as resolving agents in sorbent assays. The anionic nanoparticles adhere strongly to immobilized proteins, but not to the carbohydrate-surfaces allowing investigation of protein binding and screening of novel multivalent inhibitors. Furthermore, the use of a simple digital camera (or mobile phone) to obtain the data is shown, providing a simple ultra-low cost route to the detection of unmodified, carbohydrate-binding proteins.

Introduction

Carbohydrate–protein recognition events are responsible for a plethora of biological processes including host–pathogen interactions, immune signaling, fertilization, cancer metastasis, blood group markers and many more. Pathogenic infection is often driven by an initial carbohydrate–protein binding event such as neuraminic acid/hemagglutinin (influenza),[1,2] GM-1 ganglioside/cholera toxin (cholera)[3] and GP-120/DC-SIGN (HIV),[4] often through binding to lectins: a subset of carbohydrate-binding proteins which are neither enzymes nor antibodies. Inhibition of these binding events, to prevent infection, is termed anti-adhesion therapy and has been extensively studied for both mono- and multi-valent carbohydrate derivatives.[5–9] The binding affinity of carbohydrates to lectins (and hence their potential utility) is often enhanced by the multivalent presentation of individual carbohydrates on a polymeric/dendritic scaffold ('glycopolymer'),[10] due to the 'cluster glycoside effect',[5,11] where the polymer's binding affinity increases to an extent greater than the linear sum of the individual carbohydrates' binding affinities. The mechanisms and applications of these materials are currently under intensive investigation, but are often only focused on model plant lectins which are widely available and have well-characterized interactions. Despite many carbohydrate–protein interactions being known, there still remains a huge number of carbohydrate-mediated processes (covering recognition, uptake, trafficking and binding) which are poorly characterized and hence causes a bottle-neck in the design of new therapeutics or diagnostics.[12] For example, the structure of the oligosaccharide responsible for human sperm binding was only recently elucidated despite many years of intense study.[13] In part, this is due to the challenge associated with obtaining usable quantities of complex glycans or their derivatives, compared to other biomacromolecules (DNA, proteins) which can be recombinantly amplified and/or expressed and obtained by solid-phase chemical synthesis.[14,15] In order to evaluate carbohydrate–protein interactions, there is also a need to isolate and characterize usable quantities of the protein prior to any analysis. Standard methods for evaluating these interactions include both solution and solid phase based assays. Isothermal titration calorimetry (ITC) provides detailed information about the energetics of the binding interactions, but is time consuming, not high-throughput and can consume a relatively large amount of material and requires specialized equipment.[16] Sometimes, direct fluorescence assays probing tryptophan/tyrosine residues can be used but this is not always the case. A further problem of only probing solution interactions is that many carbohydrates are immobilized on the surface of a cell/virus, providing an adhesion site, rather than as mobile solutes. To address this,

[a]*Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK. E-mail: m.i.gibson@warwick.ac.uk; Fax: +44 (0)2476 524112*
[b]*School of Biosciences and Institute of Microbiology and Infection, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*
[c]*School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3tb20259c

Surface Plasmon Resonance (SPR) is frequently employed. SPR allows direct measurement of the binding of a biomacromolecule onto an appropriately functionalized gold surface using minimal quantities of each compound.[17,18] However, SPR equipment is specialized, often not high-throughput, is expensive and not routinely available in all laboratories. Other surface methods such as Quartz-Crystal Microbalance (QCM) and Optical Waveguide Spectroscopy (OWLS) have similar problems and benefits.[19] It should be noted that immobilization of proteins on surfaces, rather than carbohydrates is convenient but limited by the signal detection (which is essentially proportional to mass) which cannot detect small molecule binding. ELISA (Enzyme-Linked Immuno-Sorbent Assay) and fluorescence-linked assays have found wide-spread use due to their lower equipment level requirement (microplate reader) and excellent sensitivity limits.[20,21] A key challenge of sorbent assays is the need to use labelled proteins, rather than unmodified versions. Recombinant methods can introduce enzymatic or antigen tags which themselves can alter the function of the protein and direct chemical labeling of isolated proteins with fluorophores (such as fluorescein isothiocyanate) can provide detection, but gives rise to heterogeneous distributions of labels which leads to well known problems in terms of altering ligand binding-affinities.[22–24] Fluorescent labels are often not photostable presenting challenges in quantitative analysis.[25] The challenges associated with carbohydrate-immobilisation, or the generation of labelled carbohydrates should also not be underestimated.[26]

Considering that monitoring of these interactions has huge potential in the development of novel anti-adhesion therapies or diagnostics, any technology which simplifies this process, is compatible with native proteins and relies on readily available imaging platforms could have profound benefit. Gold nanoparticles (AuNPs) which display an extremely strong coloration due to their surface plasmon band (similar, but distinct from the analytical equipment discussed above) have been shown to strongly adhere to proteins by electrostatic interactions. For example, colloidal gold has been used to stain native proteins separated by electrophoresis on gels,[27] or those on microplates[28] allowing colorimetric detection. While useful, this technique has not found wide-spread use, has not been optimized in terms of particle size, nor has it been used to screen for novel inhibitors of lectin-binding. Other uses of AuNPs in imaging has also been reviewed extensively,[29–31] especially with regard to their aggregation-sensitive properties,[32] contrast agents in biological electron microscopy[33] and photothermal imaging.[34]

Here the use of AuNPs as resolving agents to allow quantification of protein absorption onto carbohydrate functionalized microplate surfaces and to use it as an accessible tool to probe multivalent inhibitors which are potential prophylactic antibacterial agents without the need for fluorescence detection is investigated. This visualization should reduce the need for expensive, complicated and low throughout analytical facilities and inaccessible, labelled, protein samples. The use of a simple digital camera to probe these interactions is also studied as an ultra-low cost, alternative.

Materials and methods

General

Concanavalin A (ConA), fluorescein isothiocyanate (FITC) labelled ConA, peanut agglutinin (PNA) from *Arachis hypogaea*, FITC labelled PNA, monosialotetrahexosylganglioside (GM-1), bovine serum albumin (BSA), cholera toxin subunit B (CTx-B), FITC labelled CTx-B, mannan, preformulated phosphate buffered saline tablets (PBS), calcium chloride ($CaCl_2$), sodium chloride (NaCl), HEPES, Tris buffer, trisodium citrate dihydrate and chloroauric acid ($HAuCl_4$) where all purchased from Sigma-Aldrich, UK. 96-well high binding microtitre plates were purchased from Greiner Bio-one. Galactose-functional polymers used in this study have been previously reported and their synthesis is summarised in the ESI.†[35] Phosphate buffered saline (PBS) was prepared using preformulated tablets dissolved in 200 mL of Milli-Q water (with resistance >18 mOhms) to give final concentrations of [NaCl] = 0.138 M, [KCl] = 0.0027 M and pH = 7.4. 10 mmol Tris buffer containing 0.1 mmol $CaCl_2$ and 0.5 mmol NaCl (pH 8, TBS) was prepared in 250 mL of Milli-Q water (with a resistance >19 mOhms), and 10 mmol HEPES buffer containing 0.1 mmol $CaCl_2$ (pH 6.5, HEPES) was prepared in 250 mL of Milli-Q water (with a resistance >19 mOhms).

Physical and analytical methods

Absorbance and fluorescence measurements were made using a BioTek Synergy HT multi-detection microplate reader and Gen5 software. Dynamic light scattering was conducted using a Nano-Zs from Malvern Instruments, UK. Scattered light was detected at 173° and the observed count rates recorded. Hydrodynamic radii (where appropriate) were determined using the manufacturer's software. Diameters are an average of 3 measurements using at least 10 scans. All protein binding curves were plotted in OriginPro (version 8.6) and the built in nonlinear curve fitting tool was used to fit a logistic curve to the data and to calculate the MIC_{50} values. MIC_{50} values in this report are defined as being the concentration of inhibitor required to reduce the binding of the lectin to the carbohydrate surface by 50% of its maximal value. The graphs showing absorbance spectra and the optimization lines were produced in Excel. Absorbance graphs were produced by plotting the absorbance measurement at 530 nm against concentration. All standard error values were calculated in Microsoft Excel and then imported into OriginPro and used to produce the error bars. A linear regression analysis for the paired cholera toxin data was performed in the open-source statistical package R (version 2.14.1).[36] Pictures of AuNP modified wells were taken using a Nikon D60 camera with automatic exposure settings from a distance of 30 cm using ambient light conditions.

Functionalization of multiwell plates

Greiner high binding 96-well plates were incubated for 16 h with 180 μL of 0.1 mg mL^{-1} mannan dissolved in PBS per well. After incubation, unbound mannan, GM-1 or BSA was removed by washing vigorously with distilled water, after which the

plates were dried and stored under an inert atmosphere at 4 °C. The same procedure was also used for GM-1 (0.1 mg mL^{-1}), bovine serum albumin (BSA) (0.1 mg mL^{-1}).

Gold nanoparticle synthesis

Sodium citrate capped gold nanoparticles were produced by chemically reducing chloroauric acid ($HAuCl_4$) with trisodium citrate dihydrate as the capping and reducing agent. An aqueous solution of trisodium citrate dihydrate (2 mL, 0.13 mmol) was added to a boiling solution of $HAuCl_4$ (35 mL, 0.35 mmol) under reflux and vigorous stirring producing a ratio of 3.5 : 1 ($HAuCl_4$ to sodium citrate). After addition of the trisodium citrate dihydrate the solution went from pale yellow to dark blue before eventually turning deep red (about 30 minutes after addition of the trisodium citrate dihydrate). After which the solution was cooled and then stored in the dark at 4 °C. Solutions were also produced using the ratios 2.5 : 1 and 1.5 : 1 following the same procedure but using different ratios of $HAuCl_4$ to trisodium citrate dihydrate.

Lectin binding assays

Lectin solutions were made up as serial dilutions in TBS from 1 mg mL^{-1} stocks apart from ConA–FITC which was made up in HEPES at 1 mg mL^{-1}. The lectin solutions were then added to 96-well plates functionalized with their target carbohydrate and incubated at 37 °C for 30 min. Unbound lectin was removed by washing extensively with buffer, followed by Milli-Q water. For FITC-labelled lectins, fluorescence was then measured at excitation/emission wavelengths of 485/528 nm. Alternatively, 50 μL of AuNP solution was added to each well and then incubated at 37 °C for 30 min. After extensive washing with Milli-Q water, the absorbance in each well was measured between 450 and 700 nm in 1 nm steps. All experiments were carried out in triplicate.

Competitive inhibition assay

96-well microtitre plates were incubated for 16 h with 180 μL of 100 μg mL^{-1} GM-1 dissolved in PBS, per well. Unbound GM-1 was removed by washing extensively with PBS and then water. Polymer solutions were made up as serial dilutions (up to 10 dilutions per sample in TBS from 1 mg mL^{-1}). 20 μL of 12.5 μg mL^{-1} CTx-B in TBS was added to 100 μL of each polymer solution to result in a final concentration of 2 μg mL^{-1} CTx-B per well. 100 μL of the polymer/CTx-B solutions were then added to the GM-1-coated wells and incubated at 37 °C for 30 min. After this time, they were extensively washed with Milli-Q water. 50 μL of AuNP solution was then added to each well before incubation at 37 °C for 30 min. After extensive washing with Milli-Q water, absorbance in each well was measured between 450 and 700 nm in 1 nm steps. All experiments were carried out in triplicate.

Digital camera-linked assay and image analysis

Microplates were prepared as described above, and the CTx-B lectin applied, washed and AuNP added following the above-described procedures. A digital photograph of the AuNP-modified plates was taken from a distance of 30 cm using ambient light and the tiff image file uploaded into the open-source image processing package ImageJ[37] (version 1.46a) where a region of interest was drawn around every well. The colour (RGB) image was then converted into a hue saturation and brightness (HSB) stack of images and the saturation image used. The regions of interest drawn on the original image were added to the saturation image using the ROI manager and average pixel intensity in each region of interest was measured using an inbuilt function in ImageJ. An example image showing the colour change achieved by gold binding to protein bound to GM-1 is shown in Fig. 6.

Results and discussion

The aim of this work was to use citrate-stabilized AuNPs as resolving agents for direct measurement of (unlabelled) protein-binding onto carbohydrate-functional surfaces, taking advantage of the extremely strong coloration of AuNPs. Here it is exploited to screen novel anti-adhesion therapies without the need for labelled proteins and to test the validity of direct 'visual' analysis of protein binding using a simple digital camera as an ultra-low cost platform.[38] The concept is outlined in Fig. 1.

A small panel of citrate-stabilized AuNPs were obtained using the standard citrate/$HAuCl_4$ reduction approach.[39,40] This method was chosen as it is simple, scalable and gives rise to charged nanoparticles, which are essential for strong non-specific protein interactions (*vide infra*). Table 1 summarises the synthesis of AuNPs used in this study which were characterized by UV-visible spectroscopy and dynamic light scattering to

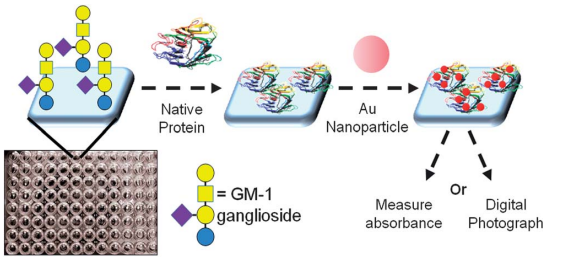


Fig. 1 Overview of the approach used to allow direct measurement of lectin–carbohydrate binding by visualization of surface-bound protein using AuNPs on microtitre plates.

Table 1 Citrate-stabilized AuNPs used in this study

Code	[Citrate] : [Gold][a]	λ_{spr}[b] (nm)	Particle size[c] (UV) (nm)	Particle size (DLS) (nm)	PDI[d]
AuNP1	3.5 : 1	526	36 nm	36	0.35
AuNP2	2.5 : 1	541	76 nm	76	0.58
AuNP3	1.5 : 1	545	70 nm	130	0.38

[a] Molar ratio of trisodium citrate to $HAuCl_4$ used in the synthesis of the nanoparticles. [b] Location of the surface plasmon peak in the UV-visible spectrum of the particles. [c] Particle size estimated by the method of Haiss *et al.*[41] [d] Polydispersity index of the particles, obtained by DLS.

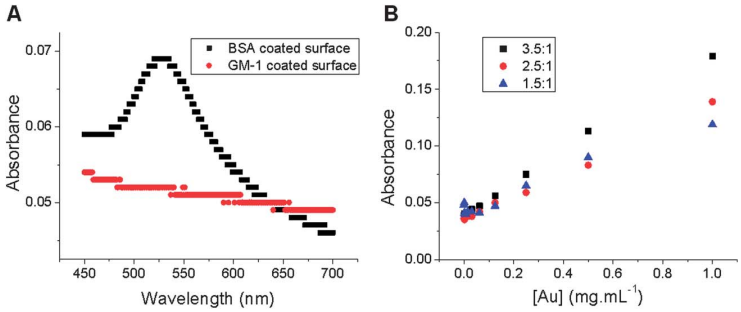


Fig. 2 Evaluation of AuNP surface interactions. (A) UV-visible spectra of BSA-coated surface (black trace) and GM-1 coated surface (red trace) following addition of **AuNP1**. (B) Absorbance intensity of protein-bound wells at 525 nm as a function of added AuNP concentration, onto BSA-coated microplates. Legend shows ratio of [citrate] : [Au] used in preparation of the nanoparticles.

provide an estimate of their dimensions, which were in the range of 30–70 nm. The polydispersities were quite large, as is common with direct citrate-only reduction methods.

The key requirement of this study is that the citrate-coated AuNPs interact with proteins that have bound to carbohydrates immobilized onto the surface of microtitre plates, providing the necessary visualization (red colour) to determine the extent of protein binding. To test the hypothesis, two different surfaces were prepared in 96-well microtitre plates designed to promote or inhibit AuNP binding. BSA was directly adsorbed onto high-binding microtitre plates to provide a 'protein rich' surface and the GM-1 ganglioside, which presents a carbohydrate-surface, was also immobilized (see Experimental section for details). These surfaces were subsequently exposed to a 0.3 mg mL^{-1} solution of **AuNP1** for 30 minutes, washed with Milli-Q water, and then their UV-Vis spectra measured, Fig. 2A. The BSA surface clearly shows absorbance at ~520 nm, attributable to AuNP absorbing to the protein surface, whereas the GM-1 surface shows no signal. The branched oligosaccharide component of GM-1 presents a highly hydrated ('non-fouling'-type[42]) surface which will not interact electrostatically with the anionically charged AuNPs. The 3 different AuNPs listed in Table 1 were subsequently evaluated with BSA surfaces as a function of AuNP concentration to determine the optimum conditions and detection limits, Fig. 2B. Across the AuNP concentration range of 0–1 mg mL^{-1}, the smallest nanoparticles (3.5 : 1 [citrate] : [Au]) gave the largest signal intensity. To maximize the signal intensity, all subsequent tests were undertaken by exposing the protein-bound surfaces to **AuNP1** at 1 mg mL^{-1} (total Au mass).

Table 2 Coatings and lectins used in this study

Coating	Carbohydrate–ligand	Lectin
Mannan	α-D, 1–6 linked mannose	Concanavalin A
GM-1	β-D-Galactose	Cholera toxin B-subunit
GM-1	β-D-Galactose	Peanut agglutinin

To evaluate the utility of AuNP as a reliable, scalable, accessible and economic alternative to standard SPR/QCM/ELISA type measurements of carbohydrate–protein interactions a representative panel of lectins were selected with, and without, fluorescent labels to allow us to validate this approach. Mannose-functional surfaces were obtained by coating microtitre plates with mannan and galactose-functional plates by coating with GM-1. Concanavalin A (ConA) was used as a mannose-specific lectin, and peanut agglutinin (PNA) and the B-subunit of the cholera toxin (CTx-B) as distinct galactose-specific lectins, Table 2. Fig. 3 shows the dose-dependent binding of PNA to GM-1 as a function of PNA concentration,

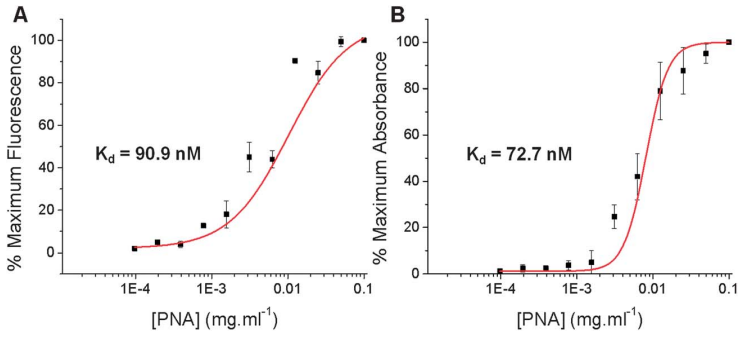


Fig. 3 Dose dependant binding curves of PNA onto GM-1 functional surfaces. (A) Fluorescence measurements (FITC – excitation/emission 485/528 nm); (B) AuNP absorbance at λ_{max}. Errors bars represent ± standard deviation from a minimum of 3 repeats.

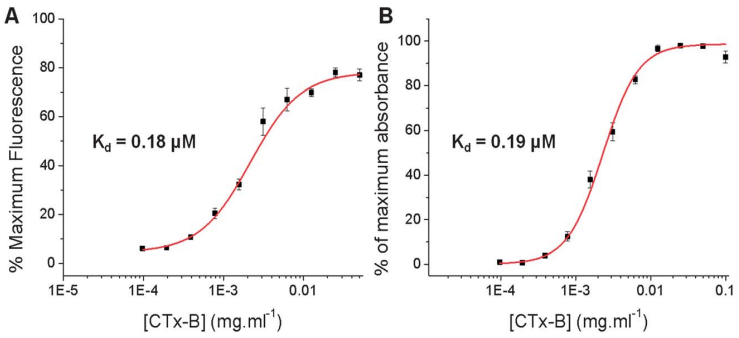


Fig. 4 Dose dependant binding curves of CTx-B onto GM-1 functional surfaces. (A) Fluorescence measurements (FITC – excitation/emission 485/528 nm); (B) AuNP absorbance at λ_{max}. Errors bars represent ± standard deviation from a minimum of 3 repeats.

with binding measured by fluorescence (A) or by addition of 50 μL of 1 mg mL^{-1} **AuNP1** (B). Both binding curves have similar profiles and give calculated dissociation constant (K_d's) of 72.7 nM (AuNP) and 90.9 nM (fluorescence) respectively. The strong agreement between the two methods shows that the AuNPs are useful and can be used to analyse complex interactions without the need for specialised equipment, or labelled proteins. Similar results were obtained when the ConA/mannan system were analysed (ESI†). Furthermore, the AuNP-resolved microtitre plates were not observed to degrade over time. After 2 weeks storage, identical data was obtained in contrast to fluorescence or enzymatic methods which suffer from photobleaching or denaturation, respectively. It should be noted that although the AuNPs will most likely bind each protein to a different extent, this does not affect the measurements shown here, which are probing equilibrium process and hence all that is required is a dose–response curve with clear start and end points.

A more clinically relevant subset of lectins for study are those secreted by pathogenic bacteria which are targets for prophylactic anti-adhesion therapy. The toxin secreted by the pathogen *Vibrio cholera* binds to GM-1 on epithelial cells and is the causative agent of cholera infection. The native cholera toxin contains 5 carbohydrate binding units ('B-subunits'). In this study a single B-subunit is used as a non-toxic analogue and the interactions probed in the same manner as PNA, Fig. 4. K_d values obtained for CTx-B by fluorescence was 0.19 μM and by AuNP 0.18 μM again showing exceptional agreement.

The data presented above shows that AuNP based measurement of protein binding is comparable to fluorescence

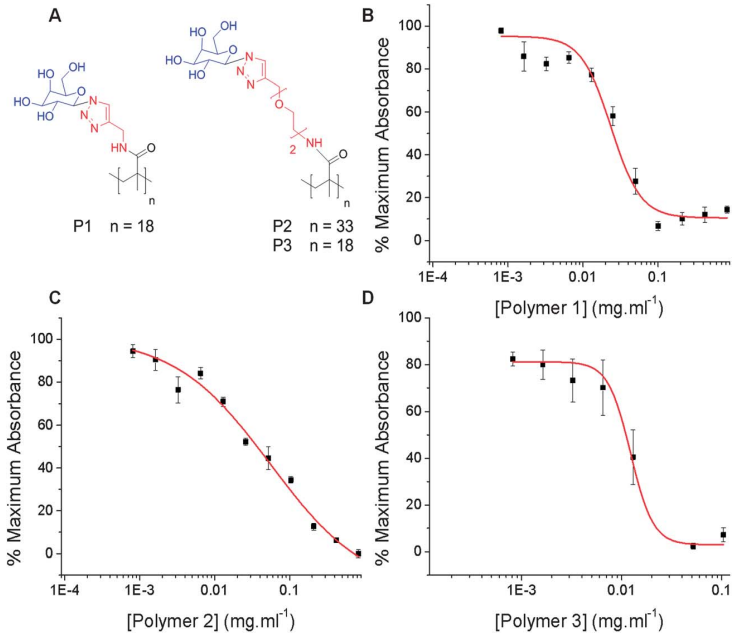


Fig. 5 Inhibition of CTx-B binding by β-D-galacto-functional glycopolymers. (A) Structure of glycopolymeric inhibitors; (B–D) inhibitory curves of polymers **1**, **2** and **3** respectively, with CTx-B, measured using AuNP method. Errors bars represent ± standard deviation from a minimum of 3 repeats.

Table 3 Inhibitory potency (MIC_{50}) of glycopolymers against CTx-B determined by gold nanoparticle colorimetric method

Polymer	DP_n [a]	Linker [b]	MIC_{50}	MIC_{50} [c] [polymer]	MIC_{50} [d] ([galactose])
P1	18	Short	56 ± 2.4 μg mL^{-1}	9.4 ± 0.4 μM	0.169 ± 0.072 mM
P2	33	Long	23 ± 18 μg mL^{-1}	1.6 ± 1.2 μM	0.055 ± 0.43 mM
P3	18	Long	24 ± 11 μg mL^{-1}	3.2 ± 1.46 μM	0.057 ± 0.027 mM

[a] Number average degree of polymerization of polymer (= galactose/chain). [b] Refers to length of spacer between polymer and carbohydrate shown in Fig. 5. [c] MIC_{50} adjusted in terms of concentration of polymer chains. [d] MIC_{50} adjusted in terms of total galactose concentration. The polymers used were reported in a previous study.[35] Error bars represent the standard deviation in the individual MIC_{50} from each individual measurement (minimum of 3 repeats).

methods. We are ultimately interested in probing for inhibition of pathogenic toxins, and therefore the Au-system was applied to a competitive binding experiment. We and others have demonstrated that synthetic glycopolymers, presenting many copies of a single carbohydrate on their side chain display high binding affinities to lectins due to the cluster glycoside effect.[35,43] Recently we demonstrated that modulation of the distance between polymer backbone and carbohydrate can improve the affinity of poly(galactoside)s for CTx-B due to an improved fit to its (relatively) deep binding site. Here, three synthetic glycopolymers which were prepared by tandem post-polymerization modification[44] with precise control over valency and linker length were assayed for their ability to inhibit CTx-B binding, Fig. 5.

P1 with a short linker showed the highest MIC_{50} value (lowest affinity) of 56 μg mL^{-1} whereas the polymers with longer linkers (**P2**/**P3**) had approximately identical MIC_{50} values of 23 and 24 μg mL^{-1} respectively which is in strong agreement with previously published observations.[35] On a per-polymer (molar) basis, **P3** was the most active, but this is biased by its higher molecular weight and hence lower molar concentration. MIC_{50} corrected to carbohydrate concentration indicates that **P3** and **P2** have identical activity and therefore in the case of CTx-B, there is no additional benefit of increasing chain length above 18 units (although further benefit of even larger polymers cannot be discounted). Compared to free galactose, all polymers were at least 100× more active highlighting the cluster glycoside effect (Table 3). This method is therefore suitable, and amenable to high-throughput screening of lectin inhibition using un-labelled proteins.

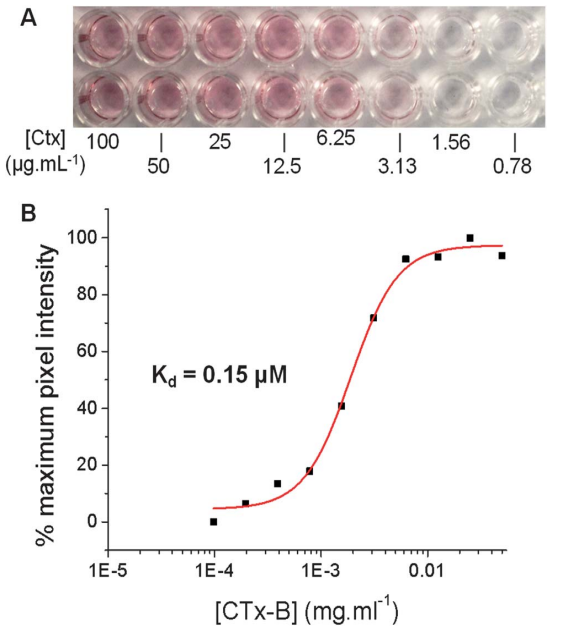


Fig. 6 Direct optical analysis of protein binding. (A) Digital photograph of GM-1 functional microplates treated with a dilution series of CTx-B and following developing with AuNPs for 30 minutes. (B) Binding curve obtained by pixel-counting direct from digital photographs.

Measuring binding using digital photography

During the above described experiments, it was observed that upon addition of AuNPs to the protein-functional plates, sufficient gold was deposited to allow visual examination of protein binding, due to the intense red colouration of AuNPs (their extinction coefficients are typically 100–1000 larger than fluorescence dyes). Fig. 6A shows a photograph of a GM-1 microtitre plate following incubation with a dilution series of CTx-B, revealed by addition of AuNPs. Strong red colouration can be seen at high concentrations of CTx-B, with the colour decreasing in line with decreasing protein concentrations. This visual read-out is potentially useful for low-cost diagnostic applications, rapid protein-binding assays in laboratories or 3rd world countries, particularly if it can be linked to a mobile phone or digital camera system. This would remove the need for a plate reader or similar whilst maintaining the ability to obtain relevant data.[38] The AuNP-exposed plates were photographed using a standard digital camera (see Experimental section) and imported into image analysis software. The average pixel intensity of each well was measured and plotted against concentration of protein added, Fig. 6B. Impressively, near identical binding curves were obtained compared to those in Fig. 5 (Ctx-B by either absorbance/fluorescence), with K_d values close to that obtained by either fluorescence or absorbance. Fig. 7 shows the correlation between average pixel intensity and both total fluorescence or absorbance. This analysis gave a strong positive correlation linear fits with R^2 values above 0.98. This offers the opportunity for evaluation of lectin binding/inhibition in laboratories lacking in dedicated glycomics facilities and would be compatible with a cellular-phone based analytical system.

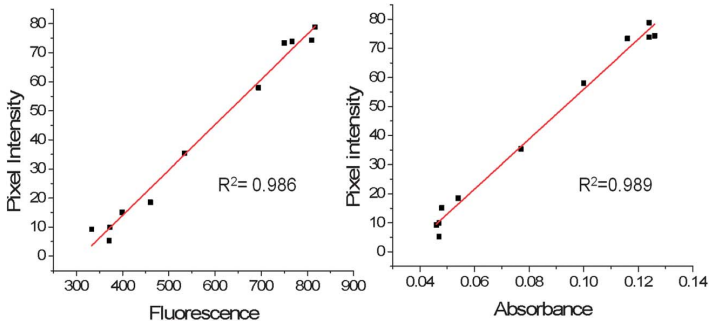


Fig. 7 Comparison of signal intensities obtained by pixel-counting methods verses AuNP and traditional fluorescence measurements.

Conclusions

In this study, we have investigated the use of citrate-coated AuNPs as resolving agents to analyse protein binding to carbohydrate functionalized surfaces, as a low-cost, accurate glycomics tool. Carbohydrate-coated surfaces were incubated with their corresponding lectins and their binding isotherms were measured by direct fluorescence-linked assays and by absorbance measurements based on gold nanoparticle absorption. Strong correlation was found between the two methods, with the gold particle method having the advantage that it removes the need to use labelled proteins and can be read with a standard microplate reader, without the need for fluorescence measurements. To demonstrate the applicability of this, a series of multivalent glycopolymers were screened for their ability to inhibit the binding of B-subunit of the cholera toxin, as a model of a pathogenic infection process. Finally, due to the extreme strong colouration of the gold-nanoparticle functionalized surfaces, it was possible to use a digital camera to simply image the microplates and use image analysis software to extract binding isotherms, which showed strong correlation to those obtained by fluorescence or absorbance measurements. This presents the opportunity to use gold nanoparticle resolving agents for label free, low cost, high-throughout evaluation of carbohydrate–protein interactions which may have diagnostic or screening applications. In particular, it allows rapid testing of inhibitors and concentrations ranges prior to more detailed biophysical analyses (*e.g.* SPR/QCM). Current work is focussed on exploiting this methodology to identify new carbohydrate-binding proteins, and to rapidly screen for novel anti-adhesion therapies.

Acknowledgements

Equipment used was supported by the Innovative Uses for Advanced Materials in the Modern World (AM2), with support from Advantage West Midlands (AWM) and part funded by the European Regional Development Fund (ERDF). MIG is a Birmingham Science City Interdisciplinary Research Fellow funded by the Higher Education Funding Council for England (HEFCE). G. B. is the Bardrick professor of microbial physiology and chemistry and holds a Royal Society Wolfson merit award. L.O. acknowledges the BBSRC funded Systems Biology DTC for a studentship. S.-J.R. acknowledges the EPSRC funded MOAC DTC for a studentship. Mathew W. Jones (now University of New South Wales) is acknowledged for providing the amino-PEG-alkyne linkers used in this study.

References

1 R. Schauer, *Glycoconjugate J.*, 2000, **17**, 485–499.
2 G. K. Hirst, *J. Exp. Med.*, 1948, **87**, 301–314.
3 T. R. Hirst, S. Fraser, M. Soriani, A. T. Aman, L. de Haan, A. Hearn and E. Merritt, *Int. J. Med. Microbiol.*, 2001, **291**, 531–535.
4 T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. van Kooyk, *Cell*, 2000, **100**, 587–597.
5 G. Mulvey, P. I. Kitov, P. Marcato, D. R. Bundle and G. D. Armstrong, *Biochimie*, 2001, **83**, 841–847.
6 A. Imberty, Y. M. Chabre and R. Roy, *Chem.–Eur. J.*, 2008, **14**, 7490–7499.
7 S. G. Spain and N. R. Cameron, *Polym. Chem.*, 2011, **2**, 60–68.
8 R. J. Pieters, *Med. Res. Rev.*, 2007, **27**, 796–816.
9 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.
10 S. G. Spain, M. I. Gibson and N. R. Cameron, *J. Polym. Sci., Part A: Polym. Chem.*, 2007, **45**, 2059–2072.
11 J. J. Lundquist and E. J. Toone, *Chem. Rev.*, 2002, **102**, 555–578.
12 X. Hong, M. Z. Ma, J. C. Gildersleeve, S. Chowdhury, J. J. Barchi, R. A. Mariuzza, M. B. Murphy, L. Mao and Z. Pancer, *ACS Chem. Biol.*, 2012, **8**, 152–160.
13 P.-C. Pang, P. C. N. Chiu, C.-L. Lee, L.-Y. Chang, M. Panico, H. R. Morris, S. M. Haslam, K.-H. Khoo, G. F. Clark, W. S. B. Yeung and A. Dell, *Science*, 2011, **333**, 1761–1764.
14 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357–2364.
15 K. T. Pilobello and L. K. Mahal, *Curr. Opin. Chem. Biol.*, 2007, **11**, 300–305.
16 S. G. Spain and N. R. Cameron, *Polym. Chem.*, 2011, **2**, 1552–1560.

17 E. A. Smith, W. D. Thomas, L. L. Kiessling and R. M. Corn, *J. Am. Chem. Soc.*, 2003, **125**, 6140–6148.
18 C. R. Becer, M. I. Gibson, R. Ilyas, J. Geng, R. Wallis and D. A. Mitchell, *J. Am. Chem. Soc.*, 2010, **132**, 15130–15132.
19 F. Hook, J. Voros, M. Rodahl, R. Jurrat, P. Boni, J. J. Ramsden, M. Textor, N. D. Spencer, P. Tengvall, J. Gold and B. Kasemo, *Colloids Surf., B*, 2002, **24**, 155–170.
20 E. Engvall and P. Perlmann, *Immunochemistry*, 1971, **8**, 871–874.
21 R. M. Lequin, *Clin. Chem.*, 2005, **51**, 2415–2418.
22 Y.-S. Sun, J. P. Landry, Y. Fei, X. Zhu, J. T. Luo, X. B. Wang and K. S. Lam, *Langmuir*, 2008, **24**, 13399–13405.
23 T. Kodadek, *Chem. Biol.*, 2001, **8**, 105–115.
24 Y. Fei, Y.-S. Sun, Y. Li, K. Lau, H. Yu, H. A. Chokhawala, S. Huang, J. P. Landry, X. Chen and X. Zhu, *Mol. BioSyst.*, 2011, **7**, 3343–3352.
25 C. Eggeling, J. Widengren, R. Rigler and C. A. M. Seidel, *Anal. Chem.*, 1998, **70**, 2651–2659.
26 K. Larsen, M. B. Thygesen, F. Guillaumie, W. G. T. Willats and K. J. Jensen, *Carbohydr. Res.*, 2006, **341**, 1209–1234.
27 R. Rohringer and D. W. Holden, *Anal. Biochem.*, 1985, **144**, 118–127.
28 C. M. Stoscheck, *Anal. Biochem.*, 1987, **160**, 301–305.
29 R. A. Sperling, P. R. Gil, F. Zhang, M. Zanella and W. Parak, *Chem. Soc. Rev.*, 2008, **37**, 1896–1908.
30 H. Jans and Q. Huo, *Chem. Soc. Rev.*, 2012, **41**, 2849–2866.
31 V. W. K. Ng, R. Berti, F. Lesage and A. Kakkar, *J. Mater. Chem. B*, 2013, **1**, 9–25.
32 S. Gupta, H. Andresen and M. M. Stevens, *Chem. Commun.*, 2011, **47**, 2249–2251.
33 C. Freese, M. I. Gibson, H.-A. Klok, R. E. Unger and C. J. Kirkpatrick, *Biomacromolecules*, 2012, **13**, 1533–1543.
34 C. Pache, N. L. Bocchio, A. Bouwens, M. Villiger, C. Berclaz, J. Goulley, M. I. Gibson, C. Santschi and T. Lasser, *Opt. Express*, 2012, **20**, 21385–21399.
35 S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew. Chem., Int. Ed.*, 2012, **51**, 7812–7816.
36 *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, 2011, http://www.R-project.org/.
37 W. S. Rasband, *ImageJ*, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/, 1997–2012.
38 A. W. Martinez, S. T. Phillips, E. Carrilho, S. W. I. Thomas, H. Sindi and G. M. Whitesides, *Anal. Chem.*, 2008, **80**, 3699–3707.
39 J. Turkevich, P. C. Stevenson and J. Hillier, *Discuss. Faraday Soc.*, 1951, **11**, 55–75.
40 G. Frens, *Nature (London), Phys. Sci.*, 1973, **241**, 20–22.
41 W. Haiss, N. T. K. Thanh, J. Aveyard and D. G. Fernig, *Anal. Chem.*, 2007, **79**, 4215–4221.
42 J. Wang, M. I. Gibson, R. Barbey, S.-J. Xiao and H.-A. Klok, *Macromol. Rapid Commun.*, 2009, **30**, 845–850.
43 M. W. Jones, S.-J. Richards, D. M. Haddleton and M. I. Gibson, *Polym. Chem.*, 2013, **4**, 717–723.
44 N. K. Singha, M. I. Gibson, B. P. Joiry, M. Danial and H.-A. Klok, *Biomacromolecules*, 2011, **12**, 2908–2913.

Appendix 2: Glycopolymers with secondary binding motifs mimic glycan branching and display bacterial lectin selectivity in addition to affinity.

Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Glycopolymers with secondary binding motifs mimic glycan branching and display bacterial lectin selectivity in addition to affinity†

Cite this: *Chem. Sci.*, 2014, **5**, 1611

M. W. Jones, L. Otten, S.-J. Richards, R. Lowery, D. J. Phillips, D. M. Haddleton and M. I. Gibson*

Received 28th October 2013
Accepted 6th February 2014

DOI: 10.1039/c3sc52982g

www.rsc.org/chemicalscience

The application of synthetic glycopolymers to anti-adhesive therapies has so far been limited by their lack of lectin specificity. Here we employ a macromolecular engineering approach to mimic glycan architecture. A new, 3-step tandem post-polymerisation methodology was developed which afforded precise control over both chain length and carbohydrate (galactose)-polymer backbone linker distance. This route also allowed a secondary binding (branched) motif to be introduced onto the linker, increasing specificity and affinity towards bacterial toxins without the need for extensive carbohydrate or organic chemistry. Sequential variation of this motif was found to dramatically alter both the affinity and the specificity of the glycopolymers towards two lectins, CTx and PNA, by up to 20-fold either *via* direct binding, or increased steric constraints. Using this method, a glycopolymer that showed increased specificity towards CTx was identified.

Introduction

Protein–carbohydrate interactions dictate the outcomes of a large and varied number of cellular recognition processes, controlling immune responses, tumour metastasis, gamete fertilisation and many more.[1] The structure, function or even identity of many glycans remains unknown. It is estimated that 50% of human proteins are glycosylated and there remains significant analytical challenges associated with the isolation and characterisation of complex glycans.[2] Proteins which recognise and process the signals associated with carbohydrates are termed lectins: carbohydrate binding proteins which are neither antibodies nor enzymes and they are widely distributed in nature.[3] Despite their role in normal physiology, lectins/glycans can also act as a potential site for infection, which can be exploited by pathogenic organisms to interface with their host. For example, pathogenic *Escherichia coli*, *E. coli*, expresses the FimH adhesin, which can bind mannose residues in the urinary tract, influenza has sialic acid binding lectin (heamaglutinins) for adhesion to erythrocytes and *Vibrio cholerae* secretes a toxin which binds to intestinal epithelial cells.[1,4–6] Conversely, HIV expresses high-mannose structures on its capsid that enables it to bind to DC-SIGN lectin on the surface of dendritic cells in the human immune system.[7] With the widespread emergence of antibiotic resistance, new interventions to prevent and detect infectious disease are urgently required.[8,9] Anti-adhesion therapy, which seeks to use compounds that have higher affinity than the pathogen for the target binding site, thus preventing the adhesion step and hence reducing the infectivity, has emerged as a promising potential treatment.[4,10–14] As this process does not involve killing the pathogen, there should be no evolutionary stress, hence reducing resistance development and could be administered prophylactically.

The binding affinity of a carbohydrate to its target lectin is typically very weak ($K_d = 10^3$–10^6 M^{-1}), limiting their use in anti-adhesion therapy. The sugars' weak affinity is overcome in nature by the presentation of multiple copies on cell surfaces which gives rise to an increase in affinity which is greater than the linear sum of the individual sugars, the so-called cluster glycoside effect.[15–19] Lee *et al.* first demonstrated that multivalent *N*-acetyllactosamine with one to four carbohydrates showed progressively increasing binding affinities, over several orders of magnitude, towards rabbit hepatocytes.[15] Kiessling and coworkers have elegantly shown that polymer architecture (linear, branched, dendritic) has profound effects on lectin binding affinity, with particular focus on their ability to cluster receptors.[20] Ambrosi *et al.* found that galacto-functional polymers have a 100-fold increase in binding affinity compared to free galactose.[21] STARFISH-based monodisperse glycodendrimers were shown by Kitov *et al.* to neutralise Shiga-like toxins, with a measured affinity over 10^6-fold greater than the monovalent carbohydrate.[22] Exploitation of the high affinity of glycopolymer–lectin interactions also has applications in biosensing.[23,24]

Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK. E-mail: m.i.gibson@warwick.ac.uk

† Electronic supplementary information (ESI) available: Glycan notation and microarray analysis; full synthetic procedures and characterisation, additional inhibition assay details, and additional binding curve details. See DOI: 10.1039/c3sc52982g

Advances in controlled (radical) polymerisation methods[11,25,26] together with the development of highly efficient and orthogonal 'click'[27] reactions has facilitated the synthesis of glycopolymers by pre- and post-polymerisation modification thus widening the chemical and architecture diversity of glycopolymers.[25,28] Despite the interest in developing synthetic glycopolymers with high affinity for their respective lectins, there has been significantly less focus placed on their specificity.[6] What is often not studied is the relative affinity of the glycopolymer for various lectins with similar specificities, which is essential to avoid unwanted therapeutic side effects or to enable precise diagnostics. An important target for multivalent anti-adhesion therapies/diagnostics is the toxin CTx, secreted by *Vibrio cholerae*, the causative agent of cholera infection which is estimated to cause over 100 000 deaths per year, and infects more than 3 million people.[29] β-Galactose functional inhibitors have been shown to have high affinity to CTx, but there is a need to avoid unwanted interactions with other mammalian lectins, such as the galectins, which also bind β-galactose and could lead to immune responses such as cytokine production.[30] Lectin targets for which selective anti-adhesive therapies have been studied include DC-SIGN/Langerhin in HIV therapy, or BmbL/DC-SIGN to treat *Burkholderia ambifaria*.[31–34] The relative affinity of a series of bivalent galactosides towards chicken galectins has also been studied.[35] Moreover, selectivity presents a challenge when identifying biological warfare agents based on lectins such as ricin.[36] Oligosaccharide-mimetic agents have also been developed with high specificity based on tuning their 3-D structure to fit the lectin binding site but without the need for total oligosaccharide synthesis.[37]

Examples of synthetic polymers that have demonstrated lectin selectivity are rare, despite the obvious benefits of their multivalent nature.[38] We have demonstrated that galactose-functional polymers can be engineered to have increased selectivity for cholera toxin B-subunit (CTxB).[39] This was achieved by modulating the distance between backbone and carbohydrate to match the relatively deep cleft of the CTx binding domain, compared to other galactose-binding lectins with shallower domains.[40] Selectivity is required here to discriminate between other pathogenic lectins (or bacteria) that bind galactose including ricin[36] or indeed dietary lectins which can reduce the inhibitor's potency as this would need to function in the intestinal tract.

In this work, we present a macromolecular engineering approach to introduce specificity into glycopolymers, inspired by glycan branching and guided by structural biology information. In particular, we were motivated to achieve selectivity without resorting to multi-step total glycan synthesis, which is non-trivial. Using a new, 3 step-tandem post-polymerisation process, secondary binding (branched) motifs are introduced to the polymer side chain, to increase specificity and affinity towards bacterial toxins.

Results and discussion

As the first step, microarray data were extracted from the Consortium for Functional Glycomics database to enable a short bioinformatics study to probe lectin specificity/affinity.[41] The relative affinity of CTx and a model galactose-binding non-pathogenic lectin, PNA (peanut agglutinin), to several oligosaccharides was measured and the most relevant results are shown in Fig. 1A (full analysis in ESI†). PNA was found to bind the disaccharide Gal-β-GalNAc 100-fold more than CTx. However, changing this to a branched oligosaccharide (GM1), which retains the Gal-β-GalNAc unit but also introduces a neuraminic acid branch results in the CTx affinity increasing approximately 100-fold, but with no change in PNA affinity. The increased binding affinity of CTx to the branched saccharide is attributable to allosteric interactions of the neuraminic acid with a secondary binding pocket within CTx, which is not present in PNA.[42] We therefore reasoned that if a secondary-binding motif could be installed on the linker between galactose and backbone on a polymer it would be possible to attenuate the binding affinity of the polymer to CTx as shown by Tran *et al.*,[43] but also selectivity towards the CTx over PNA. Fig. 1B shows the crystal structure of CTx binding to the branched glycan unit from GM1 showing both the primary and secondary binding pockets, which is simplified in Fig. 1C using standard glycan notation. Fig. 1D shows the proposed polymer, with a sufficiently long linker to penetrate the cleft in CTx and a secondary motif to target the allosteric neuraminic acid site.

To enable installation of the branched motifs in a semi-combinatorial manner, we have developed a new synthetic methodology based on three tandem post-polymerisation modifications.[19,44] This introduces large chemical diversity that is not normally found on glycans, but ensures chain length (and hence valency) homogeneity across all samples overcoming a common challenge in functional polymers, Fig. 2.

The synthetic method fulfilled the following criteria: (i) sufficient separation between backbone and carbohydrate to enable penetration into the CTx binding site; (ii) an azide group

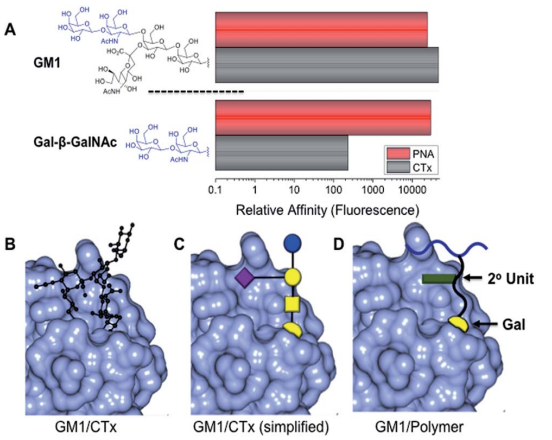


Fig. 1 (A) Glycan microarray analysis showing relative affinity of CTx and PNA to two related glycans; (B) crystal structure of CTx (blue) binding to the oligosaccharide portion of GM1; (C) CTx crystal structure with glycan drawn in standard ball/stick notation (ESI† for key); (D) synthetic polymer design concept with idealised polymer shown (schematic, not simulation).

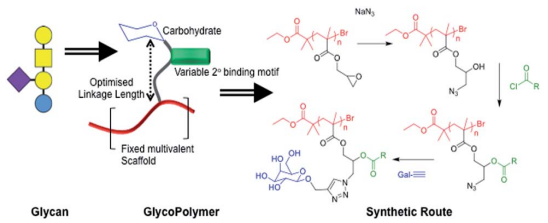


Fig. 2 Glycopolymer design principle and the newly developed synthetic route. Glycan structure is shown in standard notation (ESI†).

for subsequent glycosylation with β-D-propargyl galactose; (iii) esterification of the hydroxyl group, generated during epoxide ring-opening. Poly(glycidyl methacrylate) was synthesised by Cu(I)-mediated polymerisation to produce a well-defined polymer with a degree of polymerisation ~100 and $M_w/M_n = 1.2$. This molecular weight was targeted as our previous results have shown that above a DP of ~30, no further increase in avidity towards CTx was observed.[39] The polymer was produced by controlled radical polymerisation to ensure a lack of low molecular weight tail which would confuse the interpretation of the activity measurements (*vide infra*). Installation of the azide was achieved by addition of sodium azide in DMF at 50 °C, which simultaneously, and quantitatively, installed the necessary orthogonal handle and produced a secondary alcohol as confirmed by infrared spectroscopy (IR), Fig. 3. In the second step a range of acyl chlorides were reacted with the alcohol to install secondary motifs as confirmed by the disappearance of the OH stretch at 3400 cm^{-1} and the addition of a second carbonyl stretching frequency. The acyl chlorides were chosen based on evidence that aromatic groups can bind the sialic acid site.[43] In the final modification reaction, β-D-propargyl galactose was installed by Cu(I)-catalysed cycloaddition, which could be monitored by the reduction in the azide vibration at 2100 cm^{-1}. Table 1 summarises the polymer library obtained, the side chains installed and the calculated log *P* values of a single repeat unit of the polymer (*vide infra*). Log *P* values are included as an estimate of the relative hydrophobicity of the binding units.

With this panel of sequentially modified glycopolymers in hand, a sorbent assay was used to evaluate the affinity of the

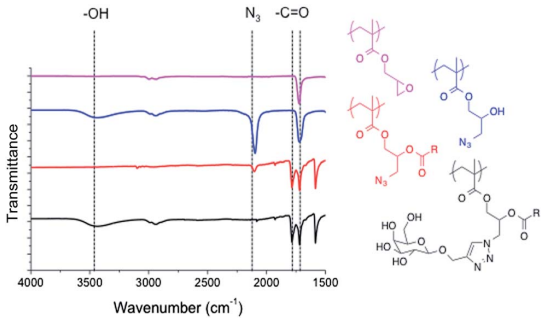


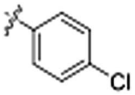
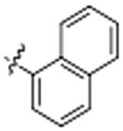
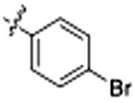
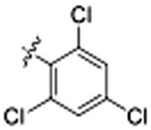
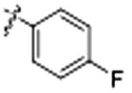
Fig. 3 Infrared analysis of the 3-stage, glycan mimetic, tandem post-polymerisation strategy used here. IR analysis was of purified product.

polymers towards each of the lectins.[24,39,45] Briefly, the glycopolymers were incubated at various concentrations with fluorescently labelled lectins. The solutions were then added to galactose-functionalised microtitre plates. The concentration of polymer required to inhibit 50% of binding to the plates was reported as the MIC_{50} value. Upon initial testing it was found that several members of the library were insufficiently soluble in buffer to be used in the assays. It was possible to solubilise the polymers in 5% (v/v) aq. DMSO, but we found this compromised the CTx-assay giving false positive results and hence only the polymers which could be directly dissolved into buffer were tested. The less soluble polymers are still shown in Table 1 to highlight the synthetic diversity achieved by this approach. As predicted, addition of the branched motifs had a dramatic influence on the inhibitory potential of the glycopolymers against both of the lectins, Fig. 4.

Fig. 4 reveals some general trends between molecular structure and lectin affinity. Three of the secondary units, **P8**, **P9** and **P10** gave rise to large 10-fold decreases in the MIC_{50} towards both lectins, compared to **P1**. Whilst these groups were relatively diverse, the common theme was that they did not contain an aromatic group, but **P8** and **P10** did contain halogenated alkanes. **P8** and **P10** side chains are significantly larger than in **P9** which suggests that their affinity modulation was not entirely due to steric constraints and may indicate that branching at the side chain increases affinity to CTx. Polymers **P11**, **P5** and **P6**, led to either no changes, or significantly increased the MIC_{50} values. Due to the size and rigidity of these functional groups, steric constraints might be crucial, preventing access of the polymer to both lectins, or limiting conformational flexibility. Polymer **P4**, which had a linear, but flexible, hexamethylene group gave modest improvements (lower MIC_{50}) in binding to both lectins. Our observations are in contrast to those of Bundle and coworkers who have observed that pendant aromatic units can enhance the binding to CTx *via* interactions with the neuraminic acid binding pocket, but they used polymers with a very low density of carbohydrate side chains and different length side chains on polydisperse scaffolds, making comparisons difficult.[43] Here we have densely packed side-chains that impose additional steric restraints. The high affinity of CTx to GM1 in nature is attributable to the intrinsic rigidity of GM1, which has also been found to be important in small-molecule GM1 mimics and is probably contributing here.[46,47] Comparison of the observed inhibitory values against the calculated partition coefficient did not reveal any obvious trend suggesting simple hydrophobic/hydrophilic interactions are not responsible (ESI†).

As indicated in the introduction, the key aim of this study was to use glycan-mimetic branching to introduce specificity/selectivity as well as affinity into synthetic glycopolymers. Analysis of the data in Fig. 4 revealed that **P1** and **P5** (chlorobenzyl) showed the most dramatic differences in terms of relative affinities for each lectin. Fig. 5 shows the relative affinity (shown as $1/MIC_{50}$ for convenience), for **P1** and **P5** against the two lectins. **P1** shows similar affinity for both PNA and CTx indicating that it cannot discriminate/select between two different galactose-binding lectins. However, addition of

Table 1 Side chains installed onto the polymers and log *P* values[a]

Code	Structure	log *P*	Code	Structure	log *P*	Code	Structure	log *P*
P1		−1.57 ± 0.61	**P5**		1.3 ± 0.7	**P9**		−0.17 ± 0.6
P2		1.96 ± 0.68	**P6**		1.48 ± 0.75	**P10**		−0.36 ± 0.82
P3		1.70 ± 0.7	**P7**		0.81 ± 0.75	**P11**		0.73 ± 0.68
P4		3.08 ± 0.6	**P8**		0.22 ± 0.63			

[a] log *P* values are calculated based on a single repeat unit of the polymer, with methyl capping groups at each chain end.

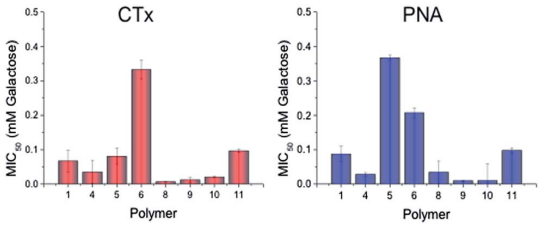


Fig. 4 Inhibition of lectin binding by synthetic glycopolymers. Values shown are the average of at least 3 measurements, and errors are the standard deviation.

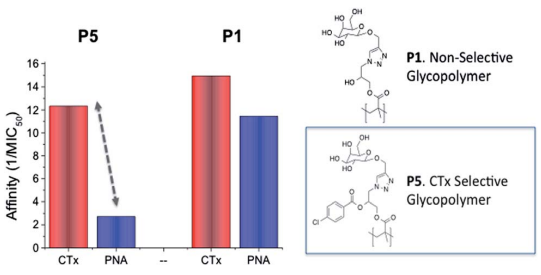


Fig. 5 Relative affinity (1/MIC_{50}) of **P1** and **P5** for CTx and PNA.

4-chlorobenzyl unit (**P5**) leads to dramatic differentiation in response to PNA and CTx, with a significant decrease in affinity towards PNA, but essentially no change in affinity to CTx. This demonstrates that **P5** displays lectin selectivity, using biomimetic macromolecular engineering, but without multistep carbohydrate chemistry. The exact mechanism of binding which leads to selectivity cannot be rationalised at this stage, but the additional bulk of the chlorobenzyl group may prevent access to the PNA binding site, but still be of correct dimensions to fit the neuraminic acid site in CTx. This will be the subject of future investigations. Furthermore, the influence of substituting the chloro- for bromo-group (**P5** – **P6**) cannot be explained. **P6** has vastly increased MIC_{50} values towards CTx relative to **P5**, but less effect on PNA binding. The additional steric bulk of a bromide might simply be too large for a good fit into the binding cleft in CTx. **P1**/**P5** also displayed similar affinities for RCA_{120}, another galactose-binding lectin indicating that the structural motifs added here only affect the target lectins (ESI†). This supports our hypothesis that the secondary motif is giving us the specificity based on interactions with the neuraminic acid binding site in CTx, which is not present in either RCA_{120} nor PNA, and hence the polymers' have decreased affinity to both of these lectins. These observations rule out non-specific hydrophobic association between polymer/lectins as this would be expected to give enhancements to all tested lectins.

Conclusions

We have demonstrated a new bio-inspired approach to mimicking glycan architecture by using macromolecular engineering, guided by structural biology, and without the need for multi-step oligosaccharide synthesis. Bioinformatics revealed that the addition of branched side chains to galacto-terminal carbohydrates could increase binding affinity to their corresponding lectins, relative to simple monosaccharides. To mimic this branched structure a new, 3 step, tandem post-polymerisation methodology was developed. This enabled precise control over not only chain length, but also carbohydrate-polymer backbone linker distance and the introduction of secondary binding (branched) motifs onto the linker. Sequential variation of this motif was found to dramatically (up to

20-fold) alter both the affinity and the selectivity of the glycopolymers towards two lectins; CTx and PNA. Using this method, a glycopolymer was identified which showed increased specificity towards CTx. Glycopolymers with high selectivity may feature in the development of sensitive and precise sensors or anti-adhesive therapies, which has so far limited the application of synthetic glycopolymers. These results show that combining structural biology tools with macromolecular chemistry enables the creation of synthetic glycans which can mimic, or outperform their natural counterparts and will find applications in anti-adhesive therapy and bimolecular sensors.

Experimental

Example acylation reaction using benzoyl chloride

Poly(2-hydroxy-3-azidopropyl methacrylate) (200 mg, 8.73 μmol) was dissolved in anhydrous THF (50 mL), along with triethylamine (0.45 mL, 3.24 mmol – 3 eq. for each polymer repeat unit). Benzoyl chloride (0.46 g, 3.24 mmol – 3 eq. for each polymer repeat unit) was dissolved in 50 mL of anhydrous DCM and added dropwise to the solution over a period of 30 minutes. Following complete addition, the solution was left to stir at room temperature for 24 hours. A further portion of triethylamine (0.45 mL, 3.24 mmol) and benzoyl chloride (0.46 g, 3.24 mmol) were added to the solution and allowed to stir for a further 24 hours. The solution was then diluted with 100 mL of DCM and quenched with 100 mL of water. The organic layer was washed with water (2 × 50 mL), dried over anhydrous $MgSO_4$, filtered and the solvent removed. The crude polymer solution was then redissolved in 50 mL of THF and twice precipitated into a 1 : 1 mixture of diethyl ether–petroleum ether. The solids were isolated by centrifugation and dried under vacuum to yield the product as a off-white powder.

Example 1,3-dipolar cycloaddition reaction of benzoyl chloride-modified polymer with galactose alkyne

Polymer (100 mg, 345.67 μmol), Cu(I)Br (4.9 mg, 34.16 μmol) and galactose alkyne (226 mg, 1.04 mmol) was dissolved in DMSO (8 mL) in a Schlenk tube. This solution was degassed by a minimum of 3 freeze–pump–thaw cycles and frozen with liquid nitrogen. The Schlenk tube was then opened, 2,2′-bipyridyl (10.8 mg, 69.15 μmol) was added and the tube re-sealed. The frozen solution was evacuated three times, back-filled with dry nitrogen and left to defrost. After stirring at ambient conditions for 4 days, the solution was diluted with distilled water and dialysed against water for 3 days. The resulting suspension was centrifuged and the supernatant was lyophilised to leave an off-white powder.

Example fluorescence-linked sorbent assay for inhibitory activity against cholera toxin B subunit (CTx)

96-well microtitre plates were incubated for 16 h with 150 μL of 100 μg mL^{-1} GCS (in 95% ethanol, 5% water and heated to 45 °C). Unbound GCS was removed by washing extensively with water. Polymer solutions were made up as serial dilutions (up to 10 dilutions per sample from either 1 mg mL^{-1} or 0.1 mg mL^{-1} in water). 10 μL of 100 μg mL^{-1} CTx–FITC in 10 mM Tris with 0.1 mM $CaCl_2$ and 0.5 mM NaCl (pH 8) was added to 90 μL of polymer solution to a final concentration of 11 μg mL^{-1}. 100 μL of the PNA/polymer solutions were then added to GCS coated wells and incubated at 37 °C for 30 min. After this the wells were extensively washed with water and fluorescence was measured at excitation/emission wavelengths of 485/528 nm. All experiments were carried out in triplicate. Percentage inhibition was compared to relative to controls of pure CTx–FITC (with no polymer).

Acknowledgements

Equipment used was supported by the Birmingham Science City (SC) Advanced Materials project, with support from Advantage West Midlands and part funded by the European Regional Development Fund. MIG was a Science City Research Fellow, supported HEFCE. SJR and RL are funded by EPSRC *via* MOAC DTC. LO is funded by BBSRC *via* Systems Biology DTC. DJP is funded by a scholarship from UoW.

Notes and references

1 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357–2364.
2 K. Marino, J. Bones, J. J. Kattla and P. M. Rudd, *Nat. Chem. Biol.*, 2010, **6**, 713–723.
3 M. Ambrosi, N. R. Cameron and B. G. Davis, *Org. Biomol. Chem.*, 2005, **3**, 1593–1608.
4 T. R. Branson and W. B. Turnbull, *Chem. Soc. Rev.*, 2013, **42**, 4613–4622.
5 T. Beddoe, A. W. Paton, J. Le Nours, J. Rossjohn and J. C. Paton, *Trends Biochem. Sci.*, 2010, **35**, 411–418.
6 R. R. Dinglasan and M. Jacobs-Lorena, *Infect. Immun.*, 2005, **73**, 7797–7807.
7 A. T. Haase, *Nature*, 2010, **464**, 217–223.
8 S. B. Levy and B. Marshall, *Nat. Med.*, 2004, **10**, S122–S129.
9 C. Dye, *Lancet*, 2006, **367**, 938–940.
10 R. J. Pieters, *Med. Res. Rev.*, 2007, **27**, 796–816.
11 S. G. Spain and N. R. Cameron, *Polym. Chem.*, 2011, **2**, 60–68.
12 G. Mulvey, P. I. Kitov, P. Marcato, D. R. Bundle and G. D. Armstrong, *Biochimie*, 2001, **83**, 841–847.
13 D. A. Rasko and V. Sperandio, *Nat. Rev. Drug Discovery*, 2010, **9**, 117–128.
14 K. I. P. J. Hidari, T. Murata, K. Yoshida, Y. Takahashi, Y.-h. Minamijima, Y. Miwa, S. Adachi, M. Ogata, T. Usui, Y. Suzuki and T. Suzuki, *Glycobiology*, 2008, **18**, 779–788.
15 Y. C. Lee, R. R. Townsend, M. R. Hardy, J. Lonngren, J. Arnarp, M. Haraldsson and H. Lonn, *J. Biol. Chem.*, 1983, **258**, 199–202.
16 J. J. Lundquist and E. J. Toone, *Chem. Rev.*, 2002, **102**, 555–578.
17 C. R. Becer, M. I. Gibson, R. Ilyas, J. Geng, R. Wallis and D. A. Mitchell, *J. Am. Chem. Soc.*, 2010, **132**, 15130–15132.
18 N. Vinson, Y. Gou, C. R. Becer, D. M. Haddleton and M. I. Gibson, *Polym. Chem.*, 2011, **2**, 107–113.
19 M. W. Jones, S.-J. Richards, D. M. Haddleton and M. I. Gibson, *Polym. Chem.*, 2013, **4**, 717–723.

20 J. E. Gestwicki, C. W. Cairo, L. E. Strong, K. A. Oetjen and L. L. Kiessling, *J. Am. Chem. Soc.*, 2002, **124**, 14922–14933.
21 M. Ambrosi, N. R. Cameron, B. G. Davis and S. Stolnik, *Org. Biomol. Chem.*, 2005, **3**, 1476–1480.
22 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.
23 Z. Shen, M. Huang, C. Xiao, Y. Zhang, X. Zeng and P. G. Wang, *Anal. Chem.*, 2007, **79**, 2312–2319.
24 L. Otten, S.-J. Richards, E. Fullam, G. S. Besra and M. I. Gibson, *J. Mater. Chem. B*, 2013, **1**, 2665–2672.
25 S. G. Spain, M. I. Gibson and N. R. Cameron, *J. Polym. Sci., Part A: Polym. Chem.*, 2007, **45**, 2059–2072.
26 S. R. S. Ting, G. Chen and M. H. Stenzel, *Polym. Chem.*, 2010, **1**, 1392–1412.
27 H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004–2021.
28 M. A. Gauthier, M. I. Gibson and H.-A. Klok, *Angew. Chem., Int. Ed.*, 2009, **48**, 48–58.
29 *World Heath Organisation Fact Sheet #107*, 2012.
30 A. Young and E. Meeusen, *Glycoconjugate J.*, 2002, **19**, 601–606.
31 L. de Witte, A. Nabatov, M. Pion, D. Fluitsma, M. A. W. P. de Jong, T. de Gruijl, V. Piguet, Y. van Kooyk and T. B. H. Geijtenbeek, *Nat. Med.*, 2007, **13**, 367–371.
32 M. Andreini, D. Doknic, I. Sutkeviciute, J. J. Reina, J. Duan, E. Chabrol, M. Thepaut, E. Moroni, F. Doro, L. Belvisi, J. Weiser, J. Rojo, F. Fieschi and A. Bernardi, *Org. Biomol. Chem.*, 2011, **9**, 5778–5786.
33 M. Thépaut, C. Guzzi, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, N. Varga, E. Chabrol, J. Rojo, A. Bernardi, J. Angulo, P. M. Nieto and F. Fieschi, *J. Am. Chem. Soc.*, 2013, **135**, 2518–2529.
34 B. Richichi, A. Imberty, E. Gillon, R. Bosco, I. Sutkeviciute, F. Fieschi and C. Nativi, *Org. Biomol. Chem.*, 2013, **11**, 4086–4094.
35 S. André, D. V. Jarikote, D. Yan, L. Vincenz, G.-N. Wang, H. Kaltner, P. V. Murphy and H.-J. Gabius, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 313–318.
36 M. Fais, R. Karamanska, S. Allman, S. A. Fairhurst, P. Innocenti, A. J. Fairbanks, T. J. Donohoe, B. G. Davis, D. A. Russell and R. A. Field, *Chem. Sci.*, 2011, **2**, 1952–1959.
37 A. Bernardi and P. Cheshev, *Chem.–Eur. J.*, 2008, **14**, 7434–7441.
38 K. C. A. Garber, K. Wangkanont, E. E. Carlson and L. L. Kiessling, *Chem. Commun.*, 2010, **46**, 6747–6749.
39 S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew. Chem., Int. Ed.*, 2012, **51**, 7812–7816.
40 B. D. Polizzotti, R. Maheshwari, J. Vinkenborg and K. L. Kiick, *Macromolecules*, 2007, **40**, 7103–7110.
41 Consortium of Functional Glycomics, http://www.functionalglycomics.org/static/index.shtml
42 W. B. Turnbull, B. L. Precious and S. W. Homans, *J. Am. Chem. Soc.*, 2004, **126**, 1047–1054.
43 H.-A. Tran, P. I. Kitov, E. Paszkiewicz, J. M. Sadowska and D. R. Bundle, *Org. Biomol. Chem.*, 2011, **9**, 3658–3671.
44 N. K. Singha, M. I. Gibson, B. P. Koiry, M. Danial and H.-A. Klok, *Biomacromolecules*, 2011, **12**, 2908–2913.
45 D. A. Sack, S. Huda, P. K. B. Negoi, R. R. Daniel and W. M. Spira, *J. Clin. Microbiol.*, 1980, **11**, 35–40.
46 A. Bernardi, L. Carrettoni, A. G. Ciponte, D. Monti and S. Sonnino, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2197–2200.
47 A. Bernardi, D. Arosio, D. Potenza, I. Sánchez-Medina, S. Mari, F. J. Cañada and J. Jiménez-Barbero, *Chem.–Eur. J.*, 2004, **10**, 4395–4406.

Appendix 3: Discrimination between lectins with similar specificities by ratiometric profiling of binding to glycosylated surfaces; a chemical 'tongue' approach.

RSC Advances

COMMUNICATION

View Article Online
View Journal | View Issue

Discrimination between lectins with similar specificities by ratiometric profiling of binding to glycosylated surfaces; a chemical 'tongue' approach†

L. Otten and M. I. Gibson*

Cite this: *RSC Adv.*, 2015, **5**, 53911

Received 12th May 2015
Accepted 12th June 2015

DOI: 10.1039/c5ra08857g

www.rsc.org/advances

Carbohydrate–lectin interactions dictate a range of signalling and recognition processes in biological systems. The exploitation of these, particularly for diagnostic applications, is complicated by the inherent promiscuity of lectins along with their low affinity for individual glycans which themselves are challenging to access (bio)synthetically. Inspired by how a 'tongue' can discriminate between hundreds of flavours using a minimal set of multiplexed sensors and a training algorithm, here individual lectins are 'profiled' based on their unique binding profile (barcode) to a range of monosaccharides. By comparing the relative binding of a panel of 5 lectins to 3 monosaccharide-coated surfaces, it was possible to generate a training algorithm that enables correct identification of lectins, even those with similar glycan preferences. This is demonstrated to be useful for discrimination between the cholera and ricin toxin lectins showing the potential of this minimalist approach for exploiting glycan complexity.

Introduction

Protein carbohydrate interactions are essential for many biological processes including cell–cell communication, fertilisation and innate immunity.[1] They are also readily exploited by pathogens during adhesion steps. These adhesion steps are mediated by carbohydrate binding proteins known as lectins. In nature, multivalent presentation of glycans at cell surfaces increases the affinity towards its binding partner which has been widely exploited to create synthetic glycomimetics, such as glyco-polymers[2] and particles.[3–5] This strategy however does not necessarily maintain or improve the selectivity complicating the design and application of multivalent glycoconjugates.[2,5–8]

Lectin interactions are mediated by the carbohydrate itself but also the linker between the carbohydrate, the cell surface and precise 3D presentation of carbohydrates on the cell surface.[6,9] Many lectins show highly specific binding to oligosaccharides but show much more promiscuous binding characteristics on a mono- and di-saccharide level. For example, peanut agglutinin (PNA) is generally described as being β-galactose specific but microarray analysis shows that it will readily bind all monosaccharides with very little difference between them.[10] The same is also true for cholera toxin, this toxin is highly specific to the GM-1 ganglioside in the body and thus is described as being galactose specific but this lectin will indiscriminately bind all monosaccharides to one degree or another.[10]

This wide variety of roles played by glycans in the body's innate processes and their prevalence in nature means the interference or detection of these interactions could have an impact in combatting infectious diseases.[11] For example, FimH is a lectin involved in the binding of uropathogenic *Escherichia coli* to mannose rich residues and is a crucial virulence factor. Cholera is caused by cell internalisation of an AB_5 toxin, mediated by the 5 lectin subunits of the toxin initiating binding to GM-1 on epithelial cells in the small intestine. Ricin is a toxic protein extracted from *Ricinus communis* seeds, it consists of one subunit responsible for cleaving an adenine residue from the 28S ribosomal RNA (thus rendering the cell incapable of protein synthesis) and one subunit responsible for binding to galactose rich residues.[12] Differences in glycosylation of cells have also been implicated in tumour cells and determining metastatic potential of cancers[13,14] and the ABO blood system is also determined by different antigenic oligosaccharides.[11,15] Serological blood groups have been implicated in individual susceptibility to many diseases and the severity of others including small pox, cholera and malaria.[15–17] As such rapid detection of lectins can aid in the early identification and prevention of diseases and also in the design of therapeutics. This broad window of binding partners means that the design of a sensor for a lectin based on glycans alone is immensely challenging.

Whilst proteomic and antibody based techniques can be used for identification of lectins these are not always suitable

Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL UK. E-mail: m.i.gibson@warwick.ac.uk

† Electronic supplementary information (ESI) is available: This includes protein preparation, surface functionalisation and LDA analysis. See DOI: 10.1039/c5ra08857g

for robust, point of care applications, and require infrastructure for preparation, storage, distribution and deployment of the sensor. Such a challenge is indeed not unique to glycobiology, and the detection of cell phenotypes, which often have dynamic surface ligand displays which change with their environment. To address this nanoparticles multiplexed biosensing has attracted much interest especially for diagnostics.[18] Rotello *et al.* have developed the use of differentially functionalised gold nanoparticles for multiplexed diagnostics. For example, 52 different mixtures of seven different proteins could be identified using just six distinct nanoparticles.[19] Gold particles coated with 3 different thiols enabled cancerous and healthy cells to be discriminated without the requirement for any specific binding epitopes.[20] Detection of pathogenic bacteria using a related system in under 5 minutes has also been demonstrated[21] as have MRI based detection of cancerous cells with differential lectin expression levels.[22] Jayawardena *et al.* have described the use of glycosylated gold nanoparticles and their characteristic shift in SPR frequencies upon protein binding to characterise lectins based on their response to a panel of sugars.[23] In this case, lectins with very different glycan specificities were used (*e.g.* concanavalin A/soybean agglutinin) and discrimination was also possible without the need for multiplexing and just using individual glycans making it a less challenging analysis.

The goal of the present research was to evaluate the use of simple and synthetically accessible mono-saccharides as multiplexed sensors to enable discrimination between different lectins which have similar binding specificities. Such a system would have widespread application especially for low-cost selective detection/monitoring of toxins.

Results and discussion

The key aim of this work was to probe the differential response of lectins to simple carbohydrates (monosaccharides), so it was essential to employ accessible (/facile) coupling chemistry. 96-multiwell plates with hydrazide functionality were used to couple a range of monosaccharides, and mixtures of different monosaccharides using an aniline catalyst at 50 °C to give glycosylated surfaces (Fig. 1A). This coupling mechanism is known to result in attachment of the monosaccharides predominantly in their ring closed (pyranose) β-anomeric form.[24] It should be noted that the presence of some acyclic species does not affect our later analysis using a training algorithm (*vide infra*). It was not possible to interrogate the functionalised polypropylene surface of the microwell plate using traditional surface analysis techniques (such as elipsometry). As an alternative to contact angle, droplet spread was measured. Hydrophobic surfaces, when viewed from above should give reduced surface coverage at equal volume, compared to a hydrophilic surface with a low contact angle. The native, and glycosylated surfaces were therefore interrogated by addition of a drop of ultra-pure water with resorufin (a dye) added to enable visualisation of the droplet spread, and subsequent image analysis. The native surfaces resulted in only 30% of the surface being covered by the drop, but the galactose functional surface resulted in droplet spreading over 50% of the surface (Fig. 1B). As a positive control, glyceraldehyde was added to generate very hydrophilic surface coating, and this resulted in spreading over ~90% of the surface, confirming this (unconventional) analytical approach. Glycan-modified surfaces should also present an uncharged, non-fouling surface, compared to the native hydrazide/polypropylene surface. Therefore, non-specific fouling (adsorption) was tested using FITC-labelled bovine serum albumin. Compared to the native surface, the glycosylated surfaces showed significantly less binding than the native surface, with no significant absorption observed at concentrations below 0.2 mg mL^{-1} again confirming the surface modification (Fig. 1C).

To highlight the challenges faced in identification and profiling of lectins with similar binding specificities, a panel of 5, fluorescently labelled, galactose (or GalNAc) binding lectins were selected, exposed to a galactose microwell plate, washed and total fluorescence measured. Fig. 2 shows the results of this, indicating that at any given concentration the total response recorded is not unique to any given lectin. Cholera toxin B subunit (CTx) gives higher binding than the others, but the absolute fluorescence intensity is obviously dependent on the concentration applied, which is not ideal for any realistic biosensory format as it requires significant prior knowledge of the solution being probed.

Considering the low information content of these single sugar assays, we proceeded to extract information for a series of Gal-binding lectins from the CFG database (consortium for functional glycomics) to a range of small mono/disaccharides (figure of this analysis included in ESI†). The CFG data revealed that any single glycan cannot predict the identity of the lectins (*i.e.* a single peak is not present) due to their inherent promiscuity. However, if many different glycans are included, there is a unique pattern of binding of each lectin to the carbohydrates (a 'barcode'). Guided by this data, we rationalised

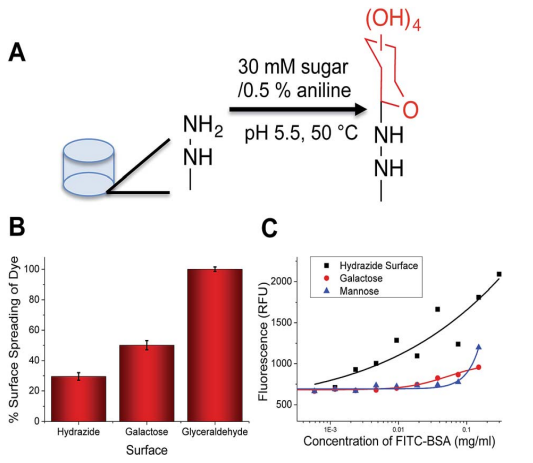


Fig. 1 Fabrication of glycosylated 96-well plates. (A) Hydrazide–carbohydrate coupling; (B) dye spreading assay showing relative hydrophilicity of surfaces; (C) non-specific binding of fluorescently-labelled bovine serum albumin onto different surfaces following 30 minutes incubation and washing.

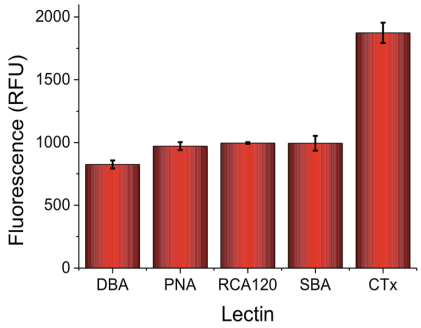


Fig. 2 Relative binding of a panel of 5 lectins to a galactose-functional surface as judged by fluorescence intensity. All lectins applied at 0.01 mg mL^{-1}, with FITC labels.

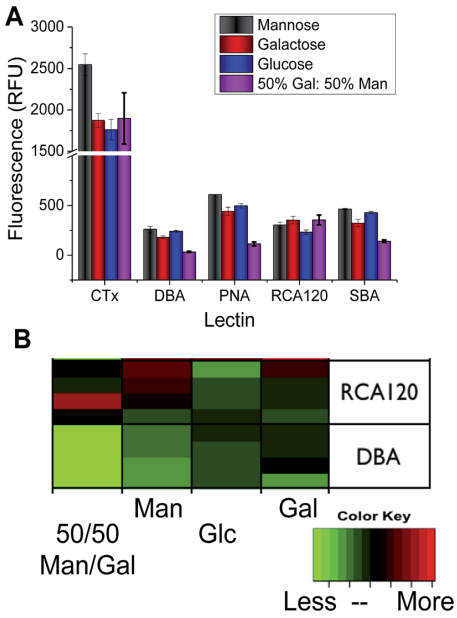


Fig. 3 (A) Relative lectin binding to glycosylated surfaces determined by fluorescence; (B) heatmap, demonstrating that each protein has a 'barcode' of responses to each glycan, each lectin displayed contains at least 4 independent replicates.

that if we could identify the 'minimum basis set' of glycans that can provide a unique barcode for each lectin, it would be possible to distinguish between these, enabling protein identification without proteomics or associated methods. Using the hydrazide coupling chemistry described above, we generated 4 differently glycosylated surfaces; Gal, Man, Glc and a 1 : 1 mixture of Gal : Man (the latter was added as in our hands this improves the resolution of our subsequent analysis. Variable density glycan mixtures are known to give non-linear responses[7]). Pleasingly, these relatively low-affinity monosaccharides produced very unique binding profiles for each lectin, as shown in Fig. 3. For example, *Ricinus communis* Agglutinin (RCA_{120}) had significantly higher binding to galactose, and the Gal/Man mixtures, than compared to Glc binding. Conversely, Soybean Agglutinin (SBA) had significantly depressed binding to the mixed surface. A summary of the relative binding of the lectins can be shown in a heat map to give a 'bar-code' which is unique to each protein.

Analysis of the individual binding of one lectin to a sugar does not give much information, but when combined together, this differential response provides sufficient information to enable a linear discriminant analysis. Linear discriminant analysis is a training algorithm that inputs a matrix of data and produces a model in which all of the categories in the initial training matrix are grouped into distinct categories based on their linear discriminant factors (which are a linear combination of the initial inputs-in this case the surfaces used). Due to the high degree of separation between categories within the model produced it allows for greater confidence in the identification of lectins responsible for binding in unknown samples when compared to the raw data alone.

Fig. 4A shows the results of a linear discriminant analysis of these lectins to the four glycosylated surfaces, revealing highly resolved groupings for each lectin. The circles around each are indicative of a 95% confidence boundary. Fig. 4B shows the LD analysis for the lectins without CTx, as when this is included the other four lectins appear more tightly bunched (but are still perfectly resolved) due the generally increased binding of CTx to all surfaces employed here. This simple, but powerful, multiplexed method enables separation and identification of lectins with similar binding profiles, but without the need for complex carbohydrates, in much the same way as a tongue has evolved to identify complex tastes based on only 5 different inputs. To test the predictive power of this, blind analysis of unknown lectin samples was also conducted, revealing 100% predictive accuracy from this training matrix.

As a final test of this sensing approach, the differentiation between two different gal-binding, pathogenic, lectins was investigated. CTx is the toxin secreted by the bacteria *Vibrio cholera*, which causes cholera and is a huge problem in developing countries and disaster zones. RCA_{120} is a surrogate for ricin, which can be weaponised as a biological warfare agent. A training algorithm was again employed, but this time the RCA_{120}/CTx solutions were applied as mixtures of the two lectins, rather than as pure protein solutions. This provides a far more challenge test, which is closer to a real world sensing

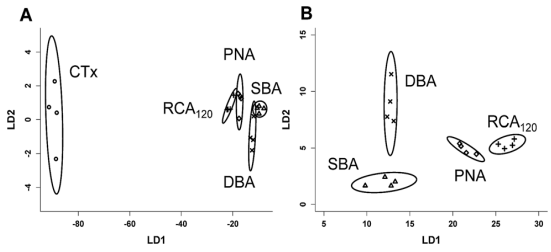


Fig. 4 Linear discriminant analysis of lectin binding to the 4 different glycosylated surfaces. (A) Lectins with CTx, and (B) lectins without CTx.

application. When CTx was present at > 50% (by mass) the LD model correctly indicated its presence, and when the RCA_{120} concentration was above 50%, this was correctly scored (see ESI† for full details and LDA graphs).

Conclusions

Here we have reported the new concept of a 'chemical tongue' for multiplexed biosensing, and discrimination between carbohydrate-binding proteins (lectins). We show that using only simple monosaccharides, which have very low intrinsic affinity and specificity, it is possible to discriminate between a panel of lectins with extremely similar binding preferences. The power of this method lies in the scalability, enabling many more (oligo)saccharides to be employed, and the use of the large glycan databases (which are freely available) to guide the design of each system. Using this approach we demonstrated that the chemical tongue can even distinguish the presence of cholera or ricin, in complex mixtures of the two lectins. Current and future work is focused on establishing the limits and scope of this method, and translating it into realistic sensory surfaces/components which would remove the need for labelled proteins.

Experimental

Full experimental detail including microplate functionalization LDA analysis is provided in the ESI.†

Acknowledgements

Equipment used was supported by the Birmingham Science City (SC) Advanced Materials project, with support from Advantage West Midlands and part funded by the European Regional Development Fund. LO is funded by BBSRC *via* Systems Biology DTC.

Notes and references

1 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357–2364.
2 M. W. Jones, L. Otten, S. J. Richards, R. Lowery, D. J. Phillips, D. M. Haddleton and M. I. Gibson, *Chem. Sci.*, 2014, **5**, 1611–1616.
3 S.-J. Richards and M. I. Gibson, *ACS Macro Lett.*, 2014, **3**, 1004–1008.
4 M. J. Marin, A. Rashid, M. Rejzek, S. A. Fairhurst, S. A. Wharton, S. R. Martin, J. W. McCauley, T. Wileman, R. A. Field and D. A. Russell, *Org. Biomol. Chem.*, 2013, **11**, 7101–7107.
5 L. L. Kiessling, J. E. Gestwicki and L. E. Strong, *Angew. Chem., Int. Ed.*, 2006, **45**, 2348–2368.
6 A. Bernardi, J. Jimenez-Barbero, A. Casnati, C. De Castro, T. Darbre, F. Fieschi, J. Finne, H. Funken, K.-E. Jaeger, M. Lahmann, T. K. Lindhorst, M. Marradi, P. Messner, A. Molinaro, P. V. Murphy, C. Nativi, S. Oscarson, S. Penades, F. Peri, R. J. Pieters, O. Renaudet, J.-L. Reymond, B. Richichi, J. Rojo, F. Sansone, C. Schaffer, W. B. Turnbull, T. Velasco-Torrijos, S. Vidal, S. Vincent, T. Wennekes, H. Zuilhof and A. Imberty, *Chem. Soc. Rev.*, 2013, **42**, 4709–4727.
7 S.-J. Richards, M. W. Jones, M. Hunaban, D. M. Haddleton and M. I. Gibson, *Angew. Chem., Int. Ed.*, 2012, **51**, 7812–7816.
8 J. L. Jimenez Blanco, C. Ortiz Mellet and J. M. Garcia Fernandez, *Chem. Soc. Rev.*, 2013, **42**, 4518–4531.
9 H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis, *Science*, 2001, **294**, 2163–2166.
10 http://www.functionalglycomics.org.
11 A. Audfray, A. Varrot and A. Imberty, *C. R. Chim.*, 2013, **16**, 482–490.
12 J. M. Lord, L. M. Roberts and J. D. Robertus, *FASEB J.*, 1994, **8**, 201–208.
13 X. Zeng, C. S. Andrade, M. L. Oliveira and X.-L. Sun, *Anal. Bioanal. Chem.*, 2012, **402**, 3161–3176.
14 Y. Kim and A. Varki, *Glycoconjugate J.*, 1997, **14**, 569–576.
15 E. Hosoi, *Journal of Medical Investigation*, 2008, **55**, 174–182.
16 D. J. Anstee, *The relationship between blood groups and disease*, 2010, pp. 4635–4643.
17 J. A. Rowe, D. H. Opi and T. N. Williams, *Curr. Opin. Hematol.*, 2009, **16**, 480–487.
18 O. R. Miranda, B. Creran and V. M. Rotello, *Curr. Opin. Chem. Biol.*, 2010, **14**, 728–736.
19 C.-C. You, O. R. Miranda, B. Gider, P. S. Ghosh, I.-B. Kim, B. Erdogan, S. A. Krovi, U. H. F. Bunz and V. M. Rotello, *Nat. Nanotechnol.*, 2007, **2**, 318–323.
20 A. Bajaj, O. R. Miranda, I.-B. Kim, R. L. Phillips, D. J. Jerry, U. H. F. Bunz and V. M. Rotello, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10912–10916.
21 X. Li, H. Kong, R. Mout, K. Saha, D. F. Moyano, S. M. Robinson, S. Rana, X. Zhang, M. A. Riley and V. M. Rotello, *ACS Nano*, 2014, **8**, 12014–12019.
22 K. El-Boubbou, D. C. Zhu, C. Vasileiou, B. Borhan, D. Prosperi, W. Li and X. Huang, *J. Am. Chem. Soc.*, 2010, **132**, 4490–4499.
23 H. S. N. Jayawardena, X. Wang and M. Yan, *Anal. Chem.*, 2013, **85**, 10277–10281.
24 K. Godula and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2010, **132**, 9963–9965.

Appendix 4: Permission for use of figures

Permission for Figure 1.5:

WOLTERS KLUWER HEALTH, INC. LICENSE TERMS AND CONDITIONS

Sep 19, 2015

This Agreement between Lucienne Otten ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	3711300787084
License date	Sep 17, 2015
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Circulation Research
Licensed Content Title	The Endothelial Glycocalyx Protects Against Myocardial Edema
Licensed Content Author	Bernard M. van den Berg, Hans Vink, Jos A.E. Spaan
Licensed Content Date	Apr 4, 2003
Licensed Content Volume Number	92
Licensed Content Issue Number	6
Type of Use	Dissertation/Thesis
Requestor type	Individual
Portion	Figures/table/illustration
Number of figures/tables/illustrations	1
Figures/tables/illustrations used	1
Author of this Wolters Kluwer article	No
Title of your thesis / dissertation	Pathogen Detection Based on Carbohydrate Adhesion
Expected completion date	Sep 2015
Estimated size(pages)	220
Requestor Location	Lucienne Otten Systems Biology DTC Senate House University of Warwick Coventry, United Kingdom CV4 7AL Attn: Lucienne Otten
Billing Type	Invoice
Billing Address	Lucienne Otten Systems Biology DTC Senate House University of Warwick Coventry, United Kingdom CV4 7AL Attn: Lucienne Otten
Total	0.00 GBP
Terms and Conditions	

Permission for Figure 6.2:

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Sep 19, 2015

This is a License Agreement between Lucienne Otten ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3711370278101
License date	Sep 17, 2015
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature
Licensed content title	The pathogenic basis of malaria
Licensed content author	Louis H. Miller, Dror I. Baruch, Kevin Marsh and Ogobara K. Doumbo
Licensed content date	Feb 7, 2002
Volume number	415
Issue number	6872
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	2
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Pathogen Detection Based on Carbohydrate Adhesion
Expected completion date	Sep 2015
Estimated size (number of pages)	220
Total	0.00 GBP
Terms and Conditions	

Appendix 5: Curriculum vitae

Lucienne Otten

Address 58 Rushmoor Drive, Coventry, CV5 8NJ
Mail L.C.Otten@warwick.ac.uk

EDUCATION

PhD Systems Biology 2012-Present
University of Warwick

My research is focussed on the development of carbohydrate microarrays using unmodified sugars for the interogation of whole organisms in order to generate a carbohydrate binding "bar-code" in order to detect pathogenicity based on adhesion patterns.

MSc Systems Biology 2011-2012
University of Warwick

This is a taught Masters which integrates mathematics, modelling and programming with biology to try and asnwer large scale biological questions.

BSc (Hons) Animal Science 2008-2011
University of Nottingham

This degree covered many biological areas but focussed mainly on the physiology of animals and various biochemical and statistical techniques.

RESEARCH EXPERIENCE

Postgraduate research: 2012-Present My PhD research lies in the combination of carbohydrate functionalised surfaces with a powerful machine learning algorithm for the detection of pathogens in the form of toxic lectins, bacterial species and parasites.

Undergraduate research:
MSc Miniprojects: 2012
Theoretical: Measuring complexity in EEG in response to speech stimuli.
Experimental: Label-free microarrays to probe host-pathogen interactions.
BSc project: 2011
Measuring the affects of various factors on rumination and methane emissions from dairy cows through statistical analysis of large data sets.

PUBLICATIONS

1. **L. Otten**, S-J. Richards, E. Fullam, G.S. Besra, M.I. Gibson *Journal of Materials Chemistry B***2013**, 1, 2665-2672. "Gold nanoparticle-linked analysis of carbohydrate-protein interactions, and polymeric inhibitors, using unlabelled proteins; easy measurements using a 'simple' digital camera."
2. M.W. Jones, **L. Otten**, S-J. Richards, R. Lowery, D.J. Phillips, D.M. Haddleton, M.I. Gibson *Chemical Science* **2014**, 5, 1611-1616. "Glycopolymers with Secondary Binding Motifs Mimic Glycan Branching and Display Bacterial Lectin Selectivity in Addition to Affinity."
3. **L. Otten** and M.I. Gibson *RSC Advances* **2015**, 5, 53911-53914 "Discrmination between lectins with similar specificities by ratiometric profiling of binding to glycosylated surfaces; a chemical 'tongue' approach"
4. **L. Otten**, E. Fullam, M.I. Gibson *in preparation*

5. **L. Otten**, S-J. Richards, D. Vlachou, M.I. Gibson *in preparation*
6. S-J. Richards, **L. Otten**, M.I. Gibson *in preparation*

TEACHING EXPERIENCE AND OUTREACH

Advanced Skills Tutor 2012-Present
The Brilliant Club, part-time

As an advanced skills tutor I designed a course based around my research which I then teach teach as university style tutorials to Key Stage 4 and 5 students in low participation schools to encourage them to go on to top Universities.

Skills: *e-tutoring, tutoring, mentoring, science communication, pedagogy*

Level 3 Award in Education and Training 2014-present
City and Guilds

I voluntarily completed this award with St John Ambulance in order to become a qualified trainer. As part of the course I was required to learn about aspects of teaching and learning theory before demonstrating my skills in a 30 minute microteach on a topic of my choice.

Skills: *teaching, learning, communication, pedagogy*

Trainer and Assessor 2014-Present
St John Ambulance, Voluntary

I train and assess volunteers for St John Ambulance to the level of Advanced First Aider and whilst volunteering as an Emergency Medical Technician I regularly mentor less experienced members in high-pressure environments at events such as the Virgin London Marathon.

Skills: *leadership, teamwork, communication, mentoring*

Outreach 2012-Present
University of Warwick, part-time

Demonstration and supervision of experiments for school age children in the Department of Chemistry

Skills: *tutoring, mentoring, science communicaiton*

GRANTS AWARDED

Fully funded PhD and MSc from the BBSRC (2011- Present)
Travel grant from the Chemistry Biology Interface Division of the RSC (2014).
Travel grant from IBCarb (2015).

SELECTED CONFERENCE CONTRIBUTIONS

Warwick Systems Biology DTC annual conference *February 2015* (oral presentation)
RSC Carbohydrate Group Autumn Meeting *September 2014* (oral presentation)
Warwick Systems Biology DTC Seminar *May 2013* (oral presentation)

REFERENCES

Available upon request