Universidade de Lisboa

Faculdade de Medicina de Lisboa

CHEMICAL TOOLS FOR THE STUDY OF N-GLYCOSYLATION IN PROTOZOAN PARASITES

Jessica Grace Morgan Bevan

Orientador: Doutor João Alexandre Guarita da Silva Rodrigues

Co-orientadores: Doutora Maria Rita Mendes Bordalo Ventura Centeno Lima

Doutor Thomas Hanscheid

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas Especialidade de Microbiologia e Parasitologia

2018

Universidade de Lisboa

Faculdade de Medicina de Lisboa

CHEMICAL TOOLS FOR THE STUDY OF N-GLYCOSYLATION IN PROTOZOAN PARASITES

Jessica Grace Morgan Bevan

Orientador: Doutor João Alexandre Guarita da Silva Rodrigues

Co-orientadores: Doutora Maria Rita Mendes Bordalo Ventura Centeno Lima

Doutor Thomas Hanscheid

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas
Especialidade de Microbiologia e Parasitologia

Júri:

Presidente: Doutor José Augusto Gamito Melo Cristino, Professor Catedrático e Presidente do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa.

Vogais:

- Doutor Fernando Siñeriz de Paz, Especialista de Reconhecido Mérito e Competência, *Research Development Manager* da OTR3 Society, Paris, França;
- Doutor Henrique Manuel Condinho da Silveira, Professor Catedrático do Instituto de Higiene e Medicina Tropical da Universidade Nova de Lisboa;
- Doutora Maria Angelina de Sá Palma, Investigadora Auxiliar da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa;
- Doutora Maria Rita Mendes Bordalo Ventura Centeno Lima, Investigadora Auxiliar do Instituto de Tecnologia Química e Biológica da Universidade Nova de Lisboa; (Co-Orientadora)
- Doutora Amélia Pilar Grases dos Santos Silva Rauter, Professora Catedrática da Faculdade de Ciências da Universidade de Lisboa;
- Doutora Alexandra Maria Moita Antunes, Investigadora Principal, Professora Auxiliar Convidada do Instituto Superior Técnico da Universidade de Lisboa.

Part of the Marie Curie ITN funded by the EU Commission (GA. 608295)

2018

1 Table of Contents

2 Acknowledgements

Firstly, I would like to express my gratitude to Dr. João A. Rodrigues for giving me the opportunity be a part of the GlycoPar ITN, for mentoring me and for enabling me to attend conferences and undertake secondments all over the world. I would also like to thank Dr Rita Ventura for accepting to be my co-supervisor, for supporting me during my time in her lab and for her scientific guidance.

I am grateful to my thesis committee; Dr. Miguel Prudêncio, Dra. Amélia Rauter and Dr. Claudio Franco for their scientific input and their time.

Thanks are also due to Dr Eva Lourenço for her constant encouragement, advice and support and for all the 6.30 am runs, gym classes and train journeys. To Dr Miguel Ferreira and Dr Eleonora Aquilini, for their support and enthusiasm and for their help in "translation of biology" to assist in my transition into Parasitology. To Professor Christopher Maycock, for the insightful conversations and sharing of his knowledge. I would like to thank the CERMAX NMR facility for their training and support in performing HR-MAS NMR.

I would like to acknowledge the contribution of Professor Mike Ferguson and Samuel Duncan for their help in designing and executing the protocols for the affinity probe enrichment of galactosyltransferase in *T. brucei* and for giving me the opportunity to work in their lab. I would also like to acknowledge Professor Markus Aebi for giving me the opportunity to work in his lab and Dr. Chia-wei Lin and Nathalie Cornillie for their invaluable insight, providing mass spectrometry analysis of *P. falciparum* glycoproteins and for teaching me how to interpret the data.

I would like to thank all past and present members of the Bioorganic Chemistry lab, especially Vanessa Miranda and Filipa Almeida for the friendly working environment and for making me laugh through the tough times. To all past and present members of Figueirêncios Lab, in particular Daniel, Fabien, Henrique, Idálio, Leonor, Mariana, Pena and Tiago for the chocolate croissant breaks and board game nights. To the members of Gonçalo Bernardes Lab, thank

you for all the good moments we have shared, especially Padma, thank you for the curry nights, gym buddy times and friendship.

To my parents, siblings and grandparents, thank you for supporting me throughout this journey, for joining me in exploring Portugal and your constant belief in me. Kandy, thank you for your patience, your encouragement, the countless Friday and Sunday evenings you spent travelling and for filling my fridge with sri racha sauce.

3 Abbreviations

Abbreviation	Full name
AFGP	Antifreeze glycoproteins
BSA	Bovine Serum Albumin
Bn	Benzyl
CID	Collision-induced dissociation
CSA	Camphorsulfonic acid
DAST	Diethylaminosulfur trifluoride
DCM	Dichloromethane
DIC	Diisopropylcarbodiimide
DIPEA	*N,N*-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DPAGT1	Dolichyl-phosphate N-acetylglucosaminephosphotransferase
DTT	Dithiothreitol
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ETD	Electron-transfer dissociation
ER	Endoplasmic Reticulum
FACE	Filter Aided Capture and Elution
FASP	Filter aided sample preparation
Gal	Galactose
GDP	Guanine Diphosphat
Glc	Glucose
GPI	Glycosylphosphatidylinositol
GSL II	Griffonia (Bandeiraea) Simplicifolia Lectin II
GALE	UDP-glucose 4-epimerase
GalT	Galactosyltransferase
GT	Glycosyltranferase
HCD	Higher-energy collisional dissociation
HOBT	Hydroxybenzotriazole
HRMAS NMR	High Resolution Magic Angle Spinning Nuclear Magnetic Resonance
NEt_3	Triethylamine
NHS	N-hydroxysuccinimide
NIS	N-Iodosuccinimide
NMR	Nuclear Magnetic Resonance
NSAF	Normalised spectral abundance factor
OST	Oligosaccharide transferase
PBS	Phosphate buffered saline
PEG	Polyethelene glycol
PNGase F	Peptide:N-Glycosidase F
Pyr.	Pyridine

RBC	Red blood cell
SDS	Sodium dodecyl sulfate
SMCC	Succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate
T. brucei	*Trypanosoma brucei*
TBAB	Tetrabutylammonium bromide
TBAI	Tetrabutylammonium iodide
TBDPS	*tert*-Butyldiphenylsilyl
TCEP	Tris(2-carboxyethyl)phosphine
TLCK	1-chloro-3-tosylamido-7-amino-2-heptone
TIC	Total ion current
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
TFA	Trifluoroacetic acid
Trityl	Triphenylmethyl
VSG	Varient surface glycoprotein
UEA	Ulex europaeus agglutini
UMP	Uridine monophosphate
UDP	Uridine diphosphate
VSG	Variant Surface Glycoprotein
WGA	Wheat Germ Agglutinin

4 Table of figures and schemes

5 Resumo

A glicosilação é a modificação pós-traducional de proteínas mais abundante em eucariotas e pode estar envolvida numa multiplicidade de funções. No entanto, as ferramentas disponíveis para estudar tipos raros ou incomuns de glicosilação, são limitadas. A N-glicosilação consiste num tipo de glicosilação onde o glicano é sempre adicionado a um resíduo de asparagina da proteína. Este tipo de glicosilação inicia-se no retículo endoplasmático onde um glicano precursor é sintetizado por glicosiltransferases, sendo posteriormente antes transferido para a proteína alvo pelo complexo enzimático da *oligosacariltransferase (OST)*. A glicoproteína é então transferida para o Golgi onde o glicano é ainda modificado por glicosidases ou glicosiltransferases. O conjunto de glicosiltransferases presentes em cada organismo sintetiza uma vasta diversidade de glicanos, muitos dos quais são específicos de cada organismo. Desta forma as vias de síntese e funções dos glicoconjugados fornecem informações importantes sobre como certos organismos patogénicos proliferam e causam doenças. Neste contexto, novas ferramentas para estudar a síntese dos glicanos e as enzimas processadoras de glicanos poderão abrir caminho para novas terapêuticas. Nesta tese são descritos três métodos químicos para explorar a N-glicosilação nos parasitas *Trypanosoma brucei* e *Plasmodium falciparum*.

O primeiro objetivo do meu trabalho de doutoramento foi o desenvolvimento de um novo método para enriquecimento seletivo de UDP-galactosil transferases, isto é glicosiltransferases que catalizam a transferência de uma molécula de galactose da UDP-Galactose (UDP-Gal) para a molécula substrato. Embora a UDP-agarose sirva para enriquecer seletivamente para glicosiltransferases a partir de misturas complexas, esta não é específica para UDP-galactosil transferases. As enzimas desta família de enzimas são responsáveis por catalizar a formação de ligações em muitos glicanos exclusivos de *T. brucei*. Desta forma, e devido à sua ausência em seres humanos, representam potencias alvos de fármacos combater a tripanossomíase. Devido à sua natureza incomum, nenhum gene homólogo deste galactosiltransferases presentes noutros eucariotas foi identificado até hoje em *T. brucei*. A primeira parte do meu projeto de doutoramento teve por objetivo sintetizar uma sonda baseada em atividade enzimática para isolar seletivamente as UDP-galactosil transferases de

T. brucei. Esta sonda baseou-se na funcionalização com análogos do nucleótido de açúcar dador desta classe de enzimas, a UDP-GalGal. Partindo de uma via de de síntese recentemente desenvolvida, a UDP-Gal e UDP-(4F)-Gal foram conjugados a uma resina de TentaGel™. Tanto quanto conhecemos, estes são os primeiros nucleótidos de açúcar acoplados a uma resina. Após um processo de otimização na qual foram testadas várias vias de síntese, a resina conjugada com a UDP-Gal foi sintetizada em nove passos a partir da metil-α-D-Gal com rendimentos de 71% até quantitativo. A resina conjugada com UDP-(4F)-Gal foi sintetizada em treze passos a partir da glicose com rendimentos 48% até quantitativo.

A metodologia de enriquecimento para galactosiltransferases utilizando a resina sintetizada foi inicialmente testada com uma galactosiltranferase comercialmente disponível, em conjunto com outras proteínas não específicas, para comprovar a sua seletividade. Uma vez comprovada a seletividade, realizou-se um ensaio para isolar galactosiltransferases a partir de lisados de *T. brucei*. A resina com UDP-GalGal não permitiu o enriquecimento em galactosiltransferases a partir dos lisados, provavelmente devido a uma baixa afinidade da UDP-Gal e à complexidade das proteínas presentes na mistura. A resina com UDP-(4F)-Gal revelou maior afinidade, permitindo enriquecer para uma galactosiltransferase recentemente descoberta em *T. brucei*: a TbGT3. Apesar destes resultados sugerirem uma boa seletividade da resina UDP-(4F)-Gal, estes ensaios beneficiariam de experiências adicionais e etapas de otimização para aumentar a capacidade de enriquecimento e permitir a identificação de novas UDP-galactosiltransferases do parasita.

A segunda parte do meu trabalho doutoral consistiu em testar o uso de pequenas moléculas inibidoras de N-glicosilação, nomeadamente de NGI-1, para estudar os efeitos da deficiência em N-glicosilação em *T. brucei*. A NGI-1 está descrita como inibidora da STT3B, a subunidade catalítica da OST, em células de mamíferos. Caso esta molécula tenha efeito em *T. brucei*, prevê-se que iniba apenas a transferência de glicanos com alto teor de manose, uma vez que estes são adicionados às glicoproteínas exclusivamente pela STT3B. A NGI-1 poderia então ser usada como uma ferramenta para a redução da glicosilação sem a necessidade de recorrer a técnicas de RNA de interferência para o gene de *STT3B* em *T. brucei*. De fato, o tratamento com NGI-1 revelou-se tóxico para *T. brucei*. No entanto, a droga apresentou uma concentração inibitória média (CI_{50}) muito alta de 75,73 µM comparativamente com as drogas

normalmente usadas para tratar tripanossomíase. No entanto foi possível observar uma redução na glicosilação com concentrações sub-tóxicas de NGI-1. A deteção de glicoproteínas por imunoblot com lectinas, sugere um efeito do fármaco na N-glicosilação, com um aumento em glicanos complexos e uma diminuição em glicanos com alto teor de manose. De forma a investigar com maior detalhe os efeitos de NGI-1 na glicosilação, foi analisadapor espectrometria de massa a glicoproteína variável de superfície (VSG) de *T. brucei*, por forma a determinar operfil exato de N-glicosilação após o tratamento com NGI-1.

O terceiro objetivo deste trabalho consistiu na identificação de glicoproteínas N-glicosiladas no parasita que causa a malária, o *Plasmodium falciparum*. Com este objetivo procurou-se enriquecer para glicoproteínas N-glicosiladas de *P. falciparum*, usando vários métodos baseados em separação por lectinas e análise por espectrometria de massa. Inicialmente, as glicoproteínas foram enriquecidas a partir de lisados de *P. falciparum* usando o protocolo FASP FACE e as lectinas GSL II e WGA. Embora os resultados da espectrometria de massa tenham indicado a presença de glicoproteínas com GlcNAc ou $GlcNAc_2$, o método utilizado não foi suficientemente preciso para determinar a sua natureza. Um analisador de massa Orbitrap foi então utilizado, o que melhorou a precisão e confirmou a presença dos glicopéptidos. Infelizmente, os péptidos encontrados não possuíam uma resolução suficientemente alta para serem identificados. Os iões para $HexNAc_2Fuc$ foram observados indicando a presença de proteínas modificadas por fucosilação, mas a sequência de tais proteínas não pôde ser identificada. Na tentativa de melhorar o passo de enriquecimento, foi testado um protocolo de separação com as lectinas GSL II e WGA acopladas a microesferas magnéticas. Houve, contudo dificuldades em conjugar estas lectina com as microesferas. Testou-se outra abordagem em que se submeteram as glicoproteínas a de *P. falciparum* a modificação química por adição de uma molécula de galactose a GlcNac com uma galactosiltransferase. Desta forma, as proteínas N-glicosiladas poderiam ser enriquecidas com a lectina ricina, que reconhece moléculas de galactose. Por análise em espectrometria de massa concluiu-se que este método de enriquecimento não foi eficaz e, portanto, métodos alternativos devem ser investigados no futuro.

Em suma, ao longo do meu trabalho doutoral foram explorados novos métodos para estudar a N-glicosilação nos parasitas *T. brucei* e *P. falciparum.* Desenvolvi uma via de síntese que permite pela primeira vez a conjugação de uma resina a um nucleótido de açucar. Em termos funcionais, a seletividade da resina funcionalizada com UDP-(4F)-Gal ficou comprovada pelo enriquecimento de uma galactosiltransferase *bona fide* recentemente caracterizada em *T. brucei*, a TbGT3. Para além disso, a nova via de síntese que desenvolvi tem a vantagem de poder ser utilizada no futuro para produzir outros nucleótidos de açúcar ligados à resina de TentaGel™.

Em *P. falciparum* foram testados diferentes métodos de separação por lectinas e identificação por espectrometria de massa, com o objetivo de estudar as glicoproteínas N-glicosiladas. Foram encontrados péptidos glicosilados com GlcNAc ou $GlcNAc_2$, embora a identificação das proteínas modificadas tenha sido impossibilitada pela falta de resolução dos péptidos produzidos.

Em suma, este trabalho permitiu o desenvolvimento de novos métodos para a descoberta de enzimas modificadoras de glicanos e de proteínas N-glicosiladas, em dois parasitas com vias de glicosilação bastante divergentes dos outros eucariotas. Estes métodos requerem ensaios futuros de otimização mas, considerando que estes parasitas possuem vias de glicosilação não-canónicas, as abordagens experimentais seguidas neste trabalho serão uma forma eficaz de descobrir e caracterizar a N-glicosilação das suas proteínas.

6 Summary

Glycosylation is the most abundant post translational modification in eukaryotic cells and can be accredited to an enormous variety of functions. Yet the tools used to study rare or unusual glycosylations are limited. N-Glycosylation, glycosylation of an asparagine residue, begins in the endoplasmic reticulum where a glycan precursor is synthesised by glycosyltransferases before it is transferred to a protein by an oligosaccharide transferase. The protein is then transferred to the Golgi where the glycan is further modified by glycosidases or glycosyltransferases. During this process, the array of glycosyltransferases present in each organism serves to produce a range of glycans, many of which are organism specific. Their formation and functions provide valuable information as to how certain parasites proliferate and cause disease. Therefore, new tools to study glycans and glycan processing enzymes could pave the way towards new therapeutics. Within this thesis, three chemical methods to explore glycosylation in *Trypanosoma brucei* and *Plasmodium falciparum* are described.

Although UDP-agarose serves to enrich glycosyltransferases from complex mixtures, it is not specific towards UDP-galactosyltransferases. This particular family of enzymes forms linkages in many unique glycans of *T. brucei* and represent interesting drug targets due to their absence in humans. Due to their unusual nature, no homologues have been identified. The majority of this PhD project was aimed at synthesising an activity based probe to selectively enrich galactosyltransferases from *T. brucei* using analogues of the donor sugar nucleotide, UDP-galactose. In a newly developed synthetic route, UDP-galactose and UDP-(4F)-galactose were attached to tentagel resin. To our knowledge, these are the first resin bound sugar-nucleotides. After initial method development with a commercially available galactosyltranferase along with other proteins, the resins were proven to bind with the desired selectivity. They were then used in an assay to enrich galactosyltransferases from *T. brucei* lysates. UDP-Galactose was unable to enrich galactosyltransferases from this complicated mixture, most likely due to a low affinity and the complexity of the proteins it was submitted to. UDP-(4F)-Galactose showed a higher affinity but was only able to enrich one galactosyltransferase: TbGT3. The assay was only performed once, therefore with repeated experiments this result may improve.

The second tool described in this thesis is the use of the small molecule inhibiter of N-glycosylation, NGI-1. From published data, this small molecule was predicted to only inhibit the transfer of high mannose glycans in *T. brucei*. In doing so, the effect of reduced glycosylation could be studied without the need for RNAi knock down of STT3B. NGI-1 was toxic to *T. brucei* but had a very high IC_{50} of 75.73 μM. Therefore, at lower concentrations the effect of the drug could be observed. By lectin blots, there appeared to be an effect of the drug on N-glycosylation with an increase in complex glycans and a decrease in high mannose. However, the results were not clear so mass spectrometry analysis of *T. brucei's* variant surface glycoprotein were sought to determine the exact N-glycosylation profile.

The third tool described is to enrich glycoproteins from *Plasmodium falciparum* lysates using various lectin based methods and mass spectrometry analysis. Initially, glycoproteins were enriched using the FASP FACE protocol and lectins GLS II and WGA. Although the mass spectrometry results indicated the presence of glycoproteins, the method used was not accurate enough to determine their nature. An Orbitrap mass analyser was then used, which improved the accuracy so that the presence of glycopeptides was confirmed. Unfortunately, these peptides were not of high enough resolution to be identified. Magnetic bead bound GSL II and WGA enrichment was tested, but there were difficulties in conjugating the lectin to the beads. Chemically modifying the glycoproteins with a galactoslytransferase so that they could be enriched with ricin (instead of GSL II and WGA) was also tested. Mass spectrometry showed that the enrichment was not successful and alternate methods must be investigated.

These new methods to study N-glycosylation in *T. brucei* and *P. falciparum* require some optimisation. However, since both parasites synthesise unique (and in the case of *P. falciparum,* disputed) N-glycans, tools such as the ones described will be the most effective way to profile their N-glycosylation.

7 Introduction

7.1 Glycosylation and the importance of glycosyltransferases

7.1.1 Glycans and the way they shape our world

A glycan, or polysaccharide, is a compound consisting of a number of linked monosaccharides. Nature synthesises an enormous variety of glycans; from the large polysaccharide, starch, containing up to 10,000 α-glucose monomers, to the simple disaccharide sucrose. The most abundant compound on earth is cellulose (polymerised β-glucose); a glycan crucial for the strength and rigidity of plant cell walls (O'Sullivan, 1997). Moreover, many cells are coated by glycosylated proteins, which can mediate cell adhesion. This is common in the case of cancer cells that express altered glycans on their surface. A number of cancerous cells express proteins glycosylated with sialyl Lewis^a and sialyl Lewis^x glycans which adhere to endothelial cell walls *via* E-selectin (Kannagi *et al.*, 2004). Anaplastic large cell lymphoma express sialic acid enabling them to adhere to galectin 8, without which growth is inhibited (Kannagi *et al.*, 2004; Suzuki, Abe and Hashimoto, 2015). In some artic fish, antifreeze glycoproteins (AFGPs) circulate in their blood allowing them to survive in sub-zero temperatures. These proteins are N-glycosylated with the disaccharide GlcNAc-Gal which allows them to behave in a non-colligative manner, disrupting ice crystal formation and inhibiting crystal growth. Without the glycans, these proteins lose their antifreeze properties (Harding, Anderberg and Haymet, 2003). Glycans can also act as a shield against the immune system for some pathogens as is the case for influenza and HIV (Baum and Cobb, 2017). Ultimately, their abundance and diversity permit them to play important roles in processes that are essential for life (Easton, 2011; Varki, 2017). Understanding how these glycans are built is important if we are to understand the many processes that they mediate.

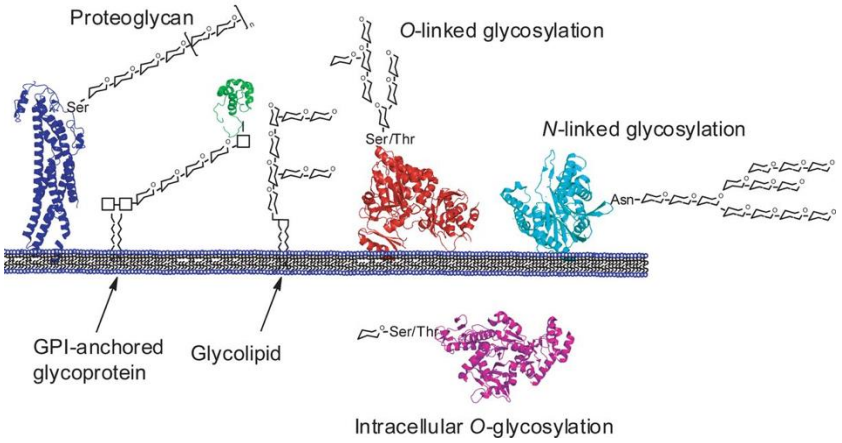


Figure 1 - Types of lipid and protein glycosylation adapted from Gloster, 2012

Glycosylation of proteins and lipids can occur *via* five main glycosylation pathways, figure 1 (Spiro, 2002; Gloster, 2012):

- **O-Glycosylation:** occurs primarily where an *N*-acetylgalactosamine is attached to the hydroxl group of a protein's serine or threonine residue. The simple O-GlcNAc modification is the most abundant posttranslational modification in metazoans and is a transient modification (Yeh and Feeney, 1996). From this *N*-acetylgalactosamine, galactosyltransferases (enzymes that transfer donor monosaccharides to acceptor structures) can build larger structures. In mammals, these structures tend to be small (three to six monosaccharides) (Spiro, 2002; Easton, 2011).
- **N-Glycosylation:** occurs when a pre-made polysaccharide structure is added to an aspargine residue of a protein in the ER when the sequon N-X-S/T is present (but not all N-X-S/Ts are glycosylated), where X is any amino acid except proline (Spiro, 2002). This pre-made polysaccharide can be modified by glycosidases or glycosyltransferases in the Golgi to create a variety of glycan structures. N-glycosylation will be discussed in more detail later in this section and is the focus of this thesis.
- **C-Mannosylation:** present when an α-mannosyl group is bound to a Trp residue *via* a carbon-carbon bond (Spiro, 2002).
- **Phosphoglycosylation:** this is when a glycan is attached to a protein *via* a phosphodiester bond on a serine of threonine residue. Their structures vary widely in terms of size and motif (Haynes, 1998).
- **Glypiation:** this is when a glycosylphosphatidylinositol moiety is attached to the C-terminal of a protein. This modification was first described in *Trypanosoma brucei* and

has a number of different roles in membrane-bound proteins of higher eukaryotes and protozoa (Ferguson *et al.*, 1988; Ferguson, 1991; Ikezawa, 2002).

Glycans can be represented by their chemical structure or by symbolic nomenclature. Although there are many types of nomenclature, the most commonly used is the Consortium for Functional Glycomics (CFG) nomenclature, figure 2 (Varki, 2017). For the remainder of my thesis, I will use this nomenclature to simplify the chemical representation of glycans.

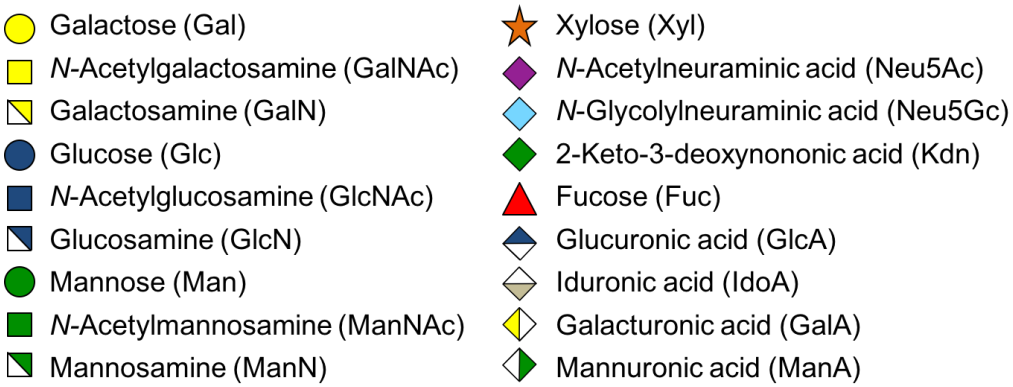


Figure 2 – CFG glycan symbol nomenclature adapted from Essentials of Glycobiology (Varki *et al.*, 2009)

Below is an example diagram of the chemical representation of a glycan and its equivalent structure in glycan symbol nomenclature (figure 3).

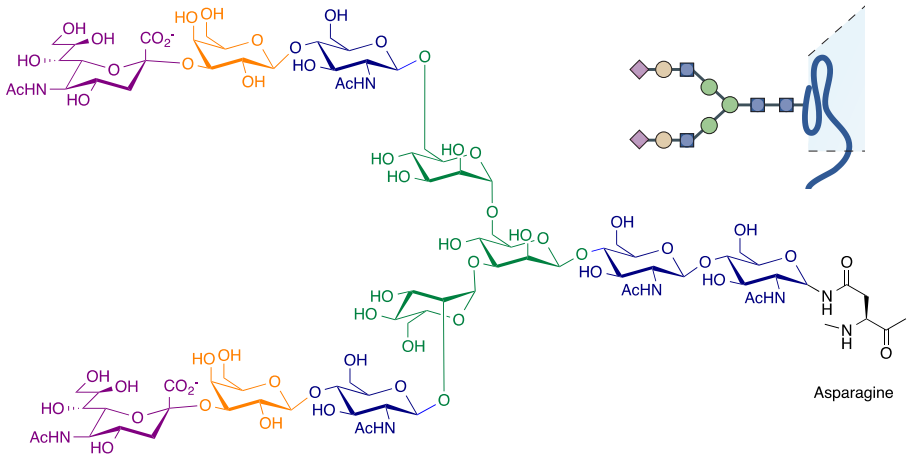


Figure 3 – Diagram displaying a typical glycan in chemical formula and glycan symbol nomenclature

7.1.1.1 Formation and diversity of N-Glyans

There are three types of N-glycans; high mannose (oligomannose), complex and hybrid and these are represented below, figure 4.

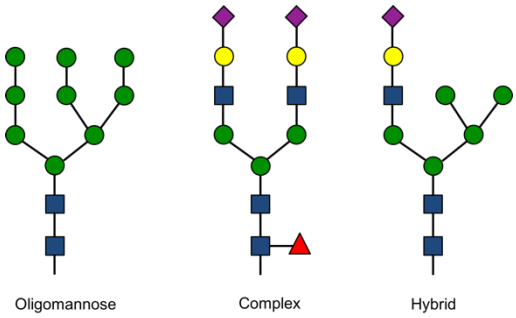


Figure 4 – Types of glycans, adapted from Essentials of glycobiology (Varki *et al.*, 2009)

High mannose glycans are characterised by containing between 5 and 9 terminal mannose residues attached to a $GlcNAc_2$ core. When the $GlcNAc_2Man_3$ core is glycosylated with any monosaccharide except mannose a complex glycan is formed. Hybrid glycans contain a mixture of terminal mannose and any other monosaccharide (Varki *et al.*, 2009).

N-glycosylation occurs in the endoplasmic reticulum (ER) where a glycan precursor is added to an asparagine residue of a protein by an oligosaccharide transferase (OST, step 9 in figure 5). A sequon is a sequence of amino acids where glycosylation can occur, most often N-linked (Varki, Cummings and Esko, 2009). The most common sequon for N-glycosylation is N-X-T followed by N-X-S, where X is any amino acid except proline (Aebi, 2013). However, not every protein containing these sequons are glycosylated. Less common sequons include N-X-C and N-X-G (Lowenthal *et al.*, 2016).

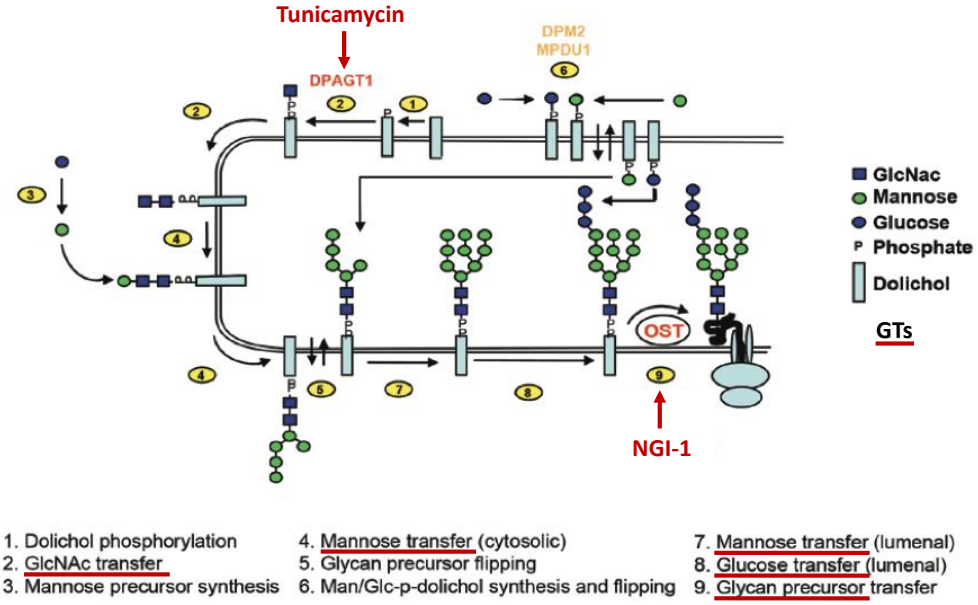


Figure 5 – Typical mechanism of N-glycosylation adapted from Juan *et al.*, 2016.

The glycan precursor is synthesised by the sequential addition of monosaccharides on to phosphorylated dolichol by glycosyltransferases (GTs). When the glycan reaches $Man_5GlcNAc_2$, a flipase translocates it from the cytosol to the lumen of the ER. Here it is further glycosylated by other GTs. Only when the glycan reaches $Glc_3Man_9GlcNAc_2$ may it be transferred to a protein by an OST (figure 5) (Varki, Cummings and Esko, 2009). From here, the correctly folded proteins move to the Golgi and misfolded folded proteins are retained (an important quality control step). The Golgi is an organelle found in most eukaryotic cells and is named after the nobel prize winning Italian scientist Camillo Golgi who identified it in 1897 (Fabene and Bentivoglio, 1998). It is comprised of two networks – the *cis* and the *trans*. Proteins from the ER first enter the *cis* network where glycosidases can remove mannose residues from or glycosyltransferases add GlcNAC to N-glycans and phosphorylation can occur. In the *trans* network, GTs can add galactose or sialic acid residues forming complex or hybrid glycans (Alberts *et al.*, 2002). This generalised mechanism can differ slightly depending on the organism.

7.1.2 Small molecule probes of N-glycosylation

The function of protein N-glycosylation has been studied mainly with pathway inhibitors. The structures of small molecule inhibitors are shown in figure 6 and their uses are described below.

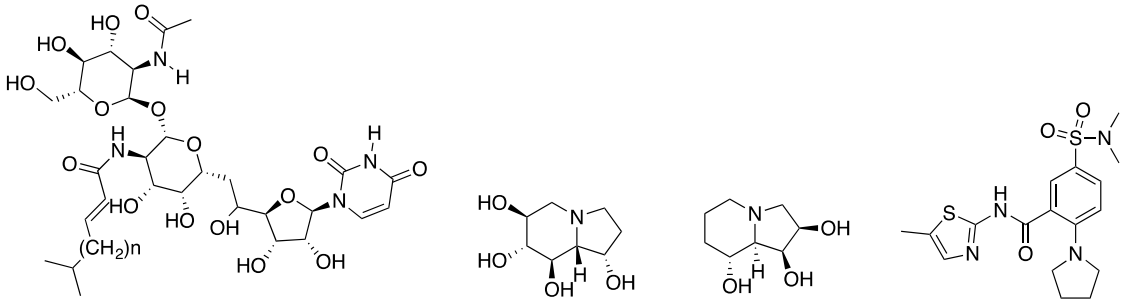


Figure 6 - Small molecule inhibitors of glycosylation

Tunicamycin is a well-known inhibitor of glycosylation that inhibits the transfer of the first GlcNAc onto phosphorylated dolichol, mediated by DPAGT1 (figure 5). Drug resistant ovarian and colorectal tumour cells were found to be more susceptible to chemotherapy when treated with tunicamycin. Studies showed that the inhibition of glycosylation disrupted the localisation and function of transporter proteins increasing their sensitivity (Wojtowicz *et al.*, 2015).

The small molecule castanospermine, found in black beans, is an inhibitor of α-glucosidases that are essential for the processing of N-glycans. The structural proteins prM and E of the dengue virus are N-glycosylated with complex glycans that require α-glucosidases for their formation. *In vitro* and *in vivo* studies of dengue with castanospermine showed that inhibiting the virus's α-glucosidases lead to the misfolding of these proteins which in turn inhibited the infection of the virus (Whitby *et al.*, 2005). There are currently no treatments or vaccines for the dengue virus therefore insights in to the effect of N-glycosylation inhibition could lead to new therapies.

Similarly, swainsonine, found in locoweed, inhibits Golgi α-mannosidase II whose inhibition leads to the production of hybrid glycans over complex. It is toxic and cattle who eat locoweed suffer a range of symptoms from ataxia and emaciation to abortion. Swainsonine was used to study the importance of glycosylation in pregnancy in cattle. Luteal steroidogenic cells (LSCs, responsible for the production of progesterone) were cultured *in vitro* with swainsonine. This showed that although there was no effect on the viability of LSCs, the change in glycosylation

did affect luteinising hormone (LH) which in turn reduced progesterone production which would cause problems in fertility (Sano *et al.*, 2015).

More recently, the small molecule named N-Linked Glycosylation Inhibitor-1 (NGI-1) was found to reduce the ability of the OST of lung cancer cells in transferring $Glc_3Man_9GlcNAc_2$ to proteins (Lopez-Sambrooks *et al.*, 2016). In mammalian cells, OST is not one single protein, but is a complex of 7 (Ribophorin II, OST48, Ribophorin I, DAD1, Stt3A or Stt3B, OST4 and N33 or 1AP) and can exist in 4 isoforms as the pairs STT3A/ STT3B and N33/1AP are mutually exclusive. Stt3 is the largest of these proteins and is the catalytic unit that transfers the oligosaccharide to the asparagine residue. The functions of the remaining proteins are less described but they are associated with protein translocation, recognition of glycosylation sites and may have oxidoreductase activity (Mohorko, Glockshuber and Aebi, 2011). NGI-1 is a reversible inhibitor of STT3 with a higher affinity for STT3B, stabilising it at high temperatures and inhibiting the growth of cancer cells. STT3A is incompletely blocked so reduced N-glycosylation remains present and the cells are viable (Lopez-Sambrooks *et al.*, 2016). Consequently, this small molecule represents a powerful tool for the study of glycosylation with its reduced toxicity compared to tunicamycin and swainsonine.

A number of monosaccharide analogues have been used to study N-glycosylation either through inhibition of GTs or by utilising the flexibility of GTs to incorporate labelled isoforms allowing protein visualisation, purification or the effects of a sugar modification (Hang *et al.*, 2004; Dube *et al.*, 2006; Breidenbach *et al.*, 2010; Gloster, 2012). Inhibitors of GTs will be described in more detail in section **7.3.1**.

If the GT has a homologue, then techniques such as protein over expression, gene knock down or knock out can help characterisation. For example, Wang *et al.* over-expressed the putative GT PtGT1 in the tobacco plant resulting in a plant that had a higher lignin content and an early flowering phenotype (Stanley and Cummings, 2009). Ramírez *et al.* characterised STT3A from *Trypanosoma brucei* by recombinantly expressing the protein, purifying it and screening its activity with a range of substrate peptides and lipid-linked oligosaccharides (LLOs). In doing so they found certain motifs were favoured and that phosphonate analogues of LLOs were competitive inhibitors (Ramírez *et al.*, 2017). Several mice knock downs or knock outs have

been made for specific GTs allowing for the confirmation of several candidates and describing their functions (Stanley, 2016). The knock down of sialyltransferase, ST6GAL1, in human pluripotent stem cells showed for the first time an association of ST6GAL1 with pluripotency and differentiation of these cells (Wang *et al.*, 2015). Rare diseases in humans caused by defects in glycosylation are classified as human congenital disorders of glycosylation or CDGs. These diseases manifest in a wide range of symptoms and severity from mental illness to muscular dystrophy, illustrating the systemic influence that the GTs have on humans (Leuven, 2002; A. D. Shaw *et al.*, 2014; Nigro and Savarese, 2014).

7.1.3 Visualising and identifying N-glycans

In order to characterise glycans or identify glycoproteins, the first step is to enrich for glycoproteins as they are often present in very low amounts. HILIC columns can be used to separate hydrophilic glycopeptides from the non-glycosylated (Gilar *et al.*, 2011). If a particular glycan motif is sought, a specific lectin can be used. Lectins are proteins that bind to glycans and each lectin has its own substrate specificity. Although they bind weakly to their substrates, they are very specific and the accumulation of these interactions between a surface with carbohydrates and a lectin is strong (Berg, Tymoczko and Stryer, 2002). Legume lectins bind to glycans found on the surface of rhizobial bacteria enabling the plant to move the bacteria to its routes where it can carry out nitrogen fixation, maintaining their symbiotic relationship (Van Rhijn *et al.*, 2001). A well-studied and commonly used lectin is ricinus communis agglutinin I. Extracted from castor beans, this lectin binds to terminal galactose residues (Lord, 1995). Biotinylated lectins can be used to visualise glycoproteins by western blot or can be bound to agarose or magnetic beads for use in pull downs. Alternatively, lectins can be used in conjunction with Filter Aided Capture and Elution (FACE). This method uses the lectin to capture peptides on top of a filter, allowing non-captured peptides to be washed away (Ongay *et al.*, 2012). These captured peptides can be released by blocking the lectin with its natural substrate. They can then be analysed by mass spectrometry with or without their glycan. For glycoprotein determination, if PNGase or Endo-H were used, there will be a detectable weight difference between the glycosylated and nonglycosylated forms of the peptide (Asn vs Asp and glycan vs a single GlcNAc respectively), see below.

To determine the structure of the glycan, they must be cleaved from the peptide. The choice of glycan cleaving method depends greatly on the type of glycan to be cleaved. For example,

Endo-H cleaves high-mannose (and some hybrid) N-glycans in the middle of the $GlcNAc_2$ core (Koide and Muramatsu, 1974). PNGase F can cleave high mannose, complex and hybrid N-glycans at the base of the glycan (between the GlcNAc and asparagine residue) only if they have the $GlcNAc_2Man_3$ core unless α(1,3)-linked core fucose is present (Tarentino, Trimble and Plummer, 1989). PNGase A is the same as PNGase F, except it can also cleave N-glycans with α-(1,3)-linked core fucose present (figure 7).

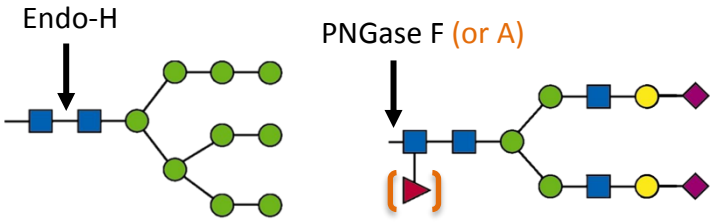


Figure 7 - Diagram to illustrate specificity of glycan cleaving enzymes

On cleavage, the asparagine residue is converted to an aspartic acid (Tarentino, Trimble and Plummer, 1989). Hydrazinolysis is a chemical method of glycan release that is non-specific and will release N and O-glycans (Takasaki, Mizuochi and Kobata, 1982). After they have been released from the peptide, their structures can be determined with monosaccharide analysis or NMR (Takasaki, Mizuochi and Kobata, 1982).

7.1.4 Glycosyltransferases: mechanisms and roles

Nature relies on subtle differences between carbohydrate building blocks to synthesise complex glycan structures. These are built by glycosyltransferases (GTs), which transfer a specific donor monosaccharide (eg. UDP-galactose) on to an acceptor, which is typically another monosaccharide, protein or glycan (Rini, Esko and Varki, 2009).

ER GTs are multi-transmembrane proteins and Golgi GTs are type II membrane proteins. There are 104 GT families classified by short sequence motifs, mostly grouped by their donor specificity (Lombard *et al.*, 2014). All glycosyltransferases contain the DXD motif – a metal ion binding site required for their activity (Rini, Esko and Varki, 2009). The largest database of carbohydrate active enzymes is the CAZy database which, in 2013, contained 119 910 GTs. However, only 1.6% of these had been characterised with biological studies (Lombard *et al.*, 2014). Now there are 346,702 GTs associated to a family and 7,563 that are not classified in the CAZy database (Lombard *et al.*, 2017).

The donor sugar is often activated by a nucleotide diphosphate and the reaction is mediated by a metal cation, for example Mn^{2+} or Mg^{2+}. The mechanism of the reaction can proceed in two ways either inverting or retaining the stereochemistry at the anomeric centre of the donor sugar. The inverted mechanism proceeds by an S_N2 attack at the anomeric centre *via* an oxonium intermediate, figure 8 (Lairson *et al.*, 2008).

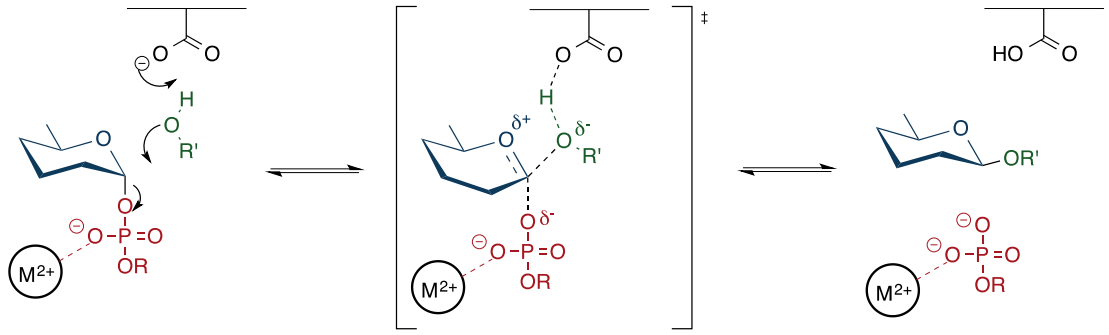


Figure 8 – Mechanism of glycosylation with stereochemistry inversion

The mechanism for the retention of the stereochemistry has been debated. However, the most widely accepted mechanism is the double displacement mechanism (figure 9) (Lairson *et al.*, 2008).

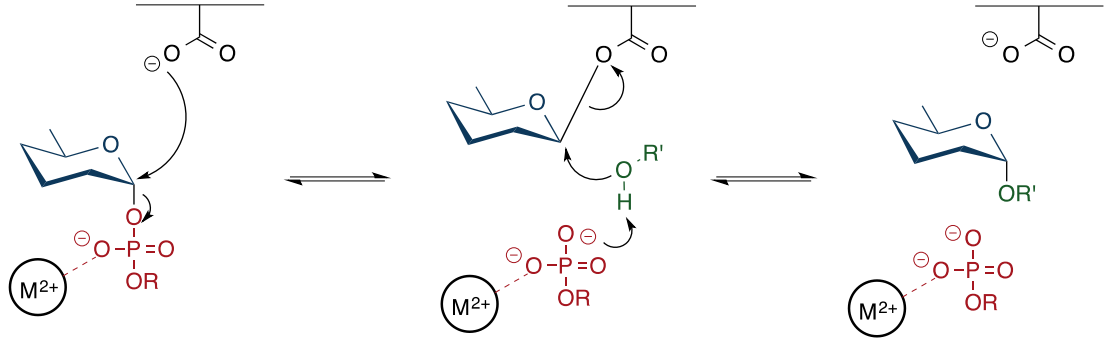


Figure 9 - Mechanism of glycosylation with stereochemistry retention.

In both the retaining and inverting mechanisms, the GT first binds to the donor. The GT can then change its conformation to bind to the acceptor (Taylor and Drickamer, 2011). Most GTs are highly specific for one donor and one acceptor. However, there are some that catalyse the same reactions or can catalyse more than one. For example, in the presence of α-lactalbumin, β-1,4-galactosyltransferase (β-1,4-GalT) catalyses the addition of galactose to glucose producing lactose. In the absence of this protein β-1,4-GalT adds galactose to *N*-acetylglucosamine (Taylor and Drickamer, 2011).

7.1.4.1 ABO blood groups and glycosyltransferases

The makeup of your personal GT repertoire will decide your blood group. The AB0 blood grouping system is based on glycans covering the surface of red blood cells. If you are 0 type, you have the basic glycan coating. If you are A, you have an extra *N*-acetylglucosamine transferase. If you are type B, you have an extra galactosyltransferase. Types AB have both of these extra GTs. These subtle differences in surface glycans impart massive differences for the recognition of red blood cells (RBCs). They are antigenic which explains why there are blood transfusion incompatibilities as reactions to antigenic glycans on the red blood cells that are non-self (Stanley and Cummings, 2009) (Stanley and Cummings, 2009). Furthermore, those with type 0 are more likely to be infected with norovirus than any other blood group and type AB or B are unlikely to develop symptoms if infected. The blood group antigens are highly expressed in the gastro intestinal tract suggesting that this effect could be due to adhesion of the virus to the glycans (Hutson *et al.*, 2002).

There are different GTs in different organisms and even between individuals, complicating their characterisation. Their existence can be predicted by observing the linkages they form using mass spectrometry. This implies that homology based approaches can only contribute to a certain point in the characterisation of predicted GTs. For those without a clear homologue there are no standard techniques.

7.2 Trypanosoma brucei

Trypanosoma brucei (*T. brucei)* is a protozoan parasite endemic to sub-Saharan Africa. It has three sub species: *T. brucei brucei, T. brucei gambiense* and *T. brucei rhodesiense* all of which are transmitted by the tsetse fly. The first (*T. b. brucei*) infects non-human vertebrates and the latter two infect humans; with *gambiense* being the most abundant and *rhodesiense* the most virulent. In cattle and horses, the resulting disease is called "nagana" taken from the Zulu word for powerless. Nagana halts the development of farming, with large economic losses in cattle production of approximately $2.5 billion per year (A. P. M. Shaw *et al.*, 2014). Although *gambiense* and *rhodesiense* do not cause disease in cattle, they are able to infect them. Thus, cattle are a large reservoir of human infecting trypanosomes and are able to promote their transmission (Kennedy, 2013; Sutherland *et al.*, 2015).

In humans, the disease is called African Sleeping Sickness or Human African Trypanosomiasis (HAT). HAT occurs in two stages; stage one infects the bloodstream and lymphatic system. In stage two (CNS stage), the parasite moves to the central nervous system infecting cerebrospinal fluid and neural tissue. Stage one presents as weight loss, chills, fatigue, headache, and lymphadopathy. These symptoms are often mistaken for malaria. Stage 2 causes severe headaches, nocturnal insomnia, hallucinations, anxiety, progressive mental deterioration and motor system disturbances. Eventually the host falls into a coma and dies (Kennedy, 2013).

Only two new treatments for HAT have been developed since the 1950s putting the total available drugs at 5. There are two strains of parasite and two stages to the disease. Each drug is only effective against one form of the disease leaving little to no choice in treatment, table 5 (Bacchi, 2009).

Stage	***Rhodesiense***	***Gambiense***
Early	Suramin	Pentamidine
Late	Melarsoprol	Eflornithine/Nifurtimox-Eflornithine

Table 1 – Drug specificity dependant on species and stage of HAT (Bacchi, 2009)

Current treatments must be administered intravenously and have extreme side effects including severe diarrhoea, organ damage and pulmonary oedema (Bacchi, 2009). Pentamidine can cause diabetes and Melarsoprol kills 5% of its patients. The combinatorial drug Nifurtimox-Eflornithine was developed with reduced side effects, although gastrointestinal and neuropsychiatric problems still occur (Franco *et al.*, 2012). All of the current treatments for HAT have ill-defined modes of action. However, we know that Eflornithine inhibits ornithine decarboxylase, Suramin targets glycolytic enzymes and Pentamidine disrupts the structure of kinetoplastic DNA (Fairlamb, 2003; Bacchi, 2009; Kennedy, 2013). *In vitro,* trypanosomes have an intrinsic circadian rhythm affecting 10% of their genes that makes them more susceptible to Suramin at certain times of day (Rijo-Ferreira *et al.*, 2017). In summary, the development of a treatment designed specifically for HAT, with reduced side effects and an easier method of administration is highly sought after.

7.3 Glycosylation in *T. brucei*

Trypanosoma *brucei* life cycle involves two proliferative stages: the procyclic form (pcf) in the tsetse fly midgut and the bloodstream slender form (bsf) in the blood of a mammal (Kennedy, 2013). In between these proliferative stages, the metacyclic and stumpy forms are produced for transmission between the two hosts, figure 10. Recently, a new form of the parasite has been described that occupies the adipose tissue of the mammalian host (Trindade *et al.*, 2016).

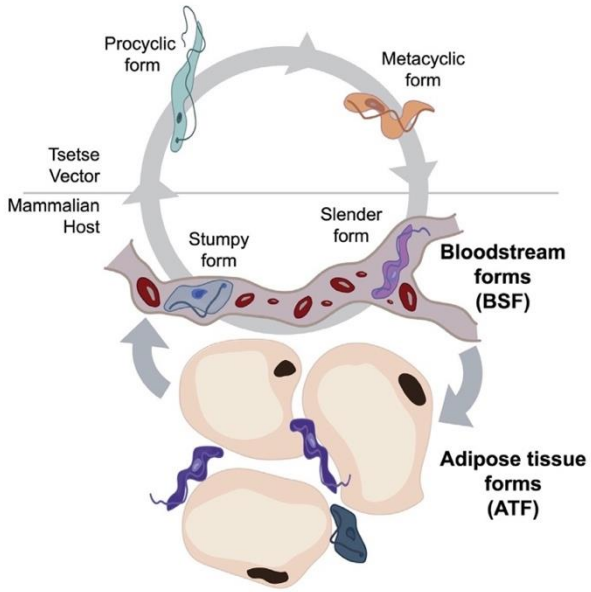


Figure 10 –Lifecycle of *T. brucei* taken from Trindade, 2016

There are a number of important glycosylated proteins and lipids in both the pcf and bsf. In the pcf, the parasite is coated in a layer of glycoproteins called procyclins. In the bsf, its surface membrane is coated in Variant Surface Glycoproteins (VSGs) (Pays *et al.*, 1989). Both procyclin and VSG (along with other proteins) are anchored to the surface membrane *via* GPI anchors – glycolipids that can be anchored to the C terminus of a protein. The first complete GPI structure was elucidated in 1988 from the VSG of *T. brucei.* Each bsf trypanosome has approximately $1x10^7$ GPI-linked VSGs on its surface and the biosynthesis of GPI is essential to bsf, but not to pcf, potentially due to its importance in the assembly of VSG (Ferguson *et al.*, 1988; Rini, Esko and Varki, 2009).

7.3.1 Surface glycoproteins and glycans of T. brucei

VSG elicits an immune response from its host, which destroys the parasite using anti-VSG antibodies. Initially, trypanosomes are cleared by IgMs then as the infection becomes established B-cells play a larger role (Stijlemans *et al.*, 2016). The parasite can change its VSG coating sequentially, a process known as antigenic variation. Trypanosomes that have switched to an alternate VSG survive and are able to proliferate until the next antibody response is triggered. This leads to peaks and troughs in the numbers of trypanosomes present in the host. This change in the VSG coating enables the parasite to evade the immune system and establish an infection, figure 11 (Marcello and Barry, 2007; Matthews *et al.*, 2015).

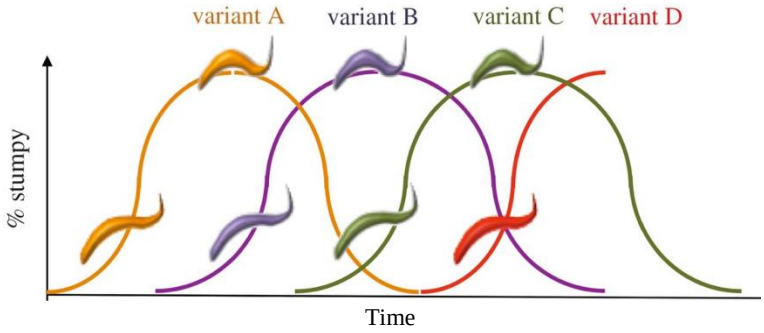


Figure 11 - Antigenic variation of *T. brucei* adapted from Matthews, 2015

The VSG coat is tightly packed, (covering 90% of the cell surface) acting as a barrier to immune factors released by the host. It has also been hypothesised that glycans on the VSG add to this barrier, filling gaps and aiding protein folding (Mehlert *et al.*, 1998). Each VSG has between 0 and 5 N-glycosylation sites linked to high mannose or complex glycans, figure 12 (Schwede *et al.*, 2011). For example, VSG MITat.1.5 has three glycosylation sites and its glycosylation was shown to be required so that the VSG can be transported to the cell surface (Ferguson *et al.*,

1986; Mehlert, Bond and Ferguson, 2002). Computer modelling of this VSG is shown in figure 12 (Mehlert, Bond and Ferguson, 2002) which demonstrates the packing and size of the N-linked glycans.

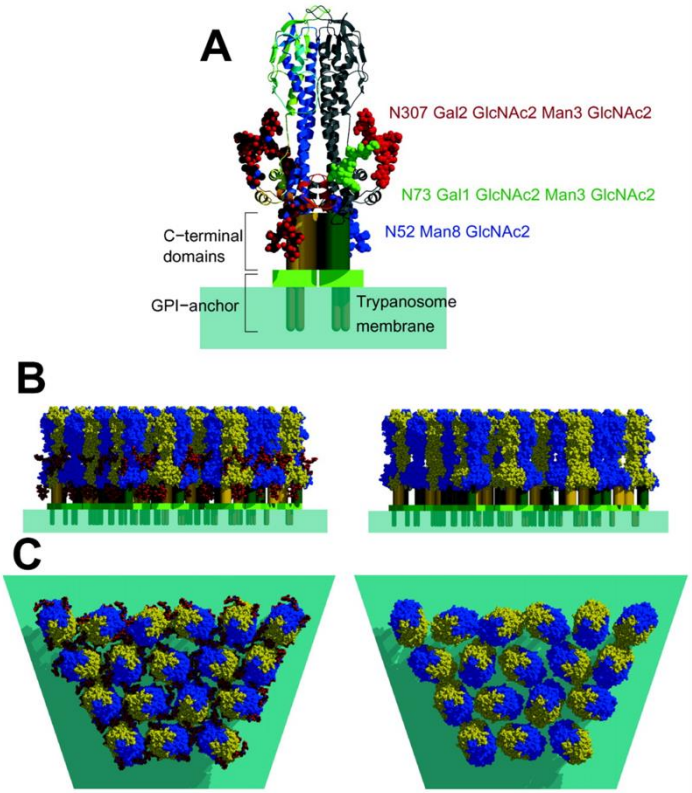


Figure 12 - Computer modelling of VSG MITat1.5 (A) VSG MITat.1.5 with attached, energy-minimised, N-linked oligosaccharides and (B and C) the cell surface coat, viewed from the side and above without (right) and with (left) attached N-linked oligosaccharides from Mehlert, 2002.

Castillo-Acosta *et al.* showed that carbohydrate binding agents are trypanocidal, inducing mutation after prolonged exposure. The mutants displayed different glycosylation patterns on their VSG and were less virulent (Castillo-Acosta *et al.*, 2013). After exposure to the lectin HHA (which binds to high mannose glycans) the parasites stopped expressing the oligosaccharide transferase TbSTT3B: an enzyme essential to the addition of high mannose glycans to proteins (Izquierdo, Schulz, *et al.*, 2009). These parasites were less infective implying that the high mannose glycans are important for virulence.

Once uptaken by the tsetse fly, VSG is replaced by procyclin – an antigenically conserved protein. There are three types of procyclin: two types of "EP" and a "GPEET". The EP forms contain between 22 and 30 Glu-Pro repeats and GPEET contains six repeats of Gly-Pro-Glu-Glu-Thr followed by three repeats of Glu-Pro. Although some forms of EP can be glycosylated,

GPEET cannot (Acosta-Serrano, 1999). Furthermore, EP only has one N-glycosylation site and is only glycosylated with one glycan: $Man_5GlcNac_2$ (Mehlert *et al.*, 1998), figure 13. The GPI anchor of procyclin differs greatly from that of VSG as it contains a long polyLacNAc side chain, (Pearson and Republic, 1989; Acosta-Serrano, 1999) figure 13.

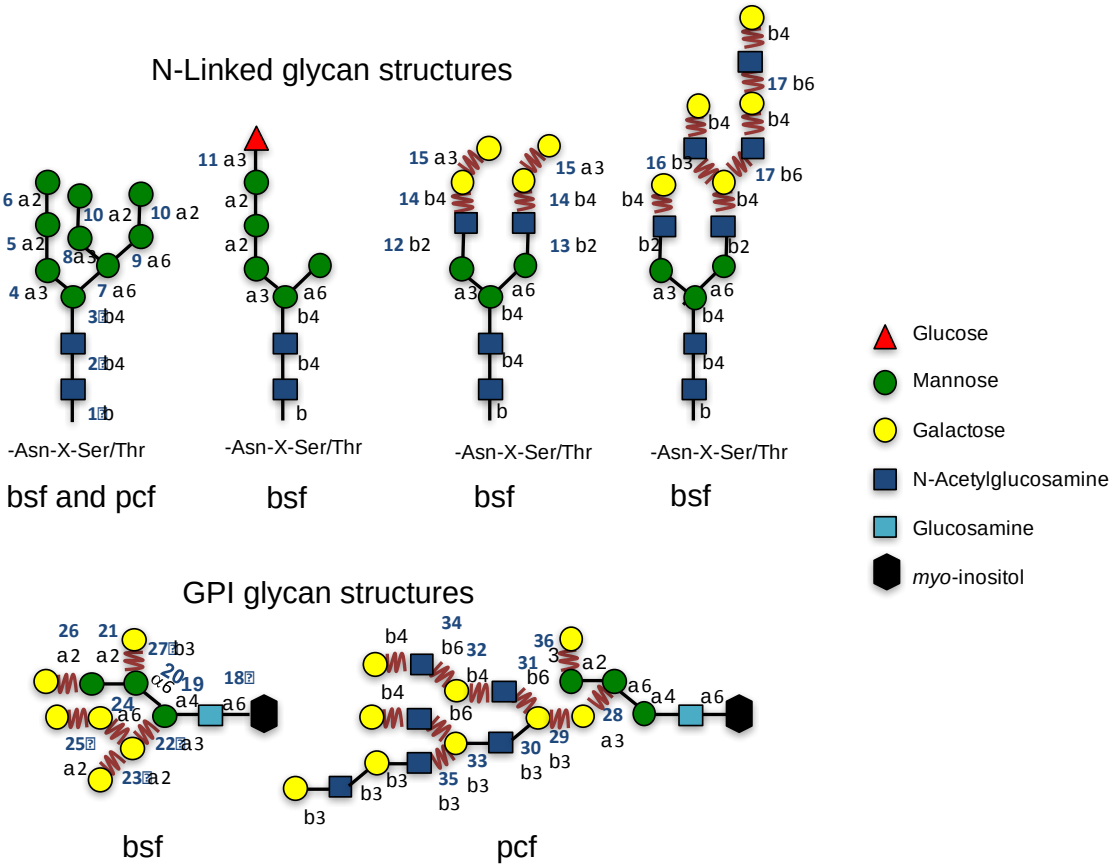


Figure 13 - Structures of VSG, procyclin and GPI anchors adapted from Izquierdo, 2009.

These glycans are built sequentially by a number of glycosyltransferases (GTs). There are at least 38 distinct glycosidic linkages in *T. brucei* glycans (as highlighted in figure 13), suggesting that there are at least 38 GTs present – probably many more. A number of these have been identified by homology, but many are unique to the parasite. No homologue has been found in *T. brucei* for the GT responsible for the linkages that are highlighted with a wavy red line in figure 13. The majority of these highlighted linkages require a galactosyltransferase and the donor UDP-Gal.

The flagellar pocket of *T. brucei* contains large, compact poly-*N*-acetyllactosamine (PolyLacNAc) glycans, with unclear roles in the parasite, figure 13. These glycans are composed of Man, Gal and GlcNAc residues and have an unusually high number of repeats compared to other known PolyLacNAcs (on average 54 repeats per glycan – twice the size of the largest

reported mammalian PolyLacNAc) (Atrih *et al.*, 2005). The flagellar pocket is the only site of endocytosis and exocytosis, therefore, due to the essentiality of the flagellar pocket to the parasite and the unusual nature of the glycans, it is proposed that their synthesis is also essential.

7.3.2 Sugar metabolism of T. brucei

Nearly all sugar metabolism occurs in the glycosome; an organelle in which glycolytic processes and purine salvage occurs (Colasante *et al.*, 2006; Haanstra *et al.*, 2008). Sugar metabolism is an important part of the glycosylation machinery. Monosaccharides must either be synthesised *de novo* or up-taken by salvage pathways and converted to nucleotide sugars. Glucose and fucose are energy stores but they are also required to synthesise other monosaccharides (Freeze and Elbein, 2009). In *T. brucei* there have been many studies surrounding its sugar metabolism, discussed below, figure 14.

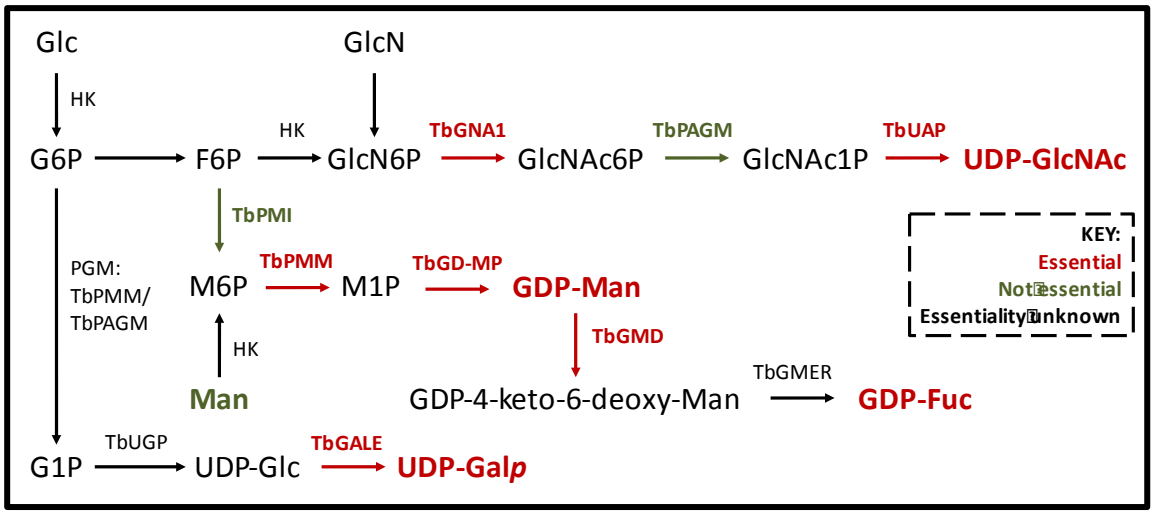


Figure 14 - Sugar metabolism of *T. brucei* adapted from Turnock and Ferguson, 2007

UDP-Gal is synthesised by the parasite from UDP-Glc using UDP-Glc-4'-epimerase (TbGALE) and is not taken up from the environment. Down regulation of TbGALE lead to cell death in both procyclic and blood stream forms, indicating that UDP-Gal is essential to *T. brucei.* Curiously, TbGALE is unable to epimerase UDP-GlcNAc to/from UDP-GalNAc unlike its human equivalent (Roper *et al.*, 2002, 2005; Urbaniak, Turnock and Ferguson, 2006).

To date, no fucosylated glycans have been identified in *T. brucei.* However, it expresses two enzymes in the GDP-Fuc biosynthesis pathway: GMD and GMER, figure 14. Procyclic TbGMD null mutants have almost no GDP-Fuc in their intracellular pools, grow slower and their flagella detached before they died. Blood stream form null mutants also grew slower and eventually

lost their flagella and died (Turnock and Ferguson, 2007; Turnock, Izquierdo and Ferguson, 2007). These mutants showed that TbGMD is essential and therefore that GDP-Fucose is also. Although knock down of *TbGMER* was not performed, it is most likely also essential.

Knockouts of *TbPMI* were only able to grow if the medium was supplemented with mannose. Equally, *T. brucei* is able to grow without mannose in the medium so long as it has TbPMI, therefore the parasite is able to use *de novo* or salvage pathways to access GDP-mannose (Kuettel *et al.*, 2012). RNAi induced knock downs in blood stream forms *in vitro* showed that TbGD-MP is also essential (Denton, Fyffe and Smith, 2010). These results show that GDP-mannose itself is essential – either because GDP-mannose itself is or because it is needed to access GDP-Fucose.

The gene *TbUAP* encodes for the enzyme that converts GlcNAc-1-P to UDP-GlcNAc and it is essential in blood stream forms *in vitro* and *in vivo.* In mammals UAP also converts GalNAc- 1-P to UDP-GalNAc, but, like TbGALE, TbUAP does not process GalNAc derived sugars (Stokes *et al.*, 2008). *TbGNA1* conditional null mutants showed that this gene is also essential. Addition of GlcN to the medium of these parasites does not rescue them suggesting that only the *de novo* pathway to UDP-GlcNAc exists in *T. brucei* (Mariño *et al.*, 2011). These results are not surprising since GlcNAc is present in all GPI anchors and N-glycans.

T. brucei phosphomutases are restricted to just TbPAGM and TbPMM who catalyse the transfer of phosphate from the C6 position to the C1 of GlcNAc and mannose respectively. The phosphomutase that would catalyse this transfer in glucose (PGM) is not present in *T. brucei;* instead, PAGM and PMM perform the role of PGM. RNAi induced knock downs of these two showed that *TbPMM* is essential but was not able to show that TbPMM is (Bandini *et al.*, 2012).

7.3.3 *T. brucei* and its large repertoire of N-glycans

The generalised mechanism of N-glycosylation was described in **7.1.1.2**, however there are some differences in *T. brucei.* The parasite has three paralogous genes for the OST: TbSTT3a, TbSTT3b and TbSTT3c. The single subunit TbSTT3a and TbSTT3b have distinct roles, but in blood stream and procyclic forms STT3c is not transcribed. STT3a transfers $Man_5GlcNAc_2$ to acceptor proteins and leads to complex glycans. STT3b transfers $Man_9GlcNAc_2$ to acceptor

proteins which leads to high mannose glycans. The substrate specificity of STT3a and STT3b is correlated with the amino acid sequence prior to the N-glycosylation sequon. STT3a recognises more acidic sequences containing aspartic acid or glutamic acid residues whereas STT3b recognises basic or neutral residues. Although the parasite can survive *in vitro* without STT3a or STT3b, it cannot survive without both and *in vivo* both are required (Izquierdo, Schulz, *et al.*, 2009).

TbALG3 is the α–mannosyltransferase that converts $Man_5GlcNAc_2$-PP-Dol to $Man_6GlcNAc_2$-PP-Dol. Null mutants of this gene clearly show the specificity of STT3a and b. The glycans that would have become $Man_9GlcNAc_2$, and thus transferred by STT3b, are stuck at $Man_5GlcNAc_2$ and are instead processed by STT3a. In the mutants, the glycans that should be high mannose are now complex in structure (Manthri *et al.*, 2008).

The gene *rft1* encodes for the protein that flips Man5GlcNAc2-PP-dolichol (M5-DLO) from the cytosol to the lumen of the ER (figure 5 in 7.1.1.1). In humans, mutations of this gene lead to glycosylation defects (Haeuptle *et al.*, 2008). Imhof *et al.* showed that knocking out *rft1* in procyclic forms of trypanosoma affects its social motility. This can clearly be seen in figure 15 where the knock out remains at the site of inoculation and the wild type was able to spread. They also showed that the knock out was unable to establish infection in the fly mid gut. This all points to the importance of glycosylation in *T. brucei* (Imhof *et al.*, 2015).

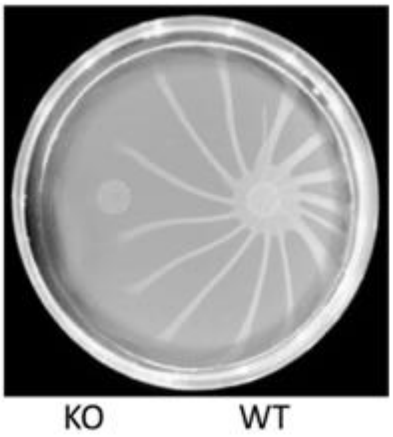


Figure 15 – Comparison of mobility in wild type (WT) and knock out (KO), (Imhof *et al.*, 2015)

Inhibition of N-glycosylation in *T. brucei* by tunicamycin *in vivo* appears to be lethal (Casero, Porter and Bernacki, 1982) although it is also toxic to the mammalian host.

7.3.4 *T. brucei* glycosyltransferases: the known and the as yet unknown

A comprehensive list of predicted GTs based on known linkages, generated by (Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J Barton, *et al.*, 2009), is available in an annex along with homologues, if they have been assigned, figure 68 and 69. From this list of linkages and associated transferases, there are 7 retaining and 5 inverting predicted galactosyltransferases whose character is unknown.

From BLAST searches of human β-1,3-*N*-acetylglucosaminyltransferase there are 21 putative UDP-Gal or UDP-GlcNAc transferases. From this list, what is now known as *TbGT8* was chosen and null mutants produced. These mutants grew normally *in vitro* and were able to infect mice. The study proved that *TbGT8* encodes for a UDP-GlcNAc: β-Gal-GPI β-1,3 GlcNAc transferase. This is involved in the synthesis of GPI-linked glycans in the procyclic form and N-linked glycans in the blood stream form of *T. brucei* (Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J Barton, *et al.*, 2009; Nakanishi *et al.*, 2014). More examples of characterized homologous GTs are below"

- *TbGT3,* the only galactosyltransferase characterised in *T. brucei,* encodes for a UDP-Gal: β-GlcNAc-GPI β-1,3-Gal transferase, expressed in both bsf and pcf of the parasite. Izquierdo *et al.* produced *TbGT3* conditional null mutants that are infectious to tsetse flies and whose growth is not affected in either bloodstream or procyclic form. In the procyclic form they observed that procyclin had a lower molecular weight because the GPI anchors were altered (Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J. Barton, *et al.*, 2009). The remaining uncharacterised galactosyltransferases are involved in the synthesis of glycans on GPI anchors, bsf VSGs and the PolyLacNAc structure.
- *TbGT11* codes for UDP-GlcNAc:α3-D-mannoside β-1,2-*N*-acetylglucosaminyl transferase. Damerow *et al.* produced bloodstream-form null mutants that grew normally *in vitro* and no change in infecting rodents was seen despite producing significantly different protein N-glycans (Damerow *et al.*, 2014).
- *TbGT15* encodes for a β-1,2-GlcNAc transferase that transfers a GlcNAc to the 6-arm of Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc. This GT does not seem to be essential and its deletion did not affect infectivity although mutants grew slower *in vitro* (Damerow *et al.*, 2016).

In conclusion, glycans play a large role in the survival of *T. brucei* in both the tsetse and mammalian host. Their structures are well characterised, but a number of the glycosyltransferases responsible for their synthesis are not known and cannot be found by homology. The majority of unique linkages are built by galactosyltransferases and are present in both life stages. We also know that UDP-Gal uptake is essential. Therefore, we used our probes that are designed to target galactosyltransferases to capture and enrich candidate GalTs, in order to characterise and better understand the mechanisms behind the production of galactose containing glycans.

7.4 Plasmodium falciparum

Plasmodium parasites are apicomplexan and cause malaria. In 2015, a staggering 212 million people worldwide were infected, of which 400,000 died. Although there are 5 species of *Plasmodium* that are able to infect humans, *Plasmodium falciparum* is the most virulent and is prevalent in sub-Saharan Africa (WHO, 2016). Malaria first presents as a fever, headache and chills before progressing to severe anaemia, respiratory distress and occasionally cerebral malaria. There are many anti-malarial drugs available but if left untreated it is fatal. Although mortality rates have dropped by 29% since 2010, malaria remains an enormous public health problem particularly in Africa where 90% of cases and 92% of deaths occurred in 2015 (WHO, 2016).

Like *Trypanosoma brucei, P. falciparum* has a complex lifecycle, figure 16. It is transmitted to humans *via* the bite of an Anopheles mosquito and in each host it changes to different life stages to adapt to the new environment. The bite of the mosquito transmits the parasite in its sporozoite form. After entering the blood of the human host, sporozoites infect liver cells replicating and maturing into merozoites. This liver stage of the disease is asymptomatic making early diagnosis difficult. Merozoites leave the liver and invade red blood cells where they undergo continuous rounds of replication as blood stage asexual forms. During this phase, some parasites can differentiate into sexual gametocytes. When a female mosquito takes a blood meal, gametocytes can be transferred to the fly's midgut where fertilisation of gametes occurs and a zygote is formed. This in turn produces an ookinete which migrates through the mosquito gut wall and forms oocysts. The oocysts contain sporozoites which migrate to the salivary glands of the mosquito, ready to be transmitted during the next blood meal. The asexual trophozoites cause the symptoms of malaria as they mature to schizonts, replicating and bursting red blood cells before reinvading and starting a new cycle (Prevention, 2016).

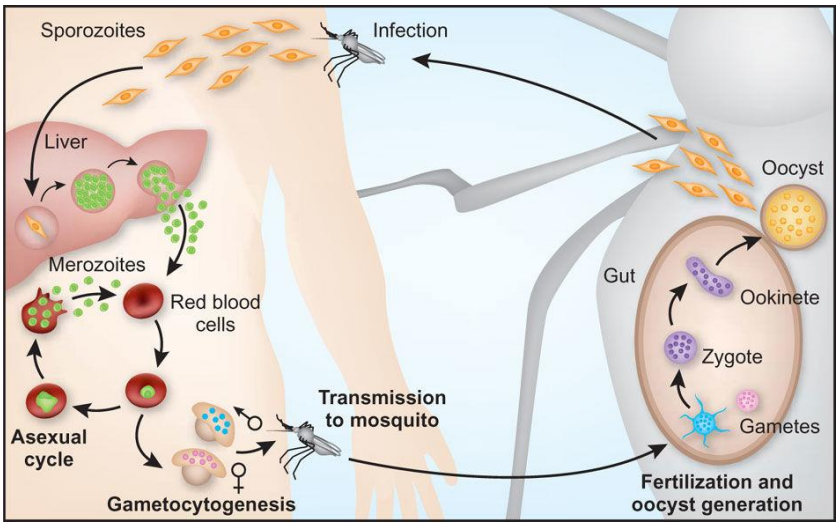


Figure 16 –Life cycle of *P. falciparum* adapted from (Pasvol, 2010)

In the 1950s, the front-line treatment for malaria was chloroquine. This compound was first discovered by Hans Andersag in Bayer's laboratories. However, Hans declared it too toxic for human use and it was not validated until 1947 (Coatney, 1963). Since then, *P. falciparum and P. vivax* have become resistant to chloroquine and nearly all other anti-malarial treatments (Wellems and Plowe, 2001). Artemisinin and its derivatives in combination therapies are used to treat chloroquine resistant malaria. These compounds were discovered by Youyou Tu during the 1960s and 70s. She used her knowledge of traditional Chinese medicine to screen thousands of plants and identified 640 that had anti-malarial activity. From these, the plant qinghao, whose use has been documented since 1596 AD, was the only one to show inhibitory effect on *P. falciparum* but the activity was very low. Youyou then realised that the extraction method they had been using could be destroying the active compound. From here, she used cold extraction techniques and was able to isolate and identify the active compound as artemisinin. The new extraction vastly improved the activity and the new anti-malarial was identified. She was awarded with the 2015 Nobel prize in Physiology or Medicine for her work (Tu, 2011).

Now there are also strains of *P. falciparum* resistant to artemisinin and there is an urgent need to produce new treatments or better yet, vaccines or a new more specific target. A number of successful vaccines for influenza and meningitis are based on glycoproteins, due to the antigenic nature of surface protein glycosylation (Adamo *et al.*, 2013). Creating a vaccine

based on a glycoprotein of *P. falciparum* could be effective, however the glycosylation of this parasite is largely unknown and a controversial topic.

7.5 Glycosylation in *P. falciparum*: To be or not to be?

Whether any type of glycosylation is present in *P. falciparum* has been debated since the early 1980s and the full picture still remains unclear. Initial evidence showed that the parasite only synthesises O-glycans (Dayal-Drager *et al.*, 1991; Dieckmann-Schuppert, Bause and Schwarz, 1993). This shifted to the acceptance that GPI anchors are the predominant glycosylation and that O-glycosylation does not exist (Gowda, Gupta and Davidson, 1997). Both the presence and absence of N-glycosylation has been reported (Dieckmann-Schuppert *et al.*, 1992; Kimura *et al.*, 1996). The uncertainty is predominantly born from the fact that the parasite is intracellular and separation of parasite proteins from host proteins is technically challenging. Metabolic labelling is also difficult as the parasite uses monosaccharides and amino acids from the host red blood cell as well as culture medium (Nirmalan, Sims and Hyde, 2004). There is also always the potential that modifications seen are not from the parasite but from induced effects on the red blood cell (Macedo *et al.*, 2010).

7.5.1 GPI anchors of Pf are its only resolved glycosylation

Glycosylphosphatidylinositol anchors (GPI anchors) are glycolipids shared by many eukaryotes. They comprise of a phosphatidylinositol linked to ethanolamine phosphate by a glycan, which in *P. falciparum* is Man_4GlcN (figure 17). They are conjugated to proteins in the ER and anchor them to the surface of lipid membranes *via* the protein's carboxyl terminal (Varki *et al.*, 2009). GPI anchors remain the only well characterised glycosylation in *P. falciparum*. A number of GPI-anchored proteins were characterised in the 90s, but it was not until 2000 that the definitive structures of *P. falciparum* GPI anchors were published (Naik *et al.*, 2000). There are at least 5 GPI structures as the length of the fatty acid tail can differ, figure 17. Naik e*t al.* also reported the presence of anti-GPI IgGs in people from malaria endemic areas of Kenya. This was significant because it indicated that GPIs are antigenic despite being at the base of a protein. Naik *et al.* went further to propose that, because GPIs are pathogenic, these antibodies can confer protection to malaria pathogenesis (Naik *et al.*, 2000).

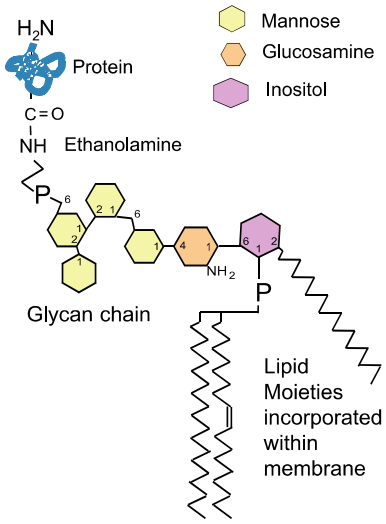


Figure 17 – General GPI anchor of *P. falciparum* adapted from (Boutlis *et al.*, 2005)

7.5.2 P. falciparum sugar metabolism

Sugar nucleotides are important as they are the donor substrates required for building glycans by glycosyltransferases. A summary of *P. falciparum* sugar metabolism is shown in figure 18, (Sanz *et al.*, 2013; Cova *et al.*, 2015). Pools of UDP-Gal, UDP-Glc, UDP-GlcNAc, GDP-Man and GDP-Fuc were found in the blood stage of *P. falciparum*. In the case of UDP-Gal, no candidate genes have been identified for GALE (to convert UDP-Glucose to UDP-Galactose) or Galactose-1-phosphate uridylyltransferase (to convert Gal1P to UDP-Gal) and so far, it has not been identified in any glycans. The authors suggest that PF3D7_0517500 is an orthologue of *L. major* UDP-sugar pyrophosphorylase which is able to produce UDP-galactose from galactose-1-phosphate. This would require a galactokinase to form galactose-1-phosphate from galactose and there is no candidate for this enzyme in the *Plasmodium* genome (Sanz *et al.*, 2013).

P. falciparum does not uptake fucose as it produces it *via* a *de novo* pathway converting GDP-Man with GMD followed by FS (figure 18). Neither enzyme is essential to the parasite and no glycoconjugates containing fucose have so far been identified (Sanz *et al.*, 2013, 2016). However, the fucosyltransferase POFUT2 is present in the genome and potential fucosylation was observed in sporozoite surface proteins CSP and TRAP (Swearingen *et al.*, 2016). POFUT2 deficient sporozoites display a reduced gliding motility but no effect was seen in blood stage forms (Lopaticki *et al.*, 2017).

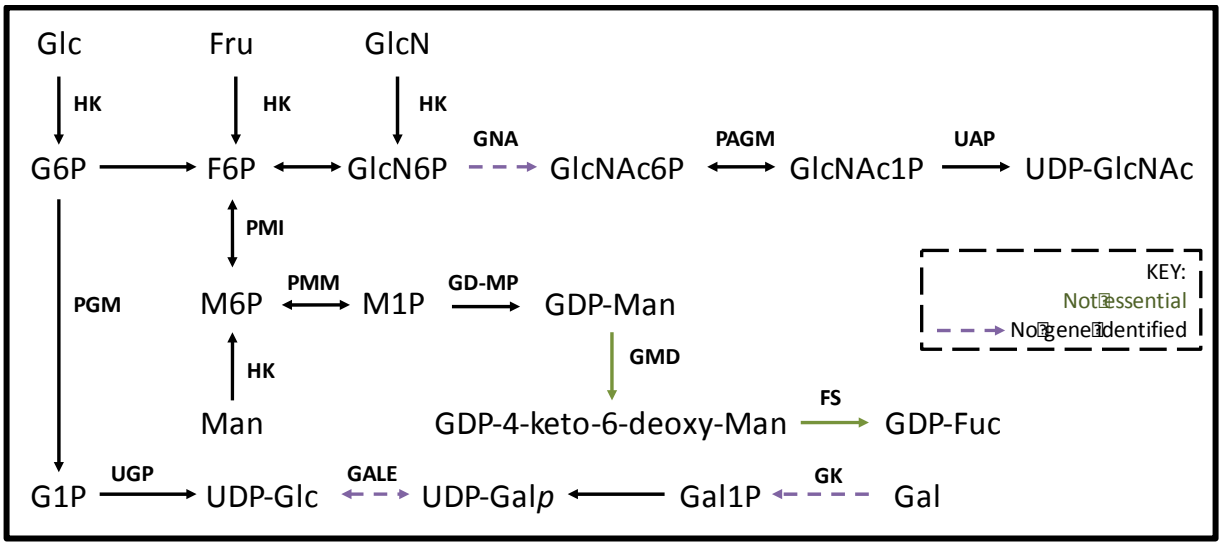


Figure 18 - Sugar metabolism of *P. falciparum* adapted from (Cova *et al.*, 2015)

P. falciparum does not synthesise or use GalNAc (Sanz *et al.*, 2013). UDP-GlcNAc is most likely made *de novo* from G6P although it could be made from a salvage pathway *via* GlcN. Both pathways require GNA to convert GlcN6P to GlcNAc6P, of which there is no candidate gene. Glucose metabolism of *P. falciparum* was reported as early at 1983 and although the parasite clearly synthesises UDP-Glc, there are no obvious glucosyltransferases present in the genome (Sanz *et al.*, 2013; Cova *et al.*, 2015). All enzymes required to produce GDP-Man both by a *de novo* or salvalge pathway are present in *P. falciparum*. Mannose is well used in the parasite's GPI anchors, but is not believed to be present in N-glycans (Naik *et al.*, 2000; Sanz *et al.*, 2013).

7.5.3 Current evidence of N-Glycosylation

In *P. falciparum*, only 3 homologues of the 14 *ALG* genes required to make the $Glc_3Man_9GlcNAc_2$ precursor of N-glycans are present in the genome: *ALG7*, *ALG13* and *ALG14*. *ALG7* is responsible for the addition of GlcNAc onto the dolicholphosphate precursor and *ALG13* and *ALG14* work as a complex to add a second GlcNAc forming GlcNAc2-PP-Dolichol. *STT3* is also present in the genome, suggesting that the parasite is able to transfer these short N-glycans to proteins, figure 19 (Samuelson *et al.*, 2005; Bushkin *et al.*, 2010).

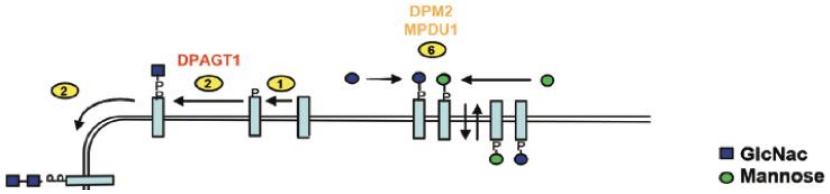


Figure 19 – Stunted N-glycosylation in *P. falciparum* adapted from Juan *et al.*, 2016.

Plasmodium's "cousin" *Giardia Lamblia* synthesises stunted N-glycans just two GlcNAcs in length (Samuelson *et al.*, 2005). The group of J. Samuleson used WGA affinity columns to pull down glycoproteins from both *P. falciparum* and *G. Lamblia*. They characterised a number of short N-glycans in *Giardia* but the poor peptide coverage meant they could not do the same for *Plasmodium* (Ratner *et al.*, 2008; Bushkin *et al.*, 2010). They use PNGaseF to show the reduction in binding from WGA on cleavage of N-glycans by this enzyme. However, PNGase only cleaves glycans when there is at least $GlcNAC_2Man_3$ present therefore should not affect these predicted short glycans. It is possible in fact that, in the case of *P. facliparum,* they observed a reduction of glycans from the host cell. However, Bushkin *et al.* were able to selectively label *Plasmodium* with GSL II whilst inside a red blood cell, indicating the presence of these short N-glycans and suggesting that they are not from the host (Bushkin *et al.*, 2010). Further evidence for the presence of N-glycans is the toxic effect that tunicamycin, the N-glycosylation inhibitor, has on the parasite which blocks its development after a 48 h cycle (Dieckmann-Schuppert, Hensel and Schwarz, 1992; Naik *et al.*, 2001). However, both these studies attributed the effect to non-specific toxicity.

7.6 Activity based proteomic profiling probes (ABPP probes)

Activity based proteomic-profiling (ABPP) uses chemical probes to target active sites of enzymes, figure 20. ABPP has a wide range of applications including finding inhibitors, target discovery and characterising enzyme active sites.

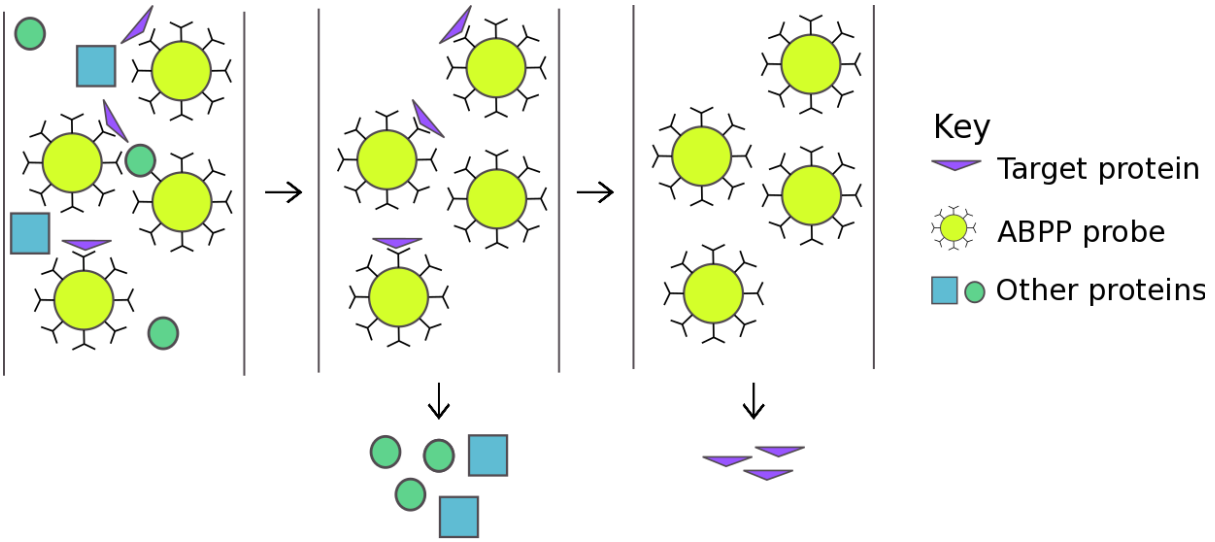


Figure 20 - Illustration of ideal activity-based probing

ABPP probes are small-molecules that bind to the active site of a given enzyme or class of enzymes. Most probes consist of a reactive group, a linker and a tag. Targeting a family of enzymes is achieved by tailoring the reactive group using transition state mimics or analogues of known substrates.

The reactive group must bind specifically and irreversibly in order to be an effective probe. The purpose of the linker is to create space between the reactive group and the tag. This reduces steric hindrance caused by the tag, which, if not reduced, could decrease binding to the target. The tag enables visualisation of the probe after reaction for characterisation. It could also be in the form of a click handle for visualisation after binding (Speers and Cravatt, 2004; Cravatt, Wright and Kozarich, 2008; Li, Overkleeft and Florea, 2012).

Hesek *et al.* developed a resin bound inhibitor of free active matrix metalloproteinases (MMP). They chose an inhibitor that was specific to MMPs, but broad enough to target a number of them. They then chemically modified it so they could tether it to an epoxy-activated sepharose resin and used it to purify MMPs from cancer patient samples. In doing so they

were able to show the presence of MMP-2 and MT1-MMP in breast carcinoma and MT1-MMP in laryngeal cancer tissue (Hesek *et al.*, 2006).

UDP-hexanolamine has long been described to enrich for glycosyltransferases. It consists of Uridine diphosphate immobilised on agarose *via* a hexanolamine chain (Barker *et al.*, 1972). However, this probe is non-specific in terms of the UDP-sugar transferase it enriches for (Khan *et al.*, 1994; Sindhuwinata *et al.*, 2010). The same group also developed a galactosyl and *N*-acetylglucosamine based resin. Galactose alone failed to capture any GTs, but the *N*-acetyl-glucosamine did, especially in the presence of UDP (Barker *et al.*, 1972). The creation of a probe that can differentiate between, for example, galactosyltransferases and *N*-acetyl-glucosamine transferases would be of great benefit to the discovery of new GTs.

We envisage using a similar method to capture glycosyltransferases by immobilising a sugar-nucleotide based inhibitor on a solid support. By incorporating both the sugar and the nucleotide and considering inhibitors that bind tightly to the active site, we hope to increase the specificity of the resin. It is my opinion that by using a tailored ABPP approach we can find uncharacterised GTs that are unique to an organism like *T. brucei*.

7.7 Design of affinity-based probes for glycosyltransferases

7.7.1 Inhibitors of glycosyltransferases

In order to design an activity based probe against a glycosyltransferase (GT) with unknown character (i.e. no homolog) we must consider their general mechanism of action and known inhibitors. The transition state of the reaction is complex, comprising of four components, making it a difficult target (see section **7.1.4 *Glycosyltransferases***). Since GTs bind to the donor first and change their conformation to then bind to the acceptor, it makes sense to design probes around the donor (Taylor and Drickamer, 2011). Furthermore, acceptors can take many complex forms therefore if the probe were modelled around the acceptor it would only be useful for one specific GT. The donor is common to all within a family. In general, the affinity for GTs to the donor nucleotide sugar lies with the nucleotide and the phosphate backbone, but the specificity lies with the sugar. This makes designing a broad affinity probe

challenging. We will be targeting galactosyltransferases (GalTs), which require a UDP-Gal donor.

In 1999, Takayama *et al.* published an extensive study on inhibitors of β-1,4 and α-1,3-galactosyltransferases (β-1,4-GalT and α-1,3-GalT) (Takayama *et al.*, 1999). The two chosen enzymes differ in mechanism; β-1,4-GalT is inverting and α-1,3-GalT is retaining. However, both use the UDP-Gal donor. After screening 56 compounds (mimics of uridine or galactose) for inhibitory action against β-1,4-GalT they found 5 strong inhibitors. These were UDP, UTP, UDP-2F-Gal and two diphosphate mimics: pyrophosphate and methylene diphosphonate. From these results, they were able to infer that the uridine moiety was important for inhibitory action, monophosphates showed little inhibitory affects and negative charge is important. Negative charge mimics the transition state promoting binding to the enzyme. They propose that UDP-2F-Gal is a good inhibitor because the fluorine could be involved in hydrogen bonding in the active site and it strengthens the glycosidic bond effectively blocking the enzyme from proceeding. To summarise, the important features of UDP-Gal in binding to β-1,4-GalT are highlighted below, figure 21.

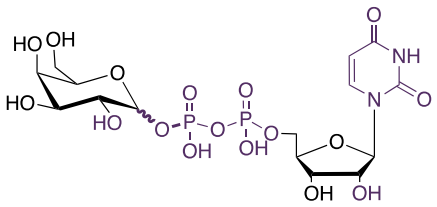


Figure 21 - Important parts of UDP-Gal for binding to β-1,4-GalT

When Takayama *et al.* tested these 5 inhibitors against α-1,3-GalT they saw that only UDP showed similar levels of inhibition. However, an iminocyclitol, despite its poor activity against β-1,4-GalT, was shown to be a highly potent inhibitor (Takayama *et al.*, 1999).

In 2006, Vidal *et al.* studied *C*-glycosyl analogues of three nucleotide diphosphate sugars, including UDP-Gal (Vidal *et al.*, 2006). They focussed on inverting GalTs that use UDP-α–Gal and synthesised the below analogue. C-glycosyls are more resistant to hydrolysis due to the relative strength of the C-C bond compared to C-O making them more practical to work with

and more stable *in vivo*. Replacing an oxygen with an ethylene forms a mimic for the S_N2 intermediate in the inverting mechanism of GalTs, which inhibits β-1,4-GalT (figure 22).

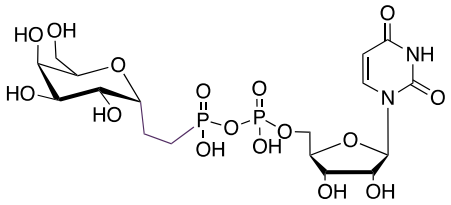


Figure 22 –Structure of UDP-Gal analogue that inhibits β-1,4-GalT

Descroix *et al.* also studied the inhibitory properties of donor analogues of galactosyltransferases (GalTs) (Descroix *et al.*, 2012; Jingqian Jiang *et al.*, 2016). They found that substituting the R group from a proton to a 5-formylthien-2-yl on uridine created a strong inhibitor of GBT, *N. meningitidis* α-1,4-GalT, and *B. taurus* α-1,3-GalT. Although several substituents inhibited the three enzymes, 5-formylthien-2-yl was the most effective. They attribute this to interactions of the moiety with a flexible loop in the enzyme. They then attempted to combine this inhibitory effect with Vidal's C-glycosidic mimic, figure 23. With these combined modifications, inhibitory action was maintained, but it was lower across all enzymes compared to the simpler 5-formylthien-2-yl. The authors do not speculate on the reason behind the reduction in activity. However, it could be that the introduction of the ethylene group changes the conformation of the uridine enough to reduce the strength of its hydrogen bond to Arg77 of the enzyme's active site.

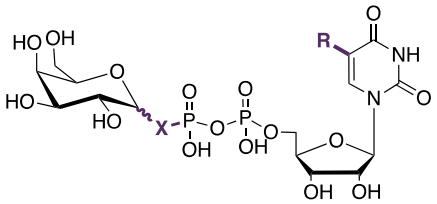


Figure 23 – UDP-Gal analogues where R = H or 5-formylthien-2-yl, and X = ethylene or O

Sujino *et al.* studied the effect of removing each hydroxyl on the galactose of UDP-Gal against three retaining αGalTs: calf thymus α–1,3-GalT, blood group B α-1,3-GalT and *Neisseria meningitidis* α-1,4-GalT (Sujino *et al.*, 2000). From this, they observed that only hydroxyls at C3 and C4 are important for recognition since all three enzymes failed to process UDP-3-deoxy and 4-deoxy-galactose, figure 24. They also noted that removing the hydroxyl at C6 only slightly affected the activity of the substrate. Persson *et al.* published the crystal structure of a retaining GalT isolated from *Neisseria meningitides* with UDP-2-deoxy-2-fluoro-galactose.

They found that O4 of the donor sugar forms hydrogen bonds with an asparagine side chain of the protein (Persson *et al.*, 2001). This interaction was also observed by Angulo *et al*. from NOE NMR studies of inverting blood group B GalT with UDP-Gal (Angulo *et al.*, 2006). These studies can explain Sujino's observations surrounding the importance of a hydroxyl at C4 as this hydrogen bond interaction may be crucial to the binding of the sugar to the enzyme's active site. Combining all of these studies, we propose that the fluorination of UDP-Gal at position C4 would increase the strength of the hydrogen bond to the GalT and would thus produce an effective and stable affinity based probe. Since position 6 is not important for recognition, conjugating a linker here should not prevent binding.

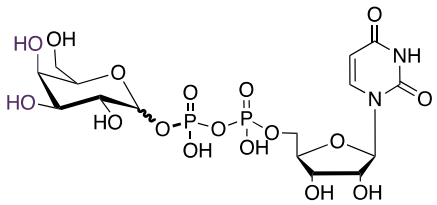


Figure 24 – Important moieties of UDP-Gal for recognition by GalTs

There have been a number of other studies related to GalT inhibitors (Burkart *et al.*, 2000; Schaefer *et al.*, 2013; Jingqian Jiang *et al.*, 2016). All these studies, along with those discussed above, target specific enzymes and none require their analogue to be immobilised. From this collection, we must design a general inhibitor that can be synthetically immobilised on a support. We know that the uridine moiety is important for recognition and modifying it can create powerful inhibitors; fluorinating C2 can inhibit β-1,4-GalT, but not α-1,3-GalT and the hydroxyl at C3 and C4 is important for retaining GalTs. Since the nature of the GalTs that we are searching for is unknown, a range of inhibitors will most likely be need. Fluorinating at C3 or C4 of UDP-Gal could create an inhibitor of retaining GalTs due to the importance of this position in recognition.

7.7.2 Resin bound sugars in solid support synthesis

Immobilising the inhibitor on a resin requires a tailored synthesis of UDP-Gal or its analogues. We suggest the introduction of the linker at C6, since this position seems to be away from the sites of recognition. Most sugars in the literature are attached to supports or resins *via* C1, therefore they require a very different strategy (Osamu Kanie and Wong, 2000; de Souza *et al.*, 2004; Jang *et al.*, 2008; Ban *et al.*, 2012). Those that couple at the C6 often do so purely as

a synthetic aid and are removed from the resin before the end of the synthesis (Xu, Springfield and Koh, 2000; López-Prados *et al.*, 2005).

Lourenço *et al.* successfully synthesised a sugar bound to tentagel MB resin *via* a succinate linker at the C6 position as a synthetic aid (Lourenço and Ventura, 2011), figure 25. This linker can be easily attached to C6 from the ring opening of succinic anhydride, leaving a carboxylic acid moiety. This can be attached to the resin with amide bond formation. For Lourenço *et al.*, the resin was used to avoid lengthy purification procedures that are often encountered in carbohydrate synthesis.

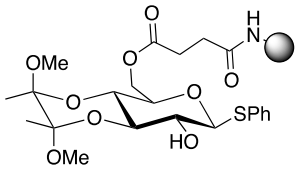


Figure 25 – Sugar bound to tentagel resin via succinic anhydride adapted from Lourenço and Ventura, 2011

Tethered sugars have been used on arrays with a PEG/ alkyl linker using either a thiol/ maleimide pairing or alkyne/ azide click chemistry (Ban *et al.*, 2012; Jianbing Jiang *et al.*, 2016). These sugars were used as acceptors in a glycosyltransferase array, therefore were tethered *via* the C1, figure 26.

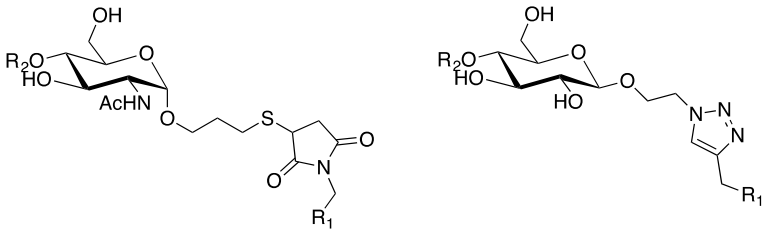


Figure 26 – Sugars linked to an array via thiol/ maleimide pairing or alkyne/ azide click chemistry, (Ban, 2012)

7.7.3 Carbohydrate chemistry

Carbohydrates are compounds consisting of carbon, hydrogen and oxygen. The simplest carbohydrates are monosaccharides, the building blocks of glycans. In 1902, Emil Fischer received the Nobel Prize for Chemistry for his outstanding work in purine and carbohydrate synthesis (Fischer, 1902). Since then, a further four Nobel prizes have been awarded for contribution to the field of carbohydrates; Walter Haworth in 1937, Carl and Gerty Cori and Bernardo Houssay in 1947 and Luis Leloir in 1970. In Fischer's prize winning speech, he said

"the abundance of substances of which animals and plants are composed, the remarkable processes whereby they are formed and then broken down again have claimed the attention of mankind of old, and hence from the early days they also persistently captivated the interest of chemists." It is the important roles that carbohydrates play in every aspect of life and their enormous abundance that has driven the work of many great scientists like Fischer.

Fischer deduced the structures of most natural monosaccharides and established the Fischer projection – a way to represent these monosaccharides in their open chain form on paper ('Rules of Carbohydrate Nomenclature', 1963). Monosaccharide backbones can be from 3 to 7 carbons long with those of length 5 called pentoses and those of length 6 hexoses. In the open form, each monosaccharide has two enantiomers which are non-superimposable mirror images of each other (D or L). However, most exist in their cyclic form. In this form, there are two anomers: α and β which are epimers at the hemiacetal carbon. Between monosaccharides with the same length carbon backbone, the differences are subtle epimerisations of hydroxyl groups anywhere except at the hemiacetal carbon. Examples of all these isomers are shown in figure 27.

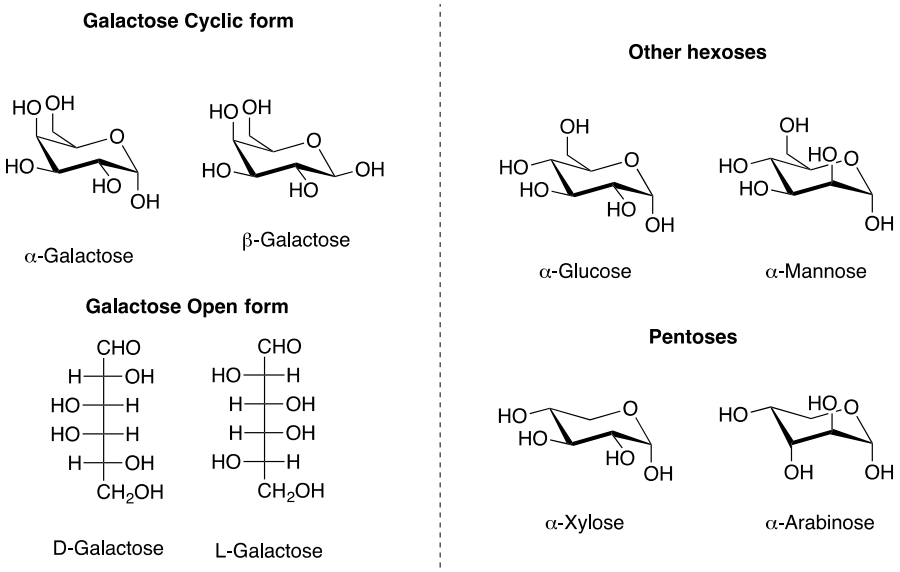


Figure 27 – The different isomers of galactose and examples of hexoses and pentoses

Monosaccharides are polar and highly soluble in water due to the high number of hydroxyl groups that can form hydrogen bonds. In nature, these monosaccharides can be bound enzymatically to produce di-, tri-, oligo- or polysaccharides. These complexes can be extracted and purified from their native organisms, but are rarely obtained in large quantities. Unnatural

carbohydrates also present interesting biological properties, but cannot be extracted from natural sources. In order to obtain large quantities of natural or unnatural carbohydrates, they must be made synthetically (Stick, 2001).

Synthesis of carbohydrates often involves extensive protection strategies in order to differentiate the near identical hydroxyl groups on the sugar. To avoid lengthy routes, some products can be obtained through chemoenzymatic methods (Hayashi *et al.*, 1997; Burkart *et al.*, 2000; Li *et al.*, 2014). However, using enzymes leads to long reaction times, the use of water as a solvent, strict control on pH and temperature and they are often expensive. More crucially, enzymes are normally used only with their naturally occurring substrates since analogues are often inhibitors of the enzyme, thus limiting their utility (Akiyaffici *et al.*, 1987).

A number of GalT inhibitors have been synthesised, as discussed. Fluorinated analogues of UDP-Gal have been synthesised as inhibitors of UDP-galactose 4-epimerase (Chapeau and Frey, 1994), Galactopyranose mutase (Burton and Boons, 1997) and Human β-1,4-Galactosyltransferase (Takaya *et al.*, 2005), along with other uses (Hayashi *et al.*, 1997; Barlow and Blanchard, 2000; Burkart *et al.*, 2000). All these studies targeted specific enzymes and did not require their analogue to be immobilised. Immobilisation on the resin will require a tailored synthesis of UDP-Gal and its analogues where the C6 position must be differentiated.

7.8 Aims

This thesis aimed to use chemical methods to study glycosylation in *Trypanosoma brucei* and *Plasmodium falciparum* under four main aims:

(i) to design a synthetic route to immobilised nucleotide sugars

(ii) to develop a strategy to characterise unknown galactosyltransferases of *Trypanosoma brucei*

(iii) to examine the effects and utility of the small molecule N-glycosylation inhibitor NGI-1 on *Trypanosoma brucei*

(iv) to characterise the N-glycans of *Plasmodium falciparum.*

To address aims (i) and (ii), activity based probes were synthesised and a protocol was developed to use these probes to enrich galactosyltransferases from a mixture of proteins. The protocol was then adapted to try and enrich galactosyltransferase from whole cell lysates of *Trypanosoma brucei.*

To address aim (iii), *Trypanosoma brucei* was cultured in the presence of NGI-1. Its growth was monitored over a range of concentrations and western blots performed to view alterations to N-glycosylation.

To address aim (iv), *Plasmodium falciparum* was cultured and a number of different lectin-based purification techniques were evaluated to purify N-glycoproteins. Mass spectrometry was then used to characterise these proteins.

8 Chapter 1

Synthesis of resin bound affinity-based galactosyltransferase probes

8. Chapter 1

Synthesis of resin bound affinity-based galactosyltransferase probes

8.1 Results and discussion

This chapter has two main aims; (i) to synthesise activity based probes based on known inhibitors of galactosyltransferases (GalTs); (ii) to develop a method to use the activity based probes to enrich galactosyltransferases from a mixture of proteins. The probes were based upon known inhibitors of GalTs and their synthesis was designed by modifying routes to similar compounds using intricate protection strategies. The enrichment protocol was developed using commercially available bovine β-1,4-galactosyltransferases (GalT) in a range of systems.

8.1.1 Author Contributions

J. G. M. Bevan, E. C. Lourenço and M. R. Ventura designed the synthetic route to the resin bound probes. J. G. M. Bevan, M. Chaves-Ferreira and J. A. Rodrigues designed the experiments to enrich β-1,4-GalT using the resin bound probes. J. G. M. Bevan performed all the reactions and experiments, data analysis and created all the figures and tables. M. R. Ventura and J. A. Rodrigues mentored and supervised the work.

8.1.2 Synthesis of resin bound glycosyltransferase probes

The synthesis of two activity based probes, UDP-galactose (UDP-Gal) and UDP-4-deoxy-4-fluoro-galactose (UDP-(4F)-Gal), will be described as well as their conjugation to tentagel resin. Although the synthesis of free UDP-Gal and UDP-(4F)-Gal have been described (Burton and Boons, 1997; Wagstaff *et al.*, 2015), this represents the first synthetic route to any nucleotide sugar attached to a solid support.

The main challenges of these syntheses are: 1) selective addition of the linker at the C6 position for subsequent coupling to the resin; 2) the sensitivity of the phosphate to acidic conditions; 3) selective fluorination at C4; 4) the low reaction rates of uridine morpholidate with galactose phosphate; 5) carrying out synthesis in water.

The PEG resin tentagel was chosen as a solid support due to its ability to swell in a wide range of solvents – important for the crossover between chemistry and biology. Its reactive end contains an amine moiety, which can be coupled to the sugar *via* an appropriate linker.

8.1.2.1 Resin bound UDP-Galactose as a glycosyltransferase probe

Starting from galactose, the synthesis of resin bound UDP-Gal was designed to consider the following:

- Hydroxyls at positions 1 and 6 must be differentiated from each other and from 2, 3 and 4 so that UDP can be added to 1 and the linker to the resin attached at 6 (Figure 28):

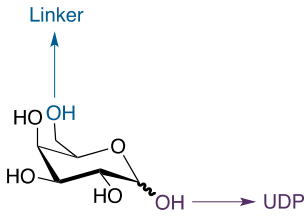


Figure 28 - Galactose

- The phosphate must be added before the sugar is attached to the resin to avoid cross reactivities:
- The UMP must be added last as it is sensitive and the reaction occurs in water.

The planned route to resin bound UDP-Gal is shown in figure 29. Orthogonal protection of C1 and C6 would produce sugar **A**. From **A**, the phosphate and the linker could be introduced producing **C**. Finally, UMP can be conjugated, producing the final sugar **D**.

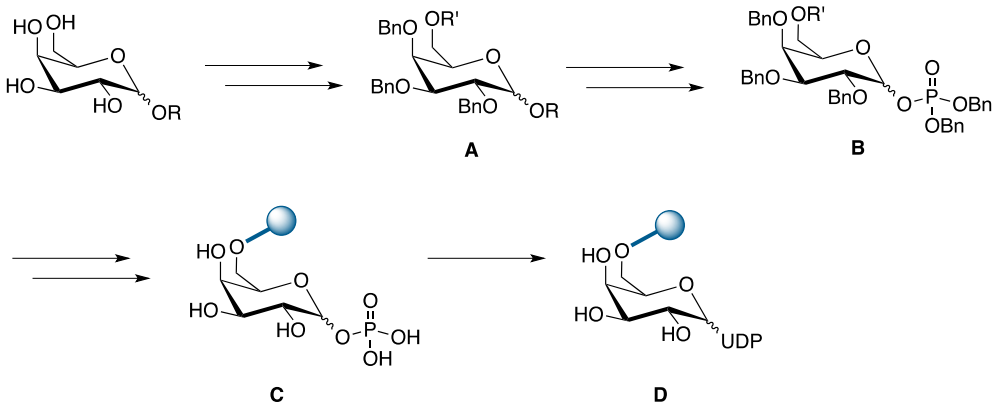
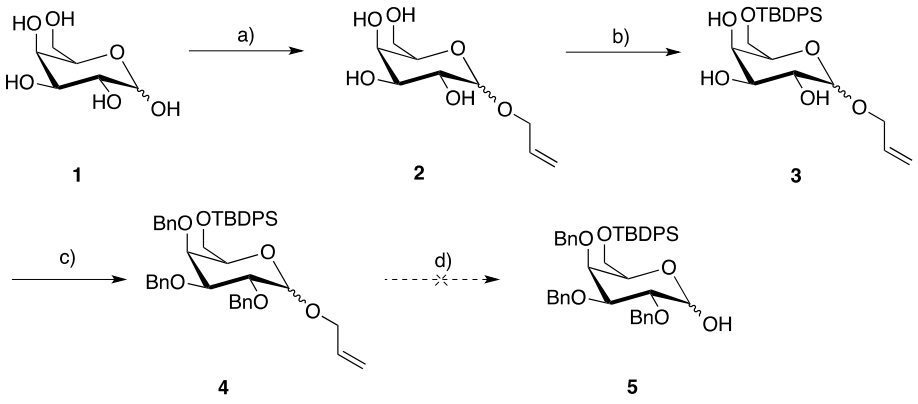


Figure 29 - Proposed general route to resin bound UDP-Gal. R = a C1 protecting group, R' = a linker.

Initially, the efficacy of the allyl and TBDPS moieties to be orthogonal protecting groups of C1 and C6 were investigated. The addition of the allyl chain is specific to C1 due to the hemiacetal nature of the anomeric hydroxyl. The bulky TBDPS is specific to C6 due to steric hindrance of the other hydroxyls (Takahashi and Nakata, 2002). Their addition leaves C2, C3 and C4 available to be protected by benzyl groups. Selective removal of either the allyl or the TBDPS would then allow for the addition of the phosphate or the linker.

Scheme 1 - Protection of galactose with an allyl chain and TBDPS

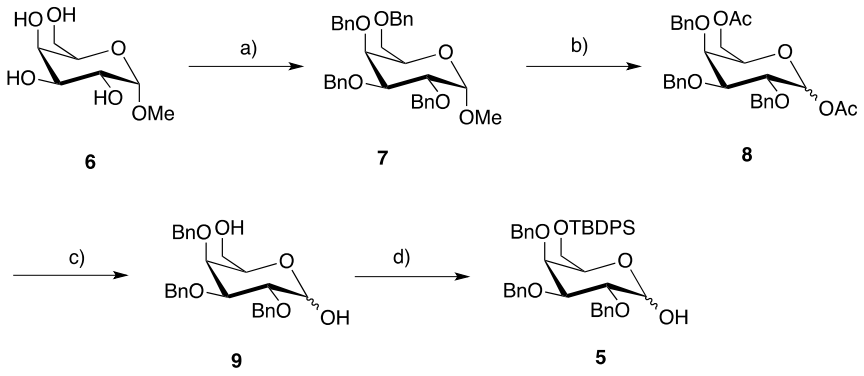


a) Alylic alcohol, CSA, 100 °C, **76%** b) TBDPSCl, DCM, Pyridine, NEt_3, **78%,**
c) BnBr, NaH, DMF, 0 °C to rt, **86%** d) $PdCl_2$, MeOH

Although the yields for reactions **a** to **c** are good to very good, this route from galactose (**1**) to sugar **5** (Scheme 1) proved to be inefficient. The deprotection of the allylic alcohol from the anomeric position of **4** to produce **5** formed two inseparable products. NMR evidence strongly

suggests that these products were the alpha and beta anomers of TBDPS migration to C4 without the allyl at C1. This is most likely due to the close proximity in space of the O4 to the O6 particular to galactose allowing for migration. The migration probably occurred during the addition of the benzyl protecting groups as these are strongly basic conditions. The mixture of anomers of the desired product could have hidden the extra product in the NMR spectra. However, once the allyl was removed, these products were then clearly visible on the TLC and the NMR showed four products. The migration of a silyl group under basic conditions is not unprecedented but this exact migration has not been reported (Stanciu and Bennett, 1974; Kira and Iwamoto, 2001). An alternative route to 5 is outlined in Scheme 2.

Scheme 2- Alternative route to sugar 5

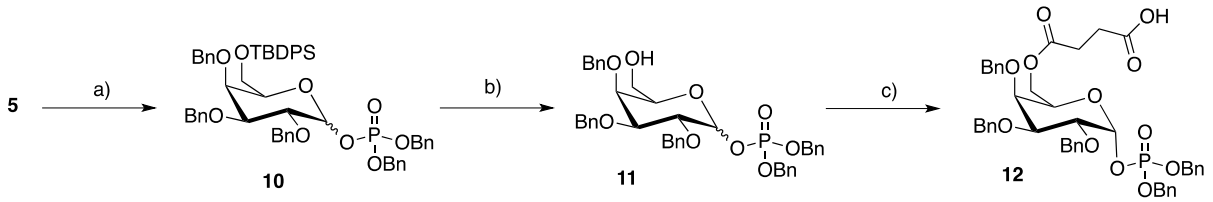


6 **7** **8** **9** **5**

a) BnBr, NaH, DMF, 0 °C to rt, **quant.** b) Ac_2O:AcOH (1:1), H_2SO_4, **81%**
c) NaOMe (0.6 eq), MeOH, 0 °C to rt, **82%** d) TBDPSCl, Imidazol, DMF, **89%**

This utilises the selectivity of the TBDPS protection of C6 over C1 as in Takahashi's synthesis of Mucocin (Takahashi and Nakata, 2002). Methyl α-glucoparanoside (**6**) was benzylated under strongly basic conditions with sodium hydride (added at 0 **°**C) in a quantitative yield (Scheme 2). Then 1-OH and 6-OH were acetylated under strongly acidic conditions to afford sugar **2** in a 82% yield (Lourenço, Maycock and Rita Ventura, 2009). They were then efficiently deprotected in a very good yield of 82%, leaving C6 to be selectively protected in a very good yield of 89% (Takahashi and Nakata, 2002). This route is a great improvement upon the previous.

Scheme 3 - Addition of phosphate to C1 and linker to C6

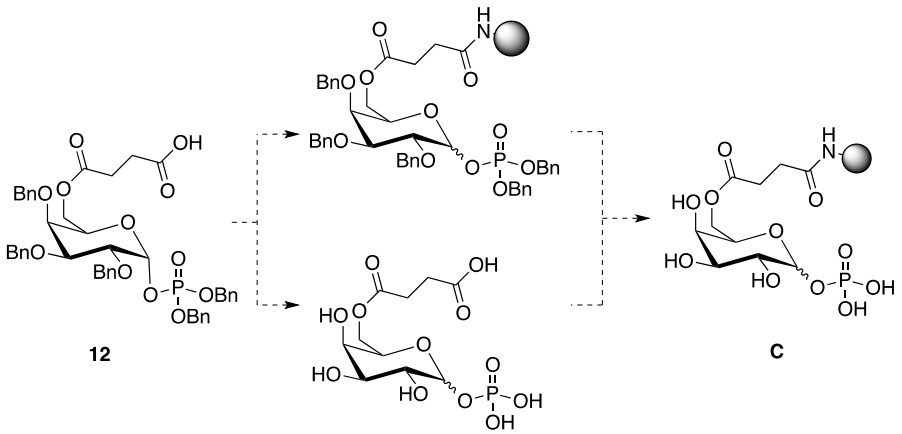


ai) $(BnO)_2PN^iPr_2$, Pyr.HCl, DCM, rt, 30 min, ii) Cumene peroxide, 0°C to rt, 1 h, **99%**
b) TBAF, THF), **76%** c) DIPEA, DMAP (cat.), Succinic anhydride, DCM, rt, **98%**

From sugar **5**, a phosphorous needed to be introduced to C1, the TBDPS removed and a linker conjugated at C6 (sugar **B** in figure 29). Dibenzyl phosphate was added *via* the phosphityl in a quantitative yield (Priestley, 1996) and the TBDPS removed in a good yield of 76% (Takahashi and Nakata, 2002) producing sugar **11**. The succinate linker was then conjugated to C6 in a near quantitative yield (scheme 5) forming sugar **12** (Lourenço and Ventura, 2011). This linker was chosen due to its previous success in the group and because it is able to form amide bonds with our chosen resin (Lourenço and Ventura, 2011). A disadvantage of this linker is that for selective conjugation to the C6 position, all other hydroxyls must be protected.

Subsequently, the sugar either had to be linked to the resin and then deprotected or deprotected and then bound to the resin (Scheme 4).

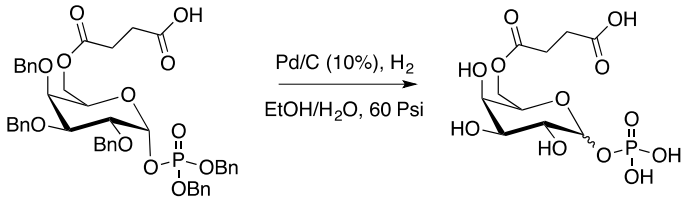
Scheme 4 - Two possible routes to theoretical sugar C



Deprotecting the sugar of its benzyl groups before linking to the resin leaves many hydroxyls free to potentially react with the resin. However, deprotecting by catalytic hydrogenation on the resin is problematic since the common reagent used is Pd/C (a solid), which would be difficult to remove. Since neither route is perfect, both were explored.

After optimisation, sugar **12** underwent hydrogenation to give **13** in a quantitative yield, Scheme 5. We saw that the solvent system of this reaction was crucial to its success and that a 50:50 mixture of EtOH:H_2O has to be used (Liu, Shen and Ichikawa, 1992). This is likely due to the solubility of the intermediate products of the reaction and of the fully protected starting material – if they are not able to dissolve they cannot react with the palladium and remain partially deprotected.

Scheme 5 - Deprotection of sugar 12



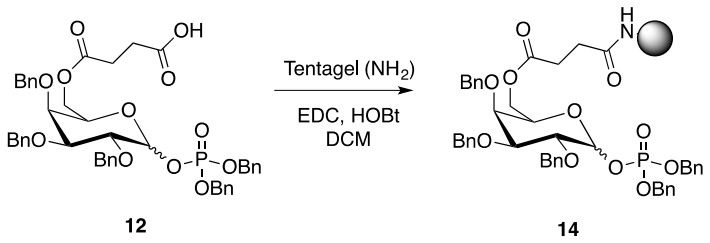
12 **13 100%**

We then tested a range of amide forming conditions to couple sugar **13** to tentagel-NH_2 resin using the coupling reagent HOBt with EDC or DIC (Lourenço and Ventura, 2011; Tonn and Meier, 2011). In order to determine that the sugar was bound to the resin, High Resolution Magic Angle Spinning NMR was used (HRMAS). This analytical technique utilises the resin's ability to swell in solvent to obtain a "semi-solid" NMR spectra. Standard NMR is unable to resolve spectra of semi-solids and solids due to dipolar coupling, quadropolar interactions and chemical shift anisotropy. The rapid movement of liquids averages out these interactions but in solids dipolar interactions and anisotropy broaden the NMR signal reducing resolution. Magic angle spinning reduces the effect of these interactions by physically spinning the sample at a mathematically defined "magic" angle thereby improving resolution (Power, 2003). In this case, HRMAS still suffers from some broadening due to the inflexible backbone of the PEG resin. In addition, the soluble arm of the resin is highly abundant in comparison to the conjugated sugars and can dominate the spectra, especially when dissolved in D_2O where it is soluble. However, HRMAS remains a powerful tool allowing characterisation of compounds conjugated to a solid support. The alternative would be cleavage of the compound and standard NMR of the cleaved product – an often-impractical solution especially when the product needs to remain on the solid support.

HRMAS of the coupling of **13** to tentagel resin showed several environments present in the ^{31}P spectra, suggesting that **13** attached *via* the phosphorous as well as the desired succinate

linker. This should have been predicted since phosphates are also able to react with amines in the presence of carbodiimides to produce phosphoramidates (Hermanson, 2008). This route was put to one side and the alternative route investigated (attaching sugar **12** to the resin followed by deprotection).

Scheme 6 - Conjugation of sugar 12 to tentagel resin



Sugar **12** underwent amide bond forming conditions with EDC, HOBt and tentagel (scheme 6) (Lourenço and Ventura, 2011). From ^{1}H, ^{13}C and ^{31}P HRMAS NMR it was clear that the resin had bound to the correct location at C6. Figure 30 shows the spectra of sugar **12** free and conjugated to the resin. The line broadening in HRMAS spectra described above can be observed here, although there are clearly peaks from the sugar on the resin. As the spectra was conducted in $CDCl_3$, the peaks associated to the soluble chain of the resin are minimised.

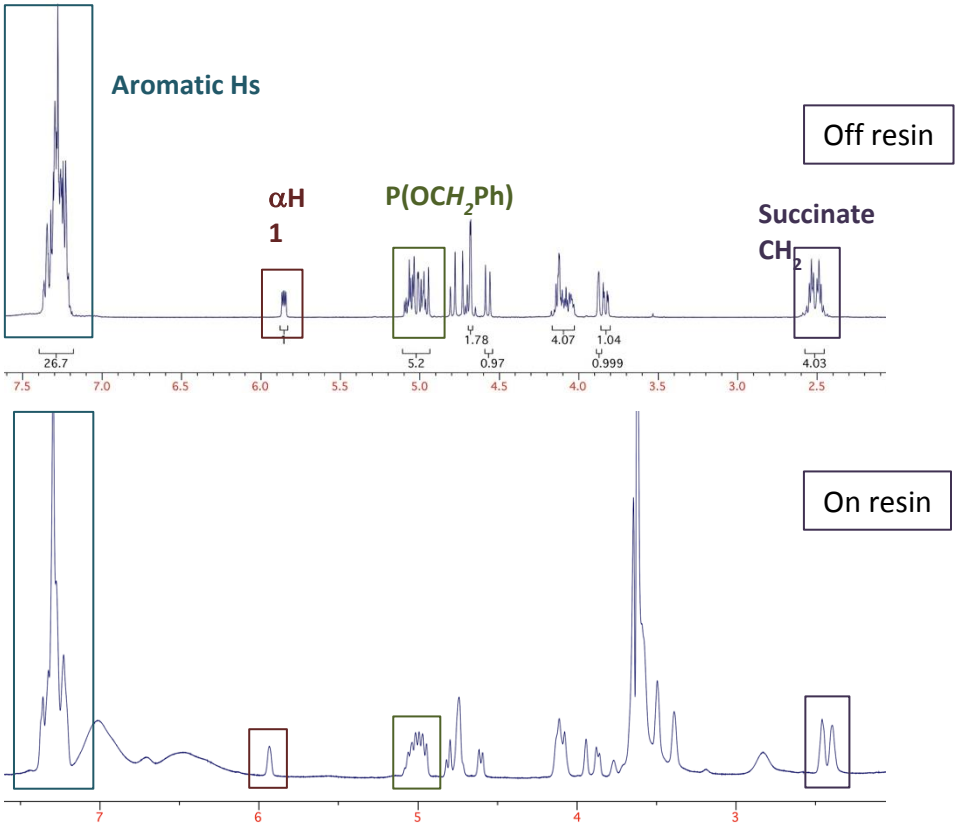


Figure 30 – ^{1}H NMR of sugar 12 and ^{1}H HRMAS of resin 14

We showed that sugar **12** can be deprotected using Pd/C in H_2O: EtOH (1:1). Therefore, these conditions were applied to the sugar on the resin. Unfortunately, there was no reaction and the palladium proved impossible to separate from the resin. Many other conditions were tried with no success (Jones, 1980; Tam, Heath and Meriifield, 1983; Bourne *et al.*, 2001; Chéry *et al.*, 2001). The main disadvantage of coupling to a solid support is that any solid phase reaction is much slower than one in liquid. With two solids (i.e. palladium on carbon and the resin) the rate is even slower since the surface area which can come together to react is very small and the chance of contact is low since their movement is more restricted than that of a liquid. Although the resin is swelled and is therefore not completely solid, the deprotection of benzyl groups must be performed before the sugar is coupled to the resin to avoid this solid-solid reaction. This route showed that the chemistry of the succinate linker and the phosphate are too similar to proceed and an alternative linker must be found.

8.1.2.2 Use of Succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) linker

The linker SMCC contains two reactive groups; an N-hydroxysuccinimide (NHS) ester and a maleimide (Figure 31) and is an FDA approved linker commonly used in bioconjugation especially in antibody-drug conjugation (Koniev and Wagner, 2015; Luo *et al.*, 2015; Yao *et al.*, 2015). These groups react towards amino and sulfhydryl groups respectively.

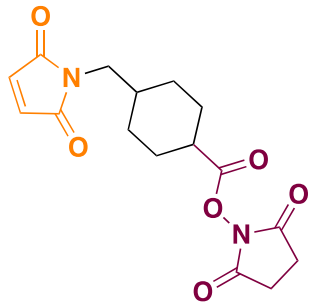


Figure 31 – SMCC, with NHS ester and maleimide groups highlighted

We chose this linker to replace the succinimide because its two reactive groups are orthogonal and should avoid overlapping reactivities with the phosphorous. A disadvantage of SMCC is that it is hydrophobic which can lead to problems with rapid clearance in antibody drug conjugation (R. Y. Zhao *et al.*, 2011). In this case, the deprotected sugar is hydrophilic and should counteract the hydrophobicity of SMCC and dissolve in water for the biological assays. Tentagel should be compatible with this linker as the NHS ester end of SMCC will link to the amine end of the tentagel resin. If a thiol can be installed at the C6 of the sugar, then it can also be linked to SMCC *via* the maleimide. The sugar in figure 32 could theoretically be attached to SMCC *via* the thiol.

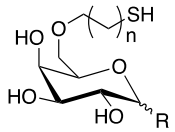
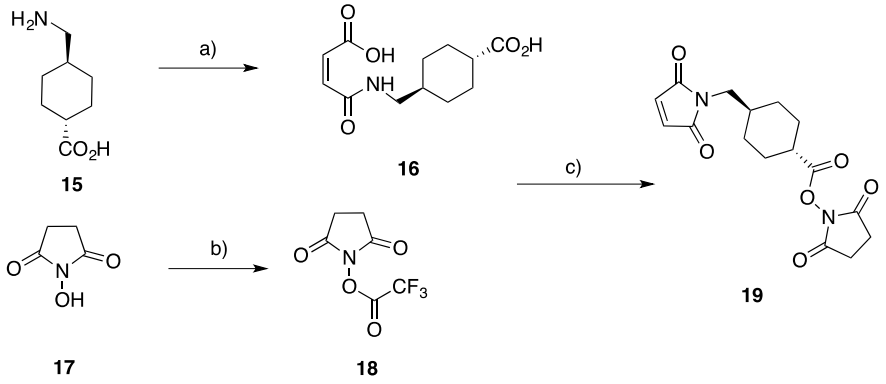


Figure 32 - Example of a sugar that could be linked to SMCC, where R is a C1 protecting group

Succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxylate (SMCC) was synthesised according to literature procedure (Scheme 7) (Paterson and Eggleston, 2008; Leonard and Brunckova, 2011).

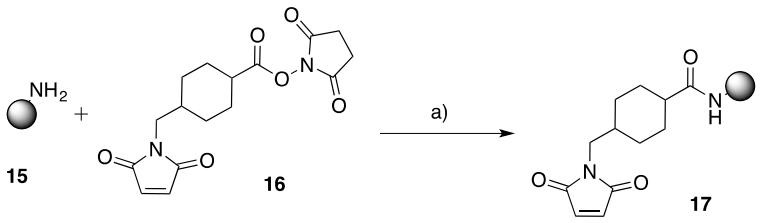
Scheme 7 - Synthesis of SMCC



a) Acetic acid, maleic anhydride, 93%, b) TFFA, 95%, c) *sym* collidine, DMF, 57%

SMCC was linked to the resin in PBS buffer and DMF (Yoshitake *et al.*, 1982; Harrison and Balasubramanian, 1998) with a loading of 76% (scheme 8). The SMCC had to be dissolved in DMF before adding it to the PBS otherwise it precipitated due to its hydrophobicity. The loading of the resin was calculated by calculating the mmol of SMCC added according to the weight increase of the resin as a percentage of the 0.3 $mmolg^{-1}$ loading of NH_2 on the unconjugated resin.

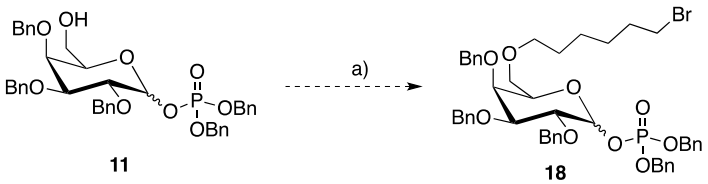
Scheme 8 - Coupling of SMCC to tentagel resin



a) PBS/DMF, **76%** loading

If an alkyl chain terminating in a bromine was coupled to C6, this bromine could be converted to a thiol, which could then be conjugated to the maleimide of the SMCC resin. Conditions to add the alkyl chain to sugar **11** were tested as shown in Scheme 9 (Dohi *et al.*, 2002).

Scheme 9 - Conditions tested to add alkyl chain to sugar 11



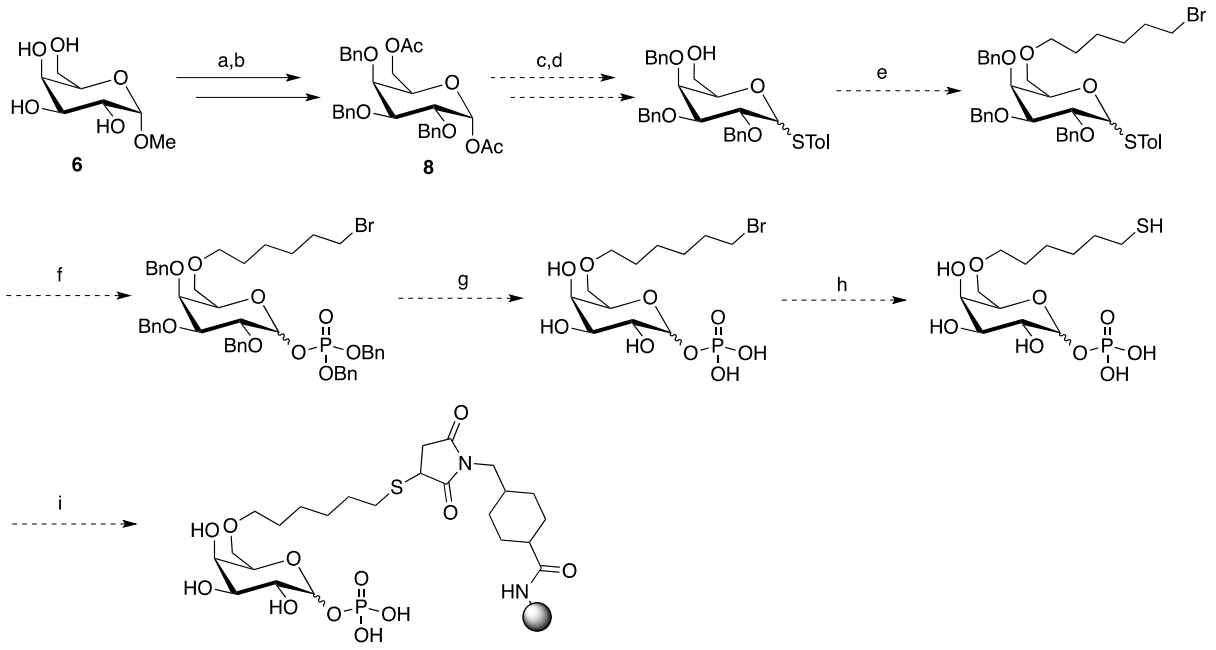
a) 1, 6 - Dibromohexane, NaH, TBAI, THF

Unfortunately, under these highly basic conditions, the benzyls on the phosphorous were removed producing a range of products, including alkylation of the phosphate. The hydride is

a hard nucleophile which could attack the phosphate displacing a benzyl ester in an S_N2 manner, the same way that fluoride is known to (Blackburn, 2006). In turn, this hydrogen could be then displaced by the alkyl halide producing the products observed.

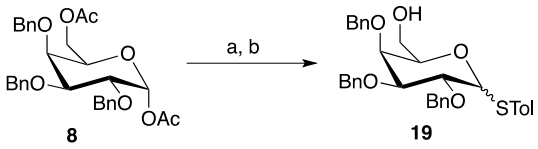
We concluded that it would be better to alkylate the sugar before introducing the phosphate. The following route was evaluated (Scheme 10). This route uses the previous steps to produce sugar **8** from sugar **6**. Then sugar **8** could be glycosylated with *p*-toluenesulphonic acid at C1. This would leave C6 differentiated for selective deprotection and available to couple to the alkyl chain linker before the phosphate is introduced to C1.

Scheme 10 – Alternative route to resin bound UDP-Gal



Donor formation with *p*-toluenethiol occurred in the very good yield of 86% (Scheme 11). The subsequent deprotection of C6 with sodium methoxide in methanol gave sugar **19** in the excellent yield of 94% (Scheme 11).

Scheme 11 – Differentiation of C6 for subsequent selective addition of alkyl chain linker

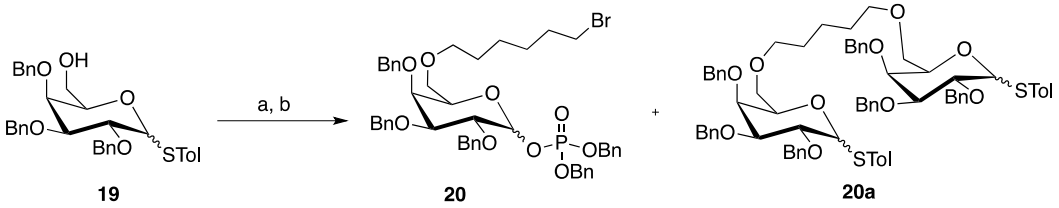


8 19

a) *p* -Toluenethiol, $BF_3.OEt_2$ DCM, 0 °C, **86%** b) NaOMe, MeOH, **94%**

With sugar **19** in hand we were able to proceed to the addition of the alkyl chain terminating in bromine (scheme 12). Dibromohexane was conjugated to C6 with sodium hydride catalysed by TBAI (Tetrabutylammonium iodide) in a yield of 68%. This reaction also gave a side product where the second bromide of the dibromohexane also coupled to **19**, but consistently in a yield lower than 8%, 20a. This side reaction is in a low yield and can be reduced by careful monitoring of the reaction so was not considered to be a problem. TBAB (Tetrabutylammonium bromide) is a common, and cheaper, catalyst used in alkylation reactions (Rong and Ruoho, 1999; Kanemaru, Yamamoto and Kadokawa, 2012), but when tested in place of TBAI there was no reaction. Both TBAI and TBAB are phase transfer catalysts but can also catalyse reactions by activating the alkyl halide (in this case alkyl bromide) by displacing the halide (Lu *et al.*, 2014). TBAI can displace alkyl bromide with the more reactive iodide whereas TBAB obviously cannot. The alkyl iodide formed is a more susceptible to attack from a nucleophile like the oxygen anion. This explains why TBAB was unable to catalyse the reaction.

Scheme 12 - Conjugation of alkyl chain linker and glycosylation



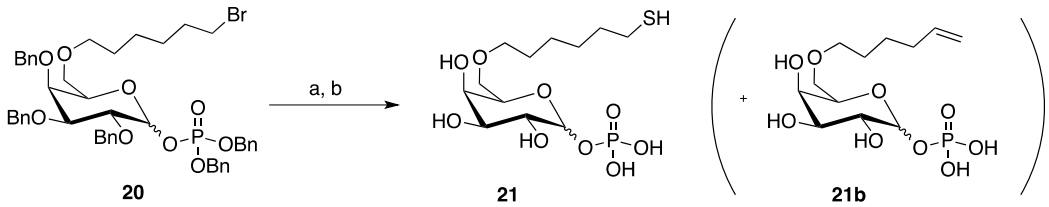
19 20 20a

a) 1,6-Dibromohexane, NaH, TBAI, DMF, **68%** b) $HOP(O)(OBn)_2$, NIS, TMSOTf, DCM, 0 °C, **85%**

Subsequent glycosylation with the acceptor dibenzyl phosphate and NIS/TMSOTf gave a very good yield of 85% with an α:β of 1:0.9 (Burton and Boons, 1997). The neighbouring benzyl group is electron rich and stabilises the oxonium ion intermediate. We found that sugar **20** could not be stored (even at -20 °C) for longer than a week as the phosphate is cleaved by acid hydrolysis.

Hydrogenolysis must be performed before the thiol is introduced to avoid poisoning the palladium catalyst. After optimisation, sugar **20** was deprotected in 30 min, quantitatively. Sugar **20** also suffers from rapid decomposition as the phosphate is sensitive to acidic conditions and the solution from the hydrogenation is acidic (pH1-2). The next step is to convert the bromine to a thiol, which can be achieved by first displacing the bromine with a thioacetate, then hydrolysing it to a thiol with a base. Once in these basic conditions, the sugar is much more stable as hydrolysis occurs in pH 1 to 4 (Foster and Overend, 1955).

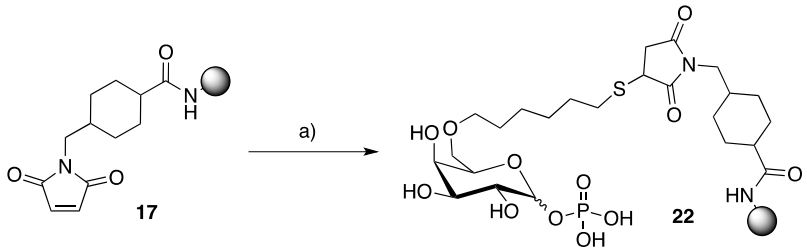
Scheme 13 - Deprotection of sugar 20 and conversion of the bromine moiety to thiol



a) Pd/C, H_2 (g), EtOH, 60 Psi, **100%**, b)i) $KSOCOCH_3$, DMF, ii) K_2CO_3, H_2O, **88%**

The reaction to add a thioacetate is quick and goes to complete conversion (Chen *et al.*, 2005). A number of bases can be used to hydrolyse the thioacetate to a thiol (Singh *et al.*, 2010). KOH gave the thiol but also the elimination side product **21b** in a 14% yield. The elimination product was also produced (11%) when attempts were made to go *via* the thiourea (Ferrell, Garces and Desmyter, 1976; Galano et al., 2013). These were only separable by using a G-10 sephadex column, therefore the alternative softer base potassium carbonate was tested. This base produced the thiol in an 88% yield and no elimination product was seen. However, over time (a few days) the thiols formed disulphide bridges between themselves. These bridges are easily broken by a reducing agent such as TCEP or DTT. Both reagents are used in excess and are difficult to separate from sugar **21** due to their solubility in water. DTT is more reactive towards maleimides as it contains a thiol and would therefore act as a competitor on conjugation of **21** to the maleimide of SMCC. TCEP is stable in a wider range of pH compared to DTT, and is a more powerful reducing agent; within 5-10 min all disulphide bonds can be broken (Burns *et al.*, 1991; Shafer, Inman and Lees, 2000). Therefore, TCEP was used to reduce disulphide bonds of **21.**

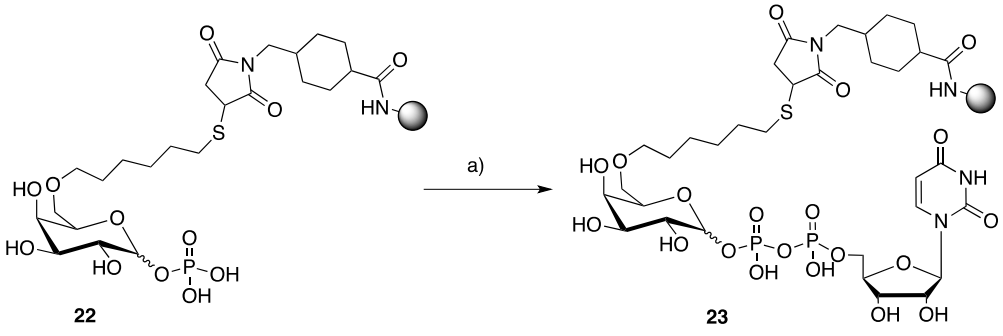
Scheme 14 - Conjugation of sugar 21 to the maleimide of SMCC on tentagel resin



Sugar **21** was coupled to the maleimide of SMCC bound to tentagel resin in H_2O/DMF (Scheme 14). The solution was acidified with K_2HPO_4 to pH 7-7.5 to allow the formation of the thiolate and the reaction allowed to stir overnight where the resin changed from yellow to pink and **22** was formed in a 71% loading (Harrison and Balasubramanian, 1998; Koniev and Wagner, 2015). The loading of the resin was calculated from amount of sugar added according to the weight increase of the resin as a percentage of the loading of SMCC on the unconjugated resin.

Uridine was activated with morpholidine to form the morpholidate. Then UMP morpholidate was coupled to the phosphate on **22** using the catalyst tetrazole. It increases the rate of reaction by protonating the morpholine group making it a better leaving group (Wittmann and Wong, 1997). This reaction is very slow, taking days to complete. Reloading the resin after 6 days with fresh UMP morpholidate and tetrazole improved the yield. Although much slower than the equivalent in-solution reaction, **23** was produced in a good loading of around 83% (Scheme 15).

Scheme 15 - Coupling of UMP with resin bound sugar 22



An exact loading was not always possible to calculate due to loss in mass of the resin after reaction. This loss in mass could be from transferring the resin, poor drying of the resin before the reaction or loss of substrate from the resin. However, by integrating the phosphate

environments of the diphosphate and the monophosphate, a ratio can be calculated and used to estimate how many have been converted to the diphosphate. This is not ideal as it assumes that no decomposition occurred. Carbohydrates can be quantified by reaction with sulphuric acid which dehydrates them to hydroxymethyl furfural. On reaction with phenol, the colour changes and this can be measured by a spectrophotometer and correlated to concentration (Nielson, 2009). Recently, a method was published that skips the phenol step and measures the concentration on a UV spectrophotometer after dehydration (Albalasmeh, Berhe and Ghezzehei, 2013). This method could be used to quantify the amount of sugar on the resin and then the ratios of the phosphate from the NMR spectra could quantify the amount of mono and diphosphate.

8.1.2.3 Resin bound fluorinated UDP-Gal

Starting from glucose, the synthesis of resin bound UDP-(4F)-Gal was designed to consider the following:

- Hydroxyls at positions 1, 4 and 6 must be differentiated from each other and from 2 and 3 so that UDP can be added to 1, fluorine at 4 and the linker to the resin attached at 6 (Figure 33). The starting sugar must be glucose due to inversion of the stereochemistry on fluorination;
- The phosphate must be added before the sugar is attached to the resin to avoid cross reactivity with the resin;
- The UMP must be added last as it is sensitive and the reaction occurs in water.

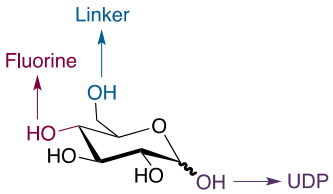


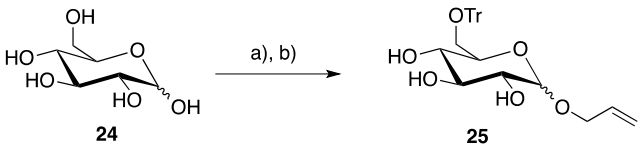
Figure 33 – Glucose

Initially, DAST was used to fluorinate position 4. DAST generates free HF in solution, which is highly corrosive and volatile but this fluorinating reagent requires the least extensive protection strategy (Card and Reddy, 1983). C1 was protected with an allyl moiety in a very good yield of 80%.

The next step was to introduce a protecting group at C6. In literature, either trityl or pivolate are chosen to protect the C6 during fluorinations (Card and Reddy, 1983; Chapeau and Frey,

1994; Hayashi *et al.*, 1997; Burkart *et al.*, 2000; Hartman and Coward, 2002; Giuffredi *et al.*, 2011; Giuffredi, Gouverneur and Bernet, 2013). When C4 of galactose is fluorinated, trityl appears to be the protecting group of choice most likely due to its robustness. The trityl moity was added to sugar **23** in a yield of 58% (scheme 16).

Scheme 16 – Protection of C1 and C6 of 24 (glucose)

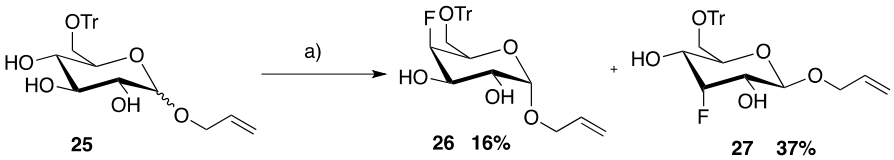


24 **25**

a) Alylic alcohol (17 eq), CSA (0.55 %), Δ, o/n, **80%** b) TrCl, DMAP, Pyridine, **58%**

Sugar **25** was then fluorinated with DAST in DCM over 3 days (Card and Reddy, 1983). This reaction was low yielding and produced two products. The desired product **26** was produced in a very low yield of 16%. Sugar **27** was isolated (fluorinated at C3) in a 37% yield (scheme 17). The two different products arise from the fact that α and β-glucose promote two different reactions. Somawardhana and Brunngraber studied the effect of the anomeric conformation on the selectivity of DAST and concluded that steric hindrance of substituted α-glucose prevented fluorination of C3 since the attack of the fluoride ion would have to approach from below on the same plane as the axial anomeric group (Somawardhana and Brunngraber, 1983). However, β-glycosides possess no such hindrance and C3 is preferentially fluorinated. Fluorination at C3 is unwanted since it produces an unnatural sugar, not galactose. Reduction of the reaction time to 1 day lead to a drastic reduction in yield to 2% for the desired product **26**.

Scheme 17 - DAST as a fluorinating agent



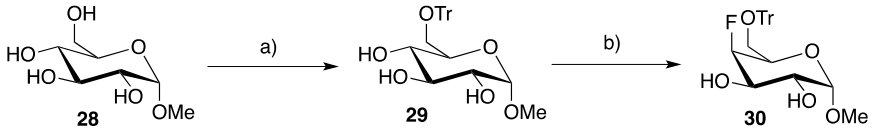
25 **26** **16%** **27** **37%**

a) DAST, DCM, 3 days

When anomerically pure methyl α-D-glucopyranoside protected at C6 with trityl, **29**, was fluorinated sugar **30** was formed in a yield of 29% (Scheme 18). Although this yield is comparable to literature, it is still low, the reaction time is very long and no starting material

can be recovered. The low yield can be explained by formation of the elimination product and decomposition of the starting material.

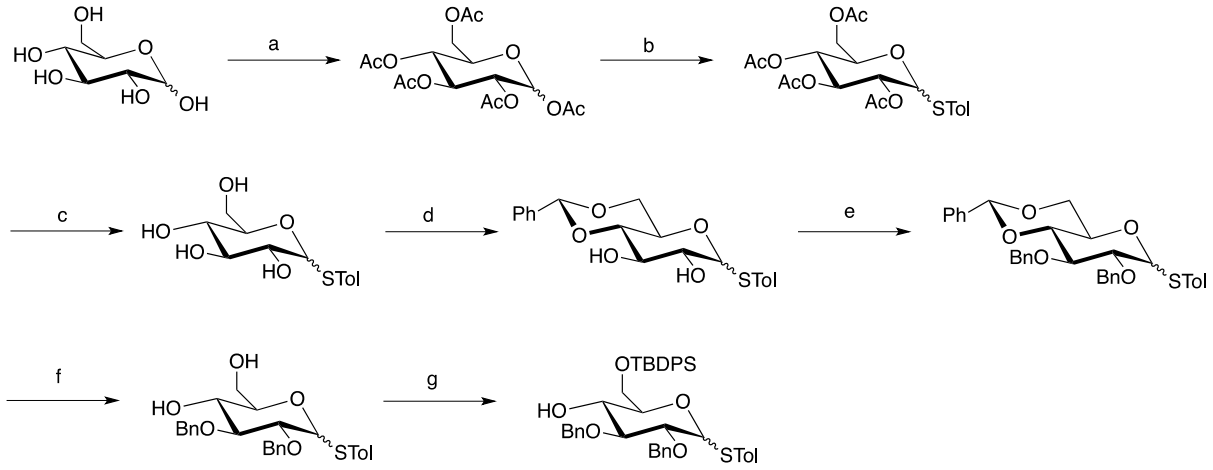
Scheme 18 – Protection of 28 and subsequent fluorination with DAST



a) TrCl, DMAP, Pyridine, **88%,** b) DAST, DCM, 3 days, **29%**

To avoid the use of DAST as the fluorinating agent, we investigated an alternate route (Koketsu *et al.*, 2011). By installing a triflate at C4, fluorine can be introduced with TBAF through displacement of the triflate with a fluoride ion. Burton *et al.* used TBAF to fluorinate the C4 position of glucose in a good yield of 74% (Burton and Boons, 1997). Since this method is not selective, the sugar must be fully protected leaving only position 4 free, requiring multiple protection steps. In literature, all steps give good to excellent yields so it was theorised that this method could improve the overall yield of the synthetic route compared to using DAST. Burton's route must be altered to suit our needs, the C6 hydroxyl must be differentiated for later coupling to dibromohexane. We proposed to introduce the TBDPS protecting group and to protect C1 with a thioester (Scheme 19). This avoids using an acetyl at C1 which was previously found to be difficult to remove. The thioester is also convenient for the glycosylation with dibenzylphosphate later.

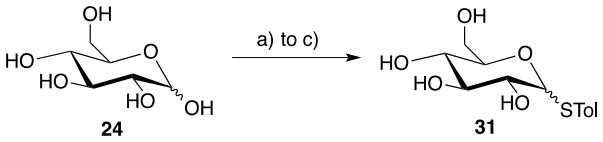
Scheme 19 - Proposed route to protected sugar for fluorination



a) AcOH:Ac_2O, H_2SO_4, b) *p* -Toluene thiol, $BF_3.OEt_2$, DCM, c) NaOMe, MeOH, d) $PhCH(OMe)_2$, CSA, DMF, 85 °C, e) BnBr, NaH, DMF, f) *p*-Toluene sulphonic acid, DCM, MeOH g) TBDMSCl, imidazole, DMF

Glucose was acetylated in a quantitative yield by stirring overnight in acetic anhydride and pyridine catalysed by DMAP. Acetylated glucose was then protected in a yield of 80% (Scheme 20). This reaction is slow, taking up to 3 days to complete. However, starting material is easily recovered for re-use. Removal of the acetyls proceeded with the excellent yield of 94% with the standard conditions of sodium methoxide in methanol.

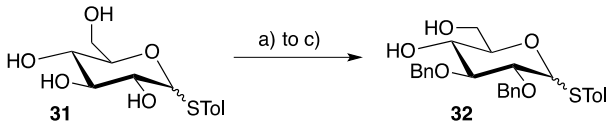
Scheme 20 - Synthesis of sugar 31 (glucose protected at C1)



a) Ac_2O, DMAP, pyridine, **quant.** b) *p* -Toluene thiol, $BF_3.OEt_2$, DCM, rt, 3 days, **80%**, c) NaOMe, MeOH, **94%**

C4 and C6 were protected with an acetal in a good yield of 70% by reacting sugar **31** with $PhCH(OMe)_2$ and a catalytic amount of CSA in DMF. The C2 and C3 of this sugar were then protected in a quantitative yield with benzyl moieties by reaction with benzylbromide in DMF with sodium hydride. The acetal could then be removed using *p*-toluene sulfonic acid in DCM/MeOH in an excellent yield of 94%, forming sugar **32** and leaving C6 open for selective protection and C4 open for fluorination.

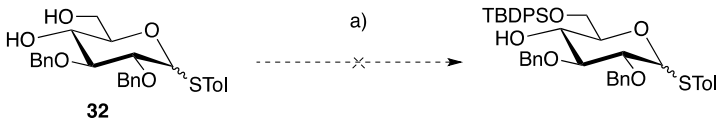
Scheme 21 – Synthesis of sugar 32



a) $PhCH(OMe)_2$, CSA, DMF, 85 °C, **70%** b) BnBr, MaH, DMF, **quant.** c) *p*-toluenesulphonic acid, DCM, MeOH, **94%**

Silyl groups are bulky and are therefore frequently used to protect the C6 of sugars that are less sterically hindered. Attempts to protect **32** with TBDPS in a range of conditions were not successful.

Scheme 22 - Protection of C6 with TBDPS

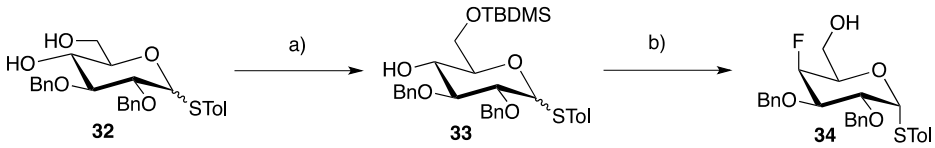


a) TBDPSCl, imidazole, DMF

When TBDPSCl was replaced with the smaller, less sterically hindered, TBDMSCl, the reaction proceeded with a quantitative yield (Scheme 23). This left C4 deprotected and free for fluorination. Using conditions adapted from Burton *et al.* the C4 was first activated with a

triflate before its displacement with a fluorine anion provided by TBAF (Burton and Boons, 1997). The addition of TBAF also simultaneously deprotects C6. After optimisation, the fluorination and deprotection of C6 afforded **34** in a 48% yield (Scheme 23). During the fluorination step, the alpha anomer was predominant. Sugar **32** is re-formed as a side product from the deprotection of C6 without fluorination of C4 or the hydrolysis of the triflate before fluorination. The C3 hydroxyl is in the same environment in the beta and the alpha anomer, therefore it is unlikely to be a case of the beta anomer forming **32** faster than the alpha (Miljkovic, 2009).

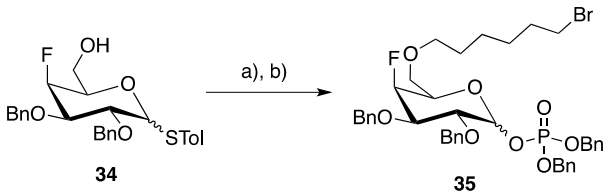
Scheme 23 – Synthesis of sugar 34



a) TBDMSCl, imidazole, DMF, **quant.**, b)i) Tf_2O, pyridine, DCM, -78 °C to rt ii) TBAF, THF, **48%**

From sugar **34** we could follow a similar route to that of resin bound UDP-Gal. Under the conditions that were applied to **19**, sugar **34** was coupled to the dibromohexane linker in a good yield of 80%. It was then glycosylated in a yield of 61% to produce sugar **35**.

Scheme 24 – Synthesis of sugar 35

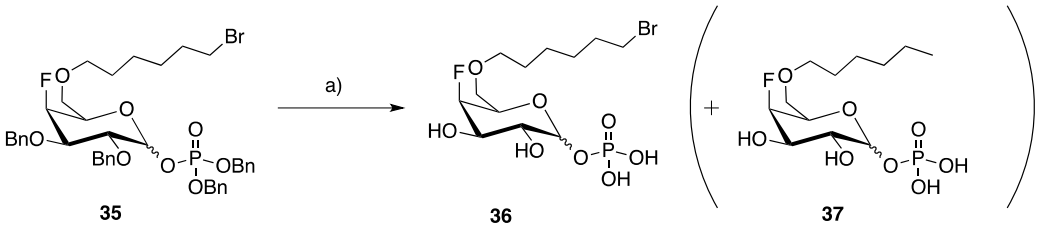


a) 1, 6 - Dibromohexane, NaH, TBAI, THF, **80%**, b) TMSOTf, $HOP(O)(OBn)_2$, NIS, DCM, 0 °C, **61%**

Burton's synthesis of UDP-(4F)-Gal is similar to this newly developed route. However, their use of the vinyl glycoside to obtain only alpha phosphate reduces their yield of glycosylation and introduces an extra step in the route. Since this selectivity wasn't necessary, our route obtained a higher overall yield: 15% over 10 steps compared to 5% over 11 steps. This is despite the higher yield Burton obtained in the fluorination step (74% compared to 48%). The smaller yield is likely due to the competing reactions of the hydrolysis of TBDMS and the fluorination of C4.

Sugar **35** was then subjected to hydrogenolysis under the conditions that were applied to the UDP-Gal derivative. The optimal reaction time was found to be 1 hour, which produced **36** in a near quantitative yield. Although bromine has the potential to poison palladium, the fact that it is a primary alkyl halide makes it difficult to hydrogenate. However, if longer reaction times are used, sugar **37** is produced where the terminal CH_2Br has been reduced to a methyl group (Scheme 25).

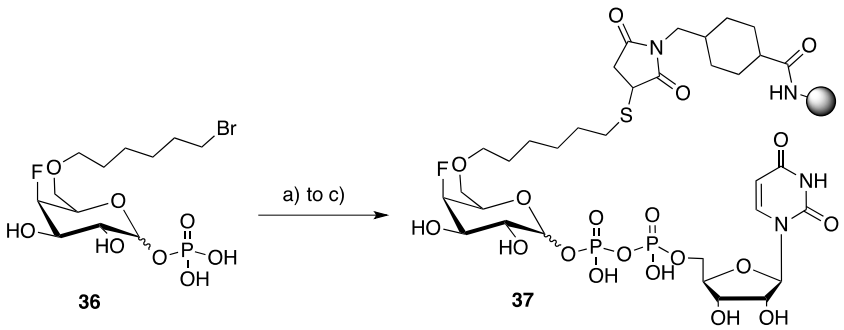
Scheme 25 – Synthesis of sugar 36 and side product 37



a) Pd/C, H_2 (g), EtOH, 60 Psi, **98%**

The bromine of sugar **36** was then converted to a thiol *via* the thioacetate in quantitative yield as previously described for sugar **20**. Resin **17** was coupled to the thiol with a loading of 78%. Subsequent coupling to uridine morpholidate catalysed by tetrazole gave **37** in an approximate yield of 38%. This was calculated as described above as an estimate of the phosphate environments.

Scheme 26 – Synthesis of resin bound sugar 37.



a)i) $KSOCOCH_3$, DMF, ii) K_2CO_3, H_2O **quant.**, b) **17**, K_2HPO_4, DMF/H_2O, **78%** c) n-octylamine, tetrazole, pyr., UMP-morpholidate

8.1.3 Assessment of resin bound UDP-Gal and UDP-(4F)-Gal as glycosyltransferase probes

8.1.3.1 Enrichment of β–1,4-galactosyltransferase with UDP-Gal based affinity probes.

Before testing if the resins are able to enrich GalTs from whole proteome extracts, we first developed a simpler system. First, the affinity of the resins to β-1,4-galactosyltransferase (β-1,4-GalT, a common, commercially available, galactosyltransferase) was tested. Then we looked at their ability to enrich β-1,4-GalT from two other proteins: variant surface glycoprotein (VSG) and BSA. VSG was chosen because in future experiments we propose to try the resin in *T. brucei* extracts and it is a highly abundant protein in *T. brucei* (Taylor and Drickamer, 2011).

A system was designed whereby the resin was placed on a column capped with a 40 μm filter, closed with parafilm and incubated with proteins. The 40 μm filter allowed the resin to be washed whilst the parafilm kept the buffer in place whilst the resin was incubated. The resin was washed and any bound proteins were eluted by boiling in laemmli buffer. The washed and eluted proteins were then run on a gel to visualise if any enrichment had occurred, figure 34.

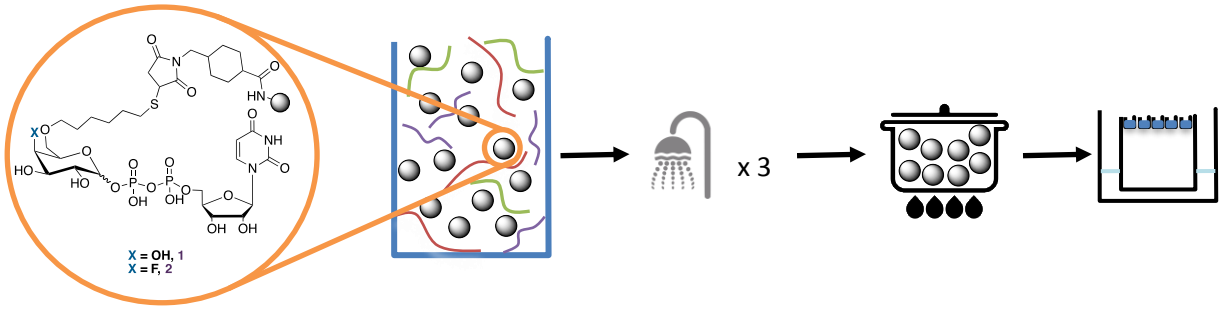


Figure 34 - Method to enrich for GalTs with an affinity resin

Initially, the amount of protein used was not sufficient to see on a gel in the washes nor the elution. Therefore, to determine the amount of protein to use, a gel was run with varying amounts of β-1,4-GalT and BSA. From this, it was clear that above 2 μg of β-1,4-GalT and above 1 μg of BSA was enough to see on a gel, see annex figure 67.

When the UDP-Gal resin or the UDP-(4F)-Gal resin is incubated with only β-1,4-GalT in a buffer containing $MnCl_2$ (to help GalT bind to the resin) it is eluted after the resin is boiled, as can be seen in lanes 5 and 10 of the gel in fig. 35. The washing lanes (2 to 4 and 7 to 9) in the gel show little protein in comparison. The control lane 6 contained half the amount of protein compared

to the amount incubated with the resins. Unfortunately, this does not seem to be enough to visualise well and more needs to be used. The multiple bands seen are most likely glycosylated forms of the GalT and aggregates of the enzyme. As a monomer, it has a molecular weight of 44 kDa and is the lowest band seen. The identity of all the bands is not known and some could be due to the resin itself.

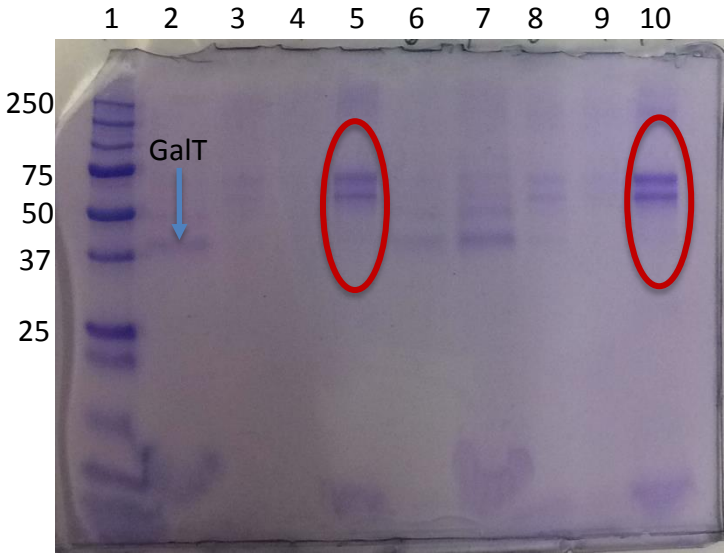


Figure 35 – Binding of β-1,4-GalT to the UDP-Gal and UDP-(4F)-Gal resins. Lane 1 = ladder, UDP-Gal resin lanes 2 to 5: 2 = wash 1, 3 = wash 2, 4 = wash 3, 5 = elution, lane 6 = control, UDP-(4F)-Gal resin lanes 7 to 10: 7 = wash 1, 8 = wash 2, 9 = wash 3, 10 = elution.

Since it was evident β-1,4-GalT was able to bind to the resin, the next experiment moved to a more complicated system to assess if it preferentially bound to β-1,4-GalT over VSG and BSA. The resins were incubated with β-1,4-GalT, VSG and BSA, washed three times and bound proteins eluted with laemmli buffer. In order to check that the resins themselves were not leaching material that caused the bands seen in previous gels, they were also boiled in laemmli without exposure to any proteins. The washes, elution and the supernatant from the boiled unexposed resins were loaded on to a gel along with a control lane of the three proteins together, figure 36.

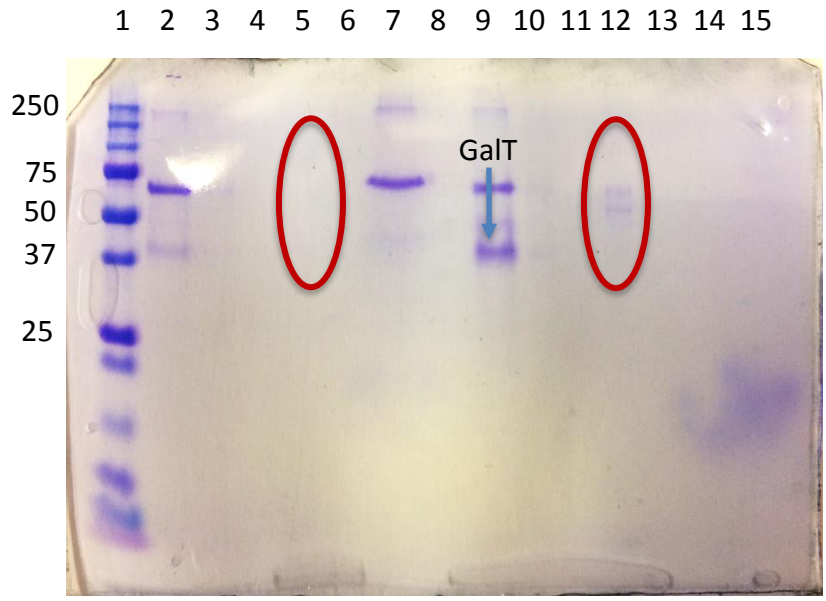


Figure 36 – Gel showing enrichment of β-1,4-GalT by UDP-Gal and UDP-(4F)-Gal. Lanes 1 = ladder, UDP-Gal resin lanes 2 to 5: 2 = wash 1, 3 = wash 2, 4 = wash 3, 5 = elution, lane 6 = blank, lane 7 = control, lane 8 = blank, UDP-(4F)-Gal resin lanes 9 to 11: 9 = wash 1, 10 = wash 2, 11 = wash 3, 12 = elution, 13 = blank, 14 = UDP-Gal resin, 15 = UDP-(4F)-Gal.

The elution lane of the UDP-(4F)-Gal resin, 12, contains the same bands that were seen in the experiment where only β-1,4-GalT was present, suggesting that the enzyme has selectively bound to the resin. The elution lane of the UDP-Gal resin, 5, does not contain any protein – perhaps the resin was blocked with non-specific proteins binding that washed off and the affinity was not high enough to bind β-1,4-GalT in the presence of other proteins. The strong band seen in W1 of both resins (2 and 9) and the control lane is VSG, which does not seem to bind strongly to the resin. Boiling the resins that were not exposed to proteins proved that they alone do not cause bands to appear (lanes 14 and 15).

This experiment was repeated omitting the controls of resin that hadn't been exposed to protein, figure 38. In parallel, the resins were also incubated with only VSG and BSA in order to assess the binding of these proteins in the absence of β-1,4-GalT, figure 37. The washes, elution and control samples of protein were run on an SDS gel and each band was extracted and tryptic digested for subsequent mass spectrometry analysis and compared against a database of proteins (figure 37 and 38).

The extracted bands for the condition without β-1,4-GalT are labelled in figure 37. In the control lane, 8, VSG is not very visible due to the poor de-staining. The two visible bands are from BSA. The proteins eluted from the UDP-Gal resin are in lane 6 (bands A3 and A4). In lane 14 are the proteins eluted from the UDP-(4F)-Gal resin (bands A9 and A10). The bands in lane 6 and 14 look remarkably similar to those eluted in the previous experiment where β-1,4-GalT was present figure 36, but by mass spectrometry only keratin was detected in these bands. They are likely to be BSA binding non-specifically to the resin. Perhaps the extraction was not sufficient to extract enough protein to detect as they were low in abundance, although their intensity is similar to or higher than bands A1, A2, A5, A6, A7 and A8 and VSG and BSA were detected in these lanes.

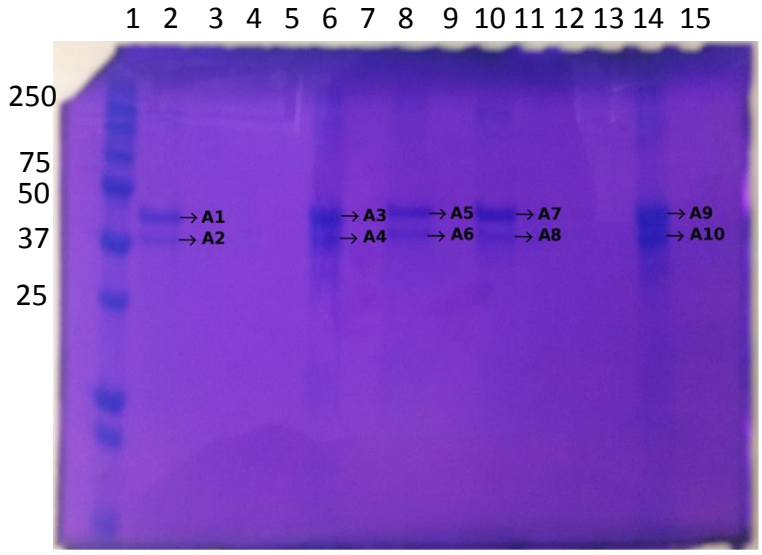


Figure 37 - Gel of UDP-Gal and UDP-(4F)-Gal incubated with VSG and BSA. Lanes 1 = ladder, UDP-Gal resin lanes 2 to 6: 2 = wash 1, 3 = wash 2, 4 = wash 3, 5 = blank, 6 = elution, lane 7 = blank, lane 8 = control, lane 9 = blank, UDP-(4F)-Gal resin lanes 10 to 14: 10 = wash 1, 11 = wash 2, 12 = wash 3, 13 = blank, 14 = elution, 15 = blank.

For the condition where β-1,4-GalT was present and for both UDP-Gal and UDP-(4F)-Gal resins, β-1,4-GalT was the major protein eluted from UDP-(4F)-Gal, figure 38. The only other protein present in the elution was keratin. For the UDP-Gal resin, β-1,4-GalT was present without a significant peptide count indicating less was bound to the UDP-Gal resin than in the UDP-(4F)-Gal experiment where it had a significant count. VSG was not present in any band and was most likely left on the gel as the de-staining was not good enough to see it and therefore could have been present in the elution. The resins were both present in excess of the 0.2 nmol of β-1,4-GalT used as they were of the order of 0.6 μmol. Therefore, the presence

of β-1,4-GalT in the first wash of each resin, lanes 2 and 10, could be due to the low affinity of the probe to the enzyme or the resin could be blocked by non-specific interactions. Some contaminants in the elution were detected in both experiments. However, since these proteins were not present in the assay, nor had high counts, we do not believe them to be true binders; they are contaminants added during sample preparation.

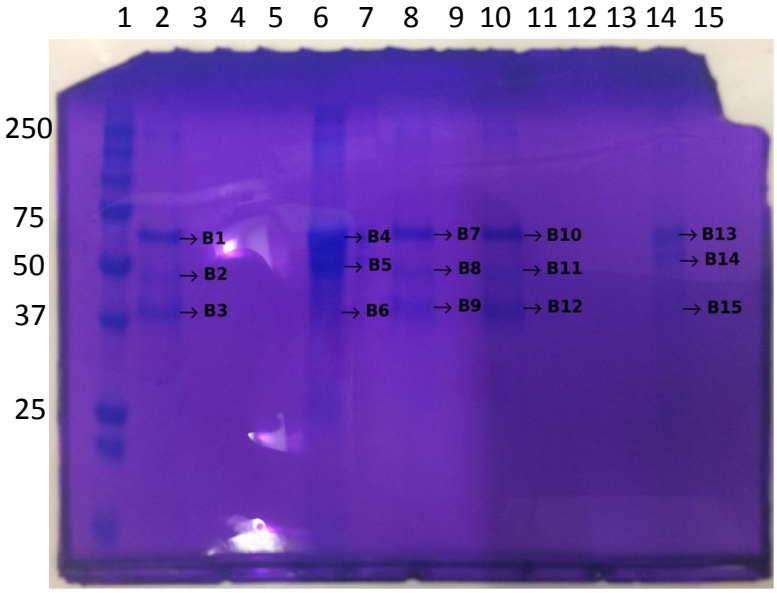


Figure 38 - Gel of UDP-Gal and UDP-(4F)-Gal incubated with β-1,4-GalT, VSG and BSA. Lanes 1 = ladder, UDP-Gal resin lanes 2 to 6: 2 = wash 1, 3 = wash 2, 4 = wash 3, 5 = blank, 6 = elution, lane 7 = blank, lane 8 = control, lane 9 = blank, UDP-(4F)-Gal resin lanes 10 to 14: 10 = wash 1, 11 = wash 2, 12 = wash 3, 13 = blank, 14 = elution, 15 = blank.

This experiment showed that the resin was able to bind to β-1,4-GalT and suggests that it is selective in binding to the enzyme over VSG and BSA. However, it isn't conclusive since VSG was not present in any extracted band – not even washes or the control. In order to improve this protocol, a protein should be added to each lane so that it can be used as a control for extraction. Furthermore, two gels should be run: one to visualise the proteins and one for extraction. The gel run for extraction should only be run enough so that the proteins only just enter the gel so they can all be extracted together as one band. This would avoid missing proteins from the extraction and differences between the extractions of individual bands. Non-specific protein binding to the resin is inevitable, but with careful experiment design these non-specific interactions can be mitigated, see **8.3.2** for description.

8.1.3.2 Limit of detection of β-1,4-galactosyltransferase

The UDP-(4F)-Gal resin performed better in the previous enrichment experiment and was therefore taken forward to decipher the limit at which it could no longer enrich for GalTs from

a mixture of proteins. It was subjected to the same enrichment conditions with β-1,4-galactosyltransferase (β-1,4-GalT) with a mixture of proteins. However, three more proteins were added to the mix to increase the complexity and six concentrations of β-1,4-GalT were trialled, table 1.

Table 2 – Amount of (β-1,4-GalT) added per condition

	Amount of β–1,4-GalT (ng)					
Condition	1	2	3	4	5	6
β-1,4-GalT	5500	550	55	5.5	0.55	0.055

The same mixtures of proteins plus β-1,4-GalT were incubated to be used as a control, without resin. The resins were then washed and the bound proteins were eluted by boiling the resin in laemmli buffer. The washes and elution were run on a gel just enough so that they entered it but did not separate. The samples were then sent to mass spectrometry. Comparing the ratio of the control proteins that were not subjected to the resin ('before") to the proteins eluted from the resin ("after"), it is possible to determine if β-1,4-GalT is enriched. Unfortunately, the mass spectrometer was unable to detect β-1,4-GalT in the control and eluted samples of conditions 4, 5 and 6 as they were too low in abundance. The mass spectrometer (4800 Plus MALDI-TOF/ TOF Analyzer, AB SCIEX, Framingham, MA, USA) is able to detect proteins in the 10-20 fmol range therefore it might have been too ambitious to try to detect below 5.5 ng of β-1,4-GalT (0.125 pmol) especially since some protein will be lost during the extraction process.

Normalised spectral abundance factor (NSAF) takes into account the length of the protein and the number of MS/MS spectra found in the mass spectrometer run. The length of the protein is taken into consideration because the longer the protein the more peptides will be produced therefore the more MS/MS spectra are found. In doing so, it can estimate the amount of protein in the sample compared to others present. Some bias has been observed for long proteins where the formula over compensates for their length and underestimates their abundance (Deng *et al.*, 2015). Most of the proteins used in this assay are around 450 amino acids in length, see table 2 for protein lengths. Ribonuclease B and Bovine Serum Albumin are of two extremes in length and their abundance should be evaluated with care. A table

summarising the NSAF counts of the proteins in each condition before and after enrichment is annex table 9.

Table 3 - Protein length of each protein used

Protein	Protein length
Bovine serum albumin	583
Variant surface glycoprotein	450
Ribonuclease B pancreatic	124
Fetuin	359
Green fluorescent protein	238
β-1,4-galactosyltransferase	402

For condition 1, 5.5 μg of β-1,4-GalT was used in a mixture of proteins where this represented 48% of the total μg of proteins, or 9% of the NSAF. After enrichment, β-1,4-GalT is clearly enriched to 23% NSAF of the proteins present in the elution, figure 39. Fetuin and BSA also seem to be enriched, although not to the same extent as β-1,4-GalT. Their binding can be attributed to these non-specific hydrophobic interactions with the backbone of the resin. Ribonuclease B is not present in the elution and the % presence of VSG and GFP are reduced in the enriched mixture. As Ribonuclease B is the shortest protein, the hydrophobic interactions it would form with the resin would be the weakest since its sequence contains less regions of hydrophobicity than the longer proteins. It is possible that the other proteins had a higher affinity for the resin than Ribonuclease B, blocking it from interacting.

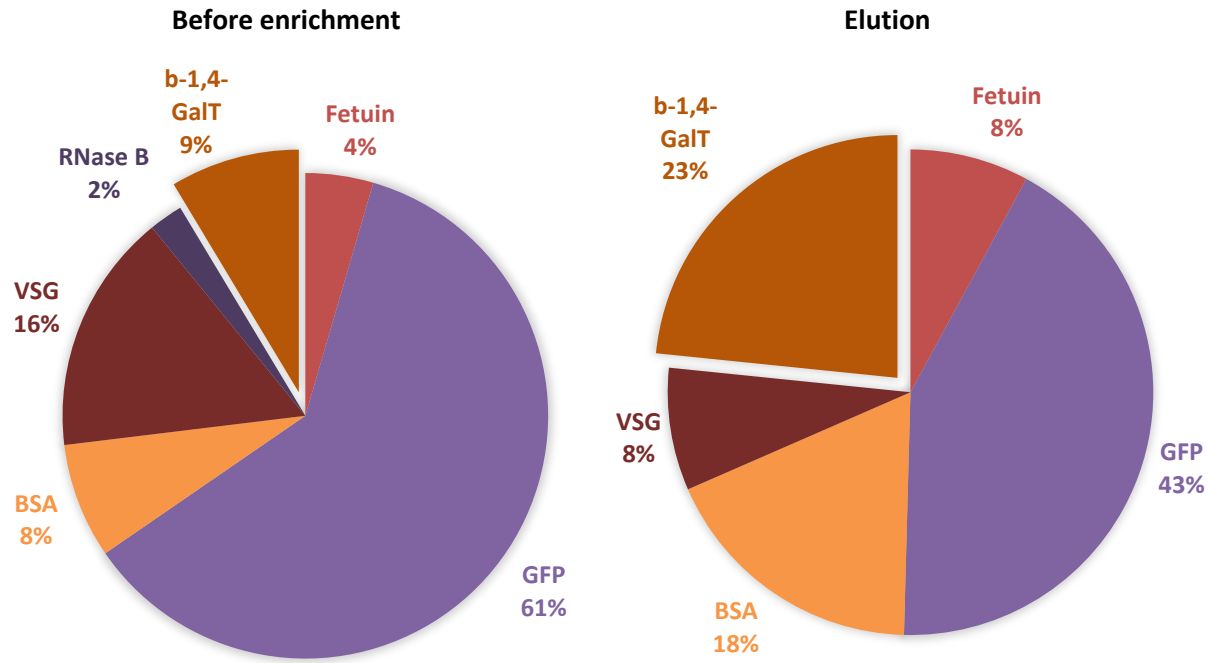


Figure 39 - Comparison of ratio of proteins present before and after enrichment of 5.5 μg of GalT

For condition 2, 0.5 μg of β-1,4-GalT was used in a mixture of proteins where this represented 8.5% of the total μg of proteins, or 13% of the NSAF. After enrichment, β-1,4-GalT is slightly less in proportion to the remaining proteins at 11% NSAF, figure 40. However, like in the previous experiment, no Ribonuclease B is present in the elution. This means that although the % β-1,4-GalT has reduced, in reality it has been enriched since there are less proteins in the elution. Fetuin and VSG are present in the elution at a higher % and BSA and GFP are reduced. It is hard to say which have been enriched since we were unable to directly quantify.

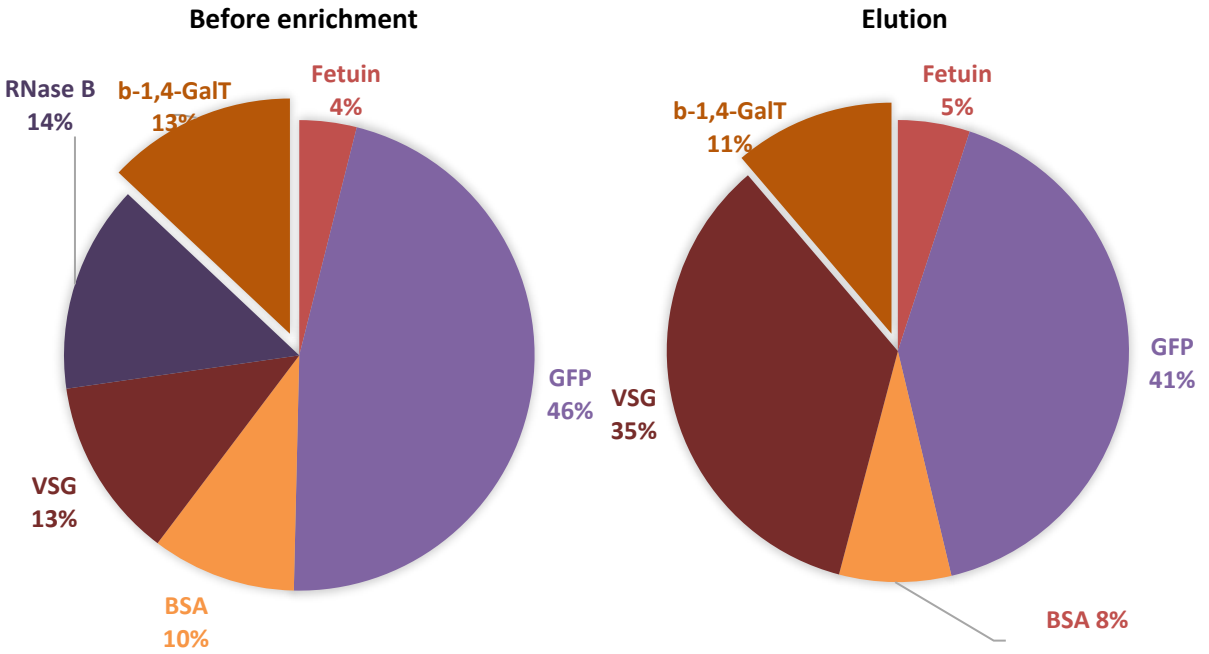


Figure 40 - Comparison of ratio of proteins present before and after enrichment of 0.55 μg of GalT

For condition 3, 55 ng of β-1,4-GalT was used in a mixture of proteins where this represented 0.9% of the total µg of proteins, or 2% of the NSAF. After enrichment, β-1,4-GalT is enriched to 5% NSAF, figure 41. Fetuin and GFP are also enriched but Ribonuclease B, VSG and BSA are reduced or remain the same. This time Ribonuclease B was not washed away. This could be because there was less β-1,4-GalT present so there was more space available for non-specific binding to the resin.

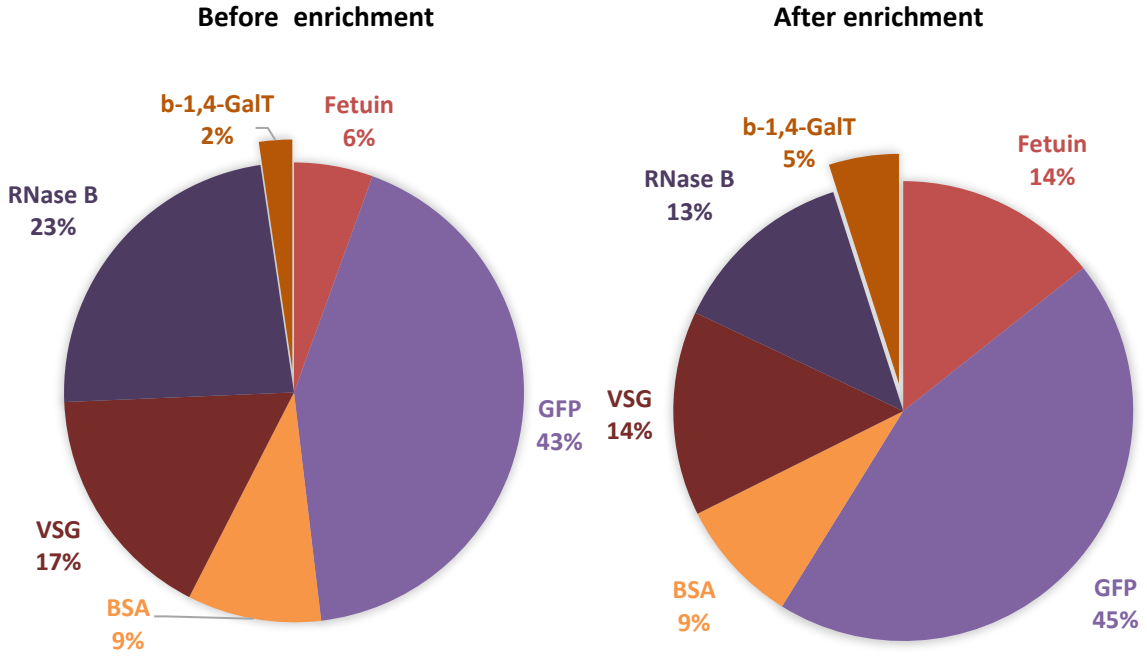


Figure 41 - Comparison of ratio of proteins present before and after enrichment of 55 ng of GalT

Although the resin clearly has an affinity for GalTs, it becomes more difficult to enrich with lower quantities of the β-1,4-GalT present compared to the other proteins. It is still enriched when representing 0.9% of total protein, but along with many other proteins that form non-specific interactions with the resin. The PEG resin will always attract non-specific binders as it has hydrophobic and hydrophilic regions. A better way to perform this experiment would be to use an internal standard that would be added to the mixture when loaded on to a gel so that the proteins could be quantified and the % recovery of β-1,4-GalT (and the other proteins) could be calculated. This standard would also account for loss of proteins during the extraction process and allow direct comparisons between lanes. The experiment also cannot distinguish between proteins that bind to the resin by interacting specifically with the UDP-(4F)-Gal or *via* non-specific interactions. Repeating the experiment in the same way but with an internal standard could help in discerning the proteins the bind non-specifically from those that are specific but the best way would be to have a system that discourages non-specific interactions.

This could be done by adding a protein that blocks the resin (to remove non-specific interactions) or by adding a small molecule that inhibits the target protein (a reduction in binding would indicate that it had been binding specifically).

8.2 Chapter summary

In this chapter, UDP-galactosyltransferase affinity based probes were synthesised based upon inhibitors of GalTs, with the aim of using these probes to enrich for galactosyltransferases which could enable their characterisation. These probes, UDP-Gal and UDP-(4F)-Gal, were attached to tentagel resin *via* an SMCC linker. This synthesis of resin bound nucleotide sugars is the first of its kind, as most sugars are bound to resins *via* the C1 and none are coupled to a nucleotide. Preliminary studies using a commercially available GalT indicated that the probes were potential affinity probes and that UDP-(4F)-Gal had a higher affinity.

8.3 Materials and methods

8.3.1 Synthesis

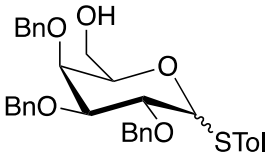
8.3.1.1 General information

Reactions were carried out under an inert atmosphere (argon), using standard techniques with oven-dried glassware. Anhydrous solvents were obtained by distillation and drying (DMF, DCM, Pyridine). All non-aqueous reactions were run using anhydrous solvents. Commercial reagents were used as supplied or purified by standard techniques where necessary. Na_2SO_4 was used as the organic phase drying reagent.

Flash column chromatography was performed using 230-400 mesh silica with the indicated solvent system according to standard techniques. Analytical thin-layer chromatography (TLC) was performed on precoated, foil-backed gel plates. Visualisation of the developed chromatogram was performed by UV absorbance (254nm). Staining of the developed chromatogram was carried out using a 5% phosphomolybdic acid in ethanol stain.

Standard nuclear magnetic resonance spectra were recorded on a Bruker AV II 400 (400 MHz) spectrometer. NMR spectra of resins were recorded on a Bruker AV II 500 (500 MHz) spectrometer equipped with a 4mm HRMAS Triple H/P/C probe-head. Chemical shifts for ^{1}H NMR spectra were recorded in $CDCl_3$ in parts per million from tetramethylsilane, the internal standard (δ = 0 ppm). For samples in DMSO-d6, MeOH-d4 and D_2O the solvent peak was the internal standard. Data were reported as follows: chemical shift [multiplicity (s = singlet, d = doublet, t = triplet, m = multiplet and br = broad), coupling constant in Hz, integration and assignment]. ^{13}C NMR and ^{31}P spectra were recorded with complete proton decoupling. Chemical shifts are reported in parts per million with the solvent resonance as the internal standard ($^{13}CDCl_3$: 77.0 ppm, MeOD: 49.2 ppm, DMSO: 39.5 ppm). NMR experiments used for the assignments of NMRs: APT, COSY and HMQC or HSQC.

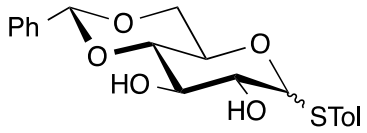
8.3.1.2 6-deoxy-2,3,4-tri-O-(phenylmethyl)-1-thio-α/β-D-Glucopyranoside (3)



Methyl galactosylpyranoside (5.0 g, 25.8 mmol) and BnBr (15.8 mL, 128.8 mmol) were dissolved in DMF (50 mL) and stirred at 0 °C. NaH (3.7 g, 154.5 mmol) was added portion-wise, the reaction allowed to warm to room temperature and stirred for 17 h. The reaction was quenched with MeOH, water added, extracted with Et_2O (3 x 50 mL), dried and concentrated *in vacuo*. Purification by flash chromatography (100% Hex to 7:3 Hex:EtOAc) afforded the product as an oil (14.7 g, **100%**) (Lourenço, Maycock and Rita Ventura, 2009). To this sugar, (4.7 g, 8.5 mmol), acetic acid (20.5 mL) and acetic anhydride (20.5 mL) were added and the reaction stirred at room temperature. Concentrated sulphuric acid (0.83 mL) was added drop wise and the reaction stirred for 26 h. The reaction was quenched with $NaHCO_3$ (60 mL) and water added until pH 7. It was extracted with EtOAc (3 x 50 mL), dried and concentrated *in vacuo.* Purification by flash column (Hex:EtOAc, 8:2) afforded the product as an oil (3.7 g, **81%**) (Lourenço, Maycock and Rita Ventura, 2009). This sugar, (19.3 g, 36 mmol) was dissolved in DCM and cooled to 0 °C. BF_3OEt_2 (10 mL, 54 mmol) and *p*-toluenethiol (9 g, 72 mmol) were added and the reaction stirred at 0 °C for 1 h. The reaction was quenched with $NaHCO_3$ (500 mL), extracted with DCM (3 x 900 mL), dried and concentrated *in vacuo*. Purification with flash chromatography afforded the product as an oil (18.6 g, **86%**). (18.6 g, 31 mmol) and NaOMe (1 g, 18.5 mmol) were stirred in MeOH (40 mL) for 1 h. The reaction was quenched with $NaHCO_3$ (100 mL), extracted with EtOAc (3 x 300 mL), dried and conc *in vacuo.* Purification by flash chromatography (100% Hex to 50:50 Hex:EtOAc) afforded product **3** as an oil (16.2 g, **94%**); α:β = 1:8.3; δH(400 MHz; $CDCl_3$, Me_4Si) 7.46-7.31 (18 H, m, Ar-H), 7.04 (2 H, d, *J* = 8.2 Hz, Ar-H), 4.97 (1 H, d, *J* = 11.7 Hz), 4.84 (1 H, d, *J* = 10.2 Hz, 1 x PhC*H*$_2$), 4.77-4.71 (4 H, m, 4 x PhC*H*$_2$), 4.64 (1 H, d, *J* = 11.7 Hz, 1 x PhC*H*$_2$), 4.59 (1 H, d, *J* = 9.7 Hz, C(1)H), 3.93 (1 H, t, *J* = 9.4 Hz, C(2)H), 3.86-3.81 (2 H, m, C(4)H + C(6)H), 3.60 (1 H, dd, *J* = 9.2, 2.7 Hz, C(3)H), 3.53-3.49 (1 H, m, C(6)H), 3.44-3.41 (1 H, m, C(5)H), 2.30 (3 H, s, Me); δC(101 MHz; $CDCl_3$, Me_4Si) 138.3 (Ar-Cq), 138.3 (Ar-Cq), 138.1 (Ar-Cq), 137.4 (Ar-Cq), 132.2 (Ar-C), 130.0 (Ar-Cq), 129.6 (Ar-C), 128.49 (Ar-C), 128.39 (Ar-C), 128.34 (Ar-C), 128.31 (Ar-C), 128.22 (Ar-C), 127.83 (Ar-C), 127.79 (Ar-C), 127.75 (Ar-C), 127.63 (Ar-C), 88.1 (C1), 84.3 (C3), 78.7 (C5), 77.6 (C2), 75.7 (Ph*C*H$_2$), 74.1

(Ph*C*H$_2$), 73.3 (C4), 73.1 (Ph*C*H$_2$), 62.3 (C6), 21.1 (*C*H$_3$); HR-MS: calcd for $C_{34}H_{36}O_5SNa^+$ [M+Na]$^+$: 579.2176 ; found: 579.2164; ν_{max}/cm^{-1} 3445.6 (br, OH).

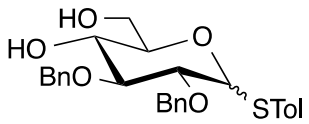
8.3.1.3 p-Tolyl 4,6-O-benzylidene-1-thio-α/β-D-Glucopyranoside (6)



Glucose (11.28 g, 62.6 mmol) was dissolved in pyridine (76.7 mL) and cooled to 0 °C. Ac_2O (60 mL) and DMAP (cat. amount) were added and the reaction stirred overnight. The reaction was quenched with $NaHCO_3$ (100 mL) and extracted with EtOAc (3 x 150 mL). The organic phase was dried and concentrated *in vacuo* afforded the product as a white powder (24.4 g, 100%) (Ermolenko and Sasaki, 2006). The sugar (490 mg, 1.25 mmol) was dissolved in DCM (10 mL) with *p*-toluenethiol (311 mg, 2.5 mmol). $BF_3.OEt_2$ (771 ul, 6.25 mmol) was added at 0°C and the reaction allowed to warm to rt before stirring for 2 d. The reaction was washed with $NaHCO_3$ (20 mL) and the aqueous phase extracted with DCM (3 x 10 mL) and the combined organic phases were dried and concentrated *in vacuo*. Purification by flash chromatography (8:2, Hexane: Ethyl acetate) gave the product as a yellow solid (445 mg, 78%) (Bruneau *et al.*, 2015; Chatterjee *et al.*, 2015). The sugar (135 mg, 0.30 mmol) and NaOMe (10 mg, 0.18 mmol) were stirred in MeOH (3 mL) for 1.5 h. The reaction was quenched with dowex, filtered and concentrated *in vacuo* to yield the product as a yellow solid (80 mg, 94%) (Bruneau *et al.*, 2015). The sugar (492 mg, 1.71 mmol), $PhCH(OMe)_2$ (520 mL, 3.44 mmol) and CSA (cat. amount) were dissolved in DMF (15 mL), the reaction mixture heated to 85 °C and stirred for 30 h. The reaction was then quenched with $NaHCO_3$ (200 mL) and the aqueous phase extracted with EtOAc (3 x 100 mL) before the combined organic layers were dried and concentrated *in vacuo.* Purification by flash chromatography (8:2, Hex:EtOAc to 100% EtOAc) afforded the product **6** as a white solid (450 mg, 70%); α:β = 1:0.5; δH(400 MHz; $CDCl_3$, Me_4Si) 7.55-7.37 (m, 11 H, Ar-H), 7.16 (dd, *J* = 7.5, 5.5 Hz, 3 H, Ar-H), 5.53 (t, *J* = 9.2 Hz, 2 H, αH1 + C*H*Ph), 4.57 (d, *J* = 9.7 Hz, 1 H, βH1), 4.40-4.26 (m, 2 H, αH5 + αH6), 4.01-3.89 (m, 1 H, αH2), 3.88-3.74 (m, 3 H, βH3 + βH4 + βH6), 3.58-3.39 (m, 4 H, αH4 + βH4 + βH2 + βH5), 3.21-3.07 (m, 1 H, O*H*), 2.89 (s, 2 H, O*H*), 2.37 (m, 5 H, α*CH*$_3$ + β*CH*$_3$); δC(101 MHz; $CDCl_3$, Me_4Si) 138.8 (Ar-C_q), 138.1 (Ar-C_q), 137.0 (Ar-C_q), 136.9 (Ar-C_q), 133.6 (Ar-C), 133.4 (Ar-C), 132.7 (Ar-C),

123.0 (Ar-C), 129.9 (Ar-C), 129.7 (Ar-C_q), 129.3 (Ar-C), 129.0 (Ar-C), 128.4 (Ar-C), 127.4 (Ar-C_q), 127.0 (Ar-C), 126.4 (Ar-C), 126.3 (Ar-C), 102.0 (α*C*HPh), 101.9 (β*C*HPh), 91.1 (αC1), 88.7 (βC1), 81.0, 80.2, 77.4, 77.0, 76.7, 74.5, 72.51 (βC2), 72.47, 72.1, 70.5, 68.7 (αC6), 68.6 (βC6), 63.8 (αC), 60.4, 21.2 (βCH_3), 21.1 (αCH_3); HR-MS: calcd for $C_{20}H_{22}O_5SNa^+$ $[M+Na]^+$: 397.1080; found: 397.1073; ν_{max}/cm^{-1} 3417.8 (br, OH).

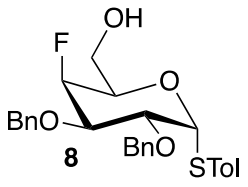
8.3.1.4 p-tolyl 2,3-di-O-benzyl-1-thio-α/β-D-Glucopyranoside (7)



Sugar **6** (450 mg, 1.19 mmol) and BnBr (360 uL, 2.99 mmol) were dissolved in DMF (10 mL) and stirred at 0 °C. NaH (85 mg, 3.57 mmol) was added portion-wise, the reaction allowed to warm to room temperature and stirred for 18 h. The reaction was quenched with MeOH (5 mL), water added (5 mL), extracted with EtOAc (3 x 15 mL), dried and concentrated *in vacuo*. Purification by preparative TLC (8:2 Hex:EtOAc) afforded the product as an oil (660 mg, 100%). The sugar (660 mg, 1.19 mmol) and *p*-TsOH (cat. amount) were dissolved in DCM:MeOH (3 mL: 15 mL) and the reaction stirred for 16 h. Another portion of *p*-TsOH was added and the reaction stirred for a further 5 h. The reaction was quenched with $NaHCO_3$ (10 mL), extracted with EtOAc (3 x 20 mL), dried and concentrated *in vacuo*. Purification by flash chromatography yielded product **7** as a white solid (520 mg, 94%); δH(400 MHz; $CDCl_3$, Me_4Si) 7.45 (4 H, d, *J* = 8.0 Hz, 4 x Ar-*H*), 7.37 (8 H, m, 8 x Ar-*H*), 7.15 (2 H, d, *J* = 7.8 Hz, 2 x Ar-*H*), 5.58 (1 H, d, *J* = 5.5 Hz, αC(1)H), 5.06 (d, *J* = 11.4 Hz), 4.99 (2 H, dd, *J* = 10.9, 5.3 Hz, 2 x CH_2Ph), 4.76 (2 H, t, *J* = 10.4 Hz, 2 x CH_2Ph), 4.69 (1 H, d, *J* = 9.4 Hz, βC(1)H), 4.26-4.22 (m, αH), 4.18-4.13 (m, αH), 3.89 (2 H, dd, *J* = 11.9, 3.2 Hz, 1 x βC(6)H + αC(2)H), 3.77 (1 H, dd, *J* = 11.7, 5.1 Hz, 1 x βC(6)H), 3.64-3.45 (3 H, m, βC(2)H + βC(3)H + βC(4)H), 3.35 (1 H, dt, *J* = 8.9, 4.4 Hz, βC(5)H), 2.93 (d, *J* = 24.5 Hz), 2.37 (3 H, s, Me); δC(101 MHz; $CDCl_3$, Me_4Si) 138.3 (βAr-C_q), 138.0 (βAr-C_q), 137.9 (βAr-C_q), 132.8 (αAr-*C*), 132.5 (βAr-*C*), 129.9 (αAr-*C*), 129.8 (βAr-*C*), 129.6 (βAr-C_q), 128.7 (βAr-*C*), 128.6 (αAr-*C*), 128.5 (αAr-*C*), 128.5 (βAr-*C*), 128.3 (βAr-*C*), 128.1 (αAr-*C*), 128.1 (βAr-*C*), 128.0 (αAr-*C*), 128.0 (βAr-*C*), 127.9 (αAr-*C*), 127.9 (βAr-*C*), 88.0 (βC1), 87.4 (αC1), 86.1 (βC3), 81.6 (αC), 80.9 (βC2), 79.6 (αC2), 79.1 (βC5), 75.4 (β*C*H_2Ph), 75.4 (β*C*H_2Ph), 72.3 (α*C*H_2Ph), 71.7

(αC), 70.4 (βC4), 70.1 (αC), 62.7 (βC6), 62.3 (αC6), 21.1 (*C*H_3); HR-MS: calcd for $C_{27}H_{30}O_5SNa^+$ $[M+Na]^+$: 489.1706; found: 489.1697; ν_{max}/cm^{-1} 3405.1 (br, OH).

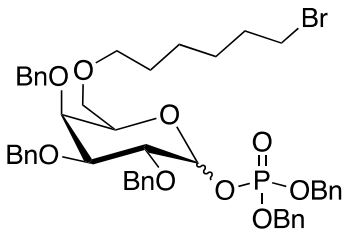
8.3.1.5 p-Tolyl 2,3-di-O-benzyl-4-deoxy-4-fluoro-1-thio-α-D-Glucopyranoside (8)



8

Sugar **7** (520 mg, 1.1 mmol), imidazole (230 mg, 3.3 mmol), and TBDMSCl (500 mg, 3.3 mmol) were dissolved in DMF (5 mL). The reaction was stirred at rt for 16 h. The reaction was quenched with water (2 mL), extracted with Et_2O (3 x 5 mL), dried and concentrated *in vacuo.* Purification by flash chromatography (100% hexane to Hex:EtOAc, 7:3) afforded the product as an oil (582 mg, 90%). This sugar (73 mg, 0.13 mmol) was dissolved in DCM (1 mL) and pyridine (0.4 mL) and cooled to -78 °C. Tf_2O (60 mL, 0.19 mmol) was added, the reaction was allowed to warm to room temperature and stirred for 1.5 h. The solvent was evaporated, DCM added (5 mL) and the organic phase washed with $NaHCO_3$ (5 mL) and brine (5 mL). The organic phase was the dried and concentrated *in vacuo*. The crude product was dissolved in THF (1 mL), TBAF was added (0.25 mL, 0.25 mmol) and the reaction was stirred for 2 h. THF was evaporated, DCM (5 mL) added and the organic phase washed with brine (5 mL). The organic phase was dried and concentrated *in vacuo* and the crude purified by preparative TLC (60:40, EtOAc:Hex). This afforded the product **8** as a yellow solid (30 mg, 48%); δH(400 MHz; $CDCl_3$, Me_4Si) 7.45 (4 H, t, *J* = 8.0 Hz, 4 x Ar-*H*), 7.39-7.31 (8 H, m, 8 x Ar-*H*), 7.13 (2 H, d, *J* = 8.1 Hz, 2 x Ar-*H*), 4.86-4.72 (5 H, m, 4 x CH_2Ph), 4.63-4.61 (1 H, m, C(1)H), 3.99-3.94 (1 H, m, 1 x C(6)H), 3.82-3.72 (2 H, m, C(2)H + 1 x C(6)H), 3.65-3.62 (1 H, m, C(3)H), 3.58-3.55 (1 H, m, C(4)H), 3.52-3.48 (1 H, m, C(5)H), 2.35 (3 H, s, Me); δC(101 MHz; $CDCl_3$, Me_4Si) 138.1 (Ar-C_q), 138.0 (Ar-C_q), 132.7 (Ar-C), 129.8 (Ar-C), 128.5 (Ar-C), 128.4 (Ar-C), 128.3 (Ar-C), 128.0 (Ar-C), 127.9 (Ar-C), 87.8, 87.0, 85.1, 81.1, 81.0 (C5), 80.9, 77.3 (C5), 77.2 (C2), 75.8 (*C*H_2Ph), 72.3, 61.48 (C6), 61.43, 21.1 (Ph-*C*H_3); δF (376 MHz; $CDCl_3$): -215.9 (dt, *J* = 51.58, 26.39); HR-MS: calcd for $C_{27}H_{29}FO_4SNa^+$ $[M+Na]^+$: 491.1663; found: 491.1659; ν_{max}/cm^{-1} 3445.8 (br, OH).

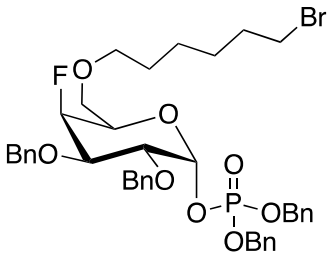
8.3.1.6 1-O-Dibenzylphosphate-2,3,4-tri-O-benzyl-6-O-hexyl-6-bromide-α/β-D-galactopyranoside (9)



Sugar **3** (8.86 g, 16 mmol) and NaH (0.77 g, 32 mmol) was stirred in DMF (5 mL) for 1 h. 1,6-di-Bromohexane (6.6 mL, 43 mmol) and TBAI (cat. amount) were added and the reaction stirred for 4 h. Another portion of NaH (0.38 g, 16 mmol) was added and the reaction stirred for a further 16 h. Water was added to the reaction, it was washed with $NaHCO_3$ (400 mL) and then extracted with EtOAc (3 x 300 mL). The organic phase was dried and concentrated *in vacuo* before purification by flash chromatography afforded the product as an oil, (7.84 g, 68%). This sugar (86 mg, 0.12 mmol) and dibenzyl hydrogen phosphate (37 mg, 0.13 mmol) were stirred in DCM (2 mL) with powdered molecular sieves for 30 min. The reaction was cooled and maintained at 0 °C. NIS (32 mg, 0.14 mmol) and TMSOTf (10 □l, 0.06 mmol) were added and the reaction stirred for 15 min before being quenched with $Na_2S_2O_3$ (10% soln, 1 mL). $NaHCO_3$ (2 mL) was added and extracted with DCM (3 x 2 mL). Purification by flash chromatography (EtOAc:Hex) afforded the product **9** as an oil (89 mg, 85%); α:β = 1:0.9; δH(400 MHz; $CDCl_3$, Me_4Si) 7.38-7.25 (25 H, m), 5.99 (1 H, dd, *J* = 6.6, 3.4 Hz, αC(1)H), 5.25 (1 H, t, *J* = 7.3 Hz, βC(1)H), 5.12-4.97 (10 H, m, C*H*$_2$Ph), 4.85-4.73 (9 H, m, C*H*$_2$Ph), 4.69-4.62 (3 H, m), 4.18-4.06 (3 H, m, αC(2)H), 3.99-3.89 (5 H, m, αC(3)H + αC(4)H + αC(6)H_2 + βC(2)H), 3.69-3.59 (2 H, m, αC(5)H + □C(3)H), 3.56-3.45 (4 H, m), 3.44-3.25 (10 H, m, 2 x C*H*$_2$Br), 1.87-1.79 (4 H, m, 2 x C*H*$_2$$CH_2$Br), 1.50-1.29 (15 H, m, 2 x OCH_2C*H*$_2$R + $O(CH_2)_3$C*H*$_2$R + 2 x $O(CH_2)_2$C*H*$_2$R); δC(101 MHz; $CDCl_3$, Me_4Si) 138.54 (ArC_q), 138.49 (ArC_q), 138.47 (ArC_q), 138.3 (ArC_q), 138.1 (ArC_q), 138.0 (ArC_q), 135.98 (ArC_q), 135.93 (ArC_q), 135.91 (ArC_q), 135.90 (ArC_q), 135.86 (ArC_q), 135.72 (ArC_q), 135.6 (ArC_q), 128.5 (Ar-C), 128.41 (Ar-C), 128.40 (Ar-C), 128.38 (Ar-C), 128.33 (Ar-C), 128.29 (Ar-C), 128.28 (Ar-C), 128.26 (Ar-C), 128.23 (Ar-C), 128.11 (Ar-C), 128.05 (Ar-C) 128.0 (Ar-C), 127.8 (Ar-C), 127.8 (Ar-C), 127.74 (Ar-C), 127.70 (Ar-C), 127.66 (Ar-C), 127.58 (Ar-C), 127.57 (Ar-C), 127.51 (Ar-C), 127.49 (Ar-C), 127.4 (Ar-C), 99.3 (d, *J* = 6.4 Hz, αC1), 96.6 (d, *J* = 8.1 Hz, βC1), 82.0 (αC5), 82.0 (d, *J* = 2.1 Hz, βC3), 79.2, 79.1, 78.7, 78.1, 75.82, 75.75, 75.2, 74.9, 74.70, 74.67, 74.61, 74.1, 73.6, 73.3, 73.04, 72.95, 72.90, 71.4, 71.3, 69.5, 69.23, 69.20, 69.18, 69.15,

69.00, 68.95, 68.87, 68.6, 33.8 (*C*H_2Br), 32.6 (*C*$H_2$$CH_2$Br), 29.4 (OCH$_2$*C*$H_2$R), 27.9 (O(CH$_2$)$_3$*C*$H_2$R), 25.3 (O(CH$_2$)$_2$*C*$H_2$R); δP(162 MHz; $CDCl_3$) -2.14 (α), -2.23 (β); HR-MS: calcd for $C_{47}H_{54}BrO_9PNa^+$ $[M+Na]^+$: 895.2581; found: 895.2580; ν_{max}/cm^{-1} 946.1 (s, P-OBn) and 1167.7 (s, P=O).

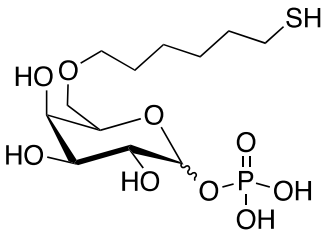
8.3.1.7 1-O-Dibenzylphosphate-2,3-di-O-benzyl-4-deoxy-4-fluoro-6-O-(hexyl-6-bromide)-α-D-Glucopyranoside (10)



Sugar **8** (107 mg, 0.23 mmol) and NaH (11 mg, 0.46 mmol) were stirred in DMF (2 mL) for 30 min. 1,6-dibromohexane (240 μl, 0.62 mmol) and TBAI (cat. amount) were added and the reaction stirred for 3 h. Further NaH was added (5 mg, 023 mmol) and the reaction stirred for a further 3 h. Water was added to the reaction, it was washed with $NaHCO_3$ (5 mL) and then extracted with EtOAc (3 x 5 mL). The organic phase was dried and concentrated *in vacuo* before purification by flash chromatography afforded the product as a white solid (115 mg, 80%). This sugar (98 mg, 0.16 mmol) with dibenzyl phosphate (48 mg, 0.17 mmol) was stirred in DCM (2 mL) with molecular sieves for 30 min. The reaction was cooled to 0 °C, NIS (42 mg, 0.19 mmol) and TMSOTf (13 μL, 0.08 mmol) were added and it was stirred for 30 min before being quenched with $Na_2S_2O_3$ (10% soln, 1 mL). $NaHCO_3$ was added (2 mL) and extracted with DCM (3 x 3 mL). Purification by flash chromatography (EtOAc:Hex) afforded product **10** as a white solid (116 mg, 95%); δH(400 MHz; $CDCl_3$, Me_4Si) 7.43-7.23 (6 H, m, Ar(H)), 5.95 (1 H, td, *J* = 3.4, 2.9 Hz, αC(1)H), 5.08 (dd, *J* = 7.6, 4.1 Hz), 5.03 (1 H, d, *J* = 7.4 Hz), 4.93-4.80 (1 H, m, C(4)H), 4.78 (1 H, d, *J* = 4.7 Hz, C(1)H), 4.11-4.07 (m), 4.03-3.97 (1 H, m, C(2)H), 3.89-3.79 (1 H, m, C(3)H), 3.60-3.56 (m), 3.46-3.44 (m), 3.43-3.37 (2 H, m, C*H*$_2$Br), 1.83 (2 H, quintet, *J* = 7.2 Hz, *C*$H_2$$CH_2$Br), 1.55-1.48 (1 H, m), 1.45-1.38 (2 H, m, C*H*$_2$(CH$_2$)$_2$Br), 1.35-1.28 (2 H, m, C*H*$_2$(CH$_2$)$_3$Br); δC(101 MHz; $CDCl_3$, Me_4Si) 137.8, 137.7, 135.8, 135.7, 128.5, 128.44, 128.38, 128.36, 128.1, 127.84, 127.77, 127.75, 96.1 (d, *J* = 184.1 Hz, αC1), 86.8 (d, *J* = 6.2 Hz, αC4), 75.1 (C2), 75.08, 75.04, 75.01, 74.97, 74.94, 73.6, 72.5, 71.5, 70.05, 69.86, 69.3 (d, *J* = 4.5 Hz, PO*C*H_2Ph), 69.1 (d, *J* = 4.2 Hz, PO*C*H_2Ph), 67.8 (d, *J* = 5.4 Hz, C6), 33.8 (*C*H_2Br), 32.6 (*C*$H_2$$CH_2$Br),

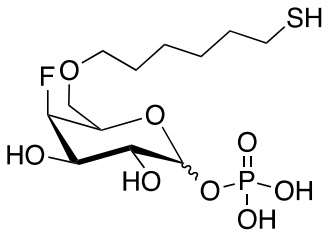
29.3, 27.9, 25.2 $CH_2(CH_2)_3Br$; δP (162 MHz; $CDCl_3$) -2.17; δF (376 MHz; $CDCl_3$) 42.0 (dt, J = 52.8, 24.5); HR-MS: calcd for $C_{40}H_{47}BrFO_8PNa^+$ $[M+H]^+$: 785.2249; found: 785.2252.

8.3.1.8 1-O-Diphosphate-2,3,4-trihydroxy-6-O-hexyl-6-thiol-α/β-D-galactopyranoside (11)



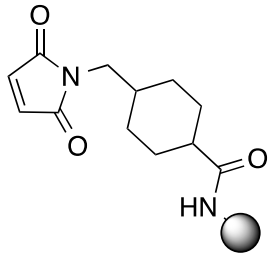
Sugar **9** (914 mg, 1.046 mmol) was dissolved in EtOH (5 mL) and added to a reactor containing Pd/C (550 mg, 0.52 mmol). The reactor was shaken under a pressure of 50 psi and an atmosphere of H_2 for 30 min. The reaction mixture was filtered through celite and washed with H_2O, MeOH and DCM. The solvents were evaporated and the residue dissolved in water (10 mL). The aqueous phase was washed with Et_2O three times (20 mL) and evaporated leaving the product as a solid (400 mg, **88%**). This sugar (347 mg, 0.81 mmol) was dissolved in DMF (5 mL) and potassium thioacetate (75 mg, 0.98 mmol) added. The reaction was stirred overnight before evaporating the DMF, adding water (5 mL) and washing with EtOAc (3 x 5 mL). The aqueous phase was dried and re-dissolved in water (3 mL) and K_2CO_3 added (329 mg, 2.37 mmol). The reaction was stirred overnight, the water was evaporated and the product purified by sephadex column (water) affording the product **11** as a solid (303 mg, quant.); α:β = 1:0.2; δH (400 MHz; D_2O): 5.41 (1 H, dd, *J* = 7.8, 3.6 Hz, C(1)H), 4.20-4.17 (1 H, m, C(5)H), 3.92-3.90 (1 H, m, C(4)H), 3.84 (1 H, dd, *J* = 10.3, 3.3 Hz, C(3)H), 3.70-3.67 (1 H, m, C(2)H), 3.59 (2 H, m, C(6)H), 3.52-3.47 (2 H, m, OCH_2CH_2), 2.47 (2 H, t, *J* = 7.2 Hz, CH_2SH), 1.54-1.49 (4 H, m, alkyl CH_2), 1.35-1.22 (4 H, m, alkyl CH_2); δC (101 MHz; D_2O): 99.0 (d, J = 5.2 Hz, βC1), 93.8 (d, J = 5.2 Hz, αC1), 71.6 (OCH_2), 69.8 (OCH_2), 69.6 (C5), 69.5 (C6), 69.4 (C4), 69.0 (C3), 69.0 (d, J = 5.6 Hz, C2), 33.2 (CH_2), 28.2 (CH_2), 27.3 (CH_2), 24.6 (CH_2), 23.8 (CH_2SH); ^{31}P NMR (162 MHz;): δ 2.70 (β), 2.66 (α); ν_{max}/cm^{-1} 3176.9 (br, OH), 952.7 (s, P-OH) and 1167.7 (s, P=O).

8.3.1.9 1-O-Di-phosphate-4-deoxy-4-fluoro-6-O-(hexyl-6-thiol)-α/β-D-galactopyranoside (12)



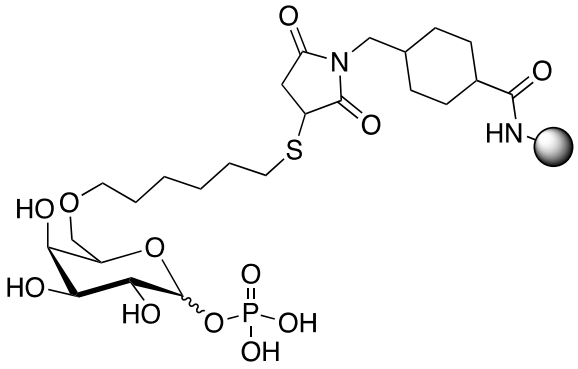
Sugar **10** (168 mg, 0.21 mg) was dissolved in EtOH (5 mL) and added to a reactor with Pd/C (70 mg, 0.07 mmol). The reactor was shaken at 50 psi under a H_2 atmosphere for 1 h before filtering through celite. The celite was washed with H_2O, MeOH and DCM. The solvents were evaporated and the residue dissolved in water (20 mL). The aqueous phase was washed with Et_2O (3 x 10 mL) and evaporated leaving the product as a solid (79 mg, **98%**). This sugar (110 mg, 0.29 mmol) was dissolved in DMF and potassium thioacetate (80 mg, 0.56 mmol) added. The reaction was stirred for 40 min before evaporating the DMF, adding water and washing with DCM. The aqueous phase was evaporated, re-dissolved in water (1 mL) and K_2CO_3 added (120 mg, 0.87 mmol). The reaction was stirred for 5 h, the water was evaporated and the product purified by sephadex column affording the product **12** as a solid (106 mg, quant.); α:β = 1:0.9; δH(400 MHz; D_2O) 5.45 (1 H, dd, *J* = 7.7, 3.4 Hz, αC(1)H), 4.84 (1 H, dd, *J* = 31.0, 2.8 Hz, αC(4)H), 4.87-4.75 (1 H, m, βC(1)H + αC(4)H), 4.27 (1 H, dt, *J* = 31.5, 6.3 Hz, βC(4)H), 4.01-3.89 (2 H, m, αC(3)H), 3.89-3.85 (1 H, m), 3.79 (1 H, dd, *J* = 10.1, 2.7 Hz), 3.75-3.74 (1 H, m), 3.72 (1 H, d, *J* = 2.7 Hz, αC(2)H), 3.70-3.63 (5 H, m), 3.52 (1 H, m, βC(2)H), 3.50 (5 H, m), 2.48 (2 H, t, *J* = 7.2 Hz, 2 x C*H*$_2$SH), 2.41-2.29 (4 H, m), 2.18-2.11 (3 H, m), 1.54-1.51 (9 H, m, CH_2), 1.29 (10 H, m, CH_2); δC(101 MHz; D_2O) 97.2 (d, *J* = 4.5 Hz, βC1), 93.8 (d, *J* = 5.1 Hz, αC1), 91.7, 90.7, 89.9, 88.9, 72.2, 71.99, 71.94, 71.62, 71.58, 71.3, 71.1, 69.08, 69.02, 68.99, 68.95, 68.55, 68.4, 68.35, 68.32, 68.29, 68.1 (d, *J* = 3.0 Hz, βC4), 32.9 (*C*H_2), 29.80 (*C*H_2), 29.77 (*C*H_2), 28.2 (*C*H_2), 27.17 (*C*H_2), 27.16 (*C*H_2), 26.3 (*C*H_2), 25.8 (*C*H_2), 24.5 (*C*H_2), 23.7 (*C*H_2SH); δF(376 MHz; D_2O) -216.5 (β), -219.2 (α); δP(162 MHz; D_2O) 2.46(β), 2.38 (α); ν_{max}/cm^{-1} 3250.5 (br, OH), 1099.1 (str, br, P=O), 975.5 (str, P-OH).

8.3.1.10 Resin bound SMCC derivative (15)



SMCC was synthesised according to a literature procedure. Observed data (**^{1}H NMR** and **^{13}C NMR**) is consistent with that reported (Leonard and Brunckova, 2011). SMCC (175 mg, 0.55 mmol) was dissolved in DMF (2 mL) and tentagel resin (813.8 mg, 0.24 mmol) was stirred in DMF (10 mL). The two were combined, PBS (2 mL of stock solution: 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 in 1 L H_2O) added, and the mixture stirred for 16 h. The resin was washed with DMF, water, MeOH then DCM and dried leaving a white resin **15** (872.8 mg, 76% loading); δH(500 MHz; $CDCl_3$) 11.62-11.58 (1 H, m, N*H*), 6.66 (br, C*H*CONR), 3.76-3.73 (br), 3.38-3.31 (br), 3.16-3.14 (br), 2.86-2.74 (br), 2.05-2.01 (br), 1.89-1.77 (br), 1.80-1.66 (br), 1.46-1.36 (br), 0.98-0.94 (br); δC(126 MHz; $CDCl_3$) 134.0, 128.1, 70.5, 70.1, 69.9, 45.6, 44.9, 43.7, 43.6, 43.5, 39.5, 39.0, 38.9, 36.9, 36.9, 36.8, 36.4, 36.3, 29.84, 29.76, 29.7, 29.6, 29.5, 28.8, 28.73, 28.68, 28.67, 28.6.

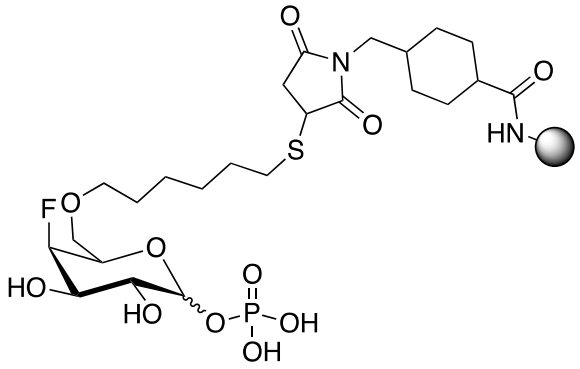
8.3.1.11 Resin bound 1-O-Diphosphate-2,3,4-trihydroxy-6-O-hexyl-6-thiol-α/β-D-galactopyranoside (16)



Sugar **11** (297 mg, 0.79 mmol) was dissolved in water (10 mL) and added to resin **15** (766.5 mg, 0.16 mmol) already swelled in DMF (5 mL). The pH was adjusted to pH 8 using K_2HPO_4 10% soln. and the reaction was stirred for 16 h. The resin was washed with DMF, water, MeOH and DCM then dried leaving a pink resin **16** (809.8 mg, 71%); δH(500 MHz; MeOD) 7.13-6.93 (br), 5.63 (br, sugar), 4.27-4.24 (br, sugar), 3.95-3.84 (br, sugar), 3.79-3.77 (br, sugar), 3.64 (s),

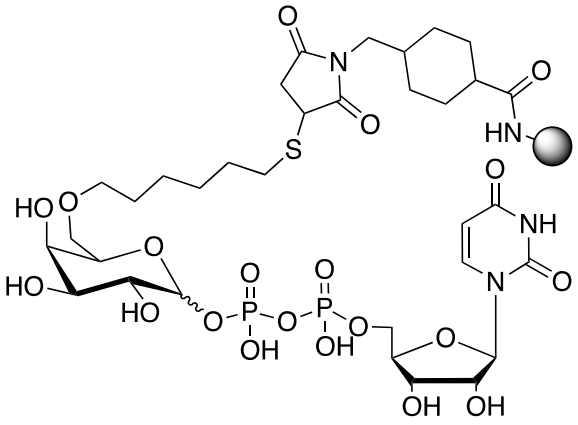
3.54-3.51 (m), 3.36-3.33 (m), 2.90-2.87 (m), 2.76-2.74 (m), 2.50 (br, sugar), 2.46 (br, sugar), 2.17-2.17 (br, sugar), 1.86-1.84 (m), 1.78-1.76 (m), 1.73-1.63 (m), 1.60-1.59 (m), 1.46-1.42 (m), 1.26 (s), 1.06-1.04 (m); δC (126 MHz; MeOD) 177.5, 166.6, 136.5 (SMCC), 128.6 (SMCC), 102.07 (sugar), 102.05 (sugar), 78.5 (SMCC), 78.3 (SMCC), 78.0 (SMCC), 70.0 (SMCC), 69.4 (sugar), 48.33, 48.15, 47.99, 47.82, 47.66, 47.65, 47.49, 47.32, 45.35, 45.27, 44.82 (SMCC), 44.70, 38.9 (SMCC), 37.08, 36.91, 29.76 (SMCC), 29.63 (SMCC), 28.96 (SMCC), 28.76 (SMCC); δP(202 MHz; MeOD) 1.5 (br).

8.3.1.12 Resin bound 1-O-Di-phosphate-4-deoxy-4-fluoro-6-O-(hexyl-6-thiol)-α/β-D-galactopyranoside (17)



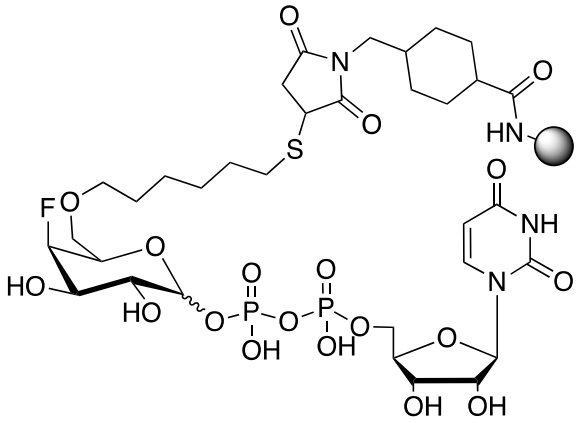
Sugar **12** (151 mg, 0.40 mmol) was dissolved in water (1 mL) and added to resin **15** (798.9 mg, 0.17 mmol) already swelled in DMF (5 mL). The pH was adjusted to pH 8 using KH_2PO_4 10% soln. and the reaction was stirred for 17 h. The resin was washed with DMF, water, MeOH and DCM then dried leaving an orange resin **17** (850.4 mg, loading 78%); δH(500 MHz; MeOD) 7.13-6.99 (br), 6.65-6.41 (br), 5.67-5.65 (br), 5.01-4.99 (br), 4.84 (br), 4.63-4.56 (br), 4.38-4.31 (br), 3.82-3.77 (br), 3.70 (br), 3.63-3.55 (br), 3.54-3.54 (br), 3.37-3.33 (br), 2.92-2.89 (br), 2.78-2.75 (br), 2.54-2.48 (br), 2.24-2.14 (br), 1.89-1.58 (br), 1.47-1.43 (br), 1.08-1.03 (br); δC(126 MHz; MeOD): δ 102.1, 69.9, 69.9, 69.3, 69.3, 67.9, 48.3, 48.3, 48.3, 48.1, 48.1, 47.9, 47.9, 47.7, 47.7, 47.6, 47.6, 47.6, 47.5, 47.4, 47.3, 47.2, 46.5, 44.6, 44.6, 38.9, 38.9, 38.9, 35.9, 29.7, 28.7, 25.4; δ δP(202 MHz; MeOD) 1.2.

8.3.1.13 Resin bound UDP-Galactose (18)



Trioctylamine (157 uL, 0.361 mmol) was added to resin **16** (478.7 mg, 0.072 mmol) in pyridine (2 mL) and the pyridine evaporated. It was re-dissolved in dry pyridine (2 mL) and again the pyridine evaporated twice more. UMP morpholidate was synthesised according to literature procedures and observed data (**^{1}H NMR** and **^{13}C NMR**) is consistent with that reported (Moffatt and Khorana, 1961). UMP morpholidate (114 mg, 0.289 mmol) and tetrazole (20 mg, 0.289 mmol) were co-evaporated with dry pyridine three times before being added to the resin mixture in pyridine (5 mL). The mixture was shaken for 6 d. The resin was then washed with DMF, H_2O, pyr., MeOH and DCM and dried. The process was repeated exactly and the mixture shaken for a further 6 d before the resin was washed with DMF, H_2O, pyr., MeOH and DCM affording a yellow resin **18** (577.4 mg); δH (500 MHz; MeOD) 8.10-8.08 (br), 7.83-7.80 (br), 7.07-7.02 (br), 6.55-6.44 (br), 6.32-6.29 (br), 6.00-5.96 (br), 5.94-5.89 (br), 5.76-5.72 (br), 5.61-5.58 (br), 5.18 (br), 4.84 (br), 4.64-4.43 (br), 4.44-4.35 (br), 4.35-4.20 (br), 4.16-4.13 (br), 4.11-4.07 (br), 3.96-3.96 (br), 3.93-3.89 (br), 3.85-3.84 (br), 3.82-3.76 (br), 3.64 (br), 3.57-3.47 (br), 3.36-3.33 (br), 3.24 (br), 3.17-3.15 (br), 3.12-3.01 (br), 2.91-2.87 (br), 2.79-2.75 (br), 2.52-2.47 (br), 2.22-2.15 (br), 1.98-1.94 (br), 1.90-1.80 (br), 1.80-1.68 (br), 1.62-1.58 (br), 1.40-1.34 (br), 1.25-1.22 (br), 1.06-1.03 (br), 0.93-0.92 (br); δC(126 MHz; MeOD) 102.2, 102.0, 101.9, 69.3, 65.8, 54.6, 47.7, 44.5, 38.8, 33.1, 31.5, 30.9, 29.6, 28.7, 24.8; δP(202 MHz; MeOD) -0.5 (monophosphate), -11.2 (diphosphate), -13.2 (diphosphate).

8.3.1.14 Resin bound 4-deoxy-4-fluoro-UDP galactose (19)



Trioctylamine (0.2 mL, 0.45 mmol) was added to resin **17** (562.0 mg, 0.09 mmol) in pyridine (5 mL) and the pyridine evaporated. It was re-dissolved in dry pyridine (5 mL) and again the pyridine evaporated twice more. UMP morpholidate (141 mg, 0.36 mmol) and tetrazole (25 mg, 0.36 mmol) were co-evaporated with dry pyridine three times before being added to the resin mixture in pyridine (5 mL). The mixture was shaken for 7 d. The resin was then washed with DMF, H_2O, pyr., MeOH and DCM and dried. The process was repeated exactly and the mixture shaken for a further 6 d before the resin was washed with DMF, H_2O, pyr., MeOH and DCM affording a brown resin **19** (537.3 mg); δH (500 MHz; MeOD): 7.13-6.99 (br), 6.65-6.41 (br), 5.67-5.65 (br), 5.01-4.99 (br), 4.84 (br), 4.63-4.56 (br), 4.38-4.31 (br), 3.82-3.77 (br), 3.70 (br),3.63-3.55 (br), 3.54-3.54 (br), 3.37-3.33 (br), 2.92-2.89 (br), 2.78-2.75 (br), 2.54-2.48 (br), 2.24-2.14 (br), 1.89-1.58 (br), 1.47-1.43 (br), 1.08-1.03 (br); δP NMR (202 MHz; MeOD): -11.2, -13.4.

8.3.1.15 Enrichment of β-1,4-galactosyltransferase with UDP-Gal based affinity probes.

UDP-Gal resin or UDP-(4F)Gal resin were incubated at 37 °C for 3 h with 30 ul of solution A (4 ul of VSG2 (1.5 ug/ul), 2 ul of BSA (2.5 ul/ug), 4.8 ul of MnCl2 (100 mM), 36 ul of buffer (50 mM HEPES, 125 mM NaCl, 10 mM CaCl2, 5% NP-40) and 43.2 ul of water) or solution B (9.6 uL of β(1,4)GalT (0.85 ug/ul), 4 ul of VSG2* (1.5 ug/ul), 2 ul of BSA (2.5 ul/ug), 4.8 ul of MnCl2 (100 mM), 36 ul of buffer (50 mM HEPES, 125 mM NaCl, 10 mM CaCl2, 5% NP-40) and 33.6 ul of water) As a control, 30 ul of solution A and B were also incubated alone at 37 °C for 3 hours. Then the resins were spun for 1 min at 13.4 rpm and the filtrate kept (W1). The collection Eppendorf was replaced with a new one and water (30 ul) was added to each resin, and the samples were spun again for 1 min at 13.4 rpm and the filtrate kept (W2). The collection

Eppendorf was replaced with a new one and water (500 ul) was added to each resin, and the samples were spun again for 1 min at 13.4 rpm and the filtrate kept (W3). W3 was spun on the speed vac until dry, then 30 ul of water and 30 ul of 2 x laemmli buffer was added. To W1, W2 and the resins, 30 ul of water and 30 ul of 2 x laemmli buffer was added. The elution corresponds to the proteins eluted when the resin is boiled in laemmli. All samples were boiled for 8 mins at 90 °C then loaded onto 15 ul well SDS page gels and run in TGS buffer. They were stained in coomassie, the bands were cut, in-gel trypsin digestion was performed and the samples sent to mass spectrometry.

		Resin used (mg)	
	µmol/ mg	**Condition A**	**Condition B**
UDP-Gal	0.205	2.4	2.3/ 0.5 µmol
UDP-(4F) Gal	0.167	2.2	2.1 /0.36 µmol

Table 4 - Table to show mg of each resin used in each condition. *VSG2 purified from *T. brucei*

8.3.2 Protocol for MALDI-TOF/TOF and Protein Identification

Samples were centrifuged and the supernatant removed to a new eppendorf for each sample. Each peptide samples were concentrated in a speedvac (Eppendorf) and resuspended in trifuoroacetic acid 0.1% (v/v). Samples were concentrated and cleaned according to the manufacturer's instructions on micro C18 column (ZipTip, Millipore, Bedford, MA, USA). The proteins were eluted in 1.5 µL directly onto the MALDI plate using the matrix α-cyano-4-hydroxycinnamic acid (α-CHCA) at 7 mg/mL prepared in ACN (50/50% v/v), trifuoroacetic acid (0.1%) and ammonium phosphate (6 mM).

Peptide mass spectrometry analyses were performed using a MALDI-TOF/TOF (4800 Plus MALDI-TOF/ TOF Analyzer, AB SCIEX, Framingham, MA, USA) in reflector positive mode (700–5000 Da). The experimental mass spectra were searched against the Uniprot protein sequence database with the MASCOT (Matrix Science, London, UK) algorithm, integrated in the GPS Explorer software (AB SCIEX, Framingham, MA, USA). The search parameters used were up to two maximum trypsin missed cleavages, mass tolerance of 50 ppm, cysteine

carbamidomethylation (fixed modification), methionine oxidation (as variable modification) and a charge state of +1.

8.3.3 Limit of detection assay

24 µL of buffer (50 mM HEPES, 125 mM NaCl, 10 mM $CaCl_2$, 5 mM NP-40) was added to 12.3 µL water, 3.2 µL of 100 mM $MnCl_2$, 1.7 µL Ribonuclease B (1 µg/ µL), 0.7 µL of GFP (2.5 µg/ µL), 2 µL of fetuin (1 µg/ µL), 1.3 µL BSA (2.5 µg/ µL) and 2 µL of VSG (1.5 µg/ µL) to make the **reaction buffer**. β–1,4-GalT (0.85 µg/ µL) in water was added according to the below table 4 by serial dilutions so that the total volume equalled 12.8 µL and 6.4 µL of each dilution was added to the UDP-(4F)-Gal resin (0.167 mmol/g) to form conditions 1 to 6. 23.6 µL of reaction buffer was added to each UDP-(4F)-Gal resin and 23.6 µL was added to 6.4 µL of the corresponding β–1,4-GalT dilution for a control. All solutions (1-6 with resin and 1-6 control) were then incubated at 37 °C for 3 h.

Table 5 - Table of conditions

	mg of resin / µL of β–1,4-GalT					
Condition	1	2	3	4	5	6
UDP(4F)Gal	3.68	3.84	3.39	3.94	4.11	4.03
β-1,4-GalT	12.8	1.28	0.128	0.0128	0.00128	0.000128
% β-1,4-GalT of total protein	48.19	8.46	0.85	0.1	0.00	0.00

After 3 h, the parafilm was removed from the samples (excluding controls) and they were spun for 1 min at 13.4 rpm and the filtrate kept – this is W1. The collection Eppendorf was replaced with a new one. Water (30 µL) was added to each resin, and the samples were spun again (1 min at 13.4 rpm) – W2. The collection Eppendorf was replaced with a new one. Water (500 µL) was added to the resins and the samples were spun again (1 min at 13.4 rpm) – W3. The collection Eppendorf was replaced with a new one. W3 was spun on the speed vac for a few hours until dry, then 30 µL of water was added. To the resins, parafilm was added to the column and 15 µL of water and 15 µL laemmli buffer. To W1, W2, W3 and the controls: 20 µL of laemmli was added. All samples were boiled for 8 mins then loaded onto two 15 x 15 µL well gels (7 µL ladder used). The gels were run for 20 mins at 50 V and for 15 min at 110 V –

just enough so that the proteins entered the gel but did not separate. The bands were cut and digested according to the Boisvert lab protocol and sent for mass spectrometry.

8.3.4 LC–MS/MS

All samples were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 µm trap (180 µm × 20 mm Waters) and a nanoAcquity HSS T3 1.8 µm C18 capillary column (75 µm × 250 mm, Waters). The trap wash solvent was 0.1% (v/v) aqueous formic acid, and the trapping flow rate was 10 µL/min. The trap was washed for 5 min before switching flow to the capillary column. The separation used a gradient elution of two solvents (solvent A: 0.1% (v/v) formic acid; solvent B: acetonitrile containing 0.1% (v/v) formic acid). The flow rate for the capillary column was 300 nL/min. Column temperature was 60 °C with the following gradient profile: Linear 2-30%B over 50 min and then linear 30-50% B over 5 min. All runs then proceeded to wash with 95% B for 5 min. The column was returned to initial conditions and re-equilibrated before subsequent injections.

The nanoLC system was interfaced to a maXis HD LC–MS/MS System (Bruker Daltonics) with a CaptiveSpray ionization source (Bruker Daltonics). Positive ESI–MS and MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition, and processing were performed using Compass 1.7 software (microTOF control, Hystar, and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage: 1450 V, dry gas: 3 L/min, dry gas temperature 150 °C, ion acquisition range: *m*/*z* 150–2000. AutoMS/MS settings were: Cycle time 3s, MS spectra rate 2 Hz, MS/MS spectra rate 2 Hz at 2500 cts and 12 Hz at 250000 cts, Quadrupole low mass 300 m/z, collision RF 1400 Vpp, Transfer time 120 ms and Collision Energy 100. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table, absolute threshold 200 counts, preferred charge states: 2–4, singly charged ions excluded. A single MS/MS spectrum was acquired for each precursor and former target ions were excluded for 0.8 min unless the precursor intensity increased 4-fold. The data was processed using MaxQuant15 version 1.3.0.5 which incorporates the Andromeda search engine. Proteins were identified by comparing against a protein sequence database from combined from UniProt and frequently observed contaminants. A MS tolerance of 6 ppm, an MS/MS tolerance at 0.5 Da and full trypsin specificity, allowing for up to two missed cleavages were specified. Fixed modifications were

set to carbamidomethylation of cysteine and oxidation of methionines. Variable modifications were set to N-terminal protein acetylation and N-pyroglutamate. Peptides were required to be at least 7 amino acids in length and a MaxQuant score >5, with false discovery rates (FDRs) of 0.01 calculated at the levels of peptides, proteins and modification sites based on the number of hits against the reversed sequence database. Protein inference and spectral counts were attained using the open source software PeptideShaker searched against a compiled database from UniProt and validated at a 1% false discovery rate.

9 Chapter 2

Affinity probe enrichment of galactosyltransferases from *Trypanosoma brucei.*

9. Chapter 2

Affinity probe enrichment of galactosyltransferases from *Trypanosoma brucei.*

9.1 Results and Discussion

This section will describe how resin bound UDP-galactose (UDP-Gal) and resin bound UDP-(4F)-Galactose (UDP-(4F)-Gal) probes were used in an attempt to enrich candidate GalTs in *T. brucei.* First, the use of Stable isotope labelling by amino acids in culture (SILAC) to distinguish between affinity binders and contaminants will be explained. Then this will be put into the context of our affinity probe protocol before the results of mass spectrometry analysis are explored. This chapter has two main aims:

(i) to use the activity based probes and the developed method from the previous chapter to search for GalTs in *T. brucei*

(ii) to characterise any GalTs found using RNAi

Using a combination of the developed enrichment protocol and Stable Isotope Labelling of Amino acids in Cell culture, the probes were used to enrich GalTs from lysates of *T. brucei.* Any enriched proteins were evaluated to determine if they were potential GalTs.

9.1.1 Author contributions

J. G. M. Bevan, M. Chaves-Ferreira, J. A. Rodrigues, S. Duncan and M. A. Ferguson designed the experiments. J. G. M. Bevan and S. Duncan performed the experiments. J. G. M. Bevan and the FingerPrints Proteomic Facility at the University of Dundee performed the data analysis. J. G. M. Bevan created all the figures and tables. J. A. Rodrigues mentored and supervised the work.

9.1.2 Stable isotope labelling by amino acids in cell culture (SILAC)

Immobilising the proposed ABPP probe on a resin will allow for characterisation of captured proteins. However, the resin itself will trap non-specific proteins and the true binders to the probe must be differentiated from these. This can be done in a couple of ways:

1) Comparing probe-bound resin (a) to resin without the probe
2) Comparing probe-bound resin plus free UDP-Gal (a) with probe-bound resin (b)

Option 1 would differentiate between those proteins that bind to the PEG resin versus those that bind to the probe specifically. In this way, we would look for proteins whose binding increases in (a) compared to (b). Option 2 is similar but the addition of free UDP-Gal would reduce the binding of proteins that are specific to UDP-Gal by competitive inhibition. For this condition, we would look for the proteins whose binding is reduced in (a) compared to (b). In order to compare the two different conditions using mass spectrometry, they must be labelled. There are a number of methods used to enable comparison between two samples and these can be split in to three categories; label free, chemical labelling and metabolic labelling.

Label free comparisons use either the abundance or volume of ion counts or spectral counts. There is no limit to the number of experiments that can be compared as there are no labels, which also reduces the cost. However, there are no formalised internal standards meaning that errors in each dataset have to be averaged with replicates and certain biases are introduced by the algorithms used to calculate abundance based on the length of the peptide (Patel *et al.*, 2000; Deng *et al.*, 2015).

Chemical labelling is often used to follow multiple time points or many experiments or for cultures that cannot be labelled metabolically. It is more accurate than label free although it is more expensive (Megger *et al.*, 2014). The number of experiments that can be compared depends on the label chosen. Isotope coded affinity tagging (ICAT) can label only three different conditions, but tandem mass tags (TMT) and Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) can label 6 and 8 respectively (Wiese *et al.*, 2007; Mertins *et al.*, 2012). Each chemical modification has its own advantages and disadvantages. ICAT labels cysteines and therefore suffers from errors caused by proteins who do not contain a cysteine or only have one in their sequence and cannot be reliably quantified (Hsu *et al.*, 2003). Dimethyl

labelling is a cheaper alternative that labels globally and can be used in triplex (Hsu *et al.*, 2003).

All chemical labels are added at a late stage and therefore are subject to errors introduced by the reaction itself. Metabolic labelling is the most sensitive method to compare samples because the labels are incorporated early on in sample preparation. Stable Isotope Labelling by Amino Acids in Cell culture (SILAC) involves labelling cells with heavy atoms through metabolic incorporation of heavy amino acids. When it was first reported, deuterated leucine was the only heavy amino acid used and it still outperformed ICAT as leucine is 20 x more abundant than cysteine (Ong *et al.*, 2002). Now, the most commonly used amino acids are heavy arginine and lysine (Lau *et al.*, 2014; Xiao, Tang and Wu, 2016). These experiments can be performed in triplex by using commercially available light, medium ($^{2}H_{4}$-lysine (Lys4) and $^{13}C_{6}$-arginine (Arg6)) and heavy ($^{15}N_{2}{}^{13}C_{6}$-lysine (Lys8) and $^{15}N_{4}{}^{13}C_{6}$-arginine (Arg10)) amino acids.

Not all organisms can incorporate these amino acids, or they metabolise them which can complicate the spectra. For example, heavy isoleucine ($^{13}C_{6}{}^{15}N_{1}$) is the only amino acid that can be used to metabolically label *Plasmodium falciparum* as it is the only amino acid that the parasite cannot produce itself, nor does it uptake it from red blood cells or metabolise it to another amino acid (Nirmalan, Sims and Hyde, 2004). Of course, this means that any peptide without an isoleucine would not be labelled and cannot be quantified. In order to follow the metabolism of glucose in *T. brucei*, Creek *et al* used heavy glucose ($^{13}C_{1}$) and in doing so were able to follow its distribution (Creek *et al.*, 2015). Although SILAC has proven to be a useful tool in proteomics, not every cell line grows efficiently in the dialysed medium required and some take too long to incorporate.

We propose to use SILAC to determine specific from non-specific binders. Trypanosomes readily and quickly incorporate heavy arginine and lysine into their proteins. This technique is well established in *T. brucei* and has been used to differentiate the glycosome proteome from the mitochondrial in order to establish *T. brucei* glycosome proteome (Guther *et al.*, 2014). It has also been used to identify phosphorylation sites in two life stages of *T. brucei* and to quantify the differences between the two stages (Urbaniak, Martin and Ferguson, 2013).

Anything that has bound in both experiments will be a non-specific binder but if there is a large increase in intensity for a specific protein in the heavy experiment compared to light then it has specifically bound to the probe, figure 42.

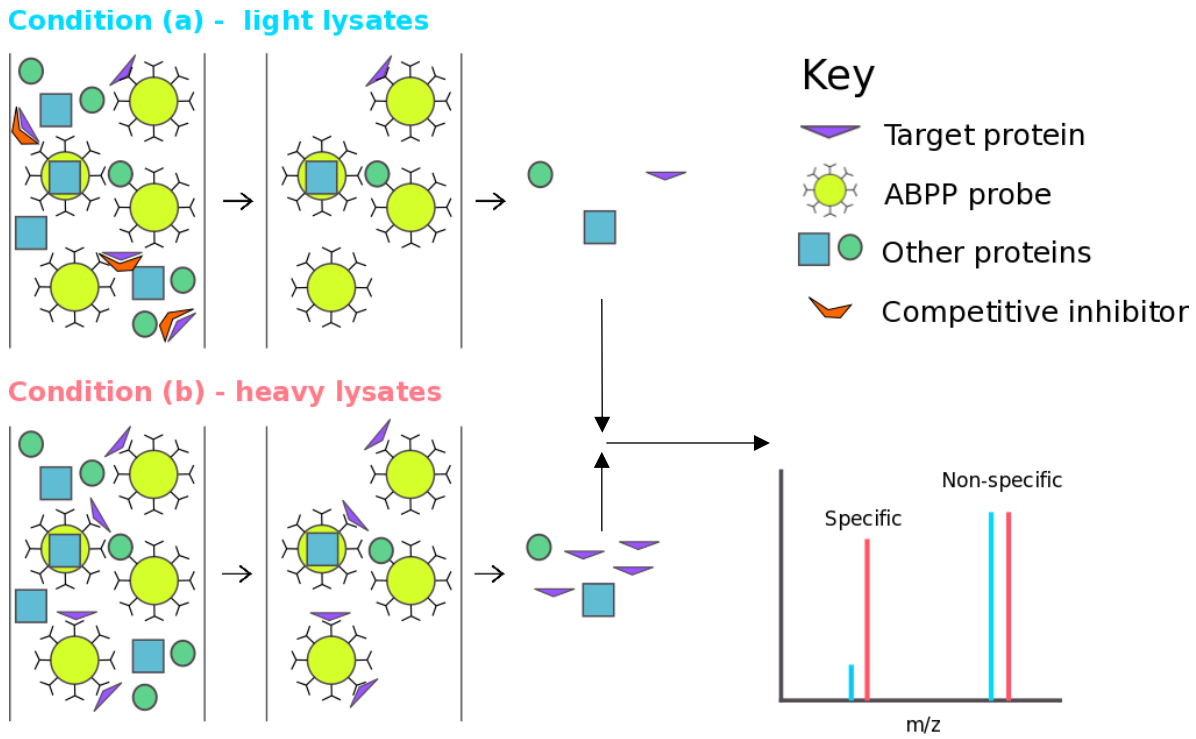


Figure 42 - Application of SILAC to affinity probes – Option 2

9.1.3 Design of affinity probe protocol

T. brucei procyclics are able to grow at a higher density compared to *T. brucei* blood stream forms therefore the SILAC experiment will be performed with procyclics to increase the amount of protein available per mL of medium. Another advantage of using procyclics is that the only characterised galactosyltransferase in *T. brucei* is present, TbGT3 – responsible for linkage **33**, figure 13. This GalT forms two glycosidic linkages on the glycan of the GPI anchor of procyclin and is a good positive control. The GalTs responsible for linkages 28,29, 32 and 36 (figure 13) will also be present. GTs can be identified using characteristics described in the section **7.1.4 *Glycosyltransferases***. If the protein sequence contains a DXD motif, a transmembrane domain and/or is listed as a putative GT of *T. brucei* then it will be considered as a candidate worth pursuing (Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J Barton, *et al.*, 2009; Rini, Esko and Varki, 2009).

Due to the specificity of each GT, they are inherently low in abundance (Taylor and Drickamer, 2011). In order to capture GalTs and reduce the amount of non-specific background, the

amount of probe used must be of a similar order to GalT abundance. If the protein is known, then this could be calculated (Milo, 2013) but we will have to estimate. To get an idea of scale, two highly abundant proteins in the proliferative stages, procyclin and VSG, have around 10^6 and 10^7 copies per cell, respectively. However, the transferrin receptor TfR, only expressed in blood stream forms, only has around 2.3 x 10^3 copies per cell (Magez and Radwanska, 2013). Assuming GalTs are of a similar low abundance of around 1 x 10^3 to 1 x 10^4 we can estimate the following:

$$200 \text{ mL of procyclics at } 2 \text{ x } 10^7 \frac{\text{cells}}{\text{mL}} = 2 \text{ x } 10^9 \text{ cells}$$

$$2 \text{ x } 10^9 \text{ cells x } 1 \text{ x } 10^3 \text{ to } 1 \text{ x } 10^4 \text{ GTs } = 2 \text{ x } 10^{12} \text{ to } 2 \text{ x } 10^{13} \text{ GT molecules}$$

$$50\% \text{ yield } = 1 \text{ x } 10^{12} \text{ to } 1 \text{ x } 10^{13} \text{ GT molecules}$$

$$\frac{GT\ molecules}{6.022 \text{ x } 10^{23}} = 1.7 \text{ x } 10^{-12} \text{ to } 1.7 \text{ x } 10^{-11} \text{ mol of GT} = 1.7\ pmol\ to\ 17\ pmol\ GT$$

There are around 10 GalTs – assuming 5 "known" (28, 29, 32, 33, 36, figure 13) plus 5 unknown:

$$max.mol\ of\ GalT\ in\ 200\ ml\ culture = 10\ x\ 17\ pmol = 170\ pmol$$

To allow a 10 fold excess of probe to GalT:

$$max.mol\ of\ probe = 10\ x\ 170\ pmol = 1.7\ nmol$$

For a 200 mL culture, considering that the density of UDP-Gal on the resin is 0.205 mmolg^{-1} and UDP-(4F)-Gal on the resin is 0.167 mmolg^{-1}:

1.7 nmol = 8 µg UDP-Gal resin

1.7 nmol = 10 µg UDP-(4F)-Gal resin

Once a protocol is optimised, 3 biological replicates should be completed and the data pooled in order to reduce errors introduced by sample handling.

9.1.1 Resin bound UDP-Galactose has a low affinity to galactosyltransferases

9.1.1.1 Trial 1

First the ability of the UDP-Gal resin to pull out GalTs from pcf lysates was tested. Pcfs were cultured until 200 mL of culture at 2x10^7 cells/mL in light and heavy medium were obtained.

The cells were washed, lysed and ultracentrifuged (see method) so that a clarified supernatant without cell debris was obtained with a concentration of $4x10^8$ cells/mL (10 mL). 0.1 mM of free UDP-Gal was added to the light lysates in order to block the interactions of GalTs with the resin.

The heavy and light supernatants were passed through 4 mm PVDF membranes on which 10 µL of swelled UDP-Gal resin sat. The PVDF was used to act as a support for the resin to allow it to be incubated with proteins. The small diameter was chosen in order for all the lysates to pass through the bed of resin without having to use a large quantity of resin. As discussed above, the more resin used the more chance there is of encouraging non-specific interactions. Unfortunately, after 3 mL of lysates (approx. $1.2x10^9$ cells) were passed through the membrane before it blocked. As these are small membranes, 10 mL of concentrated protein was too much. In order to continue, the resin was back-washed through the PVDF membrane and washed in an Eppendorf. The proteins were eluted in an Eppendorf with 2% SDS, figure 43.

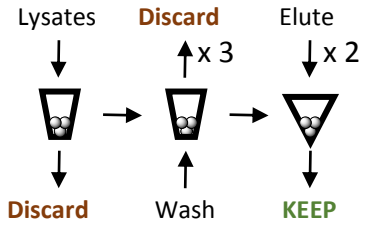


Figure 43 - Method used with UDP-Gal resin to pull out GalTs from pcf lysates

These light and heavy samples were then joined and prepared for mass spectrometry. There were 265 proteins listed as enriched in this experiment (see annex table 10 for the top 12). However, only one protein was significantly enriched (more than 5 times). This protein has an assigned function and is a catalytic subunit of the vacuolar transporter chaperone 4, with no characteristics of a GalT. There were no putative GalTs present in the captured protein list of the light nor the heavy as discerned by the lack of a DXD motif. Perhaps repeated experiments are required to reduce the background of non-specific binding and reveal those that have an affinity for the probe not the resin. However, since no GalT was present in the light or the heavy experiment is more likely due to the low volume of lysates able to pass through the resin, the low abundance of GalTs and low affinity of the GalTs to the probe. If it a matter of abundance, repeats could increase the chances of seeing a GalT but with a low affinity probe the probability this might not be feasible. The amount of free UDP-Gal used in the light

experiment would influence the resolution between heavy and light but as no GalTs were seen in the light, the amount of free UDP-Gal is not likely to be the problem. If more resin was used, this would increase the chances of capturing a GalT. It would also increase the amount of non-specific binding; therefore, many repeats would be needed to reduce the background noise. To summarise, from this experiment it is unclear if the probe has a low affinity or if the volume of lysates passed was insufficient to capture enough GalT to be seen by mass spectrometry.

9.1.1.2 Trial 2

Two 200 mL cultures of heavy pcfs and two of light were grown. To improve upon the previous experiment, this time each time a membrane blocked the resin was moved to a new membrane. In this way, all the lysates could be passed through the resin. A second method was also trialled where the resin was rolled in a falcon with the lysates for 3 hours at 4 °C. Since there were two biological replicates, these were spilt in two to be used in each method giving a total of four experiments, figure 44. To ensure that all GalTs are blocked in the light experiment, the amount of free UDP-Gal was increased to 1 mM.

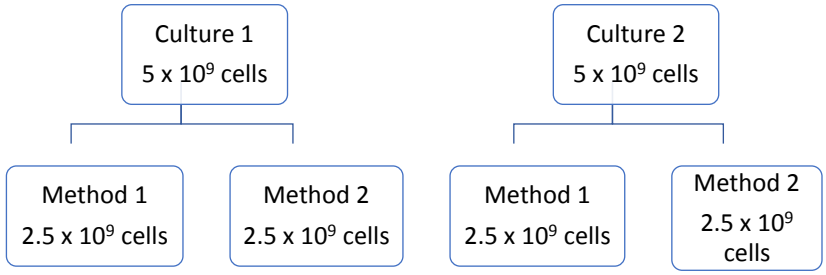


Figure 44 – Illustration of how biological replicates were split in to the two methods: passing through a membrane or rolled in a falcon.

The light and heavy eluted proteins for each of the four experiments were joined to their respective partner and sent to mass spectrometry. During initial mass spectrometry runs, the sample intensities observed were too low, so methods 1 and 2 for each culture were combined so that the biological replicates still stood true. There were over 500 proteins listed as enriched in both biological replicates, double that of the previous experiment. The increase in amount of enriched proteins can be explained by the improved technique – the full 10 mL of lysates were passed through the resin. Of these 500 enriched proteins, 10 were more than 5 x enriched (see annex table 11 for the top 20), but none of the 500 were putative GalTs.

Included in the list of enriched proteins were UDP-galactose 4-epimerase (GALE, 2.7 fold in culture 1 and 1.7 fold in culture 2) and a UDP-Gal transporter (Tb927.10.13900, 9 fold in

culture 1 and 1.5 fold in culture 2). Although they are not the focus of this study, these are both UDP-Gal processing enzymes and their presence is a promising indication that the probe is available to bind to enzymes with affinity to UDP-Gal. These two enzymes could have a higher abundance than GalTs or they could simply have a higher affinity to the probe. They each are unique enzymes (unlike the approximate 10 GalTs) therefore it is plausible that their abundance is higher increasing their chances of binding.

The affinity of the UDP-Gal probe to GalTs is clearly not high enough to enrich them when they are in such low abundance. However, the change in method from the PVDF membrane to rolling did improve the binding of the resin to UDP-Gal processing enzymes. This is probably because more lysates were exposed to the resin and for a longer time giving an increased opportunity for these enzymes to interact. Furthermore, in the PVDF method, the transfer of the resin from membrane to membrane lead to loss of resin and therefore sample. If the affinity of the resin could be improved, the "rolling" method could potentially enrich GalTs.

9.1.2 Resin bound UDP-(4F)-Galactose has an improved affinity to galactosyltransferases

UDP-(4F)-Gal was designed to have a higher affinity to GalTs than UDP-Gal since the introduction of the fluorine increases the strength of the 4-OH hydrogen bond to the enzyme, as discussed in **7.3.1**. Since the UDP-Gal resin has a low affinity it can replace the use of free UDP-Gal and can act as the control for the non-specific binders compared to the UDP-(4F)-Gal resin, figure 45.

Figure 45 – Use of UDP-Gal resin compared to UDP-(4F)-Gal resin

Using the rolling method as validated previously, the light lysates were incubated with the UDP-Gal resin and the heavy with the UDP-(4F)-Gal resin and the eluted proteins combined for mass spectrometry, see methods. Nearly 700 proteins were enriched with the UDP-(4F)-Gal resin, of which 80 were significantly enriched and these enrichments were greater than before (see annex table 12 for the top 20 proteins). This great increase in enrichment could

be because the addition of the fluorine has increased affinity of many proteins to the resin, but almost all the proteins that were enriched were not UDP-Gal processing; it therefore seems unlikely that the fluorine caused the increase in number. The increase in number of proteins enriched could be due to the improvement in technique that increased the number of proteins detected in general. In ***Trial 1*** only 453 proteins were detected. This increased to 761 for ***Trial 2*** and even further for this experiment to 1408. The technique was improved partly by practise and partly by utilising the rolling method which lead to less loss of resin. By using only one method, less resin per culture was used which reduces the amount of non-specific interactions. Of course, this increase could be anomalous and repeated experiments would be needed to confirm.

Within the 700 proteins UDP-galactose 4-epimerase was again enriched 2.5 times. None of the MaxQuant assigned enriched proteins were putative GalTs. However, GT3 (the only characterised GalT in *T. brucei)* was present in the 700. It was not assigned a H/L ratio because it and 54 other proteins were completely absent in the light experiment, according to MaxQuant. There were 4 unique peptides in the heavy experiment assigned to GT3. In order to confirm total absence of GT3 from the light experiment, extracted ion chromatographs (XIC) were performed for each of the 4 peptides. This checks the presence of the ions and ensures that the mass spectrometer did not overlook them for analysis by extracting the m/z values of the peptide across the entire chromatograph. The result is a plot that gives the intensity of the ion across the whole run. Of the four heavy peptides present, one peptide's light counterpart was absent but three clearly had light equivalents. From these three, one peptide was not overlapping any other peptide and could be quantified. The resulting ratio is 1: 3.57, L: H meaning that it is not a case of presence and absence, but GT3 was truly enriched in this experiment. The remaining two peptides overlapped others making quantification with these inaccurate. As an example, the XIC of the peptide chosen for quantification is shown in figure 46.

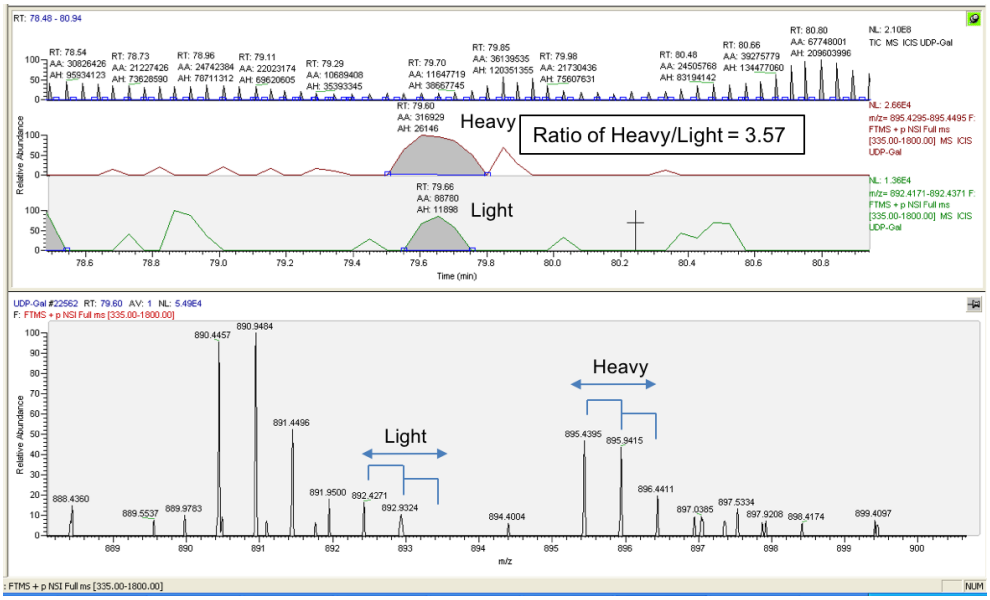


Figure 46 - XIC plot of peptide LVNTPYSESNIEDFR – Heavy m/z 895.4395, Light m/z 892.4271, RT 79.60

The top 50 enriched proteins and the 55 proteins that were present in the heavy but absent in the light experiment were checked for GalT properties. If the protein did not have an assigned function and had at least one transmembrane domain, the sequence was checked for a DXD motif. Unfortunately, no candidate GalTs were found as the few that were transmembrane proteins had no DXD motif present. There was a high number of ribosomal proteins present in the enriched list, these proteins should bind non-specifically and would reduce in number over repeated experiments.

Although the only GalT that was enriched was the already characterised GT3, this experiment showed that the probes have the potential for enriching (and allowing the identification of) GalTs from lysates. Repeating the experiment with the fluorinated probe would increase the probability of other GalTs binding and the background binding of the ribosomal proteins would be reduced. The reaction conditions could be optimised to give time for binding by increasing the time that the resin was incubated with the lysates from 3 hours to overnight.

Tentagel is an ampiphilic resin and has a reduced non-specific binding compared to purely hydrophobic solid supports who bind many proteins non-specifically. However, it can still bind non-specifically to proteins. Kim *et al.* designed a new PEG-based resin that reduced non-specific interactions by efficiently displaying small molecules on its surface. They attributed

tentagel's non-specific binding to its uneven surface and proved their smooth surfaced resin minimised non-specific interactions by reducing the sterics and allowing proteins to gain access to the affinity tag (Kim *et al.*, 2006).

Similarly, Chen *et al.* discuss the high ligand concentration of beads such as tentagel which leads to multiple interactions of one protein with many ligands. Even if this protein only has a low avidity, the multiple interactions combined would produce a high affinity that could be higher than a specific interaction. Furthermore, even if the original interaction was specific, further non-specific interactions of one protein with the hydrophobic surface of the resin would bias a study towards those proteins over the ones that could only form the original specific interaction (Chen *et al.*, 2009).

Based on these observations, the hydropathy plots of GALE and the UDP-Gal transporter were compared to that of GT3, figure 47 (Aslett *et al.*, 2010). Although there are no striking differences between the three, GT3 appears to have a region of almost neutral hydrophobicity where GALE and UDP-Gal transporter have peaks and troughs of hydrophobicity. These subtle differences could encourage binding *via* a combination of specific and non-specific interactions as suggested by Chen *et al* who propose reducing ligand loading and increasing spacing between ligands to overcome this problem (Chen *et al.*, 2009).

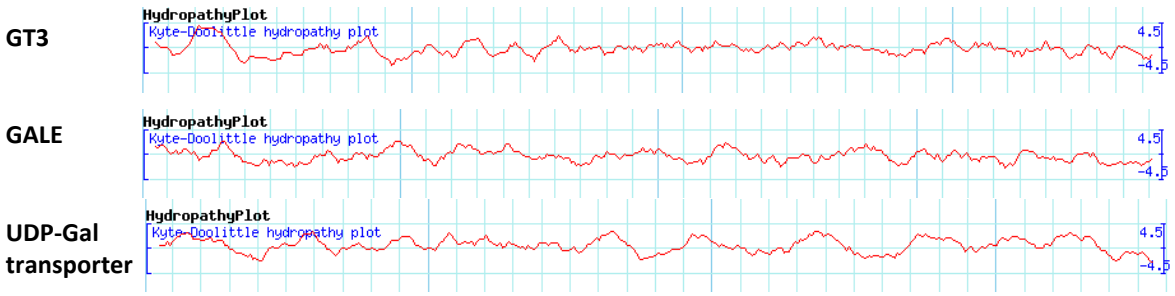


Figure 47 – Hydrophobicity plots adapted from the TriTryp database

An alternative to reducing ligand loading or spacing between ligands would be to block the resin prior to using it. Tentagel resin is often blocked by hydrophobic BSA or milk overnight which reduces non-specific binding (Nafffin *et al.*, 2003; Göksel *et al.*, 2011). This technique could be incorporated in to the protocol prior to incubation with the lysates and could reduce background binding, revealing GalTs. Finally, the affinity of the probes must be increased since

the GalTs are of very low abundance. This could be achieved by modifying the probe as discussed in **7.4**.

9.2 Chapter summary

In this chapter, we saw that, when applied to whole lysates of *T. brucei* procyclic forms, the UDP-Gal probe did not have a high enough affinity to enrich any GalTs. UDP-(4F)-Gal was able to enrich GT3, a known and characterised GalT of procylic forms. However, no other predicted GalTs were enriched. Although these results suggest good selectivity for UDP-(4F)-Gal resin, these assays would benefit from additional experiments and optimization steps to increase enrichment capacity and enable the identification of novel UDP-galactosyltransferases from the parasite. In the future, a new probe could be synthesised with an increased affinity by modifying either the sugar, the nucleotide or both, so that the low in abundance GalTs could be enriched and characterised. This work paves the way towards the synthesis other similar probes using the same synthetic approach. GlcNAc transferase are also of great importance in *T. brucei* therefore a probe based on GlcNAc could be designed to characterise the putative GlcNAc transferases as yet uncharacterised.

9.3 Materials and methods

9.3.1 Culture of procyclic form *T. brucei*

One litre of SDM-79 without L-Arg or L-Lys was made from powder with the following recipe: 800 ml MilliQ water, 1 pot of powder SILAC SDM-79 (it comes minus L-Arg and L-Lys, from Tryp Community), 2 g $NaHCO_3$ (Merck cat n. 1.06329.1000), 0.75 ml of 10 mg/ml haemin in 50 mM NaOH (haemin is Sigma cat. H5533), 10 ml Glutamax I (Invitrogen cat. 35050-38), pen-strep (Gibco 15140, diluted 1:1,000) and 150 ml Dialised and Heat inactivated FBS (PAA lot 1488). This 1L media (minus L-Arg and L-Lys) is then is divided in 2 x 500 ml, one for R0K0 and the other for R6K4. L-Arg and L-Lys were then added at R0 HCl (ordinary 12C isotope, Sigma A-5131) or R6 (heavy, $^{13}C_6$-arginine) HCl (13C isotope, Cambridge Isotope Labs cat CLM-2265-H-05) 248 mg/L and K0 2HCl (ordinary 12C isotope, Sigma L-5751) or K4 (heavy, $^{2}H_4$-lysine) 2HCl (13C isotope Cambridge Isotope Labs cat DLM-2640-05) 91 mg/L. Procyclic form *Trypanosoma brucei,* monomorphic strain 427, were grown at 28 °C without CO_2 in fully capped culture flasks in light R0K0 SDM-79. These cells, in log phase, were suspended in either SDM-79 supplemented with light lysine and arginine or SDM-79 supplemented with $^{2}H_4$-lysine (Lys4) and $^{13}C_6$-arginine (Arg6) at $2x10^6$ cells/mL and passed every 2 days (enlarging the culture each time) until 6-7 cell divisions had occurred in the heavy medium and the cultures were 200 mL at $2x10^7$ cells/mL. The cells were harvested by centrifugation and washed with ice-cold PBS three times. The cell pellet was resuspended in **lysis buffer** (2%nOG, 50 mM Tris-HCl pH 7.4, 0.1 mM TLCK, 1 μg/mL, complete protease inhibitor cocktail without EDTA and 2 mM manganese chloride) at $5x10^8$ cells/mL and incubated on ice-water for 30 mins. The suspension was clarified by ultra-centrifugations (50,000 g, 1 hr, 4 °C) and the supernatant taken for affinity purification.

9.3.2 Affinity purification on PVDF membranes with UDP-Gal resin

The affinity UDP-Gal resin was swelled in an eppendorf in lysis buffer then the excess buffer was removed. 10 μL of swelled resin (with 10 μL reach tips cut for a wider bore) was loaded on to two a 4 mm hydrophilic PVDF syringe filter membranes with a 0.45 μm pore size. The resin was washed on the membrane with 1 mL 2% SDS, 20 mM Tris-HCl pH 7.4 then 0.5 mL of lysis buffer. To the light cell lysates 1 mM of free UDP-Gal was added and to the heavy cell lysates the equivalent volume of water was added. The light lysates were then passed slowly through the PVDF membrane loaded with resin (less than 1 drop/sec). The heavy cell lysates were passed slowly through the remaining PVDF membrane loaded with resin. If the

membrane blocked, the resin was back washed into an eppendorf and loaded on to a new membrane. After all the lysates passed both membranes, the light lysate resin was back-washed into a new low bind eppendorf. This resin was washed 3 x with 0.5 mL lysis buffer with protease inhibitors and the supernatant discarded. The heavy resin was added to the light resin and washed a further 5 x with 0.5 mL lysis buffer, discarding the supernatant. **Elution buffer**, 100 μL 2% SDS, 20 mM Tris-HCl pH7.4 was added to the resins and they were incubated for 30 min at 50 °C in an incubator with shaking. They were cooled, centrifuged and the SDS-eluate from the top of the resin was transferred into a low protein bind Eppendorf. The elution was repeated and the eluate added to the same eppendorf. The resin was washed once more with elution buffer, centrifuged and the SDS-eluate was added to the same Eppendorf. This elution was taken for FASP.

9.3.3 In-falcon affinity purification with UDP-Gal resin

The affinity resin UDP-Gal was swelled in an eppendorf in lysis buffer and the excess buffer was removed. 10 μL of the swelled resin (with 10 μL reach tips cut for a wider bore) was added to two 15 mL falcons. The resins were washed with 1 mL 2% SDS, 20 mM Tris-HCl pH 7.4 then 0.5 mL of lysis buffer. To the light cell lysates 1 mM of free UDP-Gal was added and to the heavy cell lysates the equivalent volume of water was added. The lysates were added to the falcons containing UDP-Gal resin and were incubated rotating at 4 °C for 3 h. The resins were washed separately with 3 x with 0.5 mL lysis buffer with protease inhibitors and transferred to an Eppendorf together. 100 μL 2% SDS, 20 mM Tris-HCl pH7.4 (**elution buffer**) was added and the resins incubated for 30 min at 50 °C in an incubator with shaking. They were cooled, centrifuged and the SDS-eluate collected from the top of the resin into a low protein bind Eppendorf. The elution was added to the same eppendorf. The resin was washed with 100 μL of 2% SDS, 20 mM Tris-HCl, centrifuged and added to the same eppendorf. The elution was taken for FASP.

9.3.4 In-falcon affinity purification with UDP-(4F)-Gal resin

The affinity resins UDP-Gal and UDP-(4F)-Gal were swelled in an eppendorf in lysis buffer and the excess buffer was removed. 10 μl of each swelled resin (with 10 ul reach tips cut for a wider bore) was added to two 15 mL falcons. The resins were washed with 1 mL 2% SDS, 20 mM Tris-HCl pH 7.4 then 0.5 mL of lysis buffer. The light lysates were added to the falcon

containing UDP-Gal resin. The heavy cell lysates were added to the falcon containing UDP-(4F)-Gal resin. The falcons were incubated rotating at 4 °C for 3 h. The resins were washed 3 x with 0.5 mL lysis buffer with protease inhibitors and transferred to an eppendorf. The resins were combined, 100 μL 2% SDS, 20 mM Tris-HCl pH7.4 (**elution buffer**) added and incubated for 30 min at 50 °C in an incubator with shaking. They were cooled, centrifuged and the SDS-eluate collected from the top of the resin into a low protein bind Eppendorf. The elution was added to the same eppendorf. The resin was washed with 100 μL of 2% SDS, 20 mM Tris-HCl, centrifuged and added to the same eppendorf. The elution was taken for FASP.

9.3.5 Filter aided sample preparation (FASP)

Samples were prepared using a modified version of a published FASP procedure (Wiśniewski *et al.*, 2009). An Amicon 30kda column was pre-washed with 400 μL of ddH_2O. The sample was loaded on to the column 50 μL at a time with 200 μL of **UA** (8M urea in 0.1 M Tris HCl pH 8.5) and centrifuged at 14,000 g for 15 min. Another 200 μL of UA was added and centrifuged at 14,000 g for 15 min. The flow through was discarded and 100 μL of 0.05 M iodoacetamide in UA was added to the column. It was incubated in the dark for 20 min then centrifuged at 14,000 g for 10 min. The filter unit was washed twice with 100 μL of UA, each time centrifuging at 14,000 g for 15 min and discarding the flow through. It was then washed 4 x 100 μL 40 mM NH_4HCO_3, each time centrifuging at 14,000 g for 10 min and discarding the flow through. The filter unit was transferred to a new collection tube, trypsin-Lys-C added (promega, use 25:1 protein: protease, w: w) in 40 mM NH_4HCO_3 (0.4 μg trypsin/μL) and incubated overnight at 37 °C. The filter unit was centrifuged at 14,000 g for 10 min. The filter was washed twice with 40 μL of BB (40 mM Tris/HCl pH 7.6, 2mM $MnCl_2$, 2mM $CaCl_2$, 1 M NaCl) and centrifuged at 14,000 g for 10 mins. The filtrate was collected for C18 purification.

9.3.6 C18 purification

All solvents used in this purification were of mass spectrometry grade. The solutions used were as follows:

Conditioning solution: 90/10 MeOH/H_2O with 0.1% TFA (v/v/v) (10 ul TFA per 10 mL)

Equilibration solution: 0.1% TFA in H_2O (v/v)

Desalting solution: 5% MeOH/ H_2O with 0.1% TFA (v/v/v)

Elution solution: 50/50 AcCN/ H_2O with 0.1% TFA (v/v/v)

The C18 SPE cartridge was conditioned by passing 3 mL conditioning solution through the packing bed. It was then equilibrated by passing 2 mL equilibration solution through the packing bed. The sample was reconstituted in up to 1 mL of 0.1-0.5% TFA. The pH was adjusted to <3 with formic acid if needed. The sample was loaded by passing slowly (1 drop/ sec) into the cartridge. Loading was completed by passing 1 mL of equilibration solution into the cartridge. The sample was desalted by passing 1 mL desalting solution through cartridge. A new was placed below the cartridge for collection of the sample. The sample was eluted by slowly (1 drop/ sec) passing 1 mL of elution solution through cartridge. The sample was dried on a speed vac overnight.

9.3.7 Mass spectrometry

Liquid chromatography tandem mass spectrometry was performed by the FingerPrints Proteomic Facility at the University of Dundee on an LTQ Orbitrap Velos Pro mass spectrometer. Trypsin-digested peptides were separated using an UltiMate 3000 RSLC (Thermo Scientific) nanoflow LC system. Around 0.5 µg of protein was loaded with a constant flow of 5 µl/min onto an Acclaim PepMap100 nanoViper C18 trap column (100 µm inner-diameter, 2cm; Thermo Scientific). After trap enrichment, peptides were eluted onto an EASY-Spray PepMap RSLC nanoViper, C18, 2 µm, 100 Å column (75 µm, 50 cm; Thermo Scientific) with a linear gradient of 2–35% solvent B (80% acetonitrile with 0.08% formic acid, Solvent A – 0.1% formic acid) over 70 min with a constant flow of 300 nl/min and column temperature of 50 °C. The HPLC system was coupled to a linear ion trap Orbitrap hybrid mass spectrometer (LTQ-Orbitrap Velos, Thermo Scientific) via an easyspray ion source (Thermo Scientific). The spray voltage was routinely set to 1.8 kV, and the temperature of the heated capillary was set to 250 °C. Full-scan MS survey spectra (*m*/*z* 335–1800) in profile mode were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 1,000,000 ions. The fifteen most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision-induced dissociation (normalized collision energy, 35%; activation Q, 0.250; and activation time, 10 ms) in the LTQ after the accumulation of 5,000 ions. Maximal filling times were 1,000 ms for the full scans and 150 ms for the MS/MS scans. Precursor ion charge state screening was enabled, and all unassigned charge states as well as singly charged species were rejected. The lock mass option was enabled for survey scans to improve mass accuracy (Olsen, J.V. *et al.,* 2005). Data were acquired using the Xcalibur software.

9.3.8 Quantification and bioinformatics analysis

The raw mass spectrometric data was loaded into MaxQuant (version 1.5.2.8) (Cox, J., and Mann, M., (2008). *Nat Biotechnol* 26, 1367-72.) using the Andromeda search engine software (Cox, J., et al, (2011) J. Prot. Research *10* (4), 1794-1805). Enzyme specificity was set to that of trypsin, allowing for cleavage N-terminal to proline residues and between aspartic acid and proline residues. Other parameters used were:

(i) variable modifications- methionine oxidation, protein N-acetylation, gln → pyro-glu, deamidation (NQ)

(ii) fixed modifications, cysteine carbamidomethylation

(iii) database: TriTrypDB-33_TbruceiTREU927_AnnotatedProteins (170626)

(iv) labels: R0K0, R6K4

(v) MS/MS tolerance: FTMS- 10ppm , ITMS- 0.6 Da

(vi) maximum peptide length, 6

(vii) maximum missed cleavages, 2

(viii) maximum of labelled amino acids, 3

(ix) false discovery rate, 1%

Peptide ratios were calculated for each arginine- and/or lysine-containing peptide as the peak area of labelled arginine/lysine divided by the peak area of non-labelled arginine/lysine for each single-scan mass spectrum. Peptide ratios for all arginine- and lysine-containing peptides sequenced for each protein were averaged.

10 Chapter 3

N-glycosylation inhibitor NGI-1 and its uses with *T. brucei*

10. N-glycosylation inhibitor NGI-1 and its uses with *T. brucei*

10.1 Results and Discussion

This chapter had three main aims; (i) to study the effect that NGI-1 has on *T. brucei* growth; (ii) to examine the effect that NGI-1 has on *T. brucei* N-glycosylation; (iii) to investigate the utility of NGI-1 as a tool to study N-glycosylation. The parasites were grown in culture medium supplemented with NGI-1 at various concentrations and their growth monitored. Lectin blots of lysates and mass spectrometer analysis of VSG were performed to profile N-glycosylation.

10.1.1 Author Contributions

J. G. M. Bevan, M. Chaves-Ferreira and J. A. Rodrigues designed the experiments. J. G. M. Bevan and M. Chaves-Ferreira performed the experiments. J. G. M. Bevan performed the data analysis and created all the figures and tables. J. A. Rodrigues mentored and supervised the work.

10.1.2 Predicted utility of NGI-1

The small molecule inhibitor of mammalian Stt3B, NGI-1, was described in chapter 1 and is shown in figure 48. This reversible inhibitor also reduced the capabilities of mammalian Stt3A to glycosylate but did not fully inhibit it. Due to this incomplete inhibition, the cells remained viable but under-glycosylated allowing for the study of glycosylation defects within a number of mammalian cell lines (Lopez-Sambrooks *et al.*, 2016).

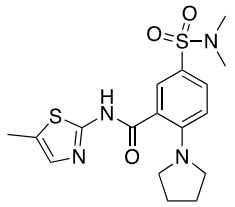


Figure 48 - NGI-1

Published knock downs of TbSTT3A and TbSTT3B showed the substrate specificities of these oligosaccharide transferases were for $Man_5GlcNAc_2$ and $Man_9GlcNAc_2$ respectively and that the parasite in blood stream forms need both OSTs for survival *in vivo. In vitro,* they could be knocked down individually without effecting the culture's growth (Izquierdo, Schulz, *et al.*, 2009). Although these knock downs provided a wealth of information about OSTs in *T. brucei*, small molecule inhibition provides unique benefits, figure 49. RNAi can be easier to monitor than small molecule inhibition since the absence of a protein is easier to visualise (western or qPCR) than the effect of its inhibition. However, small molecules are much easier to use and they can be used in conjunction with other probes including RNAi. They can also be used in varying quantities allowing observation of their effects across a range and can be developed as therapeutics if their activity is specific. Furthermore, small molecules are much faster in their inhibition than RNAi, which enables time specific experiments to be performed. Moreover, scaling up with siRNA is not efficient, but scale up with small molecules is very easy to achieve (Weiss, Taylor and Shokat, 2007).

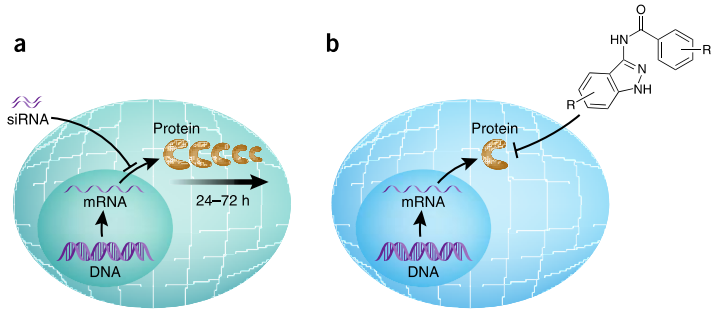


Figure 49 – siRNA (a) vs Small molecule inhibition (b), adapted from (Weiss, Taylor and Shokat, 2007)

A BLAST protein alignment showed that Stt3B of *T. brucei* has a 68.36 % similarity to the mammalian Stt3b, see annex figure 71. Unless NGI-1's target is allosteric, there is a high probability that it inhibits TbStt3B in the same way that it does for the mammalian. Lopez-Sambrooks *et al.* concluded that due to its specificity for Stt3B over Stt3A and from results of a cellular thermal shift assay that NGI-1 targets the active site of Stt3B (Lopez-Sambrooks *et al.*, 2016). Complete inhibition of TbStt3B will inhibit the transfer of oligomannose structures and lead to an increase in complex and paucimannose structures like those seen in the *TbSTT3B* knock down (Izquierdo, Schulz, *et al.*, 2009).

The variant surface glycoprotein (VSG) of monomophic *Trypanosoma brucei* 427 has two N-glycosylation sites and is called VSG221. Ans263 is glycosylated with paucimannose and complex structures which are derived from the transfer of $Man_5GlcNAc_2$ by Stt3A. Asn428 is glycosylated with oligomannose structures which are derived from the transfer of $Man_9GlcNAc_2$ by Stt3B (Manthri *et al.*, 2008). This highly abundant protein is therefore a convenient control to monitor the inhibition of Stt3A and Stt3B that has already been used in numerous studies of glycosylation in *T. brucei* (Urbaniak, Turnock and Ferguson, 2006; Manthri *et al.*, 2008; Izquierdo, Schulz, *et al.*, 2009; Izquierdo, Mehlert and Ferguson, 2012; Castillo-Acosta *et al.*, 2013). To get an overview of the effect of NGI-1 on *T. brucei* and its N-glycosylation a number of studies were performed.

*10.1.3*NGI-1 inhibits the growth of *T. brucei*

NGI-1 is soluble in DMSO up to 5 mg/mL so stock solutions were made at 10 mM (equivalent to 4 mg/mL). This lead to some limitations in terms of the concentration of NGI-1 that the cells could be grown in due to the toxicity of DMSO. DMSO only affected growth at 1%, below this the cultures grew at the same rate as a culture without DMSO, figure 50. Trypanosomes grown in 100 µM NGI-1 (1% DMSO + NGI-1) were not viable after 12 h, figure 50. This could be a combination of DMSO toxicity and the toxicity of NGI-1 on or off target. To know if it has off-target effects, a mutant resistant to NGI-1 would have to be indentified. However, this is off the scope of this project was not studied. When *T. brucei* was grown in 50 µM NGI-1 for 3 days and then transferred to medium without NGI-1, the culture reverted to normal growth showing that the inhibition if reversible, figure 50. Although the difference in trajectory between the growth and trend line from day 1 to 3 looks small from day 3 to day 4, the scale is logarithmic and this represents a rapid increase in growth.

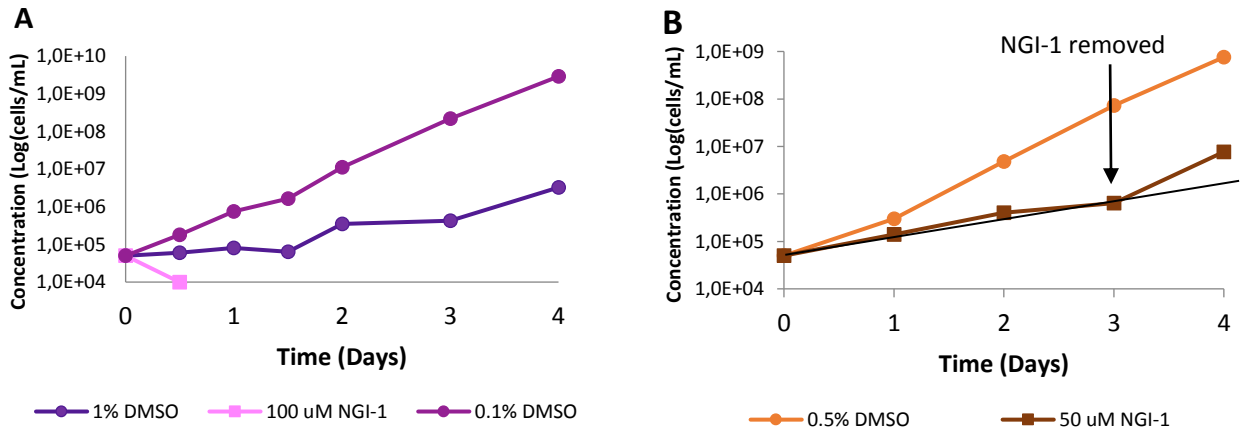


Figure 50 – A: Growth curve of *T. brucei* grown in 1% DMSO, 0.1% DMSO or 100 µM NGI-1, n = 1. B: *T. brucei* growth with 0.5% DMSO or 50 µM NGI-1 which is removed at day 3, n = 1.

T. brucei blood stream forms were grown in varying concentrations of NGI-1 dissolved in DMSO or the same concentrations of DMSO without the inhibitor, figure 52 graph A. The controls grown in the equivalent amounts of DMSO showed no reduction in growth. Increasing the concentration of NGI-1 from 10 µM to 50 µM slowed the growth of *T. brucei* with this effect being most visible after 3 days. Cultures grown with 50 µM NGI-1 often displayed the big-eye phenotype indicating the culture was stressed, figure 51 (Young and Smith, 2010).

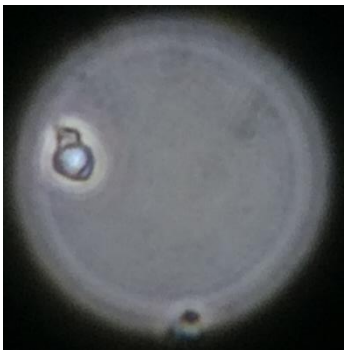


Figure 51 - Big eye phenotype observed in cultures grown in above 20 µM NGI-1

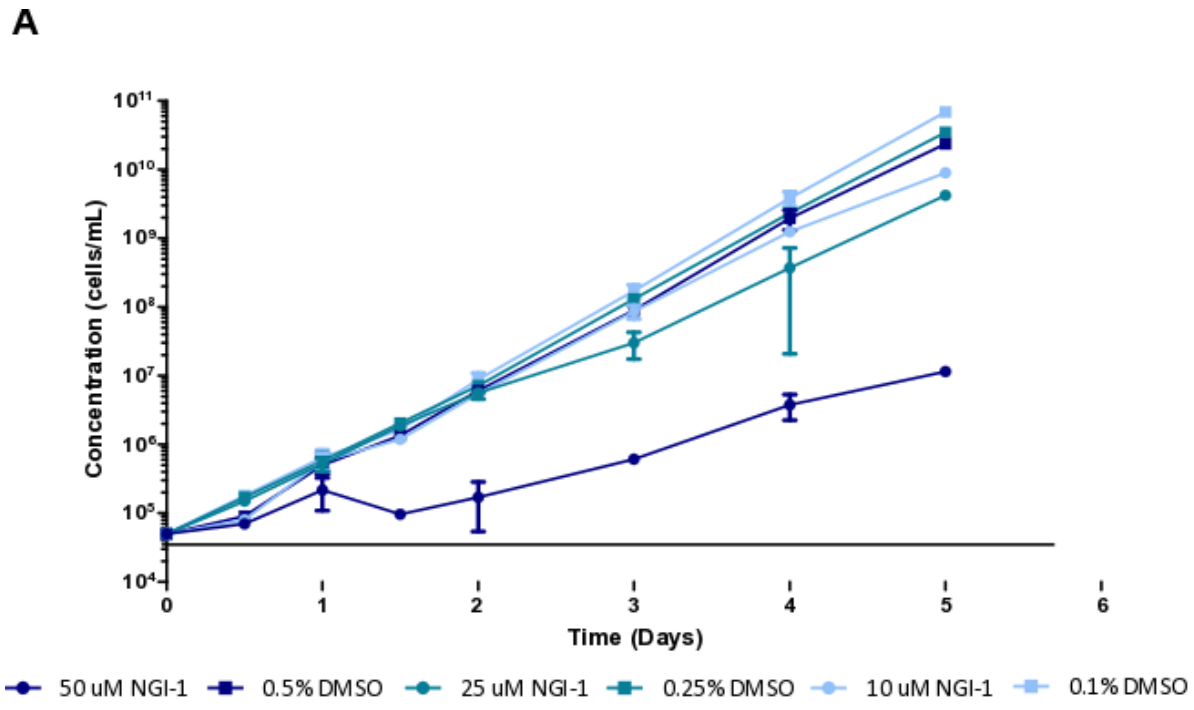


Figure 52 – A: *T. brucei* growth curves at different concentrations of DMSO and NGI-1 dissolved in DMSO. Error bars are calculated using standard error, n = 2 or 3, own data. B: Growth of the following *T. brucei* cells at 37 ℃: wild type (closed circles), TbSTT3A,B,C$^{+/-}$ - STT3BRNAi minus Tet (open squares) and TbSTT3A,B,C $^{+/-}$ - STT3BRNAi plus Tet (closed triangles) adapted from (Izquierdo, Schulz, *et al.*, 2009)**.**

Izquierdo *et al.* described a minimal growth phenotype for their Stt3B knock-down, figure 52 graph B. However, NGI-1 caused a drastic growth phenotype in RTK dependant non-small lung cancer cells even at 10 µM (Lopez-Sambrooks *et al.*, 2016). At 10 µM of NGI-1 there was little effect on the growth of *T. brucei* until day 5 where a slight decrease could be seen. The growth phenotype at higher concentrations could be due to the effect of NGI-1 on alternative pathways that cannot be predicted. It could also be that the speed of inhibition is faster with NGI-1 than that of RNAi therefore causing a more severe phenotype as the cells struggle to adapt as expected from figure 52.

The IC_{50} of NGI-1 on *T. brucei* was calculated from 6 concentrations to be 75.73 µM, figure 53. Compared to other trypanocidal like suramin (IC_{50} of 20 nM, (Boda *et al.*, 2005)), this value is high meaning that NGI-1 has a low potency for affecting growth. The IC_{50} often gives an indication of the efficacy of the drug. However, in the case of NGI-1, it could be significantly lower *in vivo* because the removal of N-glycosylation is lethal for *STT3B* knock downs *in vivo* but not *in vitro.*

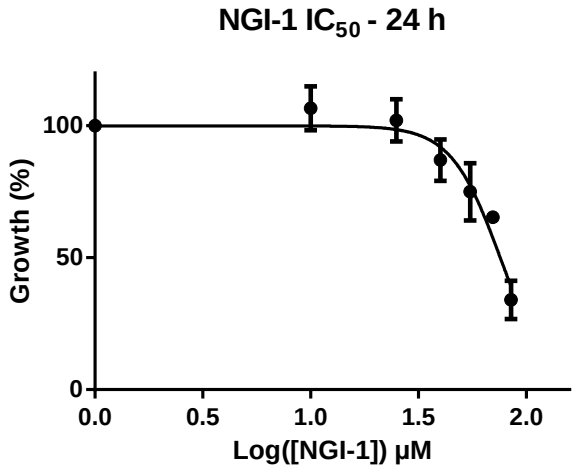


Figure 53 - IC50 plot of NGI-1 after 24 h of incubation with *T. brucei*. Results are an average of n = 3, error bars represent standard error.

10.1.4 NGI-1 appears to effect N-glycosylation

Lectin blots of cultures grown for 3 or 5 days in 50 µM NGI-1 (0.5% DMSO + NGI-1) or 0.5% DMSO were performed with both Concanavalin A (ConA) and ricin in order to visualise the effect on both types of glycosylation. ConA is a legume lectin extracted from jack beans that binds to α-D-Mannosyl and α-D-glucosyl groups and is therefore used to visualise high mannose glycans transferred by Stt3B. Ricin is also a legume lectin but is extracted from caster beans and binds to terminal galactose residues formed after the addition of $Man_5GlcNAc_2$ by

Stt3A (Liener, Sharon and Goldstein, 1986). These time points were chosen as the growth curves showed the greatest difference between the cultures from day 3, figure 53.

NGI-1 is dissolved in DMSO before being added to the medium, to improve its solubility. In blot A, figure 54, cells grown without DMSO or NGI-1 are shown in lane 1 as a negative control. From pounceau staining of the membrane, lanes 2 to 5 have a similar amount of protein present but lane 1 has less. ConA staining in lane 1 is comparable to lane 2 therefore there is no affect. Lanes 2 and 3 are from cells grown with or without NGI-1, respectively, for three days. There appears to be an increase in high mannose glycosylated proteins in the culture treated with NGI-1, the opposite of what was expected. It could be that proteins from lane 4 that are very strong in signal overflowed into lane 3, enhancing its signal and masking the effect of the drug. When looking at lanes 4 and 5 there is an obvious decrease in signal of ConA in the cells grown with NGI-1 indicating that NGI-1 successfully inhibited the Stt3B. Considering B, figure 54, ricin was used to assess if the reduction in high mannose glycans lead to an increase in complex glycans terminal in galactose. When Stt3B was knocked down, some previously high mannoses structures were replaced with complex (Izquierdo, Schulz, *et al.*, 2009). Again, lane 1 contained less protein than 2 and 3 but this time lanes 4 and 5 also had less protein compared to 2 and 3 (but similar between themselves). This makes comparisons between day 3 and day 5 difficult. Considering each day individually, there is little difference between cells grown with or without the drug. The signal on day 5 from ricin is very weak for both the control and the cells grown in NGI-1 and is difficult to interpret.

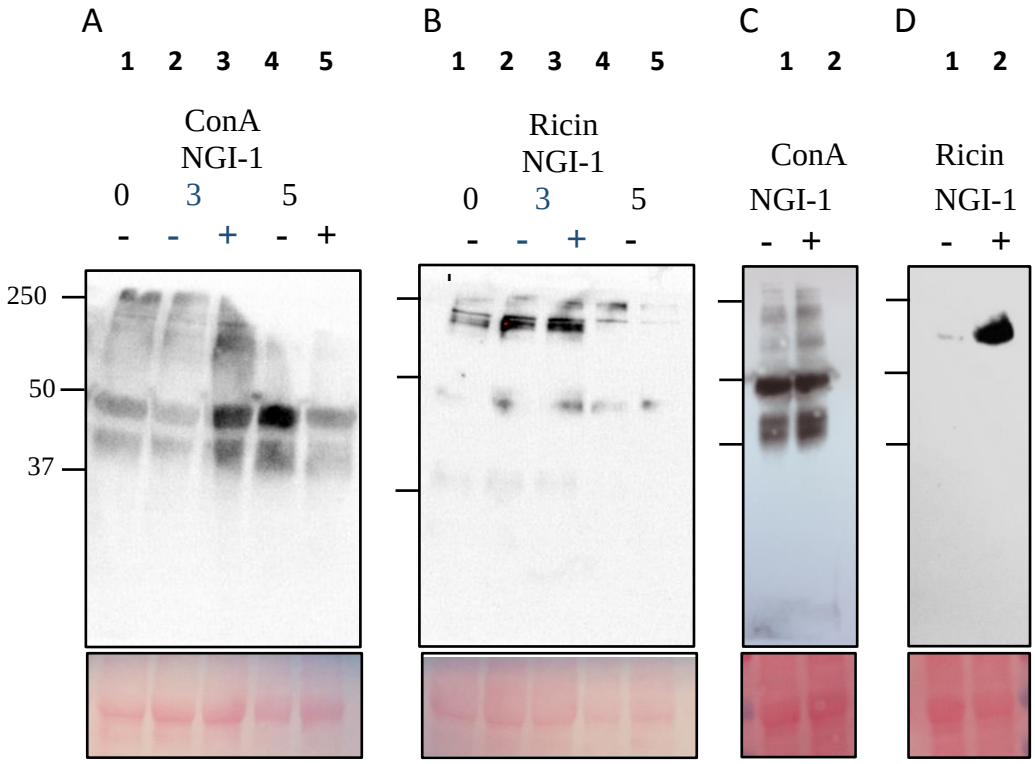


Figure 54- Cell lysate blots stained with lectin (upper panel) or pounceau red (lower panel). A: lane 1 = day 0 control, lane 2 = cultured for 3 days in 0.5 % DMSO, lane 3 = cultured for 3 days in 50 µM NGI-1, lane 4 = cultured for 5 days in 0.5 % DMSO, lane 5 = cultured for 3 days in 50 µM NGI-1 were incubated with ConA. B: lane 1 = day 0 control, lane 2 = cultured for 3 days in 0.5 % DMSO, lane 3 = cultured for 3 days in 50 µM NGI-1, lane 4 = cultured for 5 days in 0.5 % DMSO, lane 5 = cultured for 3 days in 50 µM NGI-1 were incubated with Ricin. C: lane 1 = cultured for 5 days in 0.5 % DMSO, lane 2 = cultured for 5 days in 50 µM NGI-1 were incubated with ConA. D: lane 1 = cultured for 5 days in 0.5 % DMSO, lane 2 = cultured for 5 days in 50 µM NGI-1 were incubated with ricin.

Blots C and D from figure 54 show whole lysates from cultures grown in 50 µM NGI-1 or 0.5% DMSO for five days. These conditions were repeated because of the different effects seen with the staining on ConA in A and the poor staining with Ricin in B. In C, there is no clear difference between lanes 1 and 2. In D, there is a drastic increase in signal in lane 2 compared to lane 1, suggesting a large increase in the amount of complex glycans present. This data is almost the opposite of the data from A and B.

To improve this assay, a positive control should be used. Tunicamycin is the most common positive control for the reduction of N-glycosylation and was used to compared to NGI-1 inhibition of RTK dependent cancer cell lines (Lopez-Sambrooks *et al.*, 2016). However, since this inhibitor is toxic and doesn't target high-mannose structures it isn't ideal. There are a number of other small molecule inhibitors that reduce complex and hybrid glycan formation

but non except NGI-1 reduce high mannose. Therefore, the best control would be to use a knock down of Stt3B. This control might not be the most suited since small molecular inhibition is not always equivalent to RNAi. A major difference between RNAi and small molecule inhibition is that RNAi removes the actual protein and therefore also removes any protein-protein interactions that were taking place. Therefore, by comparing the Stt3B knock down to NGI-1 the effects of protein-protein interactions, if any, could even be elucidated.

10.1.5 NGI-1's effect on variant surface glycoprotein 221

A more conclusive way to assess the effect of NGI-1 on N-glycosylation is to extract the most abundant glycoprotein, VSG. As described before, VSG221 of monomorphic *T. brucei* 427 strain contains two glycosylation sites: one high mannose and one complex. *T. brucei* blood stream forms were cultured in 50 µM NGI-1 for 3 or 5 days and their VSG was extracted and prepared for mass spectrometry analysis. The VSG is expected to lose the high mannose N-glycosylation site on exposure to NGI-1. This would lead to two potential VSGs: one with a single complex N-glycan and one with two complex N-glycans where the "empty" high mannose site has been glycosylated by Stt3A as was observed for the RNAi of Stt3B knock down (Izquierdo, Schulz, *et al.*, 2009).

VSG extracted from 3 days growth in in 50 µM NGI-1 was run on an SDS page gel. This was to visualise if the VSG extraction had been successful, to know the volume of elution of the VSG from the DE52 column and to see if any change in mass had occurred from the loss of an N-glycan, figure 55. Lanes 1 to 3 and 5 to 7 were loaded with extracted VSG from the DMSO control and NGI-1 respectively. The bands clearly show purified VSG and an elution volume of 1.5 mL was established (each lane contained 0.5 mL of eluted protein). The resolution of the gel is not sufficient to differentiate between the glycosylated forms.

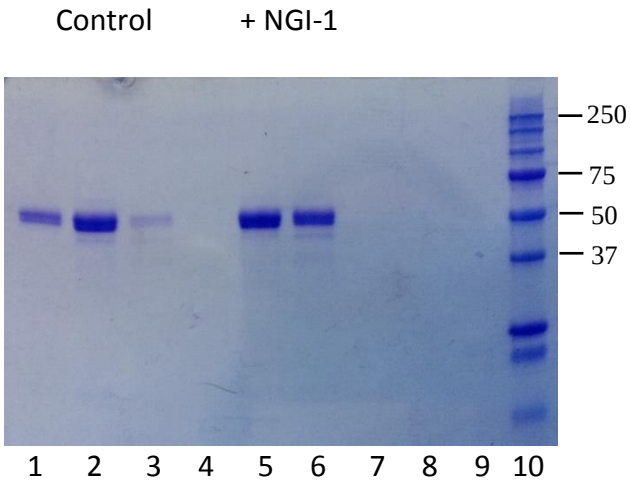


Figure 55 – Lanes 1 to 3: VSG extracted from control cells cultured for 3 days in 0.5% DMSO. Lane 4: blank. Lanes 5 to 7: VSG extracted from cells cultured for 3 days in 50 µM NGI-1. Lanes 8 and 9: blank. Lane 10: ladder.

The identity of the N-glycans present can be determined using Endo-H or PNGaseF as Endo-H cleaves high mannose N-glycans but PNGaseF cleaves complex N-glycans (Tarentino, Trimble and Plummer, 1989). By subjecting the VSG to these enzymes in turn and visualising if there is a change in mass by SDS-page the nature of the N-glycans present can be elucidated. Alternatively, these VSGs can be analysed by mass spectrometry and the exact structures of the N-glycans can be determined. This data is still in progress.

10.2 Chapter summary

The small molecule inhibitor of *STT3B*, NGI-1, proved to be an interesting tool to study the reduction of high mannose N-glycans in *T. brucei* through the inhibition of oligosaccharide transferase Stt3B. Although knock-downs of *TbSTT3B* are well characterised, small molecule inhibition is advantageous as it can be used in conjunction with RNAi or other small molecules and it is fast acting. NGI-1 effected the growth of the parasite when in concentrations of above 25 µM and was toxic in high concentrations with an IC_{50} of 75.73 µM. Lectin blots indicated that NGI-1 inhibits STT3B, but mass spectrometry analysis of the VSG will be able to confirm this. To develop NGI-1 as a therapeutic, structure activity relationship and docking studies could be performed with *TbSTT3B.* In this way, the specificity of the drug to *T. brucei's STT3B* could be increased and the toxic activity towards mammalian *STT3B* could be reduced.

10.3 Methods and materials

10.3.1 Cell culture

Blood stream form *Trypanosoma brucei* monomorphic strain 427 were grown at 37 °C in HMI-11 medium (For 1L: IMDM, 880 mL, FBS 100 mL (Gibco), hypoxanthine (stock at 13.6 mg/mL) 10 mL, bathocuproinedisulfonic acid (stock at 28.2 mg/mL) 1 mL, L-cysteine (stock at 18.2 mg/mL) 10 mL, pyruvic acid 69.2 μL and 2-mercaptoethanol (stock at 100 mM) 2 mL) and passed every day (enlarging the culture each time) until 3 or 5 days had passed and the cultures were 100 ml at $1x10^6$ cells/ml. Cultures grown in NGI-1 (AKOS) were passed into medium with fresh NGI-1 and those grown with DMSO were passed into medium with fresh DMSO. The cells were harvested by centrifugation (10 M for western blotting and 50 M for mass spectrometry).

10.3.2 IC50

Blood stream form *Trypanosoma brucei,* single marker monomorphic strain 427, were seeded at $1x10^5$ cells/ml in 6-well plates with NGI-1 at 0 μM, 10 μM, 25 μM, 40 μM, 55 μM, 70 μM or 85 μM. Each concentration was grown at 37 °C in triplicate and the cells were counted 24 h after the addition of NGI-1.

10.3.3 Western blots

The cell pellet was hypotonically lysed in 100 μL of 10 mM sodium phosphate buffer at pH 8 with 0.1 mM TLCK (Sigma, Missouri), 1 μg/mL leupeptin (Sigma, Missouri), and 1 μg/mL apronitin (Biochemica, Darmstadt) and incubated for 5 mins at 37 °C. Laemmli (Sigma, Missouri, 100 μL, 2 x conc.) was added and the solution incubated for 10 min at 99 °C. 30 μL was loaded on to a pre-cast SDS gel which was run at 60 V for 15 min then 100 V for approximately 100 min in TGS. The proteins were transferred to a nitrocellulose membrane *via* wet transfer for 1 h at 100V with 20% MeOH in TGS. The membrane was blocked for 1 h at 4 °C in 3% BSA in blocking buffer (1 L blocking buffer = 50ml TrisHCl 1M, 150ml NaCl 1M, 1 mL EDTA 1M, 500 μL NP40 and 798.5 mL H_2O). The membrane was then labelled either with ricin (1 μL in 5 mL 3% BSA in blocking buffer, stock at 1 μg/μL) or ConA (1 μL in 5 mL 3% BSA in blocking buffer, stock at 1 μg/μL) at 4 °C for 1 h. The membrane was washed with 3 x 5 mL blocking buffer for 10 min each wash. It was then labelled with ExtrAvidin-peroxidase (1 μL in

5 mL 3% BSA in blocking buffer, stock at 2 μg/μL, Sigma Missouri) at 4 °C for 1 h and then washed with 3 x 5 mL blocking buffer for 10 min each wash. The membrane was revealed with 1 ml of Immobilon western chemiluminescent HRP substrate (Millipore, Billerica) and images recorded on a chemidoc XRS+ (50 images over 300 sec).

10.3.4 VSG preparation

The cell pellet was washed twice with 50 mL TDB at 0 °C. It was transferred to an Eppendorf and washed a further three times (3 x 1 mL TDB). As much TDB as possible was removed and the pellet was re-suspended in 0.5 mL of pre-warmed 10 mM sodium phosphate buffer at pH 8 with 0.1 mM TLCK, 1 μg/mL leupeptin and 1 μg/mL appronectin and incubated for 5 mins at 37 °C. The Eppendorf was then spun for 5 min at max speed and the supernatant transferred to a fresh falcon on ice. This supernatant was applied to a mini column loaded with 300 μL of DE52 pre-equilibrated with 10 mM sodium phosphate buffer (pH 8). Once the sample was loaded on to the column, 1 mL of 10 mM sodium phosphate buffer (pH 8) was added. Another 1.5 mL of 10 m sodium phosphate buffer (pH 8) was added and the eluate was collected in an eppendorf (total 1.5 mL). This elute was loaded in 250 μL batches to an Amicon Ultra 10 kDa column and concentrated. Once loaded, the sodium phosphate buffer was removed with washes with mass spectrometry grade water (approx. 1 mL). The remaining eluate on top of the column was transferred to a low bind Eppendorf, dried and stored at -20 °C for mass spectrometry.

11 Chapter 4

Hunting for the elusive N-glycoproteins of *Plasmodium falciparum*

11. Hunting for the elusive N-glycoproteins of *Plasmodium falciparum*

11.1 Results and Discussion

This chapter has three main aims; (i) to develop a method to enrich glycoproteins from whole cell lysates of *P. falciparum*; (ii) to characterise the N-glycans present in *P. falciparum*; (iii) to identify the proteins that are N-glycosylated. Lectin affinity enrichment techniques were used to enrich glycoproteins from *P. falciparum lysates.* N-Glycans present and their associated proteins were characterised by mass spectrometry.

11.1.1 Author Contributions

J. G. M. Bevan, N. Cornillie, E. Aquilini and J. A. Rodrigues designed the experiments. J. G. M. Bevan performed all of the experiments. N. Cornillie and C. W. Lin submitted the samples of the Velos and galactose modified samples to mass spectrometry. J. G. M. Bevan, N. Cornillie and C. W. Lin performed the mass spectrometry data analysis. J. G. M. Bevan created all the figures and tables. J. A. Rodrigues mentored and supervised the work.

11.1.2 *P. falciparum* glycopeptide enrichment using the FASP FACE protocol

P. falciparum was cultured in O^+ RBCs until a high parasitemia (at least 20%) in the trophozoite form. This form contains the most proteins in the blood stage and is where there is a peak of Stt3 transcription (Roos, 2001). The red blood cells were lysed and the remaining parasites washed. The parasites were then lysed and subjected to the Filter Aided Sample Preparation (FASP) and Filter Aided Capture and Elution (FACE) protocols, figure 56.

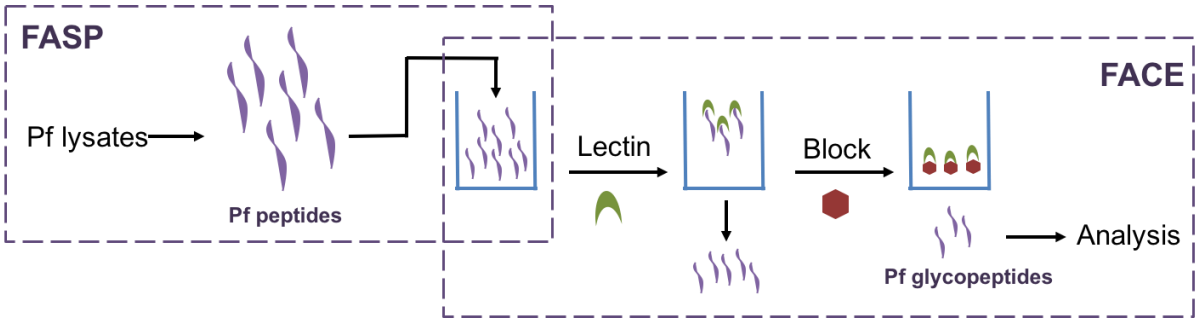


Figure 56 – Diagram of FASP and FACE protocol

This protocol enriches glycopeptides using an ultrafilter. Lectins bind to the glycopeptides, keeping them on top of the filter whilst the peptides with no affinity are washed away. The glycopeptides are then unbound by the addition of substrate with high affinity to the lectin (often a monosaccharide) and washed through the filter (Wiśniewski *et al.*, 2009; Pan *et al.*, 2011).

The lectins *Griffonia (Bandeiraea) Simplicifolia* Lectin II (GSLII) and Wheat Germ Agglutinin (WGA) were selected to enrich for *P. falciparum* glycopeptides because both lectins bind to GlcNAc residues. WGA also binds to sialic acid, both have affinity to blood groups A and B and WGA will also bind to blood group O. Both GSL II and WGA also have a lower affinity to GlcNAc compared to $GlcNAc_2$ therefore proteins conjugated to a single GlcNAc may not be enriched (Liener, Sharon and Goldstein, 1986). WGA is able to recognise GlcNAc but GSL II only recognises terminal GlcNAc, therefore together they should cover any type of GlcNAc in *P. falciparum*. Although these lectins have off-target specificities, washing steps will remove the majority of red blood cell proteins and the remaining can be distinguished from *P. falciparum* by using proteome databases. Alternative lectins are *Datura Stramonium, Lycopersicon esculentum* or *Solanum tuberosum* lectins. However, these lectins have a high affinity for oligomers of GlcNAc, over the mono or dimer, which are not expected to be present in *P. falciparum* (Liener, Sharon and Goldstein, 1986).

P. falciparum glycopeptides were eluted from WGA and GSLII with GlcNAc. These peptides were then desalted and analysed by mass spectrometry. Three fragmentation methods were used: Higher-energy collisional dissociation (HCD), Electron-transfer dissociation (ETD) and HCD with ETD triggered by oxonium ion 204 from GlcNAc. This ion is caused by the cleavage of one GlcNAc and would signal the presence of a glycopeptide.

Standard fragmentation is caused by collision induced dissociation (CID) which cleaves glycans from their peptides. HCD is fragmentation technique that keeps the glycan intact and improves the detection of low *m/z* ions enabling the detection of monosaccharides with a higher accuracy and lower background noise. It also produces what is named a Y1 ion which are from the peptide + GlcNAc making glycopeptide characterisation easier. However, it often cannot fully fragment the peptide backbone. ETD also keeps the glycan attached to the peptide. It is a complementary method that is more efficient in peptide fragmentation often elucidating different glycopeptides to HCD. A combination of the two techniques has proven to be effective and has lead to the identification of many N-glycopeptides (P. Zhao *et al.*, 2011; Singh *et al.*, 2012; Cao *et al.*, 2017).

The mass spectrometry facility performed a MASCOT search based on proteins from the PlasmoDB database using a peptide mass tolerance of 0.5 Da, searching for the HexNAc oxonium ion 204.0866 and the $HexNAc_2$ oxonium ion 407.1670. These results are summarised in table 6 and the hits are found in annex tables 14 to 17. There were a number of proteins identified as having either a HexNAc or $HexNAc_2$ from the data generated by either ETD or HCD fragmentation. However, the sequon required for N-glycosylation (NXT(S)) was absent indicating that this hit is not a true N-glycan. The HCD/ETD data gave no glycopeptides.

Performing searches using the MaxQuant software and the UniProt *P. falciparum* database on the HCD data gave a huge number of matches for O-glycosylated peptides but few for N-glycosylation. An advantage of the MaxQuant analysis is that it predicts where in the peptide the glycan is attached. This allowed for a rapid scan of those that contained sequons but for many it was not present. Those that had sequons present were manually analysed to confirm the presence of the HexNAc oxonium ion 204.0866 and the $HexNAc_2$ oxonium ion 407.1670 but to no avail.

Table 6 - Summary of results from a number of database searches against the different ionisation methods. More detailed results are in annex tables 14 to 17.

Search	Database	Raw file	Type	Hits	Sequon
MASCOT	PlasmoDb	ETD	HexNAc (N) and $HexNAc_2$ (N)	25	6
MASCOT	PlasmoDb	HCD/ETD	None	0	0
MASCOT	PlasmoDb	HCD	HexNAc (N) and $HexNAc_2$ (N)	2	1
MaxQuant	Uniprot	HCD	HexNAc (ST)	254	N/A
MaxQuant	Uniprot	HCD	HexNAc (N) and $HexNAc_2$ (N)	11	2
MaxQuant	Uniprot	HCD	HexNAc (N) and (ST)	22	7

There was also almost no overlap between all six searches: one protein in MaxQuant search of O- and N- was also present in MaxQuant N- search. This protein is ARP1, actin-related protein and the peptide that was identified has the sequence NGHLF**NTS**SEMEVVK. This peptide does contain the sequon in the middle but MaxQuant labels the first N in the sequence as the position for the modification. When the MS/MS spectra were manually analysed for the presence of the GlcNAc containing ion, no amino acids were resolved.

The fragment ions in this experiment were measured by an ion trap which is a lower accuracy and lower resolution mass analyser but it is more sensitive. The lower accuracy could be the reason behind the large number of false positives. The HexNAc ion is very low mass therefore any deviations in mass caused by the low accuracy have a large impact on their detection.

11.1.3 The orbitrap mass analyser for improved detection of *P. falciparum* glycopeptides

Samples of *P. falciparum* glycopeptides were prepared using the FACE FASP procotol described above. They were sent for analysis by mass spectrometry on an orbitrap mass analyser that has a higher accuracy and is commonly used in the identification of glycans. The resulting peptides were compared to a customised combined database of the proteomes of all genus of *Plasmodium* as well as the human proteome to exclude red blood cell contaminents on MASCOT. According to this search, all the proteins found were not glycosylated. ARP1 was again present but this time without a modification.

A Total Ion Chromatogram (TIC) shows all the ions found during the mass spectrometer scan. Analysis of this plot can provide information on the type of sample. Glycopeptides elute later in the scan therefore the profile of the TIC should have a peak at the end for an enriched sample. This is clearly the case with the sample that was run on the orbitrap, top - figure 57. In contrast, the TIC profile of the peptides that were not bound to the lectins had no such profile (bottom, figure 57), further validating the enrichment.

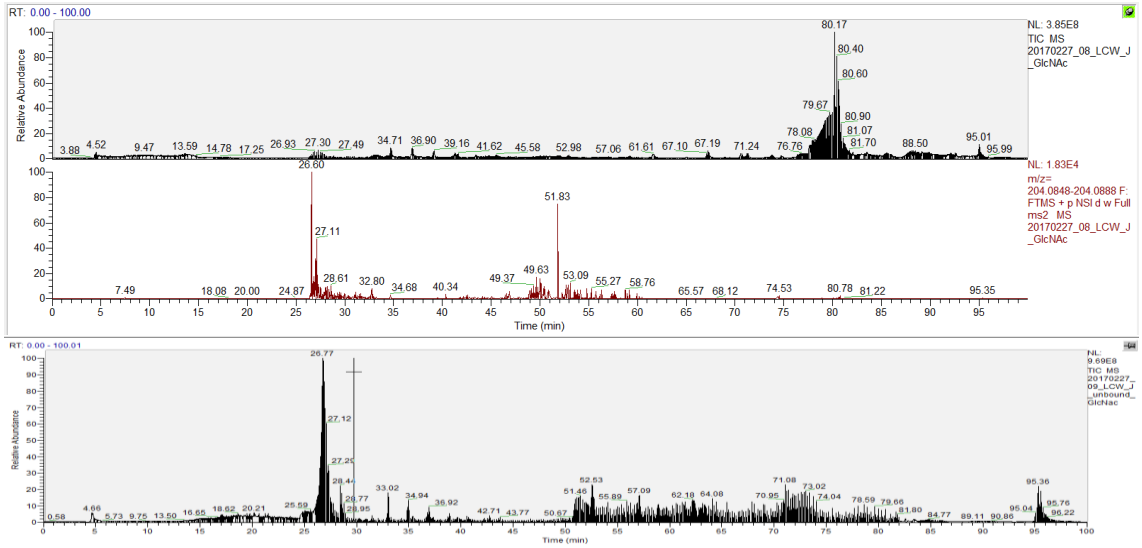


Figure 57 - Total Ion Chromatogram of enriched glycopeptides. Top: TIC of enriched glycopeptides with a search for HexNAc ion. Bottom: TIC of unbound peptides.

Manual searches with Xcalibur showed that there were a large number of glycopeptides with either a GlcNAc or $GlcNAc_2$ ion clearly present. MASCOT had not been able to identify the peptide, and therefore the protein, associated with these glycans. Interestingly, the mass difference associated with loss of a $FucGlcNAc_2$ (552) was observed in couple of peptides but the full sequence of the peptides could not be elucidated. This is particularly interesting as homologues of the GDP-Fucose *de novo* pathway, PfGMD and PfFS, and the fucosyltransferase PoFUT2 are found in *P. falciparum*'s genome (Sanz *et al.*, 2013). The surface of the parasite can be labelled with fucose binding lectin UEA which is reduced in the null mutant of PfGMD, but no fucosylated glycan has been identified so far (Sanz *et al.*, 2016).

Attempts were made to manually assign the peptides that contained a clear GlcNAc or $GlcNAc_2$ ion. To start interpretation the y1 ion (residue K at 147 or R at 175) was found. From there, mass differences were calculated between y1 ion and peaks to its right until the mass matched a peptide residue. Unfortunately, the full peptide sequence was never determined for all attempted GlcNAc or $GlcNAc_2$ peptides. The problems in identifying the peptide fragments

could be because it was not fragmented well enough, or the intensity of the fragments was not high enough to see. An example of a GlcNAc peptide is in figure 58. In this example, the oxonium ion for GlcNAc and the y1 ion for lysine are present. Attempts to work from lysine onwards gave some rare peptides that are unlikely and/or a dead end. The sequence AEE is potentially present as the mass differences match, but it was not possible to elucidate further.

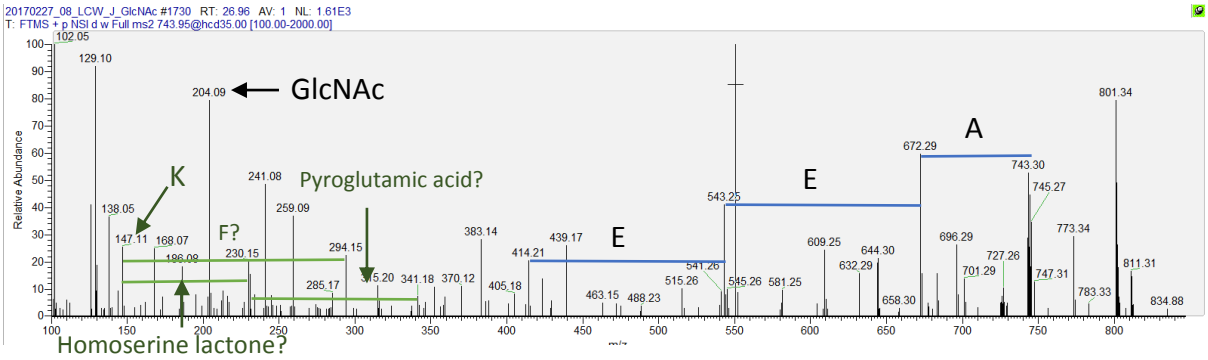


Figure 58 - Example of manual analysis of MS/MS spectra containing the 204.09 GlcNAc ion

To improve the signal from the glycopeptides, a larger culture was grown to a higher parasitemia (25%) increasing the amount of *P. falciparum* protein used in the FASP FACE protocol from 500 µg to 3.6 mg. Again GlcNAc and $GlcNAc_2$ ions were present but the increase in protein did not improve the quality of the spectra obtained. Haserick *et al.* had similar problems with the algorithm inbuilt in their peak assigning software not assigning glycopeptides in *C. parvum*. However, they were able to label their spectra manually (Haserick, Klein, *et al.*, 2017; Haserick, Leon, *et al.*, 2017).

11.1.4 Magnetic bead-bound lectins for the enrichment of *P. falciparum* glycoproteins

The FASP FACE method was proven to effectively enrich glycopeptides as observed by the TIC profile, figure 57. However, since the method uses ultra-centrifugal filters there is the possibility that some glycopeptides are lost in the washed in the step to remove non glycopeptides, that some remain bound to the lectin or that some are bound to the filter itself during both the FASP and the FACE stage. Amicon state that their 30 kDa ultra centrifugal filter has a 95% recovery of BSA when 4 mL is concentrated to 0.2 mL. However, the numerous washes involved in FASP FACE means that this ideal 95% would be lower. There are also proteins of lower Mw present in the sample which are more likely to be lost after many centrifguations.

In an effort to reduce the loss of glycopeptides, thereby improving the quality of the sample and subsequent mass spectrometry, magnetic beads bound to streptavidin were investigated. In theory, these beads can be bound to any lectin conjugated to biotin and be used in a pull downs of glycopeptides. The can be boiled to more efficiently release the bound peptides from the lectin and the number of washing steps is drastically reduced, figure 59.

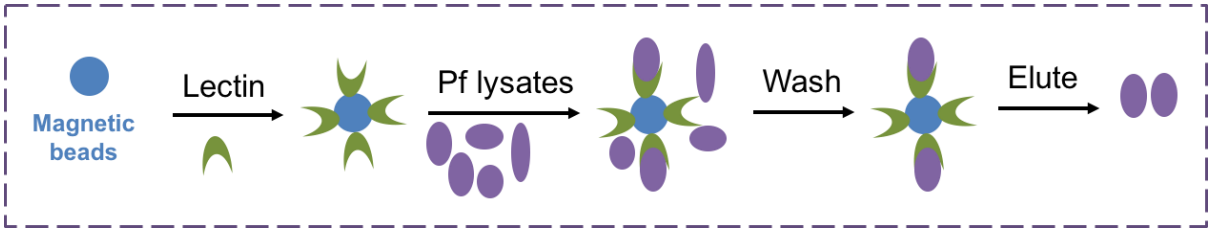


Figure 59 – Use of lectin bound to beads for enrichment of glycoproteins

Since the presence of *P. falciparum* N-glycoproteins is not yet proven, initial attempts at using magnetic beads were with *T. brucei* lysates and the lectin ricin due to the high abundance of Gal terminal N-glycans. *T. brucei* lysates were incubated with biotinylated ricin before MACs streptavidin beads were added. The beads were washed and bound proteins eluted by adding boiling laemmli buffer and loaded on to an SDS gel, gel A figure 60. The amount of protein loaded into lane 2 was a quarter of that used in the enrichment experiment. The enrichment seems to have been successful as a number of proteins appear higher in abundance in the enriched compared to the non. The same assay was therefore applied to *P. falciparum* lsyates.

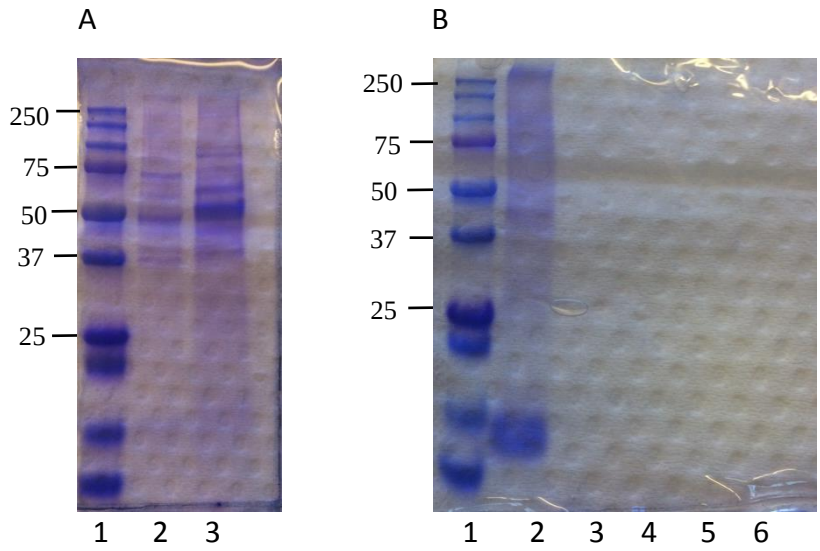


Figure 60 – A: Magnetic bead-bound ricin enrichment of *T. brucei.* Lane 1 = ladder, lane 2 = *T. brucei* lysates and lane 3 = *T. brucei* proteins eluted from the magnetic bead-bound ricin. B: Magnetic bead-bound GSLII and WGA enrichment of *P. falciparum*. Lane 1 = ladder, lane 2 = *P. falciparum* lysates, lane 3 = blank, lane 4 = *P. falciparum* proteins eluted from the magnetic bead-bound GSL II and WGA, lane 5 = *P. falciparum* proteins eluted from the magnetic bead-bound GSL II and lane 6 = *P. falciparum* proteins eluted from the magnetic bead-bound WGA.

This time GSL II (lane 5), WGA (lane 6) or both lectins (lane 4) were used to enrich the potential short GlcNAcs-proteins, gel B figure 60. Unfortunately, no proteins were seen in any elution lane (4 to 6) and the lane containing around a quarter of the proteins compared to that used in the enrichement assay only showed a poor smear with barely any protein present. There was also a need for a positive control – a protein that contains a GlcNAc for certain. BSA-GlcNAc is a commercially available protein that has been chemically labelled with a GlcNAc and is sold as a control in agarose-lectin binding assays of WGA or GSL II. It is can therefore be used as a positive control in our magnetic bead-lectin assay to determine if the lectins are not binding to *P. falciparum* lysates due to the absence of GlcNAc or if there is a problem with the lectin. Furthermore, binding the lectin to the proteins and then to the beads could lead to lectin that is not bound to any protein binding to the bead and diluting the result. Therefore, binding the lectins to the beads first was tested with the BSA-GlcNAc control. The beads were washed before the addition of the protein so that any excess lectin was removed. Both lectins were again tested together and separately, gel A figure 61. Only some high molecular weight oligomers of BSA-GlcNAc are seen in elution lanes 4 to 6. Four times more BSA-GlcNAc was used in the enrichment experiment compared to the control lane 2, therefore the small amount of high molecular weight proteins is nothing in comparison. From this experiment, it is also not possible to tell if the binding of the oligomers is due to the GlcNAc or non-specific interactions. A negative control should have been used where the lectin is not present. The band of protein at the bottom of the gel in lanes 4 to 6 is from streptavidin. It is possible that more lectin is needed than was used. The beads have a high enough loading of streptavidin to conjugated around 20 x more lectin than was used in this assay. Using more lectin could improve the binding of BSA-GlcNAc as steric blocking of the few lectins conjugated could have prevented more proteins from binding, especially as it seems like the predominant form of the protein is as oligomers.

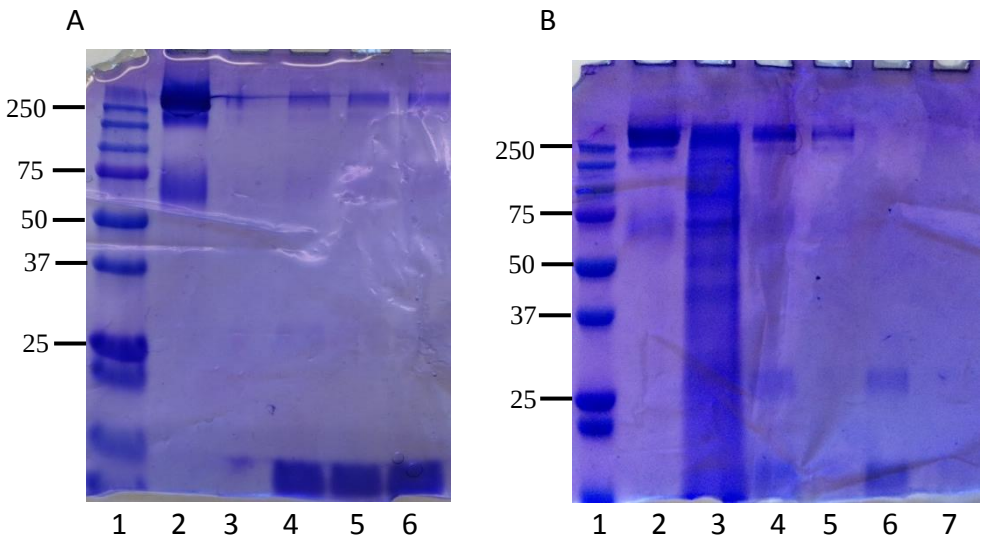


Figure 61 - A: Magnetic bead-bound GSLII and WGA enrichment of BSA-GlcNAc. Lane 1 = ladder, lane 2 = BSA-GlcNAc, lane 3 = blank, lane 4 BSA-GlcNAc eluted from magnetic bead-bound GSL II and WGA, lane 5 = BSA-GlcNAc proteins eluted from magnetic bead-bound GSL II and lane 6 = BSA-GlcNAc proteins eluted from magnetic bead-bound WGA. - B: Magnetic bead-bound GSLII and WGA enrichment of BSA-GlcNAc and *P. falciparum*. Lane 1 = ladder, lane 2 = BSA-GlcNAc, lane 3 = *P. falciparum* lysates, lane 4 BSA-GlcNAc eluted from magnetic bead-bound GSL II and WGA, lane 5 = BSA-GlcNAc proteins eluted from magnetic beads without lectin, lane 6 = *P. falciparum* proteins eluted from the magnetic bead-bound GSL II and WGA and lane 7 = *P. falciparum* proteins eluted from magnetic beads without lectin.

The assay was repeated with 10 x more lectin with both the positive control BSA-GlcNAc and fresh *P. falciparum* lysates and the negative control of beads that were not conjugated to lectins and freshly purchased Dynaeads. There was only a slight increase in BSA-GlcNAc in lane 4 where lectins were present compared to lane 5 where they are absent. This means that the interactions of the BSA-GlcNAc oligmoers is predominantly non-specfic. It could be that the high density of proteins in the oligomer sterically hinders the GlcNAc residue from binding. However, since this is sold as a standard for GSL II and WGA it would be unusual. There are no *P. falciparum* proteins present in lane 6 where lectin was present nor lane 7 where it was absent. In a way, this is a positive result since it means that the proteins do not interact with the beads in a non-specific manner. However, no glycoproteins were enriched. The band in lanes 4 and 6 just above 25 Da is biotin. GSL II has a molecular weight of 113 kDa and WGA 38 kDa. The action of boiling in laemmli buffer breaks the streptavidin bond releasing the lectins from the beads, but there are no bands of this Mw seen on the gel. Both lectins exist as dimers and although they were subjected to reducing conditions it is possible that they have remained as dimers (GSL II at 226 kDa, WGA at 76 kDa). There are bands at these molecular weights but they match those from the control lane of BSA-GlcNAc so are unlikely to be from

the lectins. It could be that the lectins needed longer than what they were given to react with the dynabeads.

11.1.5 Chemical modification of *P. falciparum* N-glycans to enable their enrichment

After trying numerous methods to pull down GlcNAc and GlcNAc2 residues from *P. falciparum* lysates without being able to identify any glycoproteins, chemically modifying the N-glycans to allow for a different pull down was considered. As previously described, β-1,4-galactosyltransferase adds a galactose to a terminal GlcNAc. If this enzyme was able to galactosylate *P. falciparum* N-glycans then they could be enriched with ricin instead of GSL II and WGA. This would serve two purposes; first ricin might improve the enrichment protocol by binding with a higher affinity to Gal residues than GSL II and WGA do to GlcNAc, secondly it provides a way to validate the GlcNAc peptides found previously. If a GalGlcNAc peptide is present in the ricin pull down and the GlcNAc present in pull down then it is almost certainly glycosylated. *R. communis* seeds contain two galactose binding lectins: ricin and *Ricinus communis agglutinin I* (RCA-I). RCA-I is a tetramer, therefore it binds with a higher avidity to terminal galactose residues than ricin (Liener, Sharon and Goldstein, 1986).

Dieckmann-schupper *et al*. used a similar technique to try and prove the presence of O-GlcNAc residues in blood stage form *P. falciparum* with radiolabeled UDP(^{14}C)Gal but they did not perform any glycoprotein enrichment. They found the amount of Gal labelled GlcNAc was almost the same in infected and non-infected erythrocytes but after many repeats the infected cells always had a higher quantity, suggesting the parasite is contributing to their presence. However, they were unable to identify any proteins with a GlcNAc modification (Dieckmann-Schuppert, Bause and Schwarz, 1993).

11.1.5.1 Using β-1,4-Galactosyltransferase to produce GalGlcNAc residues for characterisation

The reaction of β-1,4-Galactosyltransferase (β-1,4-GalT) with GlcNAc is very sensitive and requires optimisation before it can be used on a large scale with *P. falciparum.* The previously mentioned BSA-GlcNAc was used to optimise conditions. In order to visualise GlcNAc or Gal residues, dot blots were performed: samples of proteins are pipetted on to a nitrocellulose membrane and dried. The membrane can then be labelled with a biotinylated lectin and revealed by chemiluminescence. This technique allows rapid visualisation of the glycans present in the sample. By performing dot blots with GSL II and ricin of each sample the extent

of reaction with β-1,4-GalT will be determined. For all dot blots shown in this chapter, if a dot of protein is present but there is no signal or the signal is weak then it is circled. If the dot is clearly labelled, then it is not circled.

In this first dot blot are samples of BSA-GlcNAc incubated with either different amounts of β-1,4-GalT or UDP-Gal or a different buffer, figure 62. The exact conditions used can be found in annex table 18. The absence of signal in the GSL II blot is most likely because the membrane was contaminated with something that lead to a large increase in non-specific signal across the membrane. Only two conditions produced Gal-GlcNAc, giving rise to the signal at positions 5 and 6 on the ricin dot blot. These conditions used either an increase in concentration of HEPES or Tris-HCl and had no NP-40 present which may have been interfering with the enzyme's interactions.

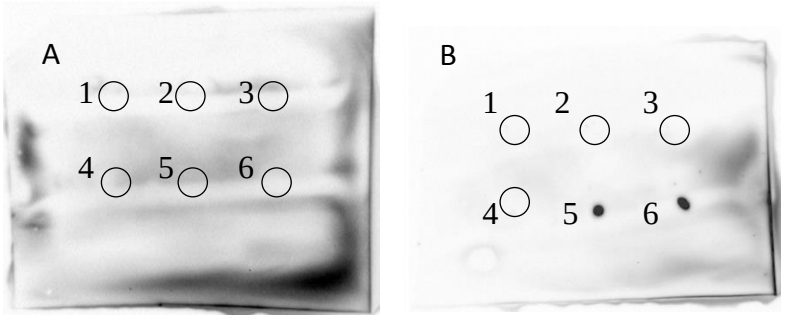


Figure 62 - Dot blots of BSA-GlcNAc revealed to β-1,4-GalT and UDP-Gal in different buffers. Blot A was exposed with GSL II and blot B was revealed with ricin. Dot 1: 1.7 µg GalT, 125 µm UDP-Gal, 4°C, o/n. Compared to dot 1: dot 2: less UDP-Gal, dot 3: less GalT and less UDP-Gal, dot 4: less time, higher temperature, dot 5: more concentrated HEPES and no NP-40, dot 6: 50 mM TrisHCl and no HEPES.

β-1,4-GalT is known to autogalactosylate and if this reaction is faster than that with the desired substrate it can drastically reduce the efficiency of reaction (Whiteheart *et al.*, 1989). Therefore, using 100 mM HEPES, the next dot blot was performed to assess if pre-incubating β-1,4-GalT before reacting it with our BSA-GlcNAc substrate would improve the efficiency of the reaction. The effect of increasing [$MnCl_2$], [UDP-Gal] and Gal-T were also assessed. Control blots were made of identical samples but incubated with added GlcNAc or Galactose and lactose for GSL II and ricin respectively. These sugars block the action of the lectin and therefore show if any labelling in the corresponding blot is specific or not. The full set of conditions are found in annex 19.

Control blots B and D show no non-specific binding of the lectins to the blots, figure 63. Blot C shows that all conditions produced galactosylated BSA-GlcNAc. However, blot A shows that for conditions 2, 6, 7 and 10 a small amount remains ungalactosyalted. Conditions 6, 7 and 10 are all pre-incubated before the addition of BSA-GlcNAc to encourage auto-galactosylation before the intended reaction ot improve efficiency. However, the fact that 3 of the 5 pre-incubated conditions still had non-galactosylated BSA-GlcNAc present suggests that the enzyme used up all of the UDP-Gal galactosylating itself. Conditions 8 and 9 had an increased amount of UDP-Gal and therefore were able to galactosylated BSA-GlcNAc as well as auto-galactosylated β-1,4-GalT. $MnCl_2$ is required for the enzyme to bind to its substrates. Adding a higher concentration of $MnCl_2$ (conditions 2 and 7) seems detrimental to the enzymes activity. $MnCl_2$ is known to be required for the binding of UDP-Gal and for stabilisation of the transition states on the formation of the glycosidic bond and the increase in activity on its addition is well described (Boeggeman and Qasba, 2002). $MnCl_2$ is slightly acidic, therefore increasing its concentration could have made the buffer too acidic and out of the optimal pH range of β-1,4-GalT.

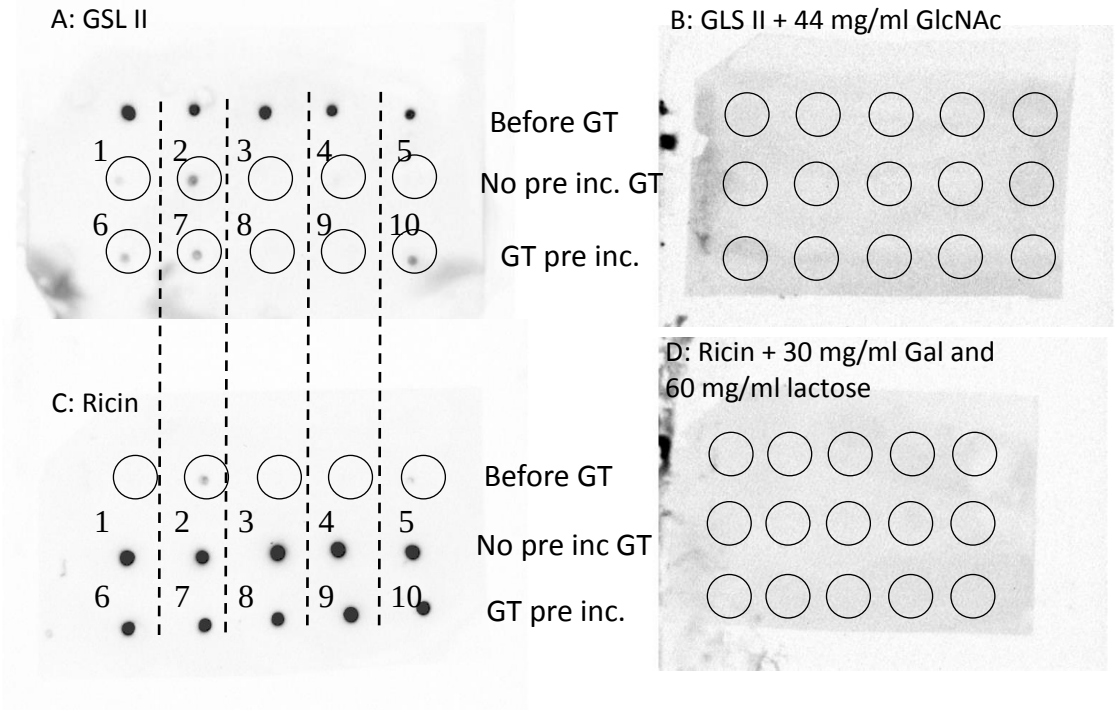


Figure 63 – Dot blots of BSA-GlcNAc exposed to β-1,4-GalT and UDP-Gal with varying [$MnCl_2$], [UDP-Gal] and [GalT], with (dots 6 to 10) and without (dots 1 to 5) pre-incubation of β-1,4-GalT with UDP-

Gal. Blots A and B were revealed with GSL II and blots C and D were revealed with ricin. Compared to conditions for dot 1: dots 2 and 7: more $MnCl_2$, dots 3 and 8: more UDP-Gal, dots 4 and 5: more UDP-Gal and more GalT and dots 5 and 10: more GalT. Dots 1 to 5 were incubated at 37 °C for an hour before addition of BSA-GlcNAc. All were incubated for an hour at 37 °C with BSA-GlcNAc.

After the addition of Gal to GlcNAc, UDP remains and is a known inhibitor of galactosyltransferases. Alkaline phosphatase is an enzyme that degrades UDP and can be used *in situ* to remove this inhibitory effect (Sujino *et al.*, 2000). A number of concentrations of alkaline phosphatase were tested but there was no major change in the conversion of BSA-GlcNAc to BSA-GlcNAcGal. Perhaps on a different scale (less β-1,4-GalT used or more substrate) alkaline phosphatase would prove more useful.

To optimise the amount of UDP-Gal and β-1,4-GalT used, two assays were preformed where the UDP-Gal or β-1,4-GalT concentrations were changed, figure 64. This was performed on a much larger scale than before (100 µg or 500 µg compared to 1 µg of BSA GlcNAc used previously) in order to mimic conditions with *P. falciparum* lysates where far more protein is present.

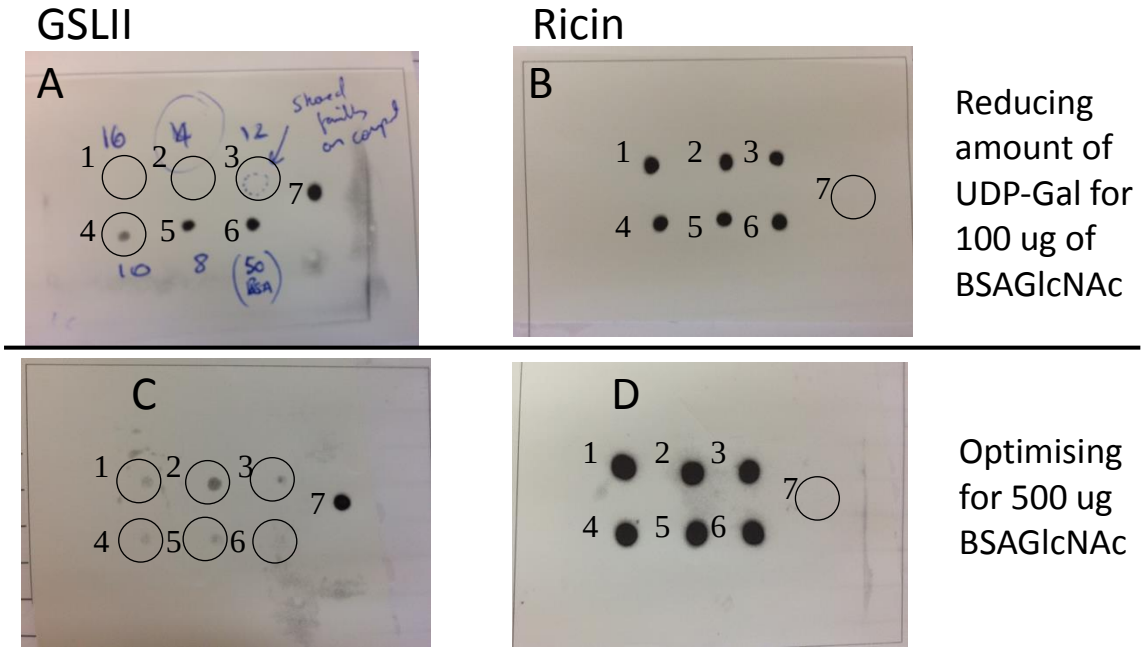


Figure 64: Blots A and C stained with GSLII. Blots B and D stained with Ricin. For blots A and B: Decreasing amount of UDP-Gal used for dots 1 to 6, no enzyme used in 7 (control) and 100 µg of BSA-GlcNAc is present in all dots. For blots C and D: 500 µg of BSA-GlcNAc is present in all dots, dots 1 to 3 used 1.7 µg of β-1,4-GalT and dots 4 to 6 used 2.5 µg of β-1,4-GalT. Dots 1 and 4 used 18 µL of UDP-Gal stock, dots 2 and 5 used 20 µL of UDP-Gal stock and dots 3 and 6 used 22 µL of UDP-Gal stock. Dot 7 was not exposed to β-1,4-GalT (UDP-Gal stock at 8.19 mM).

Blot B shows that all conditions where β-1,4-GalT was present produced BSA-GlcNAc as there is a signal for dots 1 to 6 and not for 7. However, blot A shows that at least 12 µL of the stock solution of UDP-Gal is required for complete conversion of 100 µg of BSA-GlcNAc as a faint signal is still seen in dot 3 indicating an incomplete conversion as the source of UDP-Gal was not sufficient or the concentration was not high enough for a 1 hour conversion.

Blot D shows that all conditions where β-1,4-GalT was present produced BSA-GlcNAc and in its absence, dot 7, no signal is produced. Blot C shows that almost all conditions are incomplete except dot 6. Although it had the same amounts of β-1,4-GalT as dot 4 and 5, there was more UDP-Gal present, driving the reaction to completion. Therefore, for every 500 µg of BSA-GlcNAc, 2.5 µg of β-1,4-GalT and 22 µL of UDP-Gal stock is required for the reaction to complete in 1 hour. The remaining signal seen in the other lanes could also be from non-galactosylated β-1,4-GalT rather than BSA-GlcNAc. On a dot blot it is not possible to differentiate. If the conditions are optimised so that both β-1,4-GalT and BSA-GlcNAc are fully galactosylated then the probability that all *P. falciparum* GlcNAcs are galactosylated under the same conditions is increased.

Using the optimised conditions, *P. falciparum* lysates were galactosylated and subjected to the FASP FACE conditions described previously except this time the pull down was performed with ricin in place of GSL II and WGA. It was not clear if the galactosylation would be more efficient before or after trypsinisation, both conditions were tested, figure 65.

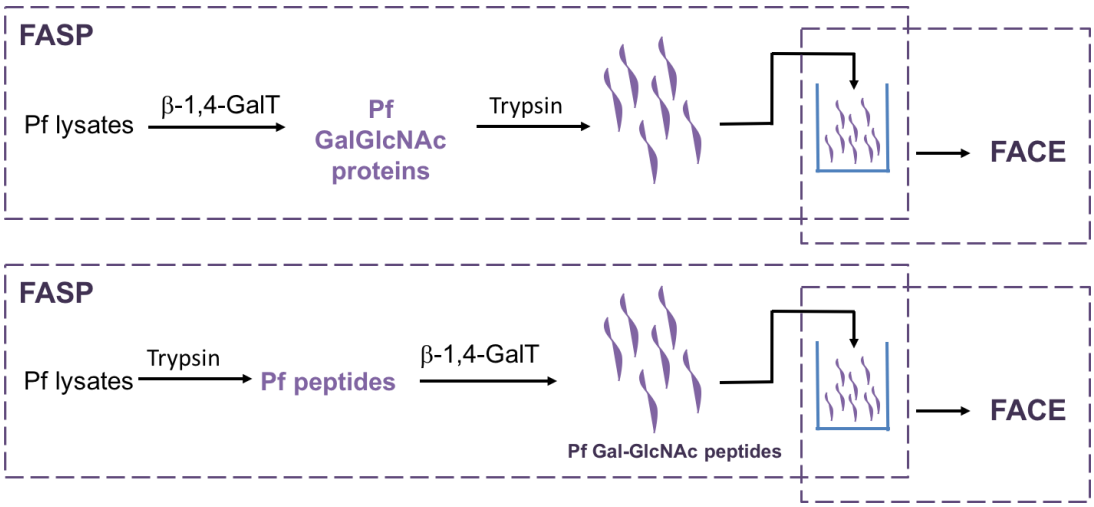


Figure 65- Illustration of the two methods by which *P. falciparum* lysates could be galactosylated. Top: galactosylation of proteins, bottom: galalcosylation of peptides.

Unfortunately, the TIC profiles of both pre- and post- trypsin galactosylation showed that no enrichment of peptides had occurred as there was no late-elution profile as seen with the GSL I and WGA pull down, figure 66.

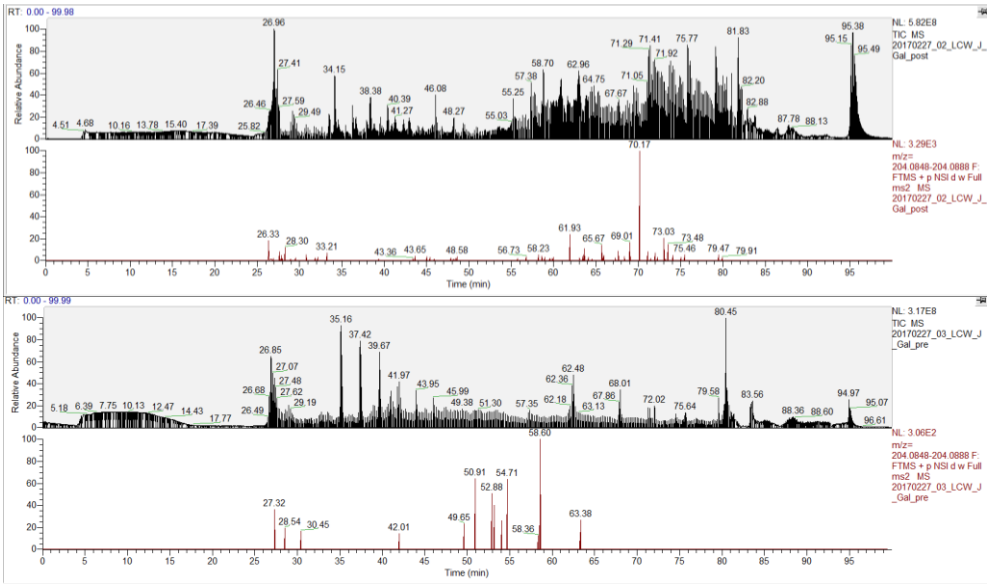


Figure 66 – *TIC profiles illustrating the poor enrichment by ricin. Top: Peptides that were galactosylated after being trypsinised. Bottom: Proteins that were galactosylated before being trypsinised.*

There are a couple of possibilities as to why the enrichment with ricin did not work. The addition of galactose might not have occurred or only occurred in small quantities. This could have been confirmed by dot blot analysis however dot blots of *P. falciparum* lysates suffer from contamination with red blood cell proteins that are labelled by both ricin and GSL II. It could be that there are simply too few, or no, GlcNAc terminating glycans to be galactosylated. This would, however, contradict our previous evidence. Perhaps ricin was not binding with enough affinity to enrich, but this seems unlikely as it is well used in pull downs of galactose containing glycoproteins from lysates (Izquierdo, Schulz, *et al.*, 2009).

11.2 Chapter summary

To identify N-glycosylated proteins in *P. falciparum* various methods based on lectin enrichment were used in order to prepare a sample for mass spectrometry analysis. Although there was evidence of truncated N-glycosylation (GlcNAc or $GlcNAc_2$), the resolution of the corresponding peptides was too poor to allow identification of the protein. Chemical modification using a GalT to theoretically produce GalGlcNAc or $GalGlcNAc_2$ followed by

enrichment with ricin was ineffective. Future work will involve improving the sample preparation so that the peptides can be resolved. This may be achieved by increasing the parasitemia to decrease the background of proteins from red blood cells.

11.3 Materials and methods

Plasmodium falciparum strain 3D7 was cultured in O^+ RBCs in RPMI medium with glutamine and HEPES added until 1 ml RBCs at between 15 and 20% parasitemia were obtained. The culture was pelleted and the pellet frozen at -20 °C. After defrosting, the pellet was washed with 3 x 40 mL PBS (centrifuging at 4,000 rpm for 10 min each wash). The pellet was then transferred to an Eppendorf and washed with a further 3 x 1 mL PBS (centrifuging at 13,000 rpm for 5 min each wash). The pellet was re-suspended in 480 μL lysis buffer with 20 μL pic (1 mL of lysis buffer = 0.4 mL 20% SDS, 0.1 mL 1 M DTT, 0.5 mL 2 % TX-114, 1 mL Tris-HCl pH 7.6 and pic = 1 pill in 400 μL H_2O). The dissolved pellet was heated at 95 °C for 1 min x 3 with 1 min pause on ice, sonicated on ice (2 min 30 sec at 10 microhertz then 30 sec pause x 3) and centrifuged for 20 min at 4 °C. The pellet was discarded and the supernatant kept for FASP.

11.3.1 Filter aided sample preparation (FASP)

Samples were prepared using a modified version of a published FASP procedure (Wiśniewski *et al.*, 2009). An Amicon 30kda column was pre-washed with 400 μL of ddH_2O. The sample was loaded on to the column (≈200 μg) before 200 μL of **UA** (8M urea in 0.1 M Tris HCl pH 8.5) was added and the column centrifuged at 14,000 g for 15 min. Another 200 μL of UA was added and the column centrifuged at 14,000 g for 15 min. The flow through was discarded and 100 μL of 0.05 M iodoacetamide in UA was added to the column. It was incubated in the dark for 20 min then centrifuged at 14,000 g for 10 min. The filter unit was washed twice with 100 μL of UA, each time centrifuging at 14,000 g for 15 min and discarding the flow through. It was then washed 4 x 100 μL 40 mM NH_4HCO_3, each time centrifuging at 14,000 g for 10 min and discarding the flow through. The filter unit was transferred to a new collection tube, trypsin-Lys-C added (promega, use 25:1 protein: protease, w: w) in 40 mM NH_4HCO_3 (0.4 μg trypsin/μL) and incubated overnight at 37 °C. The filter unit was centrifuged at 14,000 g for 10 min. The filter was washed twice with 40 μL of **BB** (40 mM Tris/HCl pH 7.6, 2mM $MnCl_2$, 2mM $CaCl_2$, 1 M NaCl) and centrifuged at 14,000 g for 10 mins. The filtrate was collected for FACE.

11.3.2 Filter aided capture and elution (FACE)

The filtrate from the FASP protocol (Pf peptides) was added to an amicon 30kda column (pre-washed with 400 μL of ddH_2O) and 36 μL GSLII (5 mg in 1 ml of 2 x BB) and 36 μL WGA (10 mg in 2 ml of 2 x BB) was added. The column was mixed at 550 rpm in a thermomix for 1 min and incubated without mixing for 60 min. The column was centrifuged at 14,000 g for 10 min and then washed with 200 μL of 1xBB with centrifuging the at 14, 000 g for 8 min three times. The flow-through from the collection tubes was put to one side (unbound and shouldn't have sugar, but kept for analysis) and the filter transferred to new collection tube. The column was washed twice with 200 μL of 1xBB followed by centrifugation at 14, 000 g for 8 min. It was then washed twice with 50 μL ABC centrifuging at 14,000 g for 8 min. The filter units were transferred to a new collection tube and 0.5 M GlcNAc in 40 ul ABC was added. The column was mixed at 550 rpm for 1 min and incubated in thermo-mixer for 1 h at room temperature. The filter units were centrifuged at 14, 000 g for 8 min and then washed twice with 40 μL ABC centrifuging the filter units at 14, 000 g for 8 min. The filtrate, containing enriched glycopeptides, was dried on a speed vac and kept for C18 purification.

11.3.3 Addition of UDP-Gal to GlcNAc residues of P. falciparum

P. falciparum peptides or lysates (500 μg) were incubated with β-1,4-GalT (2.55 μg), UDP-Galactose (25 μl of 8.19 mM stock) in 10 mM HEPES with 10 mM $MnCl_2$ and alkaline phosphatase (5 U) for 1 hour at 37 °C. The peptides were then used in the FACE protocol and the lysates in the FASP then FACE protocol.

11.3.3.1 C18 purification

All solvents used in this purification were of mass spectrometry grade. The solutions used were as follows:

Conditioning solution: 90/10 MeOH/H_2O with 0.1% TFA (v/v/v) (10 ul TFA per 10 ml)

Equilibration solution: 0.1% TFA in H_2O (v/v)

Desalting solution: 5% MeOH/ H_2O with 0.1% TFA (v/v/v)

Elution solution: 50/50 AcCN/ H_2O with 0.1% TFA (v/v/v)

The C18 SPE cartridge was conditioned by passing 3 ml conditioning solution through the packing bed. It was then equilibrated by passing 2 ml equilibration solution through the packing bed. The sample was reconstituted in up to 1 ml of 0.1-0.5% TFA. The pH was adjusted to <3 with formic acid if needed. The sample was loaded by passing slowly (1 drop/ sec) into

the cartridge. Loading was completed by passing 1 ml of equilibration solution into the cartridge. The sample was desalted by passing 1 ml desalting solution through cartridge. A new was placed below the cartridge for collection of the sample. The sample was eluted by slowly (1 drop/ sec) passing 1 ml of elution solution through cartridge. The sample was dried on a speed vac overnight.

11.3.4 Enrichment of glycopeptides using magnetic bead-bound lectins

Dynabeads (25 µL, 0.25 mg) were washed with PBS (3 x 1 mL, 7.4 pH) and then incubated in 100 µL of PBS with 25 µg of WGA or 25 µg of GSL II or 12.5 µg of each for 1 h at 4 °C. The beads were then washed with PBS (3 x 1 mL, 7.4 pH). *Plasmodium falciparum* lysates (as prepared above, 300 µL ≈ 600 µg) or BSA-GlcNAc (10 µg) were incubated in PBS (1 ml total for *P. falciparum* or 100 µL total for BSA-GlcNAc) for 1 h at 4 °C. The beads were then washed with PBS (3 x 1 mL, 7.4 pH). Bound proteins were eluted by boiling the beads in 50 µL laemmli buffer for 10 min at 95 °C and run on an SDS gel for visualisation.

11.3.5 Mass spectrometry

The samples were analyzed on a calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies, Dublin (CA), USA). The peptides were resuspended in 3% acetonitrile and 0.1% formic acid and loaded onto the self-made tip column (75 µm × 80 mm) packed with reverse phase C18 material (AQ, 3 µm 200 Å, Bischoff GmbH, Leonberg, Germany) and eluted with a flow rate of 300 nl per min by a gradient from 10 to 31.6% of solution B (99.9% ACN, 0.1% FA) applied for 28.5 min, from 31.6 to 44.8 % of solution B (99.9% ACN, 0.1% FA) applied for 25.5 min, from 44.8 to 98 % of solution B (99.9% ACN, 0.1% FA) applied for 2.5 min, and 98% B applied for 11.5 min. One scan cycle comprised of a full scan MS survey spectrum, followed by up to 10 sequential HCD MS/MS on the most intense signals above a threshold of 2000. The full-scan MS spectra (400–2000 m/z) were acquired in the FT-Orbitrap at a resolution of 60,000 at 400 m/z, while the HCD MS/MS spectra were recorded in FT-Orbitrap at a resolution of 15,000 at 400 m/z. HCD was performed with a target value of 1e5 and a collision energy from 35 V was applied. The AGC target values were 5e5 for full FTMS. For all experiments, dynamic exclusion was used with 1 repeat count, 15 s repeat duration, and 60 s exclusion duration.

The data was processed using MaxQuant15 version 1.3.0.5 which incorporates the Andromeda search engine. Proteins were identified by comparing against a protein sequence database containing *Plasmodium falciparum* and human annotated proteins and frequently observed contaminants. A peptide tolerance of 10 ppm, an MS/MS tolerance at 0.5 Da and full trypsin specificity, allowing for up to one missed cleavages were specified. Fixed modifications were set to carbamidomethylation of cysteine. Variable modifications were set to oxidation (M) and N-glycosylation (HexNAc and 2HexNAc).

12 Conclusions and future perspectives

With the field of glycobiology rapidly growing and the importance of N-glycosylation becoming widely recognised, tools to study rare glycosylation are highly sought after. This thesis has focussed on three aspects of profiling glycosylation; characterising glycosyltransferases responsible for building glycans; inhibiting an oligosaccharide that is responsible for the transfer of a glycan to proteins; characterising the proteins together with their glycans. Many pathogens require glycosylation to proliferate and cause disease. Ultimately, understanding in detail how they glycosylate could lead to new therapeutic targets.

Current methods to characterise glycosyltransferases involve finding homologous enzymes in other organisms. When this approach fails, there are no other methods to fall back on. Many pathogens synthesise unique glycans, therefore a method that can capture and subsequently characterise a glycosyltransferase involved in its synthesis could be invaluable to understanding these pathogens. These enzymes could serve as suitable drug targets in many diseases if well characterised.

In Chapter 1 of this thesis, the first synthesis of a resin bound nucleotide sugar has been described. Starting with methyl galactosylpyranoside, resin bound UDP-galactose was successfully bound to tentagel resin via and SMCC linker over 9 steps. Resin bound UDP-(4F)-galactose was then synthesised in 12 steps starting from glucose. The synthesis involved a modular approach whereby the phosphorylated sugar was produced with an alkyl chain terminating in a thiol at position 6. Tentagel resin was bound to SMCC and uridine phosphate was activated with the morpholidate. In doing so, we believe this strategy can be applied to almost any nucleotide sugar and these resin-bound nucleotide sugars could be used as probes for a wide range of enzymes. Tentagel is a versatile resin that has both hydrophobic and hydrophilic character making it the perfect support to use in chemistry and biology.

These particular UDP-Gal based resins can potentially act as affinity probes for the discovery and identification of unknown galactosyltransferases through our incorporation of the galactose moiety in the probe design. UDP-agarose is a well described affinity probe for glycosyltransferases but it is unable to distinguish between the different UDP-sugars therefore the designed probes would be of a great advantage. Initial biological assays using β-1,4-

galactosyltransferase indicate that the resin has an affinity for this UDP-Gal processing enzyme and that the affinity is higher for the fluorinated probe. However, the evidence was not conclusive as further experiments need to be performed with alternative controls. The most obvious being a control for extraction of proteins from the SDS gel. Furthermore, since the probes were designed to be an improvement upon UDP-agarose, a future experiment should be performed with the now optimised conditions to test these resins in parallel to UDP-agarose. Another interesting experiment would be to compare the resin's ability to bind to galactosyltransferases compared to other UDP-sugar processing enzymes such as α-1,3N-acetyl-galactosaminyltransferase (UDP-GalNAc) or β-1,3-N-Acetyl-Hexosaminyl-transferase (UDP-GlcNAc). This would investigate their selectivity to UDP-Galactose processing transferases. Equally, a test to see whether GDP-sugar transferases have an affinity would give invaluable insight into the probes specificity. β-1,4-galactosyltransferase is an inverting transferase and as inverting and retaining transferases operate by different mechanisms it would have been advantageous to prove the probe works in both cases. Future work could therefore involve the testing of α-1,3-GalT with the resins and analogues.

Future probe designs should focus on the inhibitors previously described in the introduction. Modification of the uridine moiety to produce strong inhibitors perhaps wouldn't retain the selectivity for galactosyltransferases and would create a general UDP-sugar transferase affinity probe. Therefore, modifications of galactose could be performed such as fluorination of C2 or introduction of the C-glycosyl moiety. Alternatively, combinations of the two modifications could be synthesised which could further increase the binding.

T. brucei synthesises an enormous array of glycans containing linkages that are unique to the parasite. Homologues of UDP-Gal/GlcNAc transferases have been identified but so far they have not been characterised. Galactosyltransferases are of particular importance to the parasite as shown by the essentiality of UDP-Gal and the prevalence of galactose in its glycans. An affinity probe that can distinguish between Gal and GlcNAc transferases and enrich galactosyltransferases for characterisation would provide a better understanding of the mechanisms behind and functions of *T. brucei* glycosylation. These enzymes are in an incredibly low abundance making affinity purification challenging. However, with the right probe design it should be possible.

In chapter 2, resin bound UDP-Gal proved unable to enrich galactosyltransferases from procyclic form *Trypanosoma brucei* lysates but it did show an affinity for GALE and a UDP-Gal transporter. Resin bound UDP-(4F)-Gal also had an affinity towards these two UDP-Gal processing enzymes. However, it also enriched GT3 3.5 fold compared to the UDP-Gal resin. This is a very positive result as GT3 is the only characterised galactosyltransferase of *T. brucei.* Disappointingly, no other enriched protein showed any GalT like properties.

The enrichment protocol with UDP-(4F)-Gal was only performed once. Therefore, with optimisation of incubation times or increased washing of the resin the protocol could be improved and the background noise of non-specifically binding proteins should be reduced. GalTs are low in abundance, therefore a reduction in background noise could reveal uncharacterised GalTs that were not present in previous trials. Furthermore, UDP-Gal was used as the control due to its proven low affinity in the previous experiments. The ideal control would be the UDP-(4F)-Gal resin with free UDP-(4F)-Gal to block any interactions of the galactosyltransferases. This is not a commercially available product therefore the next best control of UDP-Gal resin was used.

In future, an alternate strategy could be to synthesise affinity probes based on acceptors to target GalTs. For GPI anchor galactosyltransferases this would require only tri or tetra saccharide. The affinity might not be sufficient enough to bind the enzymes as they require the presence of the donor first, but perhaps with careful design or by using an irreversible inhibitor this could be successful.

GlcNAc is prevalent in many unique linkages of *T. brucei* therefore immobilised analogues of GlcNAc could make powerful probes and could help determine the specificity of the UDP-Gal/GlcNAc transferases. Future work could involve the synthesis of resin bound UDP-GlcNAc and the subsequent biological validation assays to ensure these resins have an affinity to GlcNAc-transferases. For example, 5-bromo-UTP is a known inhibitor of α-1,3-mannosyl-glycoprotein 2-β-*N*-acetylglucosaminyltransferase (Nishikawast *et al.*, 1988). They could then be used to characterise the missing *N*-acetylglucosaminyltransferases of *T. brucei*.

Moving away from biochemistry techniques, systematic knockdowns of the putative transferases or over expression could be performed. In doing so, the change in glycan structure can be determined and the identity of the transferase revealed. There are groups already performing these knock-downs, but the process is slow. Recombinant expression of these GTs could also be performed and their specificity tested in a glycan array type experiment. A major problem with recombinant expression is that the GT would not be glycosylated which can change its folding, affecting its enzymatic activity and rendering it dysfunctional.

In chapter 3, studies on the effects of NGI-1 with *T. brucei* showed that it reduces the parasites growth in blood stream forms *in vitro* and has an IC_{50} of 75.73 µM. Below 50 µM, the parasite's growth is comparable to those grown in the absence of NGI-1. Therefore, the study of NGI-1's predicted inhibition of *STT3B* was performed at low concentrations without toxic effects. Lectin blots performed on whole cell lysates indicate a change in N-glycosylation, with a reduction in high mannose structures and an increase in complex. This is in agreement with the predicated inhibition of *STT3B* that transfers the precursor to high mannose glycans. To confirm the change in glycosylation, VSG was extracted from cultures grown in the presence or absence of NGI-1. The VSGs were visualised on an SDS page gel and prepared for mass spectrometry analysis. The gel alone was not of high enough resolution to draw conclusions and the mass spectrometry analysis is still in progress. Once the mass spectrometry data has been processed, it should be clear if NGI-1 has inhibited *STT3B*.

An interesting experiment to perform would be to culture *STT3A* knock-down *T. brucei* blood stage forms with NGI-1. *In vitro,* the parasite requires Stt3a or Stt3b. If NGI-1 is specific to Stt3b, the effect of NGI-1 should be much greater when applied to the Stt3a knock-down as the mutant has a higher dependency on Stt3b. Alternatively, *in vivo* experiments could be performed in mice to try and use NGI-1 to treat a *T. brucei* infection. At low concentrations, NGI-1 is not toxic to mammalian cells nor *T. brucei in vitro.* However, *in vivo*, Stt3B is essential to the parasite. Therefore, the IC_{50} of NGI-1 *in vivo* is likely to be much lower than the 75.73 µM calculated *in vitro*. As such, NGI-1 remains an interesting small molecule with potential therapeutic value. However, mutations that reduce expression of *STT3A* and *STT3B* are known causes of congenital disorders of glycosylation in humans (Shrimal *et al.*, 2013). These

disorders are present from birth, therefore the effect of Stt3b inhibition in adults might be less severe. Although NGI-1 did not completely inhibit mammalian *STT3B*, studies would have to be performed to evaluate its toxicity to mammals *in vivo* before NGI-1 could be considered as a treatment for *Trypanosomiasis.* The huge economic burden of *Trypanosomiasis* in cattle demands more effective treatments. Targeting Stt3b with a small molecule inhibitor such as NGI-1 could lead to cheaper and more effective treatment methods.

The N-glycosylation of *Plasmodium* remains a mystery. Although there is strong evidence published that the parasite can synthesise short N-glycans, no glycoprotein with this motif has so far been characterised. From chapter 4, the mass spectrometry results from the FASP FACE protocol with GSL II and WGA were promising and showed that GlcNAc and $GlcNAc_2$ were present and successfully enriched, but without the identity of the peptide to which they were attached no more is known than before. The presence of the $GlcNAc_2Fuc$ motif suggests the use of GDP-fucose in N-glycosylation most likely as core fucosylation. Future work could involve performing a pull down using a fucose binding lectin such as CCL2, AAL or LEA to elucidate further the nature of this modification. The use of β-1,4-GalT to galactosylate was not successful probably due to the complexity of the lysates. Another approach would be to target specific proteins that have the N-X-S/T motif and are present in enriched samples and to perform an antibody based pull down of these proteins and analyse them alone (e.g. MESA and RMP1 appear often in enriched samples). Although this approach is lengthy, by combining the results from our experiments and similar published lists, the chances of enriching an N-glycoprotein should be improved. The main draw-back of this is that there are no commercial antibodies against these proteins, so proteins with already commercialised antibodies would be easier to enrich. Finding a method to produce good enough quality MS/MS to resolve the peptide structures remains a problem for the identification of glycopeptides.

In short, this work allowed the development of new methods for the discovery of carbohydrate processing enzymes and N-glycosylated proteins in two parasites with glycosylation pathways that are different from other eukaryotes. These methods require optimisation but, considering that these parasites have non-canonical glycosylation pathways, the experimental approaches followed in this work should be an effective way of discovering

and characterizing the N-glycosylation of their proteins. They can also be applied to other organisms who also display unusual glycosylation.

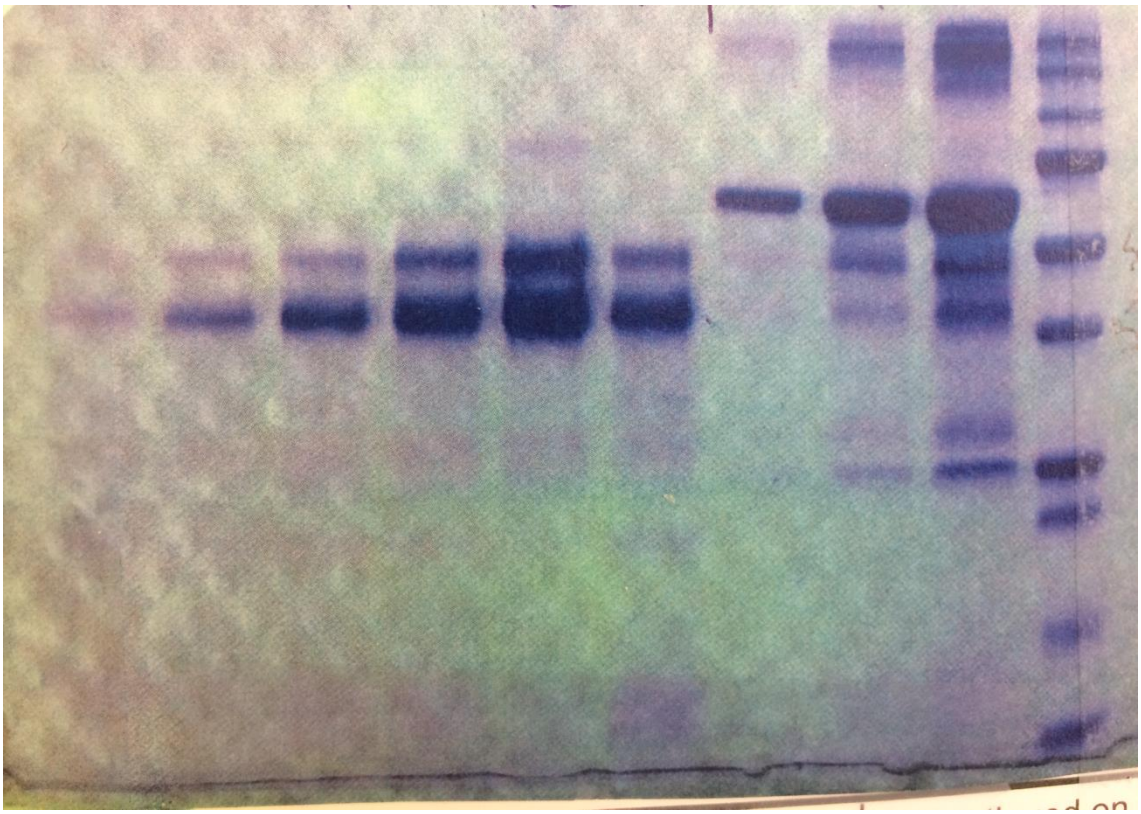
13 Annexes

Figure 67 – Gel to determine the optimal amount of β-1,4-GalT and BSA to use. *Heated at 50 °C.

----------β–1,4-GalT---------- ---------- BSA----------

μg of protein:

1 2 4 6 10 10* 1 4 10



Mass spectrometry results, Gel A:

Here each resin was incubated with VSG and BSA. If these proteins bound non-specifically to the resins, they should be detected in the elution. However, only the common contaminant Keratin was found.

Table 7 - Mass spectrometry results, Gel A (Both resins incubated with condition A)

Band	Type	Resin used	Proteins found
A1	Wash 1	UDPGal	Serum albumin, Bos taurus
A2	Wash 1	UDPGal	Variant surface glycoprotein, Trypanosoma brucei; Keratin, Homo sapiens Observed peptides, but no significant protein score: Serum albumin, Bos taurus
A3	Elution	UDPGal	Keratin, Homo sapiens
A4	Elution	UDPGal	Keratin, Homo sapiens
A5	Standard	-	Serum albumin, Bos taurus
A6	Standard	-	Variant surface glycoprotein, Trypanosoma brucei Observed peptides, but no significant protein score: Serum albumin, Bos taurus; Keratin, Homo sapiens
A7	Wash 1	UDP-4F(Gal)	Serum albumin, Bos Taurus Observed peptides, but no significant protein score: Keratin, Homo sapiens
A8	Wash 1	UDP-4F(Gal)	Variant surface glycoprotein, Trypanosoma brucei; Keratin, Homo sapiens Observed peptides, but no significant protein score: Serum albumin, Bos taurus
A9	Elution	UDP-4F(Gal)	Keratin, Homo sapiens
A10	Elution	UDP-4F(Gal)	Keratin, Homo sapiens

Mass spectrometry results, Gel B:

Here each resin was incubated with β(1,4)-GalT, VSG and BSA. If β(1,4)-GalT bound specifically to the resins, it should be detected in the elution. For the UDP-Gal resin, β(1,4)-GalT was found in the elution, but not in a large quantity. For the UDP-4F(Gal) resin, β(1,4)-GalT was found in the elution (along with the contaminant keratin. Neither VSG nor BSA were found in the elutions, supporting our hypothesis that the resins bind specifically to β(1,4)-GalT.

Table 8 – Mass spectrometry results, gel B

Band	Type	Resin used	Proteins found
B1	Wash 1	UDPGal	Serum albumin, Bos Taurus. Observed peptides, but no significant protein score: Keratin, Homo sapiens.
B2	Wash 1	UDPGal	Beta-1,4-galactosyltransferase 1, Bos taurus; Serum albumin, Bos taurus
B3	Wash 1	UDPGal	Beta-1,4-galactosyltransferase 1, Bos taurus
B4	Elution	UDPGal	Keratin, Homo sapiens
B5	Elution	UDPGal	Keratin, Homo sapien
B6	Elution	UDPGal	Keratin, Homo sapiens. Observed peptides, but no significant protein score: Beta-1,4-galactosyltransferase 1, Bos taurus
B7	Standard	-	Serum albumin, Bos taurus
B8	Standard	-	Serum albumin, Bos taurus; Beta-1,4-galactosyltransferase 1, Bos Taurus. Observed peptides, but no significant protein score: Keratin, Homo sapiens
B9	Standard	-	Beta-1,4-galactosyltransferase 1, Bos taurus. Observed peptides, but no significant protein score: Serum albumin, Bos taurus
B10	Wash 1	UDP-4F(Gal)	Serum albumin, Bos taurus
B11	Wash 1	UDP-4F(Gal)	Beta-1,4-galactosyltransferase 1, Bos taurus; Serum albumin, Bos taurus
B12	Wash 1	UDP-4F(Gal)	Beta-1,4-galactosyltransferase 1, Bos taurus; Serum albumin, Bos Taurus. Observed peptides, but no significant protein score: Keratin, Homo sapiens
B13	Elution	UDP-4F(Gal)	Keratin, Homo sapiens
B14	Elution	UDP-4F(Gal)	Keratin, Homo sapiens
B15	Elution	UDP-4F(Gal)	Beta-1,4-galactosyltransferase 1, Bos taurus; Keratin, Homo sapiens

Table 9 – Summary of NSAF counts for Control and Elution of Conditions 1 to 3

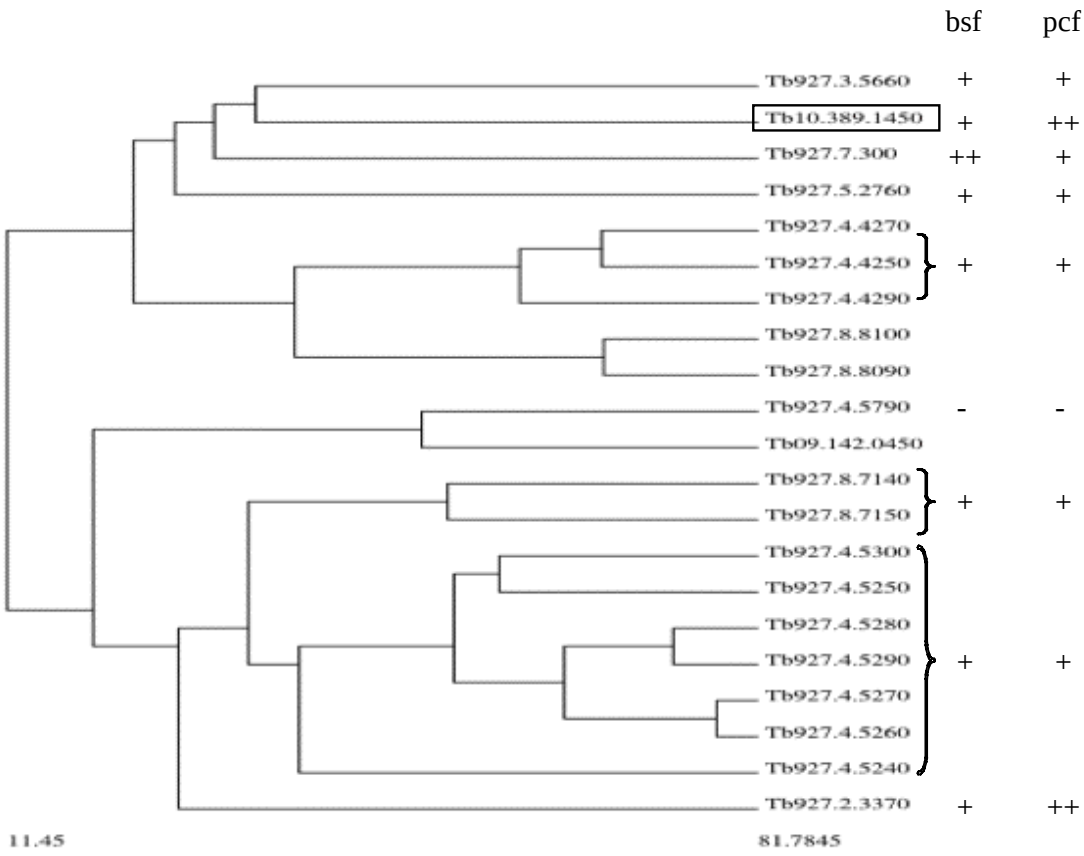
	Control			Elution		
	Spectrum Counting NSAF [fmol]					
Protein Description	1	2	3	1	2	3
Alpha-1-antiproteinase (A1AT_BOVIN)	0	388	340	0	0	0
Alpha-2-macroglobulin (A2MG_BOVIN)	242	270	233	0	0	0
Serpin A3-3 (SPA33_BOVIN)	0	0	248	0	0	0
Sesquipedalian-2 (SESQ2_BOVIN)	0	0	350	0	0	0
Alpha-2-HS-glycoprotein (FETUA_BOVIN)	**1960**	**1746**	**2683**	**2093**	**1334**	**5314**
Fetuin-B (FETUB_BOVIN)	640	1253	920	0	0	0
Green fluorescent protein (GFP_AEQVI)	**26415**	**20618**	**20582**	**11232**	**10919**	**16520**
Keratin, type II cytoskeletal 5 (K2C5_BOVIN)	0	599	210	0	0	0
Protein AMBP (AMBP_BOVIN)	0	0	288	0	0	0
Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4_BOVIN)	0	0	196	0	0	0
Lactotransferrin (TRFL_BOVIN)	684	321	202	414	0	0
Antithrombin-III (ANT3_BOVIN)	0	223	293	0	0	0
Bovine Serum albumin (ALBU_BOVIN)	**3324**	**4404**	**4555**	**4763**	**2074**	**3262**
Alpha-1B-glycoprotein (A1BG_BOVIN)	0	602	644	0	0	0
Variant Surface Glycoprotein	6961	5527	8124	2157	9178	5345
Ribonuclease pancreatic (RNAS1_BOVIN)	**981**	**6340**	**11287**	**0**	**0**	**4839**
Beta-1,4-galactosyltransferase 1 (B4GT1_BOVIN)	**3738**	**5765**	**1127**	**6179**	**2977**	**1827**
Cationic trypsin (TRY1_BOVIN)	1088	1590	2360	0	0	0
Complement C3 (CO3_BOVIN)	239	65	161	0	0	0
Chitinase-3-like protein 1 (CH3L1_BOVIN)	661	0	0	0	0	0
Inter-alpha-trypsin inhibitor heavy chain H3 (ITIH3_BOVIN)	224	0	0	0	0	0
Polyubiquitin-B (UBB_BOVIN)	121	85	0	0	0	0
Polymeric immunoglobulin receptor (PIGR_BOVIN)	227	0	0	0	0	0
Hemoglobin subunit alpha (HBA_BOVIN)	1036	0	0	1960	4998	4035
Alpha-S1-casein (CASA1_BOVIN)	856	0	0	0	0	0
Transthyretin (TTHY_BOVIN)	0	704	0	0	0	0
Alpha-1-acid glycoprotein (A1AG_BOVIN)	0	1294	0	0	0	0
Polymeric immunoglobulin receptor (PIGR_BOVIN)	0	320	0	0	0	0
Kininogen-1 (KNG1_BOVIN)	0	196	0	0	0	0
Keratin, type II cytoskeletal 79 (K2C79_BOVIN)	0	472	0	0	0	0
Keratin, type I cytoskeletal 10 (K1C10_BOVIN)	0	555	0	0	0	0
Lysozyme C, milk isozyme (LYSM_BOVIN)	0	0	0	1881	3197	0

Hemoglobin subunit beta (HBB_BOVIN)	0	0	0	1920	1632	0
Hemoglobin fetal subunit beta (HBBF_BOVIN)	0	0	0	0	1632	0
Annexin A2 (ANXA2_BOVIN)	0	0	0	0	698	0
Arginase-1 (ARGI1_BOVIN)	0	0	0	0	813	0

No.[a]	Donor	Acceptor	Anomer	Linkage	Stage[b]	Context[c]	Status[d]	Gene[e]	*T. brucei* homologue	Evidence
1	UDP-GlcNAc	Dol-P	α	N/A	bsf+pcf	N-link core	common	ALG7	Tb11.01.2220	
2	UDP-GlcNAc	βGlcNAc	β	1,4	bsf+pcf	N-link core	common	unknown		
3	GDP-Man	βGlcNAc	β	1,4	bsf+pcf	N-link core	common	ALG1	Tb10.389.0250	
4	GDP-Man	βMan	β	1,3	bsf+pcf	N-link core	common	ALG2	Tb927.4.2230	
5	GDP-Man	αMan	α	1,2	bsf+pcf	N-link core	common	unknown		
6	GDP-Man	αMan	α	1,2	bsf+pcf	N-link core	common	ALG11	Tb09.211.0860	
7	GDP-Man	αMan	α	1,6	bsf+pcf	N-link core	common	unknown		
8	Dol-P-Man	αMan	α	1,3	bsf+pcf	N-link core	common	ALG3	Tb10.70.0260	(Manthri *et al.*, 2008)
9	Dol-P-Man	αMan	α	1,6	bsf+pcf	N-link core	common	ALG12	Tb927.2.4720	(Leal *et al.*, 2004)
10	Dol-P-Man	αMan	α	1,2	bsf+pcf	N-link core	common	ALG9	Tb927.6.1140	
11	UDP-Glc	αMan	α	1,3	bsf+pcf	N-link core	common	UGT	Tb927.3.4630	
12	UDP-GlcNAc	αMan	β	1,2	bsf	N-link branch	common	GnT1		
13	UDP-GlcNAc	αMan	β	1,2	bsf	N-link branch	common	GnT2		
14	UDP-Gal	βGlcNAc	β	1,4	bsf	N-link branch	common	β4GalT1		
15	UDP-Gal	βGal	α	1,3	bsf	N-link branch	common	α1-3GalT		
16	UDP-GlcNAc	βGal	β	1,3	bsf	N-link branch	common	β3Gn-T2	several candidates	
17	UDP-GlcNAc	βGal	β	1,6	bsf	N-link branch	common	IGnT		
18	UDP-GlcNAc	inositol	α	1,6	bsf+pcf	GPI core	common	PIG-A/GPI3*	Tb927.2.1780	
19	Dol-P-Man	αGlcN	α	1,4	bsf+pcf	GPI core	common	PIG-M/GPI14*	Tb927.6.3300	
20	Dol-P-Man	αMan	α	1,6	bsf+pcf	GPI core	common	PIG-V/GPI18	Tb10.389.0300	
21	Dol-P-Man	αMan	α	1,2	bsf+pcf	GPI core	common	PIG-B/GPI10	Tb10.70.1440	(Nagamune *et al.*, 2000)
22	UDP-Gal	αMan	α	1,3	bsf	GPI branch	unique			
23	UDP-Gal	αGal	α	1,2	bsf	GPI branch	unique			
24	UDP-Gal	αGal	α	1,6	bsf	GPI branch	unique			
25	UDP-Gal	αGal	α	1,2	bsf	GPI branch	unique			
26	UDP-Gal	αMan	α	1,2	bsf	GPI branch	unique			
27	UDP-Gal	αMan	β	1,3	bsf	GPI branch	unique			
28	UDP-Gal	αMan	α	1,3	pcf	GPI branch	unique			
29	UDP-Gal	αGal	β	1,3	pcf	GPI branch	unique			
30	UDP-GlcNAc	βGal	β	1,3	pcf	GPI branch	unique		Tb10.389.1450	This study
31	UDP-GlcNAc	βGal	β	1,6	pcf	GPI branch	unique			
32	UDP-Gal	βGlcNAc	β	1,4	pcf	GPI branch	unique			
33	UDP-Gal	βGlcNAc	β	1,3	pcf	GPI branch	unique			
34	UDP-GlcNAc	βGal	β	1,6	pcf	GPI branch	unique			
35	UDP-GlcNAc	βGal	β	1,3	pcf	GPI branch	unique			
36	UDP-Gal	αMan	?	1,3	pcf	GPI branch	unique			
37	GDP-Man	Dol-P	β	N/A	bsf+pcf	Other	common	DPMS	Tb10.70.2610	(Mazhari-Tabrizi et al., 1996)
38	UDP-Glc	5-HmU	β	N/A	bsf	Other	unique			

[a] No. Refers to the glycosidic linkage number in Fig. 1
[b] Stage Refers to whether that glycosidic linkage is found in bloodstream form (bsf) and/or procyclic form (pcf) *T. brucei*
[c] Context refers to the structural context in which that glycosidic bond is formed
[d] Status refers to whether the linkage is unique to T. brucei or common to many eukaryotes
[e] Gene refers to common names for orthologues in yeast or mammals

Figure 68 - Predicted glycosyltransferases of *Trypanosoma brucei* from Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J Barton, et al., 2009



Putative UDP-Gal/UDP-GlcNAc-dependent glycosyltransferases in *T. brucei* genome. (A) Phylogenetic analysis of putative GlcNAc/Galactosyltransferases in the T. brucei genome. The box marks the gene studied in this paper. The results of Semi-quntitative RT-PCR analysis (see Figure S2) are summarised on the right. No or low expression at the mRNA level is indicated by (-). Expression at the mRNA level is indicated by (+) and greater expression by (++). (B) Multiple sequence aligment of the predicted amino acid sequences of the putative glycosyltransferases shown in panel A visualised by Jalview (Clamp et al., 2004). The putative transmembrane domains and DXD motifs of the ORFs are indicated by filled blue and green boxes respectively. The predicted amino acid sequence of Tb10.389.1450, the gene studied in this paper, is marked with an open red box. A scheme of the predicted secondary structure of this ORF, predicted by the Jnet algorithm (Cole et al., 2008, Cuff & Barton, 2000), is included and aligned at the bottom.

Figure 69 -Predicted glycosyltransferases of Trypanosoma brucei from Izquierdo, Nakanishi, Mehlert, Machray, Geoffrey J Barton, et al., 2009

Table 10 - UDP-Gal probe 1st trial: Enriched proteins

Protein IDs	Product of lead protein	Ratio H/L
Tb927.11.12220	Catalytic subunit of the vacuolar transporter chaperone 4	**5.34**
Tb927.10.10170	Predicted WD40 repeat protein	**2.84**
Tb927.10.4440 *Tb927.9.8210*	Predicted SAP domain protein	**2.70**
Tb927.4.1240	Hypothetical protein	**2.14**
Tb927.2.3080	Conserved protein	**2.13**
Tb927.2.4110	Mitochondrial-processing peptidase subunit alpha (MPPA)	**2.12**
Tb927.11.6790	Predicted WD40 repeat protein	**2.11**
Tb927.10.16120	Inosine-5-monophosphate dehydrogenase	**2.11**
Tb927.1.4550	Unspecified product	**2.10**
Tb927.7.1130 *Tb927.7.1140* *Tb927.7.1550*	Trypanothione/tryparedoxin dependent peroxidase 2 (TDPX2)	**2.05**
Tb927.3.5250	Zinc finger CCCH domain-containing protein 8 (ZC3H8)	**2.02**
Tb927.9.11900 *Tb927.4.5330*	Acyl transferase-like protein	**2.01**
Tb927.11.1820	Guanine nucleotide exchange factor subunit GCD2	**2.01**
Tb927.7.3440 *Tb927.7.3450*	I/6 autoantigen	**2.00**
Tb11.1550 *Tb927.8.4490* *Tb10.v4.0258* *Tb927.10.1510* *Tb927.8.2050*	Hypothetical protein	**2.00**
Tb927.10.11990	RNA-binding protein	**2.00**
Tb927.10.7500 *Tb927.10.8230*	Fibrillarin (NOP1)	**1.90**
Tb927.9.4190	Fatty acyl coa syntetase 1 (ACS1)	**1.88**
Tb927.4.3420	Wee1-like protein kinase	**1.88**
Tb927.9.3070	Unspecified product	**1.87**

Table 11 - UDP-Gal probe 2nd trial: Enriched proteins

Protein IDs	Product of lead protein	Ratio H/L UDP-Gal-Rep1	Ratio H/L UDP-Gal-Rep2
Tb927.11.7060	Acidocalcisomal pyrophosphatase	**9.94**	**5.01**
Tb927.11.3560	Hypothetical protein	**8.84**	**4.62**
Tb927.11.1490	Unspecified product	**6.66**	**3.70**
Tb927.11.2990	Krepb4 (krepb4)	**6.25**	**2.79**
Tb927.7.7350	Hypothetical protein	**5.99**	**3.62**
Tb927.9.5660	Unspecified product	**5.68**	**3.18**
Tb927.9.11980	Unspecified product	**5.45**	**3.99**
Tb927.11.4450	Alba-domain protein (alba2)	**5.23**	**2.91**
Tb927.1.3500	Unspecified product	**5.16**	**3.32**
Tb927.11.9330	Helicase-like protein	**5.00**	**3.51**
Tb927.3.1920	NOT5 protein (NOT5)	**4.87**	**2.78**
Tb927.10.2910	Splicing factor PTSR1 interacting protein (TSR1IP)	**4.85**	**2.28**
Tb927.8.2450	5-AMP-activated protein kinase subunit beta (AMPKB)	**4.82**	**2.04**
Tb927.9.12120 *Tb927.9.4100* *Tb927.8.1760* *Tb927.6.1300*	Unspecified product	**4.66**	**2.50**
Tb927.6.4370	Eukaryotic translation initiation factor 3 subunit 7-like protein	**4.59**	**1.84**
Tb927.10.14950	Zinc finger CCCH domain-containing protein 40 (ZC3H40)	**4.52**	**2.41**
Tb927.9.9080	Present in the outer mitochondrial membrane proteome 5 (POMP5)	**4.40**	**2.31**
Tb927.8.3330	Mitochondrial carrier protein (MCP18)	**4.13**	**2.90**
Tb927.11.12410 *Tb927.11.12420*	Kinetoplastid kinetochore protein 10 (kkt10)	**4.02**	**1.44**
Tb927.8.3860	WD-repeat containing protein	**3.92**	**-**

Table 12 –UDP-(4F)-Gal probe, enriched proteins

Protein IDs	Product of lead protein	Ratio H/L
Tb927.6.720	40S ribosomal protein L14, putative	**30.87**
Tb927.3.3320	60S ribosomal protein L13, putative	**29.91**
Tb927.3.5050	60S ribosomal protein L4	**28.10**
Tb927.11.14130	ribosomal protein L18, putative	**27.88**
Tb927.9.7620	60S ribosomal protein L11, putative	**27.75**
Tb927.10.11390	60S ribosomal protein L6, putative	**26.33**
Tb927.9.8420	QM-like protein	**24.13**
Tb927.10.15410	glycosomal malate dehydrogenase	**23.57**
Tb927.10.4120.2	60S ribosomal protein L30	**22.98**
Tb927.9.14000	60S ribosomal protein L12, putative	**21.79**
Tb927.11.7685	ribosomal protein S10p/S20	**20.98**
Tb927.9.15170	60S ribosomal protein L5, putative	**20.03**
Tb927.11.10910	40S ribosomal protein SA, putative	**19.92**
Tb927.7.1050	40S ribosomal protein S16, putative	**19.67**
Tb927.10.1100	60S ribosomal protein L9, putative	**19.36**
Tb927.8.1340	60S ribosomal protein L7a, putative	**18.90**
Tb927.9.6070	40S ribosomal protein S3, putative	**18.15**
Tb927.8.1110	40S ribosomal protein S9, putative	**17.83**
Tb927.9.15420	60S ribosomal protein L32	**17.75**
Tb927.11.6440	hypothetical protein, conserved	**17.21**

Table 13 - GT3 enrichment result

GENE	RATIO H/L	INTENSITY L	INTENSITY H
TB927.2.3370	N/A	0	1352000

```
 Alignment of Sequence_1: TbSTT3B with  Sequence_2: STT3B, Homo sapiens

Similarity : 538/787 (68.36 %)

Seq_1  1     --------------------------MTKGGKVAVTKGSAQSD-GAGEGGMSKAKSSTTFV  34
                                       :::::::  | ||:: :| :||:   |    ::
Seq_2  1     MAEPSAPESKHKSSLNSSPWSGLMALGNSRHGHHGPG-AQCAHKA-AGGAAPPK-PAPAG  57

Seq_1  35    ATGGGSLPAWALKAVSTVVSAVILIYSVHRA-YDIRLTSVRLYGELIHEFDPWFNYRATQ  93
               || ::  | :: |   |  ||:  :: |  ::||  | :  |:| ::  :: :  ::
Seq_2  58    LS-GGLSQP-A-GWQS-LLSFTILFL-AWLAGFSSRLFAVIR-FESIIHEF-DP-WF-NY  108

Seq_1  94    YLSDNGWRAFFQWYDYMSWYPLGRPVGTTIFPGMQLTGVAIHRVLEMLGRGMSINNICVY  153
              : :: ::::|:   :: : ::  | |  ::: :::  : ::: :::: :  :|::: ::
Seq_2  109   RSTHHLASHGFYEFLNWFDERAWYPLGRIVGGTVYPGLMITAGLIHWILNTLNITVHIRD  168

Seq_1  154   IPAWFGSIATVLAALIAYESSNSLSVMAFTAYFFSIVPAHLMRSMAGEFDNECVAMAAML  213
              ::: :  :  :|  ::  :  : | :  :   :::: :| : ::  ::: ::: ::   :
Seq_2  169   VCVFLAPTFSGLTSISTFLLTRELWNQGAGLLAACF-IAIVPGYISRSV-AGSFDNEGIA  226

Seq_1  214   LTFYMWVRSLRSSSSWPIGALAGVAYGYMVSTWGGYIFVLNMVAFHASVCVLLDWARGTY  273
              : :     :|:::|:::::::: : :  ::  :  :::::::| : :    :: :
Seq_2  227   IFALQFTYYLWVKSVKTGSVFWTMCCCLSYF-YM-VSAWGGYVFIINLIPLHVFVLLLMQ  284

Seq_1  274   SVSLLRAYSLFFVIGTALAICVPPVEWTPFRSLEQLTALFVFVFMWALHYSEYLRERARA  333
              ::|:::   ::: :   | ::::: :: ::  ::::::::::||:  ::| : : :|:|
Seq_2  285   RYSKRVYIAYSTFYIVGLILSMQIPFVGFQPIRTSEHMAAAGVFALLQAYAFLQYLRDRL  344

Seq_1  334   PIHSSKALQIRARIFMGTLSLLLIVAIYLFSTGYFRPFSSRVRALFVKHTRTGN-PLVDS  392
              ::     |::  ::::  ::: |:| |||::|||::|:|:|   |::: ::: : |: :|
Seq_2  345   TKQEFQTLFFLGVSLAAGAVF-LSV-IYLTYTGYIAPWSGRFYSLWDTGYAKIHIPIIAS  402

Seq_1  393   VAEHHPASNDDFFGYLHVCYNGWIIGFFFMSVSCFFHCTPGMSFLLLYSILAYYFSLKMS  452
              | || |  ::: |  : :  :  :: :::: :  :::  :: ::  || |:  :: ::::
Seq_2  403   VSEHQPTTWVS-FFFDLHILV-CTFPAGLWFCIKNINDERVFVA--LYAIS-AVYFAGVM  457

Seq_1  453   RLLLLSAPVASILTGYVVGSIVDLAADCFAASGTEHADSKEHQGKARGK-GQKEQITVEC  511
              ::|:|  : :::: :  : |:| ::   ::::: | :  :|:::: ::| :| ::::::
Seq_2  458   VRLMLTLTPVVCMLSAIAFSNVFEHY-LGDDMKRE-NPPVEDSSDEDDKRNQ-GNLYDKA  514

Seq_1  512   GCHNPFYKLWCNSFSSRL-VVGKFFVVVVLSICGPTFLGSNFRIYSEQFADSMSSPQIIM  570
              | ::::: :   ::  :|  : |:: | :| :  : ::  :: ::::::|:|: |:::::
Seq_2  515   G-KVRKHATEQEKTEEGLGPNIKSI-VTML-MLMLLMMFAVHCTWVTSNAYSSPSVVLAS  571

Seq_1  571   RATVGGRRVILDDYYVSYLWLRNNTPEDARILSWWDYGYQITGIGNRTTLADGNTWNHEH  630
               ::: | |:|||||    |:||| ||:|:||::|||||||| |: |||||:|:||||::|
Seq_2  572   YNHD-GTRNILDDFREAYFWLRQNTDEHARVMSWWDYGYQIAGMANRTTLVDNNTWNNSH  630

Seq_1  631   IATIGKMLTSPVKESHALIRHLAD-YVLIWAGYD-GSDLLKSPHMARIG---NSVYRD-I  684
              || :||:: |: :: : ::|:|:: |||:  | : |::::: ::  :::    : : : |
Seq_2  631   IALVGKAMSSNETAAYKIMRTLDVDYVLVIFGGVIGYSGDDINKFLWMVRIAEGEHPKDI  690

Seq_1  685   CSEDDPLCTQFGFYSGDFSKPTPMMQRSLLYNLHRFGTDGGKTQLDKNMFQLAYVSKYGL  744
                |:|::  |: | :::::::||: :: :::::::||| :  : ::  : ::::::   |:
Seq_2  691   R-ESDYFTPQGEFRVDKAGSPTLLNCLMYKMSYYRFGEMQLDFRTPPGFDRTRNAEI-GN  748

Seq_1  745   VKIYKVMNVSEESKAWVADPKNRKCDAPGSWICTGQYPPAKEIQDMLAKRIDYEQLEDFN  804
              ::|: ::    |     ::::: : : : ::: | :::|:::::  : | ::::::::::
Seq_2  749   KDIKFKHLE--EAFTSEHWLVRIYKVKAPDNRETLDHKPRVTNIFPKQKYLSKKTTKRKR  806

Seq_1  805   RRNRSDAYYRAYMRQMG---  821
              :  ::: : :   ::::
Seq_2  807   GYIKNKLVFKKGKKISKKTV  826
```

Figure 71 – Sequence aligment of TbStt3b and *Homo Sapien* Stt3b

Table 14 - Peptides with N-Glycans from ETD trap run, MASCOT analysis with PlasmoDB database

Gene ID	Protein Description	Peptide Sequence	Modification
PF3D7_1442600	TRAP-like protein,TREP	NIHHKGNVNSETNEMT **NNS** GTQNIISNEQFEK	3 HexNAc (N);3 HexNAc(2) (N);Oxidation (M);2 Phospho (ST)
PF3D7_1448300	conserved Plasmodium protein, unknown function	ESDNND **NTS** LYLS **NHS** TMAQMDNDK	3 HexNAc (N);HexNAc(2) (N);Oxidation (M);2 Phospho (ST)
PF3D7_1222300	endoplasmin, putative (GRP94)	LGVDNNLK	HexNAc (N)
PF3D7_0819800	conserved Plasmodium protein, unknown function	EIQIENDKK	HexNAc (N)
PF3D7_0519600	zinc finger protein, putative	EQKQNEEKENNEVNK	HexNAc (N)
PF3D7_1462100	conserved protein, unknown function	NELNEIEK	HexNAc (N)
PF3D7_1011500	conserved membrane protein, unknown function	QIENKEKQNQK	HexNAc (N)
PF3D7_1147000	sporozoite asparagine-rich protein (SLARP)	EINNDIQK	HexNAc (N)
PF3D7_1465200	conserved protein, unknown function	LIQEENIK	HexNAc (N)
PF3D7_1325900	conserved protein, unknown function	NIGNNLIIEKK	HexNAc (N)
PF3D7_0104300	ubiquitin carboxyl-terminal hydrolase 1, putative (UBP1)	ENNEAMDFYKSVLIEE **NNS** M **NIS** K	HexNAc (N);2 Phospho (ST)
PF3D7_1132400	conserved membrane protein, unknown function	NNVHTF **NMS** TSYSINDANFMKTKK	HexNAc (N);HexNAc(2) (N);Phospho (ST)
PF3D7_1127000	protein phosphatase, putative	IMKNPNVK	HexNAc (N);Oxidation (M)
PF3D7_1127000	protein phosphatase, putative	IMKNPNVK	HexNAc (N);Oxidation (M)
PF3D7_1228800	conserved protein, unknown function	LFVNGTTENIK	HexNAc (N);Phospho (ST)
PF3D7_0412000	conserved protein, unknown function	**NVS** SFNHNSETFK	HexNAc(2) (N)

Table 15 - - Peptides with N-Glycans from ETD trap run, MASCOT analysis with PlasmoDB database

Gene ID	Protein Description	Peptide Sequence	Modification
PF3D7_1350100	lysine-tRNA ligase, putative (KRS1)	ITMFLTNKNSIK	HexNAc (N); Phospho (ST)
PF3D7_0928200	onserved Plasmodium protein, unknown function	RGGSFENT **NYT** NDCLLYIYKE	HexNAc (N); Phospho (ST)

Table 16 - Peptides with N-Glycans from HCD, MaxQuant analysis with UniProt database search for O- and N-Glycans

Proteins	Modified Sequence	Modifications
PF3D7_0807800	_(he)NDENKLET(he)SK_	HexNAc (ST),HexNAc (N)
PF3D7_1017300	_(he)**NDS**NSNSHSS(he)**NGS**(he)SK_	2 HexNAc (ST),HexNAc (N)
PF3D7_1041300	_(he)NGAT(he)GEKS(he)K_	2 HexNAc (ST),HexNAc (N)
PF3D7_0103800	_(he)NGHLF**NTS**SEMEVVK_	HexNAc (N)
PF3D7_1116000	_(he)NGIDS(he)FKGK_	HexNAc (ST),HexNAc (N)
PF3D7_0418000	_(he)NIDIDLLCDMIKRDVTYS(he)CK_	HexNAc (ST),HexNAc (N)
PF3D7_1226500	_(he)NIEMS(he)NLANS(he)LVR_	2 HexNAc (ST),HexNAc (N)
PF3D7_1020600	_(he)NIES(he)FFNYM(ox)KS(he)K_	Oxidation (M),2 HexNAc (ST),HexNAc (N)
PF3D7_1145600	_(he)NINTNVRNTFK_	HexNAc (N)
PF3D7_0425800	_(he)NKDILSDNPK_	HexNAc (N)
PF3D7_1457400	_(he)NKETEFS(he)SILKSFNTHK_	HexNAc (ST),HexNAc (N)
PF3D7_0503500	_(he)N**NDS**(he)VHREDK_	HexNAc (ST),HexNAc (N)
PF3D7_0619800	_(he)NNEIHYK_	HexNAc (N)
PF3D7_0415200	_(he)NNKQN**NVT**(he)HK_	HexNAc (ST),HexNAc (N)
PF3D7_0410800	_(he)NNLHMNS(he)FNK_	HexNAc (ST),HexNAc (N)
PF3D7_1431100	_(he)NNPKCKIDPK_	HexNAc (N)
PF3D7_0307700	_(he)NNY**NM**(ox)**S**DKQNIPS(he)IDEK_	Oxidation (M),HexNAc (ST),HexNAc (N)
PF3D7_1322200	_(he)NS(he)DNT(he)EGNSK_	2 HexNAc (ST),HexNAc (N)
PF3D7_1010200	_(he)NT(he)KESIDCK_	HexNAc (ST),HexNAc (N)
PF3D7_0805300	_(he)NVETCTS(he)SS(he)K_	2 HexNAc (ST),HexNAc (N)
PF3D7_1025900	_(he)NYM**NES**(he)STQTS(he)TETSSK_	2 HexNAc (ST),HexNAc (N)
PF3D7_1013600	_(he)NYREILRS(he)HK_	HexNAc (ST),HexNAc (N)

Table 17 - Peptides with N-Glycans from HCD, MaxQuant analysis with UniProt database, search for N-glycans

Gene Name	Modified Sequence	Modifications
PF11_0246	_(he)NDNTGNNHLDIGK_	HexNAc(N)
PFL1240c	_(he)NEDKENVVNINIQK_	HexNAc(N)
PFL1800w	_(he)NFYIIKKM(ox)FK_	Oxidation(M),HexNAc(N)
PFA_0190c	he)NGHLF **NTS** SEMEVVK_	HexNAc(N)
PF08_0118	_(he)NHVM(ox)NSINKK_	Oxidation(M),HexNAc(N)
PFL1475w	_(he)NIHEYKK_	HexNAc(N)
Rh3	he)NIM(ox)NELM(ox)RK_	Oxidation(M),HexNAc(N)
PF11_0467	_(he)NINTNVRNTFK_	HexNAc(N)
PF11_0226	_(he)NM(ox)N **NST** EK_	Oxidation(M),HexNAc(N)
PFD0555c	_(he)NNIVNKNNELK_	HexNAc(N)
PFL1990c	_(he)NTPHLFNNQQK_	HexNAc(N)

Table 18 - Standard buffer = 20 mM HEPES, 45 mM NaCl, 4 mM CaCl2 and 2% NP-40. All conditions contained 6 mM MnCl2, and 1 µg BSA-GlcNAc.

Sample	Buffer	GalT (µg)	UDP-Gal (µM)	Temp (°C)	Time
1	Standard	1.70	**125**	4	o/n
2	Standard	1.70	62.5	4	o/n
3	Standard	**0.85**	62.5	4	o/n
4	Standard	1.70	**125**	**37**	1 h
5	**100 mM HEPES**	1.70	62.5	37	1 h
6	**50 mM TrisHCl**	1.70	62.5	37	1 h

Table 19 - Optimisation of UDP-gal and GalT quantities.

Sample	1	2	3	4	5	6	7	8	9	10
$MnCl_2$ (mM)	6.7	**33.3**	6.7	6.7	6.7	6.7	**33.3**	6.7	6.7	6.7
UDP-Gal (mM)	0.55	0.55	**1.10**	**1.10**	0.55	0.55	0.55	**1.10**	**1.10**	0.55
GalT (µg)	0.85	0.85	0.85	**1.70**	**1.70**	0.85	0.85	0.85	**1.70**	**1.70**
Conditions	37 °C for 1 h					Pre-incubation then 37 °C for 1 h				

14 References:

Acosta-Serrano, A. (1999) 'The Procyclin Repertoire of Trypanosoma brucei. IDENTIFICATION AND STRUCTURAL CHARACTERIZATION OF THE GLU-PRO-RICH POLYPEPTIDES', *Journal of Biological Chemistry*, 274(42), pp. 29763–29771. doi: 10.1074/jbc.274.42.29763.

Adamo, R., Nilo, A., Castagner, B., Boutureira, O., Berti, F. and Bernardes, G. J. L. (2013) 'Synthetically defined glycoprotein vaccines: current status and future directions', *Chemical Science*, 4(8), p. 2995. doi: 10.1039/c3sc50862e.

Aebi, M. (2013) 'N-linked protein glycosylation in the ER', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1833(11), pp. 2430–2437. doi: 10.1016/j.bbamcr.2013.04.001.

Akiyaffici, A., Simon, S., Whitesides, M. and Akiyama, A. (1987) 'Enzymes in organic synthesis', 23(7), pp. 645–654.

Albalasmeh, A. A., Berhe, A. A. and Ghezzehei, T. A. (2013) 'A new method for rapid determination of carbohydrate and total carbon concentrations using UV spectrophotometry', *Carbohydrate Polymers*. Elsevier Ltd., 97(2), pp. 253–261. doi: 10.1016/j.carbpol.2013.04.072.

Alberts, B., Johnson, A., Lewis, J. and Al., E. (2002) *Biology of the Cell*. 4th edn.

Angulo, J., Langpap, B., Blume, A., Biet, T., Meyer, B., Rama Krishna, N., Peters, H., Palcic, M. M. and Peters, T. (2006) 'Blood group B galactosyltransferase: Insights into substrate binding from NMR experiments', *Journal of the American Chemical Society*, 128(8), pp. 13529–13538. doi: 10.1021/ja063550r.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Logan, F. J., Miller, J. A., Mitra, S., Myler, P. J., Nayak, V., Pennington, C., Phan, I., Pinney, D. F., Ramasamy, G., Rogers, M. B., Roos, D. S., Ross, C., Sivam, D., Smith, D. F., Srinivasamoorthy, G., Stoeckert, C. J., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G. and Wang, H. (2010) 'TriTrypDB: A functional genomic resource for the Trypanosomatidae', *Nucleic Acids Research*, 38, pp. 457–462. doi: 10.1093/nar/gkp851.

Atrih, A., Richardson, J. M., Prescott, A. R. and Ferguson, M. A. J. (2005) 'Trypanosoma brucei glycoproteins contain novel giant poly-N-acetyllactosamine carbohydrate chains.', *The Journal of biological chemistry*, 280(2), pp. 865–871. doi: 10.1074/jbc.M411061200.

Bacchi, C. J. (2009) 'Chemotherapy of human african trypanosomiasis.', *Interdisciplinary perspectives on infectious diseases*, 2009, pp. 1–5. doi: 10.1155/2009/195040.

Ban, L., Pettit, N., Li, L., Stuparu, A. D., Cai, L., Chen, W., Guan, W., Han, W., Wang, P. G. and Mrksich, M. (2012) 'Discovery of glycosyltransferases using carbohydrate arrays and mass spectrometry.', *Nature chemical biology*. Nature Publishing Group, 8(9), pp. 769–773. doi: 10.1038/nchembio.1022.

Bandini, G., Mariño, K., Güther, M. L. S., Wernimont, A. K., Kuettel, S., Qiu, W., Afzal, S., Kelner, A., Hui, R. and Ferguson, M. A. J. (2012) 'Phosphoglucomutase is absent in Trypanosoma brucei and redundantly substituted by phosphomannomutase and phospho-N-acetylglucosamine mutase', *Molecular Microbiology*, 85(3), pp. 513–534. doi: 10.1111/j.1365-2958.2012.08124.x.

Barker, R., Olsen, K. W., Shaper, J. H. and Hill, R. L. (1972) 'Agarose Derivatives a Galactosyltransferase * of Uridine Diphosphate for the Purification of and', *The Journal of biological chemistry*, (22), pp. 7135–7147.

Barlow, J. N. and Blanchard, J. S. (2000) 'Enzymatic synthesis of UDP- (2-deoxy-2-fluoro) - D -galactose and substrate activity with UDP-galactopyranose mutase', 328, pp. 473–480.

Baum, L. G. and Cobb, B. I. (2017) 'The Direct and Indirect Effects of Glycans on Immune Function', *Glycobiology*, (310), pp. 1–19. doi: 10.1093/glycob/cwx036.

Berg, J., Tymoczko, J. and Stryer, L. (2002) *Biochemistry*. 5th edn. New York: W H Freeman.

Blackburn, G. M. (2006) *Nucleic Acids in Chemistry*. Royal Society of Chemistry.

Boda, C., Enanga, B., Courtioux, B., Breton, J. C. and Bouteille, B. (2005) 'Trypanocidal activity of methylene blue: Evidence for in vitro efficacy and in vivo failure', *Chemotherapy*, 52(1), pp. 16–19. doi: 10.1159/000090236.

Boeggeman, E. and Qasba, P. K. (2002) 'Studies on the metal binding sites in the catalytic domain of β 1 , 4-galactosyltransferase', *Glycobiology*, 12(7), pp. 395–407.

Bourne, G. T., Golding, S. W., McGeary, R. P., Meutermans, W. D. F., Jones, a., Marshall, G. R., Alewood, P. F. and Smythe, M. L. (2001) 'The development and application of a novel safety-catch linker for BOC-based assembly of libraries of cyclic peptides', *Journal of Organic Chemistry*, 66(15), pp. 7706–7713. doi: 10.1021/jo010580y.

Boutlis, C. S., Anstey, E. M. R. N. M., Souza, J. B. De, Street, C. and Wit, L. (2005) 'Glycosylphosphatidylinositols in Malaria Pathogenesis and Immunity : Potential for Therapeutic Inhibition and Vaccination', pp. 145–185.

Breidenbach, M. A., Gallagher, J. E. G., King, D. S., Smart, B. P., Wu, P. and Bertozzi, C. R. (2010) 'Targeted metabolic labeling of yeast N-glycans with unnatural sugars', *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp. 3988–93. doi: 10.1073/pnas.0911247107.

Bruneau, A., Roche, M., Hamze, A., Brion, J.-D., Alami, M. and Messaoudi, S. (2015) 'Stereoretentive Palladium-Catalyzed Arylation, Alkenylation, and Alkynylation of 1-Thiosugars and Thiols Using Aminobiphenyl Palladacycle Precatalyst at Room Temperature', *Chem. Eur. J.*, 21(23), pp. 8375–8379. doi: 10.1002/chem.201501050.

Burkart, M. D., Vincent, S. P., Düffels, A., Murray, B. W., Ley, S. V and Wong, C. H. (2000) 'Chemo-enzymatic synthesis of fluorinated sugar nucleotide: useful mechanistic probes for glycosyltransferases.', *Bioorganic & medicinal chemistry*, 8, pp. 1937–1946. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11003139.

Burns, J. A., Butler, J. C., Moran, J. and Whitesides, G. M. (1991) 'Selective reduction of disulfides by tris(2-carboxyethyl)phosphine', *The Journal of Organic Chemistry*, 56(9), pp. 2648–2650. doi: 10.1021/jo00008a014.

Burton, A. and Boons, G. (1997) 'Preparation of fl uorinated galactosyl nucleoside diphosphates to study the mechanism of the enzyme galactopyranose mutase †', *J. Chem. Soc., Perkin Trans. 1*, pp. 2375–2382.

Bushkin, G. G., Ratner, D. M., Cui, J., Banerjee, S., Duraisingh, M. T., Jennings, C. V, Dvorin, J. D., Gubbels, M., Robertson, S. D., Steffen, M., Keefe, B. R. O., Robbins, P. W. and Samuelson, J. (2010) 'Suggestive Evidence for Darwinian Selection against Asparagine-Linked Glycans of Plasmodium falciparum and Toxoplasma gondii □ †', 9(2), pp. 228–241. doi: 10.1128/EC.00197-09.

Cao, L., Qu, Y., Zhang, Z., Wang, Z., Prykova, I., Wu, S., Group, E. A. and Chicago, N. (2017) 'Intact glycopeptide

characterisation using mass spectrometry', *Expert Rev Proteomics*, 13(5), pp. 513–522. doi: 10.1586/14789450.2016.1172965.Intact.

Card, P. J. and Reddy, G. S. (1983) 'Fluorinated carbohydrates. 2. Selective fluorination of gluco- and mannopyranosides. Use of {2-D} {NMR} for structural assignments', *The Journal of Organic Chemistry*, 48(4), pp. 4734–4743. doi: 10.1021/jo00172a054.

Casero, R. A., Porter, C. W. and Bernacki, R. J. (1982) 'Activity of tunicamycin against Trypanosoma brucei in vitro and in vivo', *Antimicrobial Agents and Chemotherapy*, 22(6), pp. 1008–1011. doi: 10.1128/AAC.22.6.1008.

Castillo-Acosta, V. M., Vidal, A. E., Ruiz-P??rez, L. M., Van Damme, E. J. M., Igarashi, Y., Balzarini, J. and Gonz??lez-Pacanowska, D. (2013) 'Carbohydrate-binding agents act as potent trypanocidals that elicit modifications in VSG glycosylation and reduced virulence in Trypanosoma brucei', *Molecular Microbiology*, 90(4), pp. 665–679. doi: 10.1111/mmi.12359.

Chapeau, M.-C. and Frey, P. A. (1994) 'Synthesis of UDP-4-deoxy-4-fluoroglucose and UDP-4-deoxy-4-fluorogalactose and their Interactions with Enzymes of Nucleotide Sugar Metabolism', *The Journal of Organic Chemistry*, 59(11), pp. 6994–6998. doi: 10.1021/jo00102a024.

Chatterjee, D., Paul, A., Rajkamal, R. and Yadav, S. (2015) 'Cu(ClO4)2.6H2O catalyzed solvent free per-O-acetylation and sequential one-pot conversions of sugars to thioglycosides', *RSC Adv.* Royal Society of Chemistry, 5, pp. 29669–29674. doi: 10.1039/C5RA03461B.

Chen, J., Spear, S. K., Huddleston, J. G. and Rogers, R. D. (2005) 'Polyethylene glycol and solutions of polyethylene glycol as green reaction media', *Green Chemistry*, 7(2), p. 64. doi: 10.1039/b413546f.

Chen, X., Tan, P. H., Zhang, Y. and Pei, D. (2009) 'NIH Public Access', *J. Comb. Chem*, 11(4), pp. 604–611. doi: 10.1002/humu.21020.Gene.

Chéry, F., Rollin, P., De Lucchi, O. and Cossu, S. (2001) 'Phenylsulfonylethylidene (PSE) Acetals: A Novel Protective Group in Carbohydrate Chemistry', *Synthesis*, 2001(2), pp. 0286–0292. doi: 10.1055/s-2001-10817.

Coatney, G. R. (1963) 'Pitfalls in a discovery: the chronicle of chloroquine.', *The American journal of tropical medicine and hygiene*, 12(November), pp. 121–128. doi: 10.4269/ajtmh.1963.12.121.

Colasante, C., Ellis, M., Ruppert, T. and Voncken, F. (2006) 'Comparative proteomics of glycosomes from bloodstream form and procyclic culture form Trypanosoma brucei brucei', *Proteomics*, 6(11), pp. 3275–3293. doi: 10.1002/pmic.200500668.

Cova, M., Rodrigues, J. A., Smith, T. K. and Izquierdo, L. (2015) 'Sugar activation and glycosylation in Plasmodium', *Malaria journal*. BioMed Central, 14(1), pp. 1–10. doi: 10.1186/s12936-015-0949-z.

Cravatt, B. F., Wright, A. T. and Kozarich, J. W. (2008) 'Activity-based protein profiling: from enzyme chemistry to proteomic chemistry.', *Annual review of biochemistry*, 77, pp. 383–414. doi: 10.1146/annurev.biochem.75.101304.124125.

Creek, D. J., Mazet, M., Achcar, F., Anderson, J., Kim, D. H., Kamour, R., Morand, P., Millerioux, Y., Biran, M., Kerkhoven, E. J., Chokkathukalam, A., Weidt, S. K., Burgess, K. E. V., Breitling, R., Watson, D. G., Bringaud, F. and Barrett, M. P. (2015) 'Probing the Metabolic Network in Bloodstream-Form Trypanosoma brucei Using Untargeted Metabolomics with Stable Isotope Labelled Glucose', *PLoS Pathogens*, 11(3), pp. 1–25. doi: 10.1371/journal.ppat.1004689.

Damerow, M., Graalfs, F., Güther, M. L. S., Mehlert, A., Izquierdo, L. and Ferguson, M. A. J. (2016) 'A gene of the β3-glycosyltransferase family encodes N-acetylglucosaminyltransferase II function in Trypanosoma brucei', *Journal of Biological Chemistry*, 291(26), pp. 13834–13845. doi: 10.1074/jbc.M116.733246.

Damerow, M., Rodrigues, J. A., Wu, D., Güther, M. L. S., Mehlert, A. and Ferguson, M. A. J. (2014) 'Identification and functional characterization of a highly divergent N-acetylglucosaminyltransferase I (TbGnTI) in Trypanosoma brucei.', *The Journal of biological chemistry*, 289(13), pp. 9328–9339. doi: 10.1074/jbc.M114.555029.

Dayal-Drager, R., Hoessli, D. C., Decrind, C., Del Guidice, G., Lambert, P. H. and Nasir-ud-Din (1991) 'Presence of O-glycosylated glycoproteins in the Plasmodium falciparum parasite', *Carbohydrate Research*, 209(C), pp. 5–8. doi: 10.1016/0008-6215(91)80172-J.

Deng, N., Li, Z., Pan, C. and Duan, H. (2015) 'FreeQuant: A Mass Spectrometry Label-Free Quantification Software Tool for Complex Proteome Analysis', *Scientific World Journal*. Hindawi Publishing Corporation, 2015. doi: 10.1155/2015/137076.

Denton, H., Fyffe, S. and Smith, T. K. (2010) 'GDP-Mannose pyrophosphorylase is essential in the bloodstream form of Trypanosoma brucei', *Biochemical journal*, 425(3), pp. 603–614.

Descroix, K., Pesnot, T., Yoshimura, Y., Gehrke, S. S., Wakarchuk, W., Palcic, M. M. and Wagner, G. K. (2012) 'Inhibition of galactosyltransferases by a novel class of donor analogues.', *Journal of medicinal chemistry*, 55(5), pp. 2015–2024. doi: 10.1021/jm201154p.

Dieckmann-Schuppert, A., Bause, E. and Schwarz, R. T. (1993) 'Studies on Oglycans of Plasmodium-falciparum-infected human erythrocytes Evidence for O-GlcNAc and O-GlcNAc-transferase in malaria parasites', *European Journal of Biochemistry*, 216(3), pp. 779–788. doi: 10.1111/j.1432-1033.1993.tb18198.x.

Dieckmann-Schuppert, A., Bender, S., Odenthal-Schnittler, M., Bause, E. and Schwarz, R. T. (1992) 'Apparent lack of N-glycosylation in the asexual intraerythrocytic stage of Plasmodium falciparum', *European Journal of Biochemistry*, 205(2), pp. 815–825. doi: 10.1111/j.1432-1033.1992.tb16846.x.

Dieckmann-Schuppert, A., Hensel, J. and Schwarz, R. T. (1992) 'Studies on the effect of Tunicamycin on erthyrocytic stages of Plasmodium falciparum', *Biochemical Society transactions*, p. 992.

Dohi, H., Nishida, Y., Furuta, Y., Uzawa, H., Yokoyama, S. I., Ito, S., Mori, H. and Kobayashi, K. (2002) 'Molecular design and biological potential of galacto-type trehalose as a nonnatural ligand of shiga toxins', *Organic Letters*, 4(3), pp. 355–357. doi: 10.1021/ol017135+.

Dube, D. H., Prescher, J. A., Quang, C. N. and Bertozzi, C. R. (2006) 'Probing mucin-type O-linked glycosylation in living animals.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(13), pp. 4819–24. doi: 10.1073/pnas.0506855103.

Easton, R. (2011) 'Glycosylation of Proteins – Structure , Function and Analysis', *Life Science*, (48), pp. 1–5.

Ermolenko, L. and Sasaki, N. A. (2006) 'Diastereoselective synthesis of all eight l-hexoses from L-ascorbic acid.', *The Journal of organic chemistry*, 71(2), pp. 693–703. doi: 10.1021/jo0521192.

Fabene, P. F. and Bentivoglio, M. (1998) '1898-1998: Camillo Golgi and "the Golgi": One hundred years of terminological clones', *Brain Research Bulletin*, 47(3), pp. 195–198. doi: 10.1016/S0361-9230(98)00079-3.

Fairlamb, A. H. (2003) 'Chemotherapy of human African trypanosomiasis: current and future prospects', *Trends*

in Parasitology, 19(11), pp. 488–494. doi: 10.1016/j.pt.2003.09.002.
Ferguson, M. A. J. (1991) 'Glycosyl- phosphatidylinositol Membrane Anchors: The Tale of a Tail'.
Ferguson, M. A. J., Duszenko, M., Lamont, G. S., Overaths, P. and Cross, G. A. M. (1986) 'Biosynthesis of Trypanosoma brucei Variant Surface Glycoproteins', *Journal of Biological Chemistry*, 261(1), pp. 356–362.
Ferguson, M. A. J., Homans, S. W., Dwek, R. A. and Thomas, W. (1988) 'Glycosyl-Phosphatidylinositol Moiety that Anchors Trypanosoma brucei Variant Surface Glycoprotein to the Membrane Linked references are available on JSTOR for this article : Glycosyl- Phosphatidylinositol Moiety That Anchors Trypanosoma brucei Variant Sur', 239(4841), pp. 753–759.
Ferrell, W. J., Garces, A. and Desmyter, E. A. (1976) 'SYNTHESIS AND PROPERTIES OF 35 S, 14C AND 3 H LABELED S-ALKYL GLYCEROL ETHERS AND DERIVATIVES', 16.
Fischer, E. (1902) 'Syntheses in the purine and sugar group', *Nobel Lectures*, Chemistry(1901–1921), pp. 21–35. Available at: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1902/fischer-lecture.pdf.
Foster, A. B. and Overend, W. G. (1955) 'Carbohydrate phosphates', pp. 61–85.
Franco, J. R., Simarro, P. P., Diarra, A., Ruiz-Postigo, J. A., Samo, M. and Jannin, J. G. (2012) 'Monitoring the use of nifurtimox-eflornithine combination therapy (NECT) in the treatment of second stage gambiense human African trypanosomiasis', *Research reports in tropical medicine*, 3, pp. 93–101.
Freeze, H. and Elbein, A. (2009) 'Essentials of Glycobiology', in. Cold Spring Harbor Laboratory Press, p. Chapter 4.
Galano et al., J. (2013) 'Improved Flavodoxin Inhibitors with Potential Therapeutic E ff ects against Helicobacter pylori Infection', *journal of Medical Chemestry*, 56(15), pp. 6248–6258. doi: 10.1021/jm400786q.
Gilar, M., Yu, Y.-Q., Ahn, J., Xie, H., Han, H., Ying, W. and Qian, X. (2011) 'Characterization of glycoprotein digests with hydrophilic interaction chromatography and mass spectrometry.', *Analytical biochemistry*. Elsevier Inc., 417(1), pp. 80–8. doi: 10.1016/j.ab.2011.05.028.
Giuffredi, G. T., Gouverneur, V. and Bernet, B. (2013) 'Intramolecular OH×××FC hydrogen bonding in fluorinated carbohydrates: CHF is a better hydrogen bond acceptor than CF2', *Angewandte Chemie - International Edition*, 52, pp. 10524–10528. doi: 10.1002/anie.201303766.
Giuffredi, G. T., Jennings, L. E., Bernet, B. and Gouverneur, V. (2011) 'Facile synthesis of 4-deoxy-4-fluoro-α-d-talopyranoside, 4-deoxy-4-fluoro-α-d-idopyranoside and 2,4-dideoxy-2,4-difluoro-α-d-talopyranoside', *Journal of Fluorine Chemistry*, 132, pp. 772–778. doi: 10.1016/j.jfluchem.2011.05.017.
Gloster, T. M. (2012) 'Development of inhibitors as research tools for carbohydrate-processing enzymes.', *Biochemical Society transactions*, 40(5), pp. 913–928. doi: 10.1042/BST20120201.
Göksel, H., Wasserberg, D., Möcklinghoff, S., Araujo, B. V. and Brunsveld, L. (2011) 'An on-bead assay for the identification of non-natural peptides targeting the Androgen Receptor-cofactor interaction', *Bioorganic and Medicinal Chemistry*, 19(1), pp. 306–311. doi: 10.1016/j.bmc.2010.11.019.
Gowda, D. C., Gupta, P. and Davidson, E. A. (1997) 'Glycosylphosphatidylinositol anchors represent the major carbohydrate modification in proteins of intraerythrocytic stage Plasmodium falciparum', *Journal of Biological Chemistry*, 272(10), pp. 6428–6439. doi: 10.1074/jbc.272.10.6428.
Guther, M. L. S., Urbaniak, M. D., Tavendale, A., Prescott, A. and Ferguson, M. A. J. (2014) 'High-Con fi dence

Glycosome Proteome for Procyclic Form Trypanosoma brucei by Epitope-Tag Organelle Enrichment and SILAC Proteomics', *Journal of proteome research*, 13, pp. 2796–2806.

Haanstra, J. R., van Tuijl, A., Kessler, P., Reijnders, W., Michels, P. A. M., Westerhoff, H. V, Parsons, M. and Bakker, B. M. (2008) 'Compartmentation prevents a lethal turbo-explosion of glycolysis in trypanosomes.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), pp. 17718–23. doi: 10.1073/pnas.0806664105.

Haeuptle, M. A., Pujol, F. M., Neupert, C., Winchester, B., Kastaniotis, A. J., Aebi, M. and Hennet, T. (2008) 'Human RFT1 deficiency leads to a disorder of N-linked glycosylation Human RFT1 deficiency leads to a disorder of N-linked glycosylation', *Biochemistry*, 82(3), pp. 600–606. doi: 10.1016/j.ajhg.2007.12.021.

Hang, H. C., Yu, C., Pratt, M. R. and Bertozzi, C. R. (2004) 'Probing glycosyltransferase activities with the Staudinger ligation.', *Journal of the American Chemical Society*, 126(1), pp. 6–7. doi: 10.1021/ja037692m.

Harding, M. M., Anderberg, P. I. and Haymet, A. D. J. (2003) '"Antifreeze" glycoproteins from polar fish', *European Journal of Biochemistry*, 270(7), pp. 1381–1392. doi: 10.1046/j.1432-1033.2003.03488.x.

Harrison, J. G. and Balasubramanian, S. (1998) 'Synthesis and hybridization analysis of a small library of peptide- oligonucleotide conjugates', *Nucleic Acids Research*, 26(13), pp. 3136–3145.

Hartman, M. C. T. and Coward, J. K. (2002) 'Synthesis of 5-Fluoro N-Acetylglucosamine Glycosides and Pyrophosphates via Epoxide Fluoridolysis : Versatile Reagents for the Study of Glycoconjugate Biochemistry', (18), pp. 10036–10053.

Haserick, J. R., Klein, J. A., Costello, C. E. and Samuelson, J. (2017) 'Cryptosporidium parvum vaccine candidates are incompletely modified with O -linked- N - acetylgalactosamine or contain N-terminal N -myristate and S - palmitate', *PloS one*, pp. 1–22.

Haserick, J. R., Leon, D., Samuelson, J. and Costello, C. E. (2017) 'N-glycans and glycoproteins of Cryptosporidium parvum', *Molecular & Cellular Proteomics*, pp. 617–638.

Hayashi, T., Murray, B. W., Wang, R. and Wong, C. (1997) 'A Chemoenzymatic Synthesis of UDP- (2-deoxy-2-fluoro) - galactose and Evaluation of its Interaction with Galactosyltransferase', 5(3), pp. 497–500.

Haynes, P. A. (1998) 'Phosphoglycosylation: A new structural class of glycosylation?', *Glycobiology*, 8(1), pp. 1–5. doi: 10.1093/glycob/8.1.1.

Hermanson, G. T. (2008) 'Bioconjugate Reagents', *Bioconjugate Techniques*, pp. 214–233. doi: 10.1016/B978-0-12-370501-3.00003-5.

Hesek, D., Toth, M., Meroueh, S. O., Brown, S., Zhao, H., Sakr, W., Fridman, R. and Mobashery, S. (2006) 'Design and Characterization of a Metalloproteinase Inhibitor-Tethered Resin for the Detection of Active MMPs in Biological Samples', *Chemistry and Biology*, 13(4), pp. 379–386. doi: 10.1016/j.chembiol.2006.01.012.

Hsu, J. L., Huang, S. Y., Chow, N. H. and Chen, S. H. (2003) 'Stable-Isotope Dimethyl Labeling for Quantitative Proteomics', *Analytical Chemistry*, 75(24), pp. 6843–6852. doi: 10.1021/ac0348625.

Hutson, A. M., Atmar, R. L., Graham, D. Y. and Estes, M. K. (2002) 'Norwalk Virus Infection and Disease Is Associated with ABO Histo – Blood Group Type', *Journal of Infectious Diseases*, 185, pp. 1335–1337.

Ikezawa, H. (2002) 'Glycosylphosphatidylinositol (GPI)-anchored proteins.', *Biological & pharmaceutical bulletin*, 25(April), pp. 409–417. doi: 10.1248/bpb.25.409.

Imhof, S., Vu, X. L., Bütikofer, P. and Roditi, I. (2015) 'A glycosylation mutant of Trypanosoma brucei links social motility defects in vitro to impaired colonisation of tsetse in vivo', *Eukaryotic Cell*, 14(6), p. EC.00023-15. doi: 10.1128/EC.00023-15.

Izquierdo, L., Mehlert, A. and Ferguson, M. A. J. (2012) 'The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases.', *Glycobiology*, 22(5), pp. 696–703. doi: 10.1093/glycob/cws003.

Izquierdo, L., Nakanishi, M., Mehlert, A., Machray, G., Barton, G. J. and Ferguson, M. A. J. (2009) 'Identification of a glycosylphosphatidylinositol anchor-modifying beta1-3 N-acetylglucosaminyl transferase in Trypanosoma brucei.', *Molecular microbiology*, 71(2), pp. 478–491. doi: 10.1111/j.1365-2958.2008.06542.x.

Izquierdo, L., Nakanishi, M., Mehlert, A., Machray, G., Barton, G. J. and Ferguson, M. a J. (2009) 'Identification of a glycosylphosphatidylinositol anchor-modifying ??1-3 N-acetylglucosaminyl transferase in Trypanosoma brucei', *Molecular Microbiology*, 71(2), pp. 478–491. doi: 10.1111/j.1365-2958.2008.06542.x.

Izquierdo, L., Schulz, B. L., Rodrigues, J. A., Güther, M. L. S., Procter, J. B., Barton, G. J., Aebi, M. and Ferguson, M. A. J. (2009) 'Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases', *The EMBO Journal*, 28(17), pp. 2650–2661. doi: 10.1038/emboj.2009.203.

Jang, K.-S., Chung, W.-J., Kim, H.-K., Kim, Y.-G., Lee, Y.-S. and Kim, B.-G. (2008) 'Selective removal of anti-α-Gal antibodies from human serum by using synthetic α-Gal epitope on a core-shell type resin', *Biotechnology and Bioprocess Engineering*, 13(4), pp. 445–452. doi: 10.1007/s12257-008-0141-1.

Jiang, J., Kanabar, V., Padilla, B., Man, F., Pitchford, S. C., Page, C. P. and Wagner, G. K. (2016) 'Uncharged nucleoside inhibitors of β-1,4-galactosyltransferase with activity in cells', *Chem. Commun.* Royal Society of Chemistry, 52(20), pp. 3955–3958. doi: 10.1039/C5CC09289B.

Jiang, J., Kuo, C.-L., Wu, L., Franke, C., Kallemeijn, W. W., Florea, B. I., van Meel, E., van der Marel, G. A., Codée, J. D. C., Boot, R. G., Davies, G. J., Overkleeft, H. S. and Aerts, J. M. F. G. (2016) 'Detection of Active Mammalian GH31 α-Glucosidases in Health and Disease Using In-Class, Broad-Spectrum Activity-Based Probes', *ACS Central Science*, p. acscentsci.6b00057. doi: 10.1021/acscentsci.6b00057.

Jones, W. H. (1980) *Catalysis in organic synthesis*.

Kanemaru, M., Yamamoto, K. and Kadokawa, J. I. (2012) 'Self-assembling properties of 6-O-alkyltrehaloses under aqueous conditions', *Carbohydrate Research*. Elsevier Ltd, 357, pp. 32–40. doi: 10.1016/j.carres.2012.05.014.

Kannagi, R., Izawa, M., Koike, T., Miyazaki, K. and Kimura, N. (2004) 'Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis', *Cancer Science*, 95(5), pp. 377–384. doi: 10.1111/j.1349-7006.2004.tb03219.x.

Kennedy, P. G. (2013) 'Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness).', *The Lancet. Neurology*. Elsevier Ltd, 12(2), pp. 186–194. doi: 10.1016/S1474-4422(12)70296-X.

Khan, S. H., Duus, J. O., Crawley, S. C., Palcic, M. M. and Hindsgaul, O. (1994) 'Acceptor-substrate recognition by N-acetyl- glucosaminyl-transferase-V: Role of the mannose residue in betaDGlcNAc(1-2)alphaDMan(1-6)betaDGlcOR', *Tetrahedron Asymmetry*, 5(12), pp. 2415–2435.

Kim, D., Lee, H. Y., Kim, H., Kim, H., Lee, Y. and Park, S. B. (2006) 'Quantitative Evaluation of Hi Core Resin for

the Nonspecific Binding of Proteins by On-Bead Colorimetric Assay', *J. Comb. Chem*, 8, pp. 280–285.
Kimura, E. A., Couto, A. S., Peres, V. J., Casal, O. L. and Katzin, A. M. (1996) 'N-linked glycoproteins are related to schizogony of the intraerythrocytic stage in Plasmodium falciparum', *Journal of Biological Chemistry*, 271(24), pp. 14452–14461. doi: 10.1074/jbc.271.24.14452.
Kira, M. and Iwamoto, T. (2001) *Silyl migrations*.
Koide, N. and Muramatsu, T. (1974) 'Endo-beta-N-acetylglucosaminidase acting on carbohydrate moieties of glycoproteins. Purification and properties of the enzyme from Diplococcus pneumoniae.', *The Journal of biological chemistry*, 249(15), pp. 4897–4904.
Koketsu, M., Kuwahara, M., Sakurai, H. and Ishihara, H. (2011) 'First Synthesis of a Trisaccharide of Glycosylkaemferide : A Resistance Factor in Carnations First Synthesis of a Trisaccharide of Glycosylkaemferide : A Resistance', (November), pp. 37–41. doi: 10.1081/SCC-120027259.
Koniev, O. and Wagner, A. (2015) 'Developments and recent advancements in the field of endogenous amino acid selective bond forming reactions for bioconjugation', *Chemical Society Reviews*. Royal Society of Chemistry, 44, pp. 5495–5551. doi: 10.1039/C5CS00048C.
Kuettel, S., Wadum, M. C. T., Güther, M. L. S., Mariño, K., Riemer, C. and Ferguson, M. A. J. (2012) 'The de novo and salvage pathways of GDP-mannose biosynthesis are both sufficient for the growth of bloodstream-form Trypanosoma brucei', *Molecular Microbiology*, 84(2), pp. 340–351. doi: 10.1111/j.1365-2958.2012.08026.x.
Lairson, L. L., Henrissat, B., Davies, G. J. and Withers, S. G. (2008) 'Glycosyltransferases: structures, functions, and mechanisms.', *Annual review of biochemistry*, 77, pp. 521–555. doi: 10.1146/annurev.biochem.76.061005.092322.
Lau, H.-T., Suh, H. W., Golkowski, M. and Ong, S.-E. (2014) 'Comparing SILAC-and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics', *Journal of Proteome Research*, 13, pp. 4164–4174. doi: 10.1021/pr500630a.
Leonard, N. M. and Brunckova, J. (2011) 'In situ formation of N-trifluoroacetoxy succinimide (TFA-NHS): One-pot formation of succinimidyl esters, N-trifluoroacetyl amino acid succinimidyl esters, and N-maleoyl amino acid succinimidyl esters', *Journal of Organic Chemistry*, 76(21), pp. 9169–9174. doi: 10.1021/jo201686e.
Leuven, K. U. (2002) 'Congenital Disorders of Glycosylation : A Review', 52(5), pp. 618–624. doi: 10.1203/01.PDR.0000031921.02259.35.
Li, N., Overkleeft, H. S. and Florea, B. I. (2012) 'Activity-based protein profiling: an enabling technology in chemical biology research.', *Current opinion in chemical biology*, 16(1–2), pp. 227–233. doi: 10.1016/j.cbpa.2012.01.008.
Li, T., Wen, L., Williams, A., Wu, B., Li, L., Qu, J., Meisner, J., Xiao, Z., Fang, J. and Wang, P. G. (2014) 'Chemoenzymatic synthesis of ADP-d-glycero-β-d-manno-heptose and study of the substrate specificity of HldE.', *Bioorganic & medicinal chemistry*. Elsevier Ltd, 22(3), pp. 1139–1147. doi: 10.1016/j.bmc.2013.12.019.
Liener, I. E., Sharon, N. and Goldstein, I. J. (1986) *The Lectins: Properties, Functions and Applications in Biology and Medicine*.
Liu, J., Shen, G. and Ichikawa, Y. (1992) 'Overproduction of CMP-sialic acid synthetase for organic synthesis', *Journal of the American Chemical Society*, 114(17), pp. 3901–3910. Available at:

http://pubs.acs.org/doi/pdf/10.1021/ja00036a044.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2014) 'The carbohydrate-active enzymes database (CAZy) in 2013', *Nucleic Acids Research*, 42(D1), pp. 490–495. doi: 10.1093/nar/gkt1178.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B. (2017) *Carbohydrate Active Enzymes database*. Available at: http://www.cazy.org/ (Accessed: 31 August 2017).
Lopaticki, S., Yang, A. S. P., John, A., Scott, N. E., Lingford, J. P., O'Neill, M. T., Erickson, S. M., McKenzie, N. C., Jennison, C., Whitehead, L. W., Douglas, D. N., Kneteman, N. M., Goddard-Borger, E. D. and Boddey, J. A. (2017) 'Protein O-fucosylation in Plasmodium falciparum ensures efficient infection of mosquito and vertebrate hosts', *Nature Communications*. Springer US, 8(1), p. 561. doi: 10.1038/s41467-017-00571-y.
López-Prados, J., Cuevas, F., Reichardt, N.-C., de Paz, J.-L., Morales, E. Q. and Martín-Lomas, M. (2005) 'Design and synthesis of inositolphosphoglycan putative insulin mediators.', *Organic & biomolecular chemistry*, 3, pp. 764–786. doi: 10.1039/b418041k.
Lopez-Sambrooks, C., Shrimal, S., Khodier, C., Flaherty, D. P., Rinis, N., Charest, J. C., Gao, N., Zhao, P., Wells, L., Lewis, T. A., Lehrman, M. A., Gilmore, R., Golden, J. E. and Contessa, J. N. (2016) 'Oligosaccharyltransferase Inhibition Induces Senescence in RTK-Driven Tumor Cells', *Nature chemical biology*, 12(12), pp. 1023–1030. doi: 10.1016/j.celrep.2016.09.087.Roles.
Lord, J. M. (1995) 'Mutational Analysis of the Ricinus Lectin B-chains', *Journal of Biological Chemistry*. doi: 10.1074/jbc.270.35.20292.
Lourenço, E. C., Maycock, C. D. and Rita Ventura, M. (2009) 'Synthesis of potassium (2R)-2-O-α-d-glucopyranosyl-(1→6)-α-d-glucopyranosyl-2,3-dihydroxypropanoate a natural compatible solute', *Carbohydrate Research*. Elsevier Ltd, 344(15), pp. 2073–2078. doi: 10.1016/j.carres.2009.06.037.
Lourenço, E. C. and Ventura, M. R. (2011) 'Synthesis of Potassium (2R)-2-O-α-D-Mannopyranosyl-(1→2)-α-D-glucopyranosyl-2,3-dihydroxypropanoate: A Naturally Compatible Solute', *European Journal of Organic Chemistry*, 2011(33), pp. 6698–6703. doi: 10.1002/ejoc.201100934.
Lowenthal, M. S., Davis, K. S., Formolo, T., Kilpatrick, L. E. and Phinney, K. W. (2016) 'Identification of Novel N-Glycosylation Sites at Noncanonical Protein Consensus Motifs', *Journal of Proteome Research*, 15(7), pp. 2087–2101. doi: 10.1021/acs.jproteome.5b00733.
Lu, H., Yin, H., Wang, A., Shen, J., Yan, X., Liu, Y. and Zhang, C. (2014) 'O-alkylation of disodium salt of diethyl 3,4-dihydroxythiophene-2,5-dicarboxylate with 1,2-dichloroethane catalyzed by ionic type phase transfer catalyst and potassium iodide', *Korean Journal of Chemical Engineering*, 31(1), pp. 45–49. doi: 10.1007/s11814-013-0208-3.
Luo, Q., Chung, H. H., Borths, C. J., Janson, M., Wen, J., Joubert, M. K. and Wypych, J. (2015) 'Structural Characterization of a Monoclonal Antibody (mAb) - Maytansinoid Immunoconjugate', *Analytical Chemistry*, p. acs.analchem.5b03709. doi: 10.1021/acs.analchem.5b03709.
Macedo, C. S. De, Schwarz, R. T., Todeschini, A. R., Previato, J. O. and Mendonça-previato, L. (2010) 'Overlooked post-translational modifications of proteins in Plasmodium falciparum : N - and O -glycosylation - A review', 105(December), pp. 949–956.

Magez, S. and Radwanska, M. (2013) *Trypanosomes and Trypanosomiasis*. Springer Science & Business Media.

Manthri, S., Güther, M. L. S., Izquierdo, L., Acosta-serrano, A. and Ferguson, M. A. J. (2008) 'Deletion of the TbALG3 gene demonstrates site-specific N-glycosylation and N-glycan processing in Trypanosoma brucei', *Glycobiology*, 18(5), pp. 367–383. doi: 10.1093/glycob/cwn014.

Marcello, L. and Barry, J. D. (2007) 'From silent genes to noisy populations - Dialogue between the genotype and phenotypes of antigenic variation', *Journal of Eukaryotic Microbiology*, 54(1), pp. 14–17. doi: 10.1111/j.1550-7408.2006.00227.x.

Mariño, K., Güther, M. L. S., Wernimont, A. K., Qiu, W., Hui, R. and Ferguson, M. A. J. (2011) 'Characterization, localization, essentiality, and high-resolution crystal structure of glucosamine 6-phosphate N-acetyltransferase from Trypanosoma brucei', *Eukaryotic Cell*, 10(7), pp. 985–997. doi: 10.1128/EC.05025-11.

Matthews, K. R., Mcculloch, R., Morrison, L. J. and Matthews, K. R. (2015) 'The within-host dynamics of African trypanosome infections'.

Megger, D. A., Pott, L. L., Ahrens, M., Padden, J., Bracht, T., Kuhlmann, K., Eisenacher, M., Meyer, H. E. and Sitek, B. (2014) 'Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines', *Biochimica et Biophysica Acta - Proteins and Proteomics*. Elsevier B.V., 1844(5), pp. 967–976. doi: 10.1016/j.bbapap.2013.07.017.

Mehlert, A., Bond, C. S. and Ferguson, M. A. J. (2002) 'The glycoforms of a Trypanosoma brucei variant surface glycoprotein and molecular modeling of a glycosylated surface coat.', *Glycobiology*, 12(10), pp. 607–612. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12244073.

Mehlert, A., Zitzmann, N., Richardson, J. M., Treumann, A. and Ferguson, M. A. J. (1998) 'The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of Trypanosoma brucei.', *Molecular and biochemical parasitology*, 91(1), pp. 145–152. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9574932.

Mertins, P., Udeshi, N. D., Clauser, K. R., Mani, D., Patel, J., Ong, S., Jaffe, J. D. and Carr, S. A. (2012) 'iTRAQ Labeling is Superior to mTRAQ for Quantitative Global Proteomics and Phosphoproteomics', *Molecular & Cellular Proteomics*, 11(6), p. M111.014423. doi: 10.1074/mcp.M111.014423.

Miljkovic, M. (2009) *Carbohydrates: Synthesis, Mechanisms and Stereoelectronic Effects*.

Milo, R. (2013) 'What is the total number of protein molecules per cell volume? A call to rethink some published values', *BioEssays*, 35(12), pp. 1050–1055. doi: 10.1002/bies.201300066.

Moffatt, J. G. and Khorana, H. G. (1961) 'Nucleoside Polyphosphates. X. The Synthesis and Some Reactions of Nucleoside-5' Phosphoromorpholidates and Related Compounds. Improved Methods for the Preparation of Nucleoside-5' Polyphosphates', *Journal of the American Chemical Society*, 83(7), pp. 649–658. doi: 10.1021/ja01464a034.

Mohorko, E., Glockshuber, R. and Aebi, M. (2011) 'Oligosaccharyltransferase: The central enzyme of N-linked protein glycosylation', *Journal of Inherited Metabolic Disease*, 34(4), pp. 869–878. doi: 10.1007/s10545-011-9337-1.

Nafffin, J. L., Han, Y., Olivos, H. J., Reddy, M. M., Sun, T. and Kodadek, T. (2003) 'Immobilized Peptides as High-Affinity Capture Agents for Self-Associating Proteins', *Acta Biochimica Polonica*, 10, pp. 251–259. doi:

10.1016/S.
Naik, R. S., Branch, O. H., Woods, A. S., Vijaykumar, M., Perkins, D. J., Nahlen, B. L., Lal, A. A., Cotter, R. J., Costello, C. E., Ockenhouse, C. F., Davidson, E. A. and Gowda, D. C. (2000) 'Glycosylphosphatidylinositol anchors of Plasmodium falciparum: molecular characterization and naturally elicited antibody response that may provide immunity to malaria pathogenesis.', *The Journal of experimental medicine*, 192(11), pp. 1563–76. doi: 10.1084/jem.192.11.1563.
Naik, R. S., Venkatesan, M., Gowda, D. C. and Plasmodium, D. C. (2001) 'Plasmodium falciparum : The Lethal Effects of Tunicamycin and Mevastatin on the Parasite Are Not Mediated by the Inhibition of N-Linked Oligosaccharide Biosynthesis', 114, pp. 110–114. doi: 10.1006/expr.2001.4616.
Nakanishi, M., Karasudani, M., Shiraishi, T., Hashida, K., Hino, M., Ferguson, M. A. J. and Nomoto, H. (2014) 'TbGT8 is a bifunctional glycosyltransferase that elaborates N-linked glycans on a protein phosphatase AcP115 and a GPI-anchor modifying glycan in Trypanosoma brucei.', *Parasitology international*. Elsevier B.V., 63(3), pp. 513–518. doi: 10.1016/j.parint.2014.01.007.
Nielson, S. S. (2009) *Food Analysis Laboratoy Manual*, *The Journal of nutrition*. doi: 10.3329/jbau.v7i1.4985.
Nigro, V. and Savarese, M. (2014) 'Genetic basis of limb-girdle muscular dystrophies : the 2014 update', pp. 1–12.
Nirmalan, N., Sims, P. F. G. and Hyde, J. E. (2004) 'Quantitative proteomics of the human malaria parasite Plasmodium falciparum and its application to studies of development and inhibition.', *Molecular microbiology*, 52(4), pp. 1187–99. doi: 10.1111/j.1365-2958.2004.04049.x.
Nishikawast, Y., Peggs, W., Paulsenll, H. and Schachters, H. (1988) 'Control of Glycoprotein Synthesis', *the Journal of Biological Chemistry*, 263(17), pp. 8270–8281.
O'Sullivan, A. C. (1997) 'Cellulose: the structure slowly unravels', *Cellulose*, 4(3), pp. 173–207. doi: Chemistry and Materials Science.
Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002) 'Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics', *Molecular & Cellular Proteomics*, 1(5), pp. 376–386. doi: 10.1074/mcp.M200025-MCP200.
Ongay, S., Boichenko, A., Govorukhina, N. and Bischoff, R. (2012) 'Glycopeptide enrichment and separation for protein glycosylation analysis', *Journal of Separation Science*, 35(18), pp. 2341–2372. doi: 10.1002/jssc.201200434.
Osamu Kanie, G. G. and Wong, C. H. (2000) 'Removal of benzyl protecting groups from solid-supported compounds by hydrogenolysis using palladium nanoparticles', *Angewandte Chemie - International Edition*, 39, pp. 4545–4547. doi: 10.1002/1521-3773(20001215)39:24<4545::AID-ANIE4545>3.0.CO;2-V.
Pan, S., Chen, R., Aebersold, R. and Brentnall, T. A. (2011) 'Mass Spectrometry Based Glycoproteomics--From a Proteomics Perspective', *Molecular & Cellular Proteomics*, 10(1), p. R110.003251-R110.003251. doi: 10.1074/mcp.R110.003251.
Pasvol, G. (2010) 'Protective hemoglobinopathies and Plasmodium falciparum transmission.', *Nature genetics*. Nature Publishing Group, 42(4), pp. 284–5. doi: 10.1038/ng0410-284.
Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C. and Scrivens, J. H. (2000) 'A

comparison of labelling and label-free mass spectrometry-based proteomics approaches', *Journal of Proteome Research*, 8(7), pp. 3752–3759. doi: 10.1021/pr900080y.

Paterson, M. J. and Eggleston, I. M. (2008) 'Convenient preparation of N-maleoyl amino acid succinimido esters using N-trifluoroacetoxysuccinimide', *Synthetic Communications*, 38(July 2015), pp. 303–308. doi: 10.1080/00397910701750151.

Pays, E., Coquelet, H., Pays, A., Tebabi, P. and Steinert, M. (1989) 'Trypanosoma brucei: posttranscriptional control of the variable surface glycoprotein gene expression site.', *Molecular and cellular biology*, 9(9), pp. 4018–4021. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=362464&tool=pmcentrez&rendertype=abstract.

Pearson, T. W. and Republic, F. (1989) 'Procyclin Gene Expression and Loss of the Variant Surface Glycoprotein during Differentiation of Trypanosoma Brucei', 108(February), pp. 737–746.

Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G. and Strynadka, N. C. (2001) 'Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs.', *Nature structural biology*, 8(2), pp. 166–175. doi: 10.1038/84168.

Power, W. P. (2003) 'High Resolution Magic Angle Spinning - Applications to Solid Phase Synthetic Systems and Other Semi-Solids', *Annual Reports on NMR Spectroscopy*, 51(3), pp. 261–295. doi: 10.1016/S0066-4103(03)51005-5.

Prevention, C. for disease control and (2016) *Biology of malaria*, *Global Health Division of parasitic diseases and malaria*.

Priestley, N. D. (1996) 'Large Scale Synthesis of Cyclodiphospho-D-glycerate', *J. Org. Chem.*, 61(16), pp. 5697–5700.

Ramírez, A. S., Boilevin, J., Biswas, R., Gan, B. H., Janser, D., Aebi, M., Darbre, T., Reymond, J.-L. and Locher, K. P. (2017) 'Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs', *Glycobiology*, 27(6), pp. 1–11. doi: 10.1093/glycob/cwx017.

Ratner, D. M., Cui, J., Steffen, M., Moore, L. L., Robbins, P. W. and Samuelson, J. (2008) 'Changes in the N - Glycome , Glycoproteins with Asn-Linked Glycans , of Giardia lamblia with Differentiation from Trophozoites to Cysts □ †', 7(11), pp. 1930–1940. doi: 10.1128/EC.00268-08.

Van Rhijn, P., Fujishige, N. A., Lim, P. O. and Hirsch, A. M. (2001) 'Sugar-binding activity of pea lectin enhances heterologous infection of transgenic alfalfa plants by Rhizobium leguminosarum biovar viciae.', *Plant physiology*, 126(1), pp. 133–144. doi: 10.1104/pp.126.1.133.

Rijo-Ferreira, F., Pinto-Neves, D., Barbosa-Morais, N. L., Takahashi, J. S. and Figueiredo, L. M. (2017) 'Trypanosoma brucei metabolism is under circadian control', *Nature Microbiology*, 2(March), p. 17032. doi: 10.1038/nmicrobiol.2017.32.

Rini, J., Esko, J. and Varki, A. (2009) *Essentials of Glycobiology.* 2nd edn.

Rong, Y. and Ruoho, A. E. (1999) 'A New Synthetic Approach to Salmeterol', *Synthetic Communications*, 29(12), pp. 2155–2162. doi: 10.1080/00397919908086211.

Roos, D. S. (2001) 'PlasmoDB : An integrative database of the Plasmodium falciparum genome . Tools for

accessing and analyzing finished and unfinished sequence data', *Nuc*, 29(1), pp. 66–69.
Roper, J. R., Güther, M. L. S., MacRae, J. I., Prescott, A. R., Hallyburton, I., Acosta-Serrano, A. and Ferguson, M. A. J. (2005) 'The suppression of galactose metabolism in procylic form Trypanosoma brucei causes cessation of cell growth and alters procyclin glycoprotein structure and copy number', *Journal of Biological Chemistry*, 280(20), pp. 19728–19736. doi: 10.1074/jbc.M502370200.
Roper, J. R., Guther, M. L. S., Milne, K. G. and Ferguson, M. A. J. (2002) 'Galactose metabolism is essential for the African sleeping sickness parasite Trypanosoma brucei.', *Proceedings of the National Academy of Sciences of the United States of America*, 99(9), pp. 5884–5889. doi: 10.1073/pnas.092669999.
'Rules of Carbohydrate Nomenclature' (1963) *The Journal of Organic Chemistry*, 28(2), pp. 281–291. doi: 10.1021/jo01037a001.
Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R. and Robbins, P. W. (2005) 'The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(5), pp. 1548–53. doi: 10.1073/pnas.0409460102.
Sano, M., Hashiba, K., Higashi, M. and Okuda, K. (2015) 'Alteration of glycan structures by swainsonine affects steroidogenesis in bovine luteal cells', *Theriogenology*. Elsevier Inc, 84(5), pp. 827–832. doi: 10.1016/j.theriogenology.2015.05.026.
Sanz, S., Bandini, G., Ospina, D., Bernabeu, M., Mariño, K., Fernández-Becerra, C. and Izquierdo, L. (2013) 'Biosynthesis of GDP-fucose and other sugar nucleotides in the blood stages of Plasmodium falciparum', *Journal of Biological Chemistry*, 288(23), pp. 16506–16517. doi: 10.1074/jbc.M112.439828.
Sanz, S., López-Gutiérrez, B., Bandini, G., Damerow, S., Absalon, S., Dinglasan, R. R., Samuelson, J. and Izquierdo, L. (2016) 'The disruption of GDP-fucose de novo biosynthesis suggests the presence of a novel fucose-containing glycoconjugate in Plasmodium asexual blood stages', *Scientific Reports*. Nature Publishing Group, 6(July), p. 37230. doi: 10.1038/srep37230.
Schaefer, K., Sindhuwinata, N., Hackl, T., Ko, M. P., Niemeyer, F. C., Palcic, M. M., Peters, T. and Meyer, B. (2013) 'A Nonionic Inhibitor with High Specificity for the UDP-Gal Donor Binding Site of Human Blood Group B Galactosyltransferase: Design, Synthesis, and Characterization', *J. Med. Chem.*, 56, pp. 2150–2154.
Schwede, A., Jones, N., Engstler, M. and Carrington, M. (2011) 'The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes', *Molecular and Biochemical Parasitology*. Elsevier B.V., 175(2), pp. 201–204. doi: 10.1016/j.molbiopara.2010.11.004.
Shafer, D. E., Inman, J. K. and Lees, A. (2000) 'Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with maleimide and alpha-haloacyl groups: anomalous elution of TCEP by gel filtration.', *Analytical biochemistry*, 282(1), pp. 161–164. doi: 10.1006/abio.2000.4609.
Shaw, A. D., Tiwari, Y., Kaplan, W., Heath, A., Mitchell, P. B., Schofield, P. R. and Fullerton, J. M. (2014) 'Characterisation of Genetic Variation in ST8SIA2 and Its Interaction Region in NCAM1 in Patients with Bipolar Disorder', 9(3). doi: 10.1371/journal.pone.0092556.
Shaw, A. P. M., Cecchi, G., Wint, G. R. W., Mattioli, R. C. and Robinson, T. P. (2014) 'Mapping the economic benefits to livestock keepers from intervening against bovine trypanosomosis in Eastern Africa', *Preventive*

Veterinary Medicine. Elsevier B.V., 113(2), pp. 197–210. doi: 10.1016/j.prevetmed.2013.10.024.
Shrimal, S., Ng, B. G., Losfeld, M. E., Gilmore, R. and Freeze, H. H. (2013) 'Mutations in STT3A and STT3B cause two congenital disorders of glycosylation', *Human Molecular Genetics*, 22(22), pp. 4638–4645. doi: 10.1093/hmg/ddt312.
Sindhuwinata, N., Munoz, E., Munoz, F. J., Palcic, M. M., Peters, H. and Peters, T. (2010) 'Binding of an acceptor substrate analog enhances the enzymatic activity of human blood group B galactosyltransferase', *Glycobiology*, 20(6), pp. 718–723. doi: 10.1093/glycob/cwq019.
Singh, A., Dahanayaka, D. H., Biswas, A., Bumm, L. A. and Halterman, R. L. (2010) 'Molecularly ordered decanethiolate self-assembled monolayers on Au(111) from in situ cleaved decanethioacetate: An NMR and STM study of the efficacy of reagents for thioacetate cleavage', *Langmuir*, 26(16), pp. 13221–13226. doi: 10.1021/la100103k.
Singh, C., Zampronio, C. G., Creese, A. J. and Cooper, H. J. (2012) 'Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N‐Linked Glycoproteins'.
Somawardhana, C. W. and Brunngraber, E. G. (1983) 'Stereo and regio selectivity of Diethylaminosulfur trifluoride as a fluorinating reagent for methyl glycosides', *Carbohydrate Research*, 121, pp. 51–60.
de Souza, A. C., Halkes, K. M., Meeldijk, J. D., Verkleij, A. J., Vliegenthart, J. F. G. and Kamerling, J. P. (2004) 'Synthesis of Gold Glyconanoparticles: Possible Probes for the Exploration of Carbohydrate-Mediated Self-Recognition of Marine Sponge Cells.', *European Journal of Organic Chemistry*, 2004(21), pp. 4323–4339. doi: 10.1002/ejoc.200400255.
Speers, A. E. and Cravatt, B. F. (2004) 'Chemical strategies for activity-based proteomics.', *Chembiochem : a European journal of chemical biology*, 5(1), pp. 41–47. doi: 10.1002/cbic.200300721.
Spiro, R. G. (2002) 'Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds', *Glycobiology*, 12(4), p. 43R–56R. doi: 10.1093/glycob/12.4.43R.
Stanciu, C. and Bennett, J. R. (1974) 'Preliminary communications.', *British medical journal*, 1(5903), pp. 312–313. doi: 10.3109/00016926209171741.
Stanley, P. (2016) 'What Have We Learned from Glycosyltransferase Knockouts in Mice?', *Journal of Molecular Biology*. Elsevier B.V., 428(16), pp. 3166–3182. doi: 10.1016/j.jmb.2016.03.025.
Stanley, P. and Cummings, R. (2009) *Essentials of Glycobiology*. 2nd edn. Cold Spring Harbor Laboratory Press.
Stick, R. V. (2001) *Carbohydrates - The Sweet Molecules of Life*. ACADEMIC PRESS.
Stijlemans, B., Caljon, G., Van Den Abbeele, J., Van Ginderachter, J. A., Magez, S. and De Trez, C. (2016) 'Immune evasion strategies of Trypanosoma brucei within the mammalian host: Progression to pathogenicity', *Frontiers in Immunology*, 7, pp. 1–14. doi: 10.3389/fimmu.2016.00233.
Stokes, M. J., Güther, M. L. S., Turnock, D. C., Prescott, A. R., Martin, K. L., Alphey, M. S. and Ferguson, M. A. J. (2008) 'The synthesis of UDP-N-acetylglucosamine is essential for bloodstream form Trypanosoma brucei in vitro and in vivo and UDP-N-acetylglucosamine starvation reveals a hierarchy in parasite protein glycosylation', *Journal of Biological Chemistry*, 283(23), pp. 16147–16161. doi: 10.1074/jbc.M709581200.
Sujino, K., Uchiyama, T., Hindsgaul, O., Seto, N. O. L., Wakarchuk, W. W. and Palcic, M. M. (2000) 'Enzymatic

synthesis of oligosaccharide analogues: Evaluation of UDP-Gal analogues as donors for three retaining α-galactosyltransferases', *Journal of the American Chemical Society*, 122(7), pp. 1261–1269. doi: 10.1021/ja990964u.

Sutherland, C. S., Yukich, J., Goeree, R. and Tediosi, F. (2015) 'A Literature Review of Economic Evaluations for a Neglected Tropical Disease: Human African Trypanosomiasis ("Sleeping Sickness")', *PLOS Neglected Tropical Diseases*, 9, p. e0003397. doi: 10.1371/journal.pntd.0003397.

Suzuki, O., Abe, M. and Hashimoto, Y. (2015) 'Sialylation by b-galactoside a-2,6-sialyltransferase and N-glycans regulate cell adhesion and invasion in human anaplastic large cell lymphoma', *International Journal of Oncology*, 46(3), pp. 973–980. doi: 10.3892/ijo.2015.2818.

Swearingen, K. E., Lindner, S. E., Shi, L., Shears, M. J., Harupa, A., Hopp, C. S., Vaughan, A. M., Springer, T. A., Moritz, R. L., Kappe, S. H. I. and Sinnis, P. (2016) 'Interrogating the Plasmodium Sporozoite Surface: Identification of Surface-Exposed Proteins and Demonstration of Glycosylation on CSP and TRAP by Mass Spectrometry-Based Proteomics', *PLoS Pathogens*, 12(4), pp. 1–32. doi: 10.1371/journal.ppat.1005606.

Takahashi, S. and Nakata, T. (2002) 'Total synthesis of an antitumor agent, mucocin, based on the "chiron approach"', *Journal of Organic Chemistry*, 67(l), pp. 5739–5752. doi: 10.1021/jo020211h.

Takasaki, S., Mizuochi, T. and Kobata, A. (1982) 'Hydrazinolysis of Asparagine-Linked Sugar Chains to Produce Free Oligosaccharides', *Methods in Enzymology*, 83, pp. 263–268. doi: 10.1016/0076-6879(82)83019-X.

Takaya, K., Nagahori, N., Kurogochi, M., Furuike, T., Miura, N., Monde, K., Lee, Y. C. and Nishimura, S. (2005) 'Rational Design , Synthesis , and Characterization of Novel Inhibitors for', *Journal of medicinal chemistry*, 48, pp. 6054–6065.

Takayama, S., Chung, S. J., Igarashi, Y., Ichikawa, Y., Sepp, A., Lechler, R. I., Wu, J., Hayashi, T., Siuzdak, G. and Wong, C. H. (1999) 'Selective inhibition of beta-1,4- and alpha-1,3-galactosyltransferases: donor sugar-nucleotide based approach.', *Bioorganic & medicinal chemistry*, 7, pp. 401–409. doi: 10.1016/S0968-0896(98)00249-1.

Tam, J. P., Heath, W. F. and Meriifield, R. B. (1983) 'SN2 deprotection of synthetic peptides with a low concetration of HF in dimethyl sulfide: evidence and application in peptide synthesis', *J. Am. Chem. Soc*, 105, pp. 6442–6455. doi: 10.1021/ja00359a014.

Tarentino, A. L., Trimble, R. B. and Plummer, T. H. (1989) 'Enzymatic approaches for studying the structure, synthesis, and processing of glycoproteins.', *Methods in cell biology*, 32, pp. 111–139. doi: 10.1016/S0091-679X(08)61169-3.

Taylor, M. E. and Drickamer, K. (2011) *Introduction to Glycobiology*. Oxford University Press.

Tonn, V. C. and Meier, C. (2011) 'Solid-phase synthesis of (poly)phosphorylated nucleosides and conjugates', *Chemistry - A European Journal*, 17(35), pp. 9832–9842. doi: 10.1002/chem.201101291.

Trindade, S., Rijo-Ferreira, F., Carvalho, T., Pinto-Neves, D., Guegan, F., Aresta-Branco, F., Bento, F., Young, S. A., Pinto, A., Van Den Abbeele, J., Ribeiro, R. M., Dias, S., Smith, T. K. and Figueiredo, L. M. (2016) 'Trypanosoma brucei Parasites Occupy and Functionally Adapt to the Adipose Tissue in Mice', *Cell Host and Microbe*, 19(6), pp. 837–848. doi: 10.1016/j.chom.2016.05.002.

Tu, Y. (2011) 'The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine', *Nature Medicine*.

Nature Publishing Group, 17(10), pp. 1217–1220. doi: 10.1038/nm.2471.

Turnock, D. C. and Ferguson, M. A. J. (2007) 'Sugar nucleotide pools of Trypanosoma brucei, Trypanosoma cruzi, and Leishmania major', *Eukaryotic Cell*, 6(8), pp. 1450–1463. doi: 10.1128/EC.00175-07.

Turnock, D. C., Izquierdo, L. and Ferguson, M. A. J. (2007) 'The de novo synthesis of GDP-fucose is essential for flagellar adhesion and cell growth in Trypanosoma brucei', *Journal of Biological Chemistry*, 282(39), pp. 28853–28863. doi: 10.1074/jbc.M704742200.

Urbaniak, M. D., Martin, D. M. A. and Ferguson, M. A. J. (2013) 'Global quantitative SILAC phosphoproteomics reveals differential phosphorylation is widespread between the procyclic and bloodstream form lifecycle stages of Trypanosoma brucei.', *Journal of proteome research*, 12(5), pp. 2233–2244. doi: 10.1021/pr400086y.

Urbaniak, M. D., Turnock, D. C. and Ferguson, M. A. J. (2006) 'Galactose starvation in a bloodstream form Trypanosoma brucei UDP-glucose 4???-epimerase conditional null mutant', *Eukaryotic Cell*, 5(11), pp. 1906–1913. doi: 10.1128/EC.00156-06.

Varki, A. (2017) 'Biological roles of glycans', *Glycobiology*, 27(1), pp. 3–49.

Varki, A., Cummings, R. and Esko, J. D. (2009) 'Essentials of Glycobiology 2nd Edition', in, p. Chapter 8.

Varki, A., Cummings, R., Esko, J. D. and Al., E. (2009) *Essentials of Glycobiology 2nd Edition*. 2nd edn. Cold Spring Harbor Laboratory Press.

Vidal, S., Bruyere, I., Malleron, A., Auge, C. and Praly, J. P. (2006) 'Non-isosteric C-glycosyl analogues of natural nucleotide diphosphate sugars as glycosyltransferase inhibitors', *Bioorganic and Medicinal Chemistry*, 14(21), pp. 7293–7301. doi: 10.1016/j.bmc.2006.06.057.

Wagstaff, B. a., Rejzek, M., Pesnot, T., Tedaldi, L. M., Caputi, L., O'Neill, E. C., Benini, S., Wagner, G. K. and Field, R. a. (2015) 'Enzymatic synthesis of nucleobase-modified UDP-sugars: Scope and limitations', *Carbohydrate Research*. Elsevier Ltd, 404, pp. 17–25. doi: 10.1016/j.carres.2014.12.005.

Wang, Y.-C., Stein, J. W., Lynch, C. L., Tran, H. T., Lee, C.-Y., Coleman, R., Hatch, A., Antontsev, V. G., Chy, H. S., O'Brien, C. M., Murthy, S. K., Laslett, A. L., Peterson, S. E. and Loring, J. F. (2015) 'Glycosyltransferase ST6GAL1 contributes to the regulation of pluripotency in human pluripotent stem cells.', *Scientific reports*. Nature Publishing Group, 5(August), p. 13317. doi: 10.1038/srep13317.

Weiss, W. A., Taylor, S. S. and Shokat, K. M. (2007) 'Recognizing and exploiting differences between RNAi and small-molecule inhibitors.', *Nature chemical biology*, 3(12), pp. 739–744. doi: 10.1038/nchembio1207-739.

Wellems, T. E. and Plowe, C. V (2001) 'Chloroquine-resistant malaria', *Journal of Infectious Diseases*, 184(6), pp. 770–776. doi: 10.1086/322858.

Whitby, K., Pierson, T. C., Geiss, B., Engle, M., Zhou, Y., Doms, R. W., Diamond, S., Lane, K. and Diamond, M. S. (2005) 'Castanospermine , a Potent Inhibitor of Dengue Virus Infection In Vitro and In Vivo Castanospermine , a Potent Inhibitor of Dengue Virus Infection In Vitro and In Vivo', *Journal of Virology*, 79(14), pp. 8698–8706. doi: 10.1128/JVI.79.14.8698.

Whiteheart, S. W., Passaniti, A., Reichner, J. S., Holt, G. D., Haltiwanger, R. S. and Hart, G. W. (1989) 'Glycosyltransferase Probes', *Analytical Methods*, 179(1981), pp. 82–95.

WHO (2016) *World malaria report 2016*. doi: 978 92 4 151171 1.

Wiese, S., Reidegeld, K. A., Meyer, H. E. and Warscheid, B. (2007) 'Protein labeling by iTRAQ: A new tool for

quantitative mass spectrometry in proteome research', *Proteomics*, 7(3), pp. 340–350. doi: 10.1002/pmic.200600422.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Wiśniewski, J. R., Zougman, A., Nagaraj, N. and Mann, M. (2009) 'Universal sample preparation method for proteome analysis. Supplemental', *Nature methods*, 6(5), pp. 359–363. doi: 10.1038/nmeth.1322.

Wittmann, V. and Wong, C. H. (1997) '1H-tetrazole as catalyst in phosphomorpholidate coupling reactions: Efficient synthesis of GDP-fucose, GDP-mannose, and UDP-galactose', *Journal of Organic Chemistry*, 62(7), pp. 2144–2147. doi: 10.1021/jo9620066.

Wojtowicz, K., Januchowski, R., Nowicki, M. and Zabel, M. (2015) 'Inhibition of protein glycosylation reverses the MDR phenotype of cancer cell lines', *Biomedicine and Pharmacotherapy*. Elsevier Masson SAS, 74, pp. 49–56. doi: 10.1016/j.biopha.2015.07.001.

Xiao, H., Tang, G. X. and Wu, R. (2016) 'Site-Specific Quantification of Surface N-Glycoproteins in Statin-Treated Liver Cells', *Analytical Chemistry*, 88(6), pp. 3324–3332. doi: 10.1021/acs.analchem.5b04871.

Xu, W., Springfield, S. a. and Koh, J. T. (2000) 'Highly efficient synthesis of 1-thioglycosides in solution and solid phase using iminophosphorane bases', *Carbohydrate Research*, 325(3), pp. 169–176. doi: 10.1016/S0008-6215(99)00327-4.

Yao, Y., Yu, L., Su, X., Wang, Y., Li, W., Wu, Y., Cheng, X., Zhang, H., Wei, X., Chen, H., Zhang, R., Gou, L., Chen, X., Xie, Y., Zhang, B., Zhang, Y., Yang, J. and Wei, Y. (2015) 'Synthesis, characterization and targeting chemotherapy for ovarian cancer of trastuzumab-SN-38 conjugates', *Journal of Controlled Release*. Elsevier B.V., 220, pp. 5–17. doi: 10.1016/j.jconrel.2015.09.058.

Yeh, Y. and Feeney, R. E. (1996) 'Antifreeze Proteins: Structures and Mechanisms of Function', *Chemical Reviews*, 96(2), pp. 601–618. doi: 10.1021/cr950260c.

Yoshitake, S., Imagawa, M., Ishikawa, E., Nitsu, Y., Urushizaki, I., Nishiura, M., Kanazawa, R., Kurosaki, H., Tachibana, S., Nakazawa, N. and Ogawa, H. (1982) 'Mild and efficient conjugation of Rabbit Fab' and Horseradish peroxidase using a maleimide compound and its use for enzyme immunoassay.', *Journal of Biological Chemistry*, 92(5), pp. 1413–1424.

Young, S. A. and Smith, T. K. (2010) 'The essential neutral sphingomyelinase is involved in the trafficking of the variant surface glycoprotein in the bloodstream form of Trypanosoma brucei', *Molecular Microbiology*, 76(6), pp. 1461–1482. doi: 10.1111/j.1365-2958.2010.07151.x.

Zhao, P., Viner, R., Teo, C. F., Boons, G., Horn, D. and Wells, L. (2011) 'Combining High-Energy C-Trap Dissociation and Electron Transfer Dissociation for Protein O-GlcNAc Modification Site Assignment', *Journal of proteome research*, 10, pp. 4088–4104.

Zhao, R. Y., Wilhelm, S. D., Audette, C., Jones, G., Leece, B. A., Lazar, A. C., Goldmacher, V. S., Singh, R., Kovtun, Y., Widdison, W. C., Lambert, J. M. and Chari, R. V. J. (2011) 'Supporting Information for Synthesis and Evaluation of Hydrophilic Linkers for Antibody- Maytansinoid Conjugates', 20(c), pp. 3606–3623.