THE UNIVERSITY
of EDINBURGH

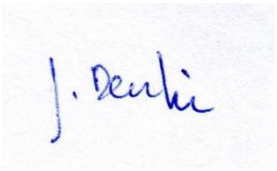
A spoonful of sugar helps the medicine go down: Biomanufacture in glycoengineered *Pichia pastoris* of the potentially therapeutic recombinant glycoprotein factor H

John Devlin

PhD

2017

Unless stated otherwise, the work described in this thesis is my own work and has not been submitted for in whole or in part for a degree of other qualification at this or any other university.



John Devlin

Abstract

Glycoengineering is a technology that could improve protein therapeutics. While protein glycosylation in general enhances solubility and stability, and reduces aggregation, immunogenicity and proteolysis, specific kinds of glycosylation may also be critical. For example, capping of glycans with N-acetylneuraminic acid (Neu5Ac) maximises circulatory half-life in humans. Moreover, some glycans directly participate in molecular recognition and other aspects of glycoprotein function.

Glycoproteins produced by non-human mammalian cells carry glycans capped by N-glycolyl-neuraminic acid rather than Neu5Ac. Yet production in human cell lines is costly and slow, requires specialist facilities, produces low yields and is subject to additional regulations. Hence there is a case for glycoengineering alternative expression systems capable of rapid, low-cost, high-yield glycoprotein production. This report focuses on the glycoengineering of *Pichia pastoris*, a yeast, to produce recombinant human glycoprotein factor H (FH) bearing human-like glycans.

FH is a potent down-regulator of the complement system. Mutations and SNPs in FH result in autoimmune diseases such as atypical haemolytic ureamic syndrome and age-related macular degeneration (AMD). Recombinant FH is an enticing therapeutic candidate for treating AMD, but high doses are required since FH is abundant (200-300 mg l^{-1}) in normal human serum. Human FH (155 kDa), with eight sites of N-linked glycosylation and 40 disulphides, is a challenging target for recombinant production. Yet FH was previously expressed to 10s of milligrams in *P. pastoris*.

In this study, methods were established to confirm that human plasma-derived (h)FH carries predominantly N-linked diantennary disialylated complex-type glycans, with monosialylated diantennary structures and triantennary structures in fucosylated and non-fucosylated forms, contributing to glycan heterogeneity. Functional comparison of native hFH, enzymatically desialylated (DeSia-) hFH and deglycosylated recombinant *P. pastoris*-produced (DeGly-r)FH showed that DeSia-hFH had the lowest affinity for complement protein C3b, its key target. Moreover, DeSia-hFH binds C3d, an opsonic C3b-breakdown product, whereas native hFH does not. DeSia-hFH had an improved ability to accelerate decay of the C3 convertase (an enzyme that cleaves C3 to C3b) compared to native hFH, but neither was as good as DeGly-rFH in this respect. In contrast, DeGly-rFH had reduced cofactor activity (for factor I-mediated degradation of C3b) compared to native hFH whereas DeSia-hFH did not have reduced cofactor activity. These data suggest that sialylation of FH glycans may play a role in stabilising a conformation of circulating FH that is not fully effective, consistent with specificity for self-surfaces and resistance to bacterial hijack.

Aiming eventually to produce human-like glycosylated FH in glycoengineered *P. pastoris*, the SuperMan 5 strain served as a starting point. While conventional strains of *P. pastoris* put hypermannosylated N-linked glycans on proteins, glycans on SuperMan 5-produced FH were shown to contain just five mannose (Man) residues. In further glycoengineering, and following unsuccessful efforts to use inABLE technology for this purpose, commercially available (GlycoSwitch) vectors were used to introduce genes encoding the glycosyltransferase enzymes N-acetylglucosamine (GlcNAc) transferase I (GnTI) and galactose (Gal) transferase. These catalysed the formation of a hybrid-type glycan containing an N-acetyllactosamine (Gal-β(1,4)-GlcNAc (LacNAc)) antennae on a five-mannose glycan. Then two more GlycoSwitch plasmids, containing genes encoding α-Mannosidase II (ManII) and GnTII, were introduced into *P. pastoris* to catalyse the formation of a second LacNAc antennae. MALDI-TOF analysis found the glycosylation of this strain to be heterogeneous,

containing the humanised diantennary digalactosyl glycan as well as other endogenous yeast glycans. This strain was designated SuperGal. Large-scale expression of rFH with terminally galactosylated complex-type glycans (Gal-rFH) in SuperGal yielded 100s of milligrams of purified Gal-rFH. Yeast-type glycans were enzymatically removed from rFH and the remaining complex-type humanised glycans were sialylated with a recombinant bacterial α(2,6)-sialyltransferase from *Photobacterium sp.* expressed in *E.coli*.

Purified sialylated (Sia-) and non-sialylated (Gal-) rFH expressed in SuperGal were functionally characterised *in vitro* using SPR-based assays. In C3b-binding assays Sia-rFH had lower affinity compared to Gal-rFH. Both bound with lower affinity than DeGly-rFH. A similar pattern of binding affinity was seen for C3d. In C3 convertase decay-acceleration assays, all rFH glycoforms performed equally well and had greater activity than hFH. Conversely, Sia- and Gal-rFH were shown to perform equally as well as hFH in CA assays, while all three versions outperformed DeGly-rFH. However, *in vivo* complement activity assay carried out in a FH-knockout mouse model showed that humanisation of the glycosylation of rFH did not significantly improve activity compared to DeGly-rFH. In addition, analysis of the circulatory half-life of rFH showed that humanisation did not improve half-life.

Further engineering steps will be required to increase the complex-type glycan site occupancy on rFH with a view to improving circulatory half-life and efficacy. However, this study represents a significant step forward in developing a therapeutically useful source of rFH.

Lay Summary

Genes encode proteins and proteins carry out the intra- and extracellular functions for life. Some proteins, especially those that occupy the extracellular environment (e.g. in the blood), are decorated with sugar chains called glycans. Glycans confer several advantages to proteins in the extracellular environment including protection and signal recognition – e.g. signalling to the immune system that the glycosylated protein is safe.
An immune pathway regulatory protein called factor H (FH) protects self-surfaces from auto-immune attack. However, mutations in the gene which encodes FH can lead to autoimmune disease which could result in blindness and kidney defects. A potential therapy for patients with a deficiency in FH is providing them with a source of functional FH.

Using genetic engineering it is possible to produce high quantities of human proteins in micro-organisms such as yeasts. The main disadvantage to using yeast to produce human proteins with glycans is that yeast attach the wrong kind of glycan to proteins. The glycans on yeast derived human FH would signal to the human immune system that the protein is not safe, resulting in the proteins destruction.

This study first focused on characterising the glycans on human FH purified from human blood to answer the questions what do the glycans look like and what role do they have in the function of human FH? Then, this study focused on changing the glycosylation pathway of yeast to humanise the glycans attached to FH. Finally, the function of yeast produced FH with human glycans was characterised in binding studies and in FH deficient mice to answer the question: Does yeast produced FH with human glycans function the same as human blood derived FH?

Acknowledgements

I would first like to thank my supervisors Prof. Paul Barlow and Dr. Ian Fotheringham for giving me the opportunity to study for a PhD.
I would like to thank the Biotechnology and Biological Sciences Research Council for funding this Collaborative Awards in Science and Engineering (CASE) PhD studentship.
I would like to express my gratitude to Mrs. Harveen Erskine and Dr. Franck Escalettes for their support and guidance throughout.
Thanks go to Dr. Eliza Makou, Dr. Andy Herbert, Dr. Yi Yang and Dr. Kevin Marchbank for their assistance.
I would like to thank the team at Ingenza for welcoming me into the company and making me feel part of the team from day one. I would especially like to thank the Bio 1 team Allan, Jack and Roxann for their enjoyable company.
Special thanks go to Atanas for his entertaining chats in the pub about science.
Last, but by no means least, my thanks and eternal gratitude to Dr. Harriet Gordon for her endless support and encouragement at all times, without which I could never have completed my PhD.

Contents

List of Figures

List of Tables

Abbreviations

ADH1p	alcohol dehydrogenase 1 promoter
AGP	α_1-acid glycoprotein
ALG	altered in glycosylation
aoWR	N-α-aminooxyacetyltryptophanylarginine methyl ester
AOX1	alcohol oxidase I
AOX1t	AOX1 terminator
AP	alternative pathway of complement
ATP	adenosine triphosphate
BEDS	bicine-NaOH ethylene glycol dimethyl sulfoxide sorbitol
BMGY	buffer complex glycerol media
BMMY	buffer complex methanol media
BSA	bovine serum albumin
c	ten times concentrated
CA	cofactor activity
CAZy	Carbohydrate Active enZyme database
CCP	compliment control protein
CDP	cytidine diphosphate
CFH	complement factor H gene
CI	confidence interval
CMP	cytidine monophosphate
ConA	concanavalin A
CP	classic pathway of complement
CSS	CMP-sialic acid synthase
CST	CMP-sialic acid transporter
CTP	cytidine triphosphate
CV	column volume
d	distilled
DAA	decay acceleration activity
dd	double distilled
DeGly-rFH	endoglycosidase H_f enzymatically deglycosylated rFH
DeSia-hFH	α(2-3,6,8)-neuraminidase enzymatically desialylated human FH
DNA	deoxyribonucleic acid
ds	double stranded
DTT	dithiothreitol
ECL	*Erythrina cristagalli* lectin
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay
Endo H_f	endoglycosidase H_f
EPO	erythropoietin
ER	endoplasmic reticulum
FB	factor B
FD	factor D
FH	factor H
FI	factor I
FT	flowthrough
GAGs	glycosaminoglycans

Gal	galactose
GAL7t	galactose-1-phosphate uridylyltransferase terminator
Gal-rFH	rFH carrying galactose terminal glycans
GalT	galactosyltransferase
GAPp	glyceraldehyde-3-phosphate dehydrogenase promoter
Gd	glycodelin
Glc	glucose
GlcNAc	N-acetylglucosamine
GlcNAc-rFH	rFH carrying N-acetylglucosamine terminal glycans
GNE	UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase
GnTI	N-acetylglucosaminyltransferase I
GnTII	N-acetylglucosaminyltransferase II
hFH	human factor H
His_6	polyhexahistidine-tag
HRP	horseradish peroxidase
Ig	immunoglobulin
ILV5p	acetohydroxy acid isomeroreductase promoter
IPTG	isopropylthio-β-galactoside
K_D	equilibrium dissociation constant
LacNAc	N-acetyllactosamine
LB	lysogeny broth
LO	linker oligonucleotide
LP	lectin pathway of complement
MALDI	matrix assisted laser desorption/ionisation
Man	mannose
ManI	α-mannosidase I
ManII	α(1,2)-mannosidase
ManNAc-6-P	N-acetylmannosamine-6-phosphate
MBP	mannose binding protein
MCS	multiple cloning site
MPGNII	Membranoproliferative glomerulonephritis II
MW	molecular weight
Neu5Ac	N-acetylneuraminic acid
OD_{600}	optical density at 600 nm wavelength
OST	oligosaccharyltransferase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pEcPh-6ST	*E. coli* expression plasmid encoding Ph-6ST
Ph-6ST	*Photobacterium sp.* α(2,6)-sialyltransferase
PMSF	phenylmethylsulphonyl fluoride
PNGase F	peptide-N-glycosidase F
PO	part oligonucleotide
PVDF	polyvinylidene fluoride
RE	restriction endonuclease
rFH	recombinant factor H
R_{max}	theoretical maximum response
RNase B	ribonuclease B
rpm	revolutions per minute
RU	response unit
SDS	sodium dodecyl sulphate

SE	standard error
Sia-rFH	rFH carrying sialic acid terminal glycans
SNA	*Sambucus nigra* agglutinin
SOC	super optimal with catabolite repression
SPP	sialylated-9-phosphate phosphatase
SPR	surface plasmon resonance
SPS	N-acetylneuraminate-9-phosphate synthase
SPZ	sorbitol phosphate zymolase
t	truncated
TEFp	translation elongation factor 1-α promoter
THAP	2',4',6'-trihydroxyacetophenone monohydrate
TMB	3,3',5,5'-tetramethylbenzidine
TOF	time of flight
TP	truncated part
TPS1t	α,α-trehalose-phosphate synthetase terminator
U	enzyme unit of activity
W	wash
YPD	yeast peptone dextrose
YPDS	yeast peptone dextrose sorbitol
χ^2	Chi squared

Glossary

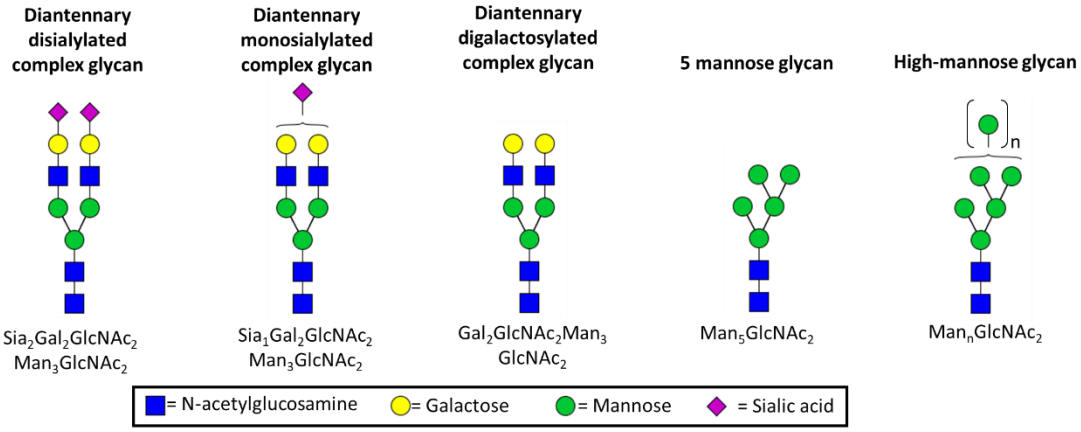


Figure G.1: Structures of the most prevalent glycans in this study. Graphic structures (non-reducing terminus top, reducing terminus bottom) and short hand notation (written non-reducing terminus left to reducing terminus right) are represented.

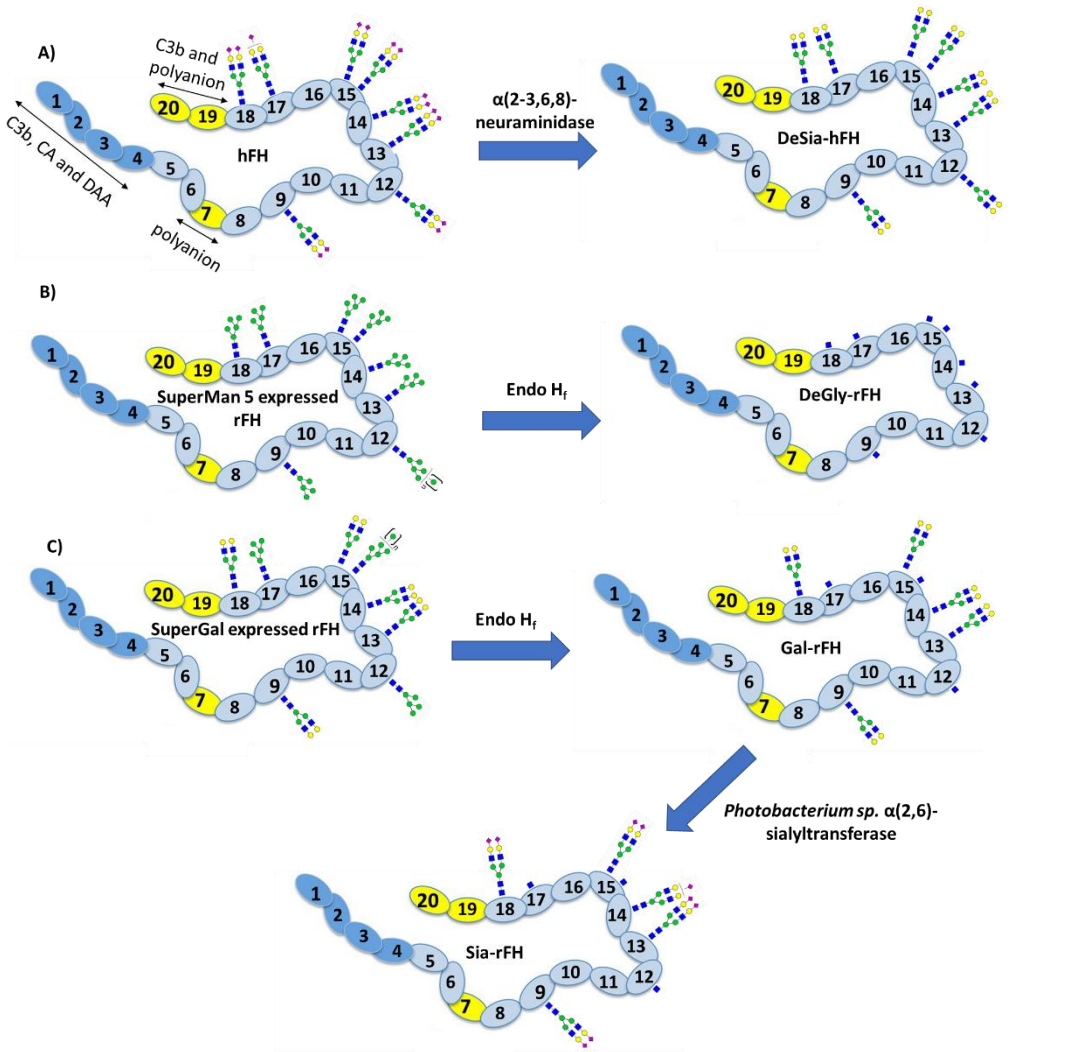


Figure G2: Figure of the different glycoforms of factor H (FH) produced in this study. Graphic representation of A) human (h)FH before and after (DeSia-hFH) enzymatic desialylation with α(2-3,6,8)-neuraminidase. B) recombinant (r)FH expressed in commercially available glycoengineered *Pichia pastoris* strain SuperMan 5 before and after (DeGly-rFH) enzymatic deglycosylation with endoglycosidase (Endo) H_f. C) rFH expressed in *P. pastoris* strain SuperGal, glycoengineered to produce humanised diantennary galactose terminal glycans, before and after (Gal-rFH) enzymatic deglycosylation with Endo H_f. Endo H_f treated Gal-rFH was then sialylated with recombinant, His-tagged *Photobacterium sp.* α(2,6)-sialyltransferase expressed in *E. coli*.

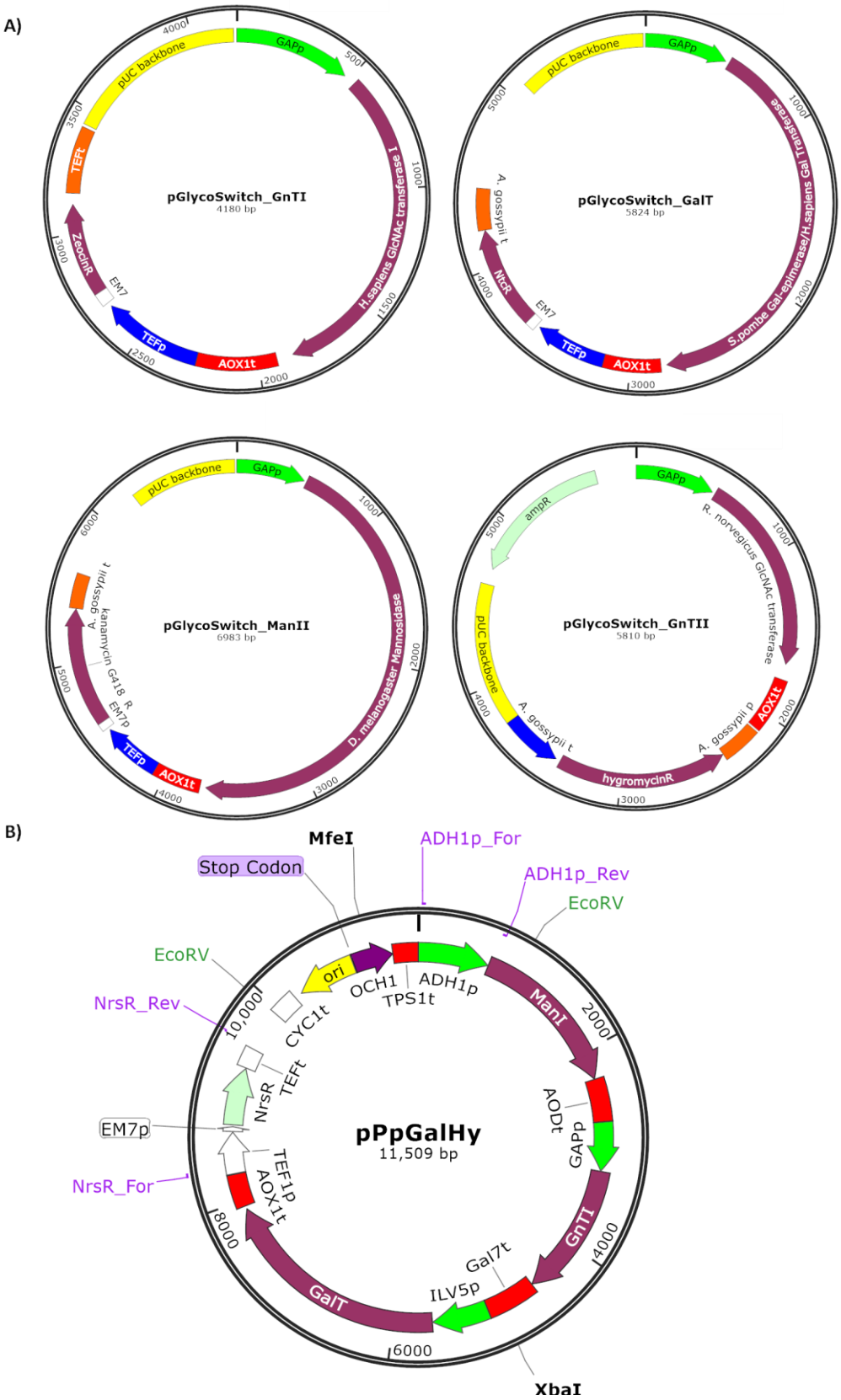


Figure G3: Figure of the *P. pastoris* expression shuttle vector plasmids used in this study. A) Commercially available GlycoSwitch plasmids which, when used to transform *P. pastoris* strain SuperMan 5, encode enzymes for the biosynthesis of humanised diantennary complex-type galactose terminal glycans in *P. pastoris*. B) *P. pastoris* expression shuttle vector plasmid generated in this study using the proprietary inABLE DNA assembly technique. This plasmid (pPpGalHy) contains multiple genes on one plasmid that encode enzymes for the biosynthesis of an intermediate hybrid-type galactose terminal glycan.

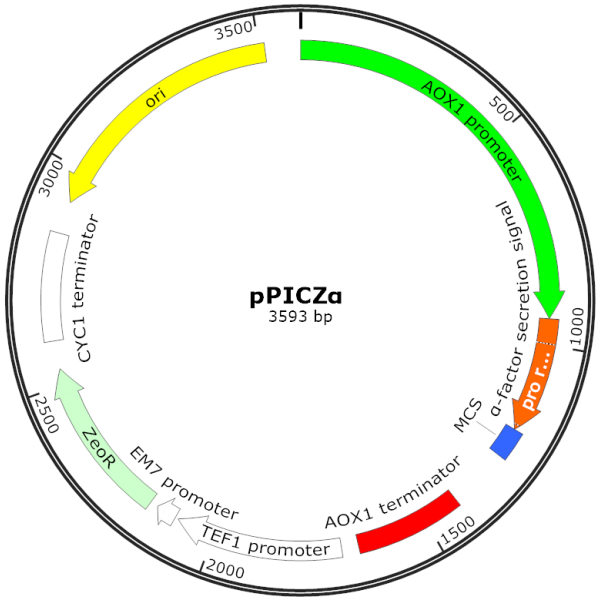


Figure G4: Figure of the commercially available *P. pastoris* expression shuttle vector plasmids pPICZα. In a previous study, Dr. Heather Kerr cloned the human *CFH* gene sequence into the multiple cloning site (MCS) and subsequently transformed the expression plasmid into *P. pastoris* strain SuperMan 5. This FH expressing SuperMan 5 strain was subsequently used in this study to modify the glycosylation pathway.

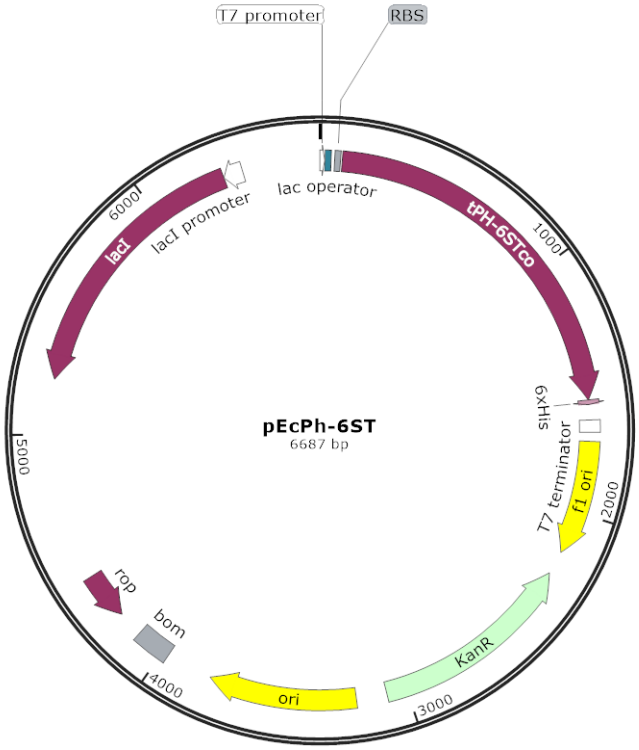


Figure G5: Figure of the commercially available pET24 *E. coli* expression plasmid modified, in this study, by cloning a bacterial *Photobacterium sp.* α(2,6)-sialyltransferase gene sequence (tPH-6STco) in frame with a His_6-tag DNA sequence. The plasmid was transformed into *E. coli* expression strain BL2, the encoded enzyme PH-6ST was expression and purified by Nickel(ii) immobilised metal affinity chromatography and used to sialylated Gal-rFH.

Chapter 1

Introduction

1.1 N-linked Glycosylation

Glycans are short, usually branched, chains of monosaccharides covalently conjugated to each other by glycosidic linkages. One terminus of a glycan is called the reducing end; this is the part of the glycan conjugated to either protein or lipid. The reducing end is so named because, when liberated from the protein or lipid conjugate, the terminal monosaccharide is a hemiacetal and thus can be reduced (by, for example, Tollen's reagent). At the other terminus of the glycan is the non-reducing end, so called because the oxygen attached to C1 (C2 in the case of sialic acid) (fig. 1.1) forms part of the glycosidic linkage with the preceding monosaccharide in the chain and, therefore, cannot be reduced.

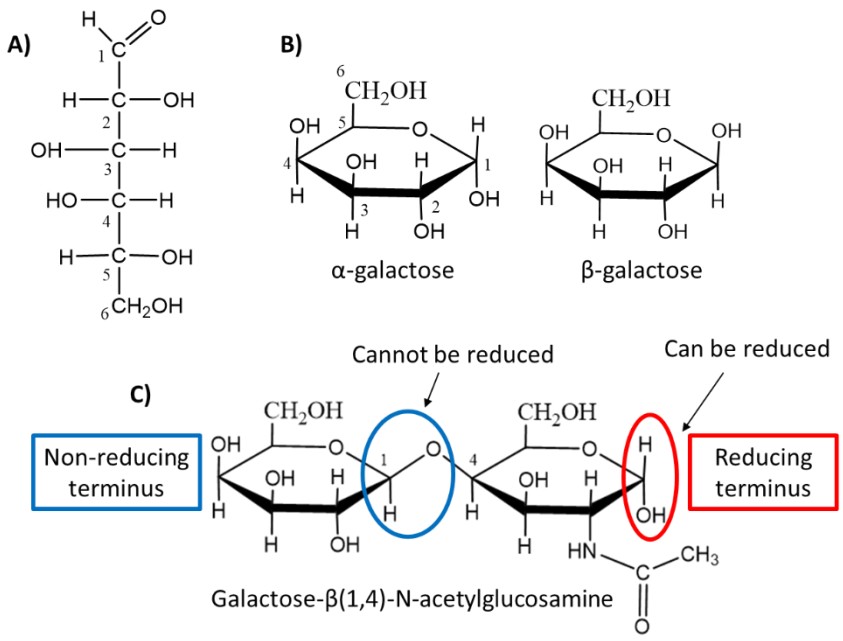


Figure 1.1: A) Linear Fischer projection and B) Haworth projection of galactose. Carbons are labelled from the first carbon in the backbone starting nearest the carbonyl carbon. In B) α- and β-anomers are identified from the position of the anomeric centre (C1) relative to the anomeric reference atom (C5) – for example, in α-galactose the C1 hydroxyl group is *trans* relative to the C5 carbon CH_2OH group. Conversely, in β-galactose the C1 OH is *cis* to the C5 CH_2OH group. C) Haworth projection of the disaccharide β(1,4)-N-acetyllactosamine. The convention is to write from left to right the non-reducing terminus to the reducing terminus, so called because the reducing terminus has a hemiacetal group that can be reduced. The glycosidic linkage is β(1,4) because it forms between the C1 carbon of galactose and the C4 carbon of N-acetylglucosamine and the bond is *cis* relative to the C5 CH_2OH group on galactose.

The convention for writing glycans is from non-reducing end to reducing end, the equivalent of 5' to 3' for nucleic acids, or N-terminus to C-terminus for polypeptides. In addition, each carbon in the monosaccharide ring structure is numbered. For example, in a linear Fischer projection of galactose (fig. 1.1 A) the carbonyl group is placed at the top of the chain and C1 is the first carbon at the top of the chain.
The anomericity of a monosaccharide and the carbon number are key factors for assigning glycosidic linkages. The convention for writing glycosidic linkages is based on the two carbons engaged in the linkage; the anomericity is written first followed by the carbon number of the non-reducing end, then a comma followed by the carbon number of the preceding monosaccharide. For example, figure 1.1 C shows galactose with a β(1,4)-glycosidic linkage with N-acetylglucosamine, Galactose-β(1,4)-N-acetylglucosamine. Galactose (Gal) is at the non-reducing terminus and in the β anomeric configuration, N-acetylglucosamine (GlcNAc) is at the reducing terminus and the two are linked via the C1 carbon of Gal to the C4 carbon of GlcNAc.

1.1.1 N-linked glycan biosynthesis: The early processing steps are conserved in all eukaryotes

As far back as the 1950s it was identified that asparagine was the N-linked glycan attachment site on glycosylated proteins (Cunningham *et al*., 1957 and, Johansen *et al*, 1958). In the 1960s, N-acetylglucosamine was identified as the first monosaccharide linking the glycan to asparagine (Johansen *et al*., 1961 and, Nuenke and Cunningham, 1961). The glycan is pre-synthesised on a lipid anchor before *en bloc* transfer onto asparagine (fig. 1.2) (Chapman *et al*., 1979 and, Lehle, 1980). The enzymes involved in glycan biosynthesis were discovered from studies of *Saccharomyces cerevisiae ALG* (altered in glycosylation) genes (Huffaker and Robbins, 1982, Snider *et al*., 1982, Runge *et al*., 1984, Runge and Robbins, 1986).

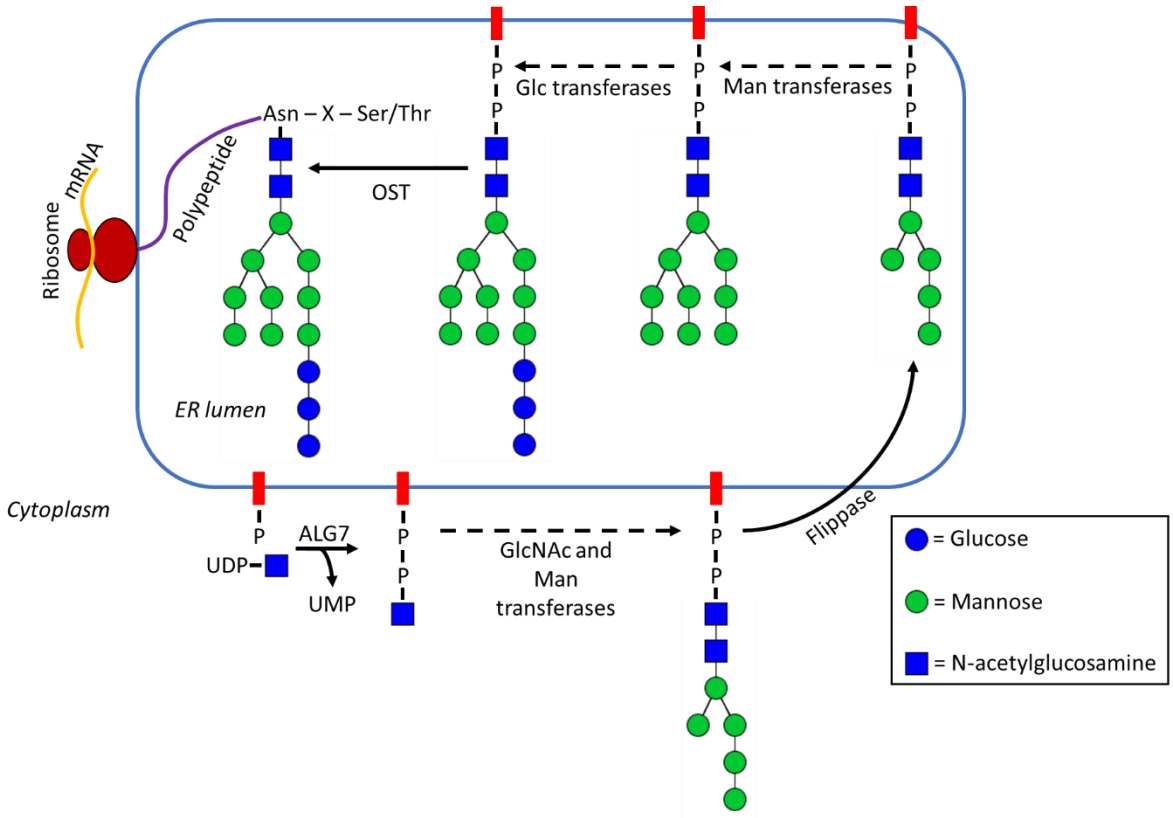


Figure 1.2: Synthesis of the precursor glycan on dolichol phosphate (red square) and transfer onto asparagine in the sequence Asn-X-Ser/Thr (where X is any amino acid except proline) of the nascent polypeptide chain in the endoplasmic reticulum. GlcNAc is N-acetylglucosamine, Man is mannose and Glc is glucose.

The substrates for glycosyltransferase enzymes are nucleotide sugars and these are synthesised in the nucleus or cytoplasm. The negative charge of nucleotide sugars prevents passive diffusion into the ER and golgi, where they are required for the bulk of glycan biosynthesis. Therefore, as well as nucleotide synthetases, glycosylation is reliant upon nucleotide sugar transporters. These are transmembrane antiporters that couple import of nucleotide sugars into the ER and golgi against export of nucleoside monophosphate products of glycosyltransferase (Caffaro and Hirshberg, 2006).

In all eukaryotes, N-linked glycans are initially biosynthesised on a lipid-like membrane anchor, called dolichol phosphate (Dol-P), on the cytoplasmic face of the endoplasmic reticulum (ER) (fig. 1.2). Glycan biosynthesis is initiated by the transfer of phosphorylated GlcNAc from uridinyl-phosphate-GlcNAc to Dol-P to form dolichol pyrophosphate GlcNAc (Dol-P-P-GlcNAc). A second GlcNAc is covalently attached via a β(1,4)-glycosidic bond to the GlcNAc of Dol-P-P-GlcNAc to form the chitobiose disaccharide core common to all N-linked glycans (fig. 1.3). The chitobiose core is further extended by the sequential addition of three mannose (Man) residues to form the trimannosyl core, $Man_3GlcNAc_2$-P-P-Dol. The trimannosyl-core is extended further by the covalent addition of six Man residues and

three glucose (Glc) residues, to form $Glc_3Man_9GlcNAc_2$-P-P-Dol, and during this process the dolichol linked precursor glycan is flipped into the ER lumen.

The multi-subunit protein complex oligosaccharyltransferase (OST) catalyses the cleavage of the GlcNAc-P bond and transfer of the precursor glycan *en bloc* to the γ-amido group of an asparagine residue in the glycosylation consensus sequence Asn-X-Ser/Thr, where X is any amino acid except proline, of a nascent polypeptide chain as it is being translated into the ER lumen. Trimming, re-addition followed by re-trimming of the three terminal Glc residues by α-glucosidases is an important quality-control check point for monitoring correct folding of glycoproteins in the ER. The glycan is trimmed further in the ER by α-mannosidase I (ManI) to form an eight-Man glycan, $Man_8GlcNAc_2$.

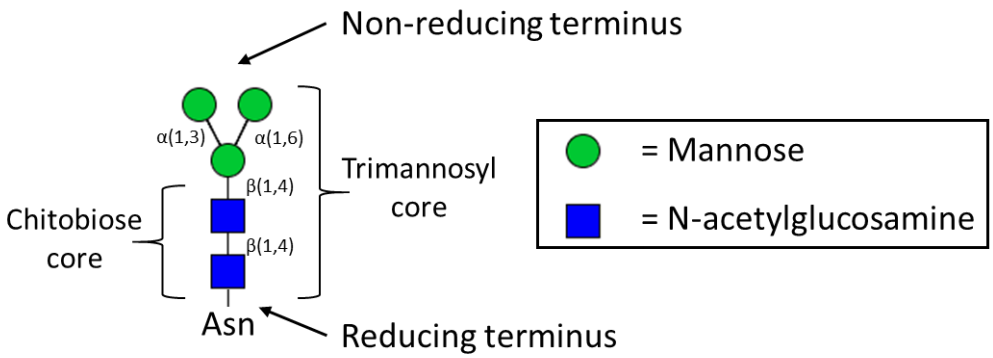


Figure 1.3: Core glycan structure common to all N-linked glycans. The chitobiose core is composed of two covalently bonded GlcNAc connected by a β(1,4)-glycosidic linkage. Extension of the chitobiose core by three mannoses create the trimannosyl core. Myriad glycosyltransferases catalyse the incorporation and extension of the trimannosyl core to create glycans of enormous diversity in monosaccharide and glycosidic-linkage composition.

1.1.2 N-linked glycan biosynthesis: The late processing steps are responsible for N-linked glycan diversity

In all eukaryotes, the initial stages of glycosylation in the ER are conserved. The glycosylation pathways of different eukaryotes diverge upon entering the Golgi apparatus. In mammals, the $Man_8GlcNAc_2$ glycan that leaves the ER has three Man residues removed, by a golgi resident ManI, to form $Man_5GlcNAc_2$. This is the key intermediate for the production of hybrid- and complex-type glycans (fig. 1.4).

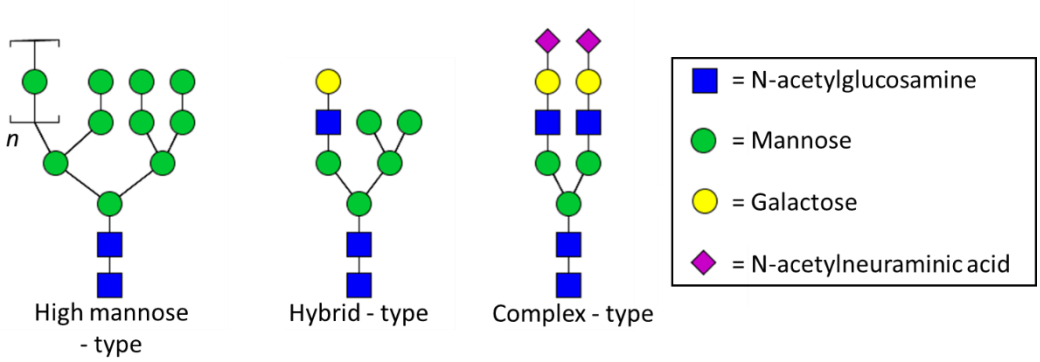


Figure 1.4: N-linked glycans are composed of three major types: high mannose-, hybrid- and complex-type. The high mannose-type glycan can be written in text, from non-reducing to reducing terminus, as $Man_nGlcNAc_2$, hybrid-type as $Gal_1GlcNAc_1Man_5GlcNAc_2$ and the complex-type glycan as $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$

A medial-golgi resident enzyme, GlcNAc transferase I (GnTI) catalyses the transfer of GlcNAc onto the α(1,3)-Man (fig. 1.3) of the trimannosyl core of $Man_5GlcNAc_2$. The resulting $GlcNAc_1Man_5GlcNAc_2$ is the precursor for hybrid-type glycans and is converted to a complex-type glycan by removal of two Man residues to expose the trimannosyl core α(1,6)-Man (fig. 1.3), by a trans-golgi resident ManII, to form $GlcNAc_1Man_3GlcNAc_2$. GnTII then catalyses the transfer of GlcNAc onto the exposed trimannosyl core α(1,6)-Man to form the complex-type glycan, $GlcNAc_2Man_3GlcNAc_2$.

In the trans-golgi, galactosylation and sialylation by galactosyltransferases (GalT) and sialyltransferases (SiaT), respectively, further extend the antennae to form hybrid-type and diantennary complex-type terminally sialylated glycans, $Sia_1Gal_1GlcNAc_1Man_5GlcNAc_2$ and $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$, respectively (fig. 1.4).

In mammals, the most common N-linked glycan is complex-type. As well as the diantennary glycan described above, additional glycan-processing enzymes can facilitate the addition of one or two more antennae to form tri- or tetra-antennary glycans, and fucosyltransferases can add α(1,6)-fucose to the asparagine linked GlcNAc in the glycan chitobiose core (fig. 1.4).

In yeast, however, N-linked glycans are mannosylated. A golgi-resident mannosyltransferase, encoded by the yeast gene *OCH1*, catalyses the addition of a single Man to the $Man_8GlcNAc_2$ glycan that leaves the ER. The $Man_9GlcNAc_2$ product is the substrate for further mannosylation, by myriad golgi resident mannosyltransferases, yielding, in *Pichia pastoris* (*Komagataella pastoris*), high-mannose glycans containing 9-15 mannose and, in *S.cerevisiae*, hypermannosylated glycans of up to 50 mannose residues (fig. 1.5).

In insects, processing of N-linked glycans removes all but the trimannosyl core mannose to yield paucimannose glycans which can be either α(1,3)- or α(1,6)-fucosylated (fig. 1.4). As well as producing high-mannose and paucimannose glycans, plants can also produce galactose terminal hybrid- and complex-type glycans with α(1,3)-fucosylation and the incorporation of xylose (fig. 1.5).

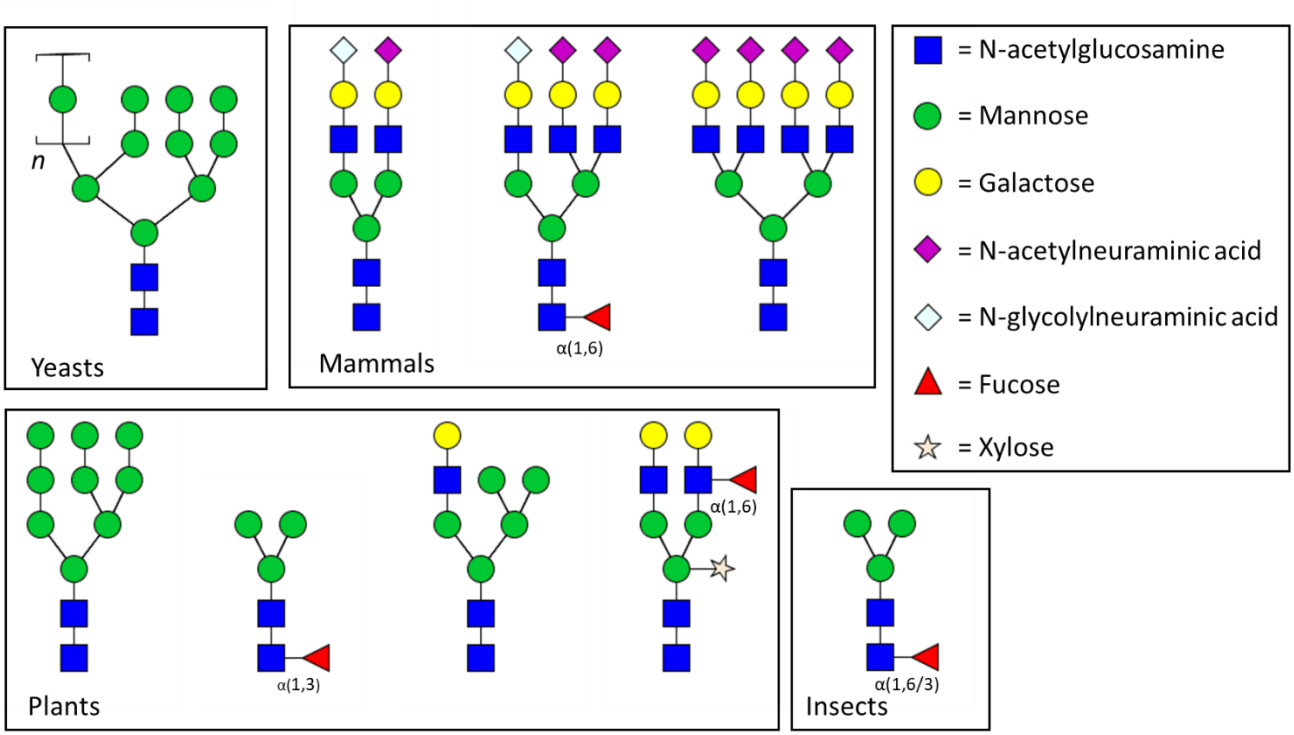


Figure 1.5: Glycan diversity in different groups of eukaryotes. The trimannosyl core is common to all N-linked glycans and extension of this basic structure leads to the wealth of diversity in N-linked glycan structures. *n* any number from 0 - 50.

1.1.3 Glycosylation has intrinsic and extrinsic functions

The function of glycosylation is twofold. Firstly, glycans confer properties on protein conjugates that are irrespective of the type of glycan, these are termed intrinsic functions. As already mentioned, in the ER addition of glucose to, and removal of glucose from, the nascent glycoprotein is key to the calnexin/calreticulin protein-folding quality-control checkpoint (Hebert *et al*., 1996). Other intrinsic effects of glycosylation include enhanced protein solubility, thermal stability, avoidance of aggregation, avoidance of cross-linking and resistance to proteolysis (Oh-eda *et al*., 1990, Delorme *et al*., 1992, Wang *et al*., 1996, Duarte-Vázquez *et al*., 2003).

In a study comparing five glycoproteins from a range of organisms, from yeast to mammals, with either N- or O-linked glycosylation, the total glycan content rather than the specific type of glycosylation conferred thermal stability and resistance to aggregation (Wang *et al*., 1996). Therefore, it is likely that these intrinsic effects, especially enhanced protein stability and improved survival were the main evolutionary drive for glycosylating proteins, particularly those that occupy the harsh vacuole or extracellular environments.

In addition to the intrinsic effects, glycosylation serves extrinsic effects. In mammals, the many processing enzymes and the non-template driven, spatio-temporal nature of glycan biosynthesis results in substantial glycan heterogeneity and complexity. The fact that this complexity is evolutionarily maintained, and that glycans are predominantly present on the surface of cells and in the extracellular environment, points to key functions for glycans. These are now thought to include information encoding and signalling. This idea is supported by the fact that several congenital disorders of N-linked glycosylation exist (Jaekan, 2011). These are caused not by a lack of glycosylation, but by an inability to process complex-type glycans. Indeed, complex-type glycan structures are important for lymphocyte-endothelial attachment and rolling (Imai *et al*., 1991 and, Somers *et al*., 2000), cell-cell interaction and signalling (Crocker, 2002), endocytosis and immunity (Ashwell and Harford, 1982, Schwartz, 1984, Fukuda *et al*., 1989 and, Taylor and Drickamer, 1993).

It is this immunity aspect of the extrinsic glycan function, especially accelerated clearance of therapeutic glycoproteins bearing aberrant glycans, together with the intrinsic functions, that is most relevant to the biopharmaceutical industry. The intrinsic effects of glycans allow for longer storage of proteins and at higher concentrations than non-glycosylated proteins. However, therapeutic glycoproteins bearing aberrant glycans are cleared rapidly from circulation. Therefore, glycoproteins present a unique challenge to the biopharmaceutical industry. Simply removing the glycans from therapeutically useful glycoproteins is not effective at enhancing efficacy. For example, studies on erythropoietin (EPO) which has three N-linked glycosylation sites, show that the *in vitro* activity is increased upon deglycosylation but the *in vivo* activity is decreased (Delorme *et al*., 1992). This reduced *in vivo* activity is likely due to accelerated clearance of the deglycosylated protein.

An additional challenge to the biotherapeutic industry is recombinant production of therapeutically useful glycoproteins in expression systems. As was seen in the case of EPO, proteins lacking human-like terminally sialylated glycans are generally cleared rapidly. There are two primary routes of glycoprotein clearance in humans, both of which target aberrant, non-sialylated glycans. First, the asialoglycoprotein receptor, which is located on the extracellular face of hepatocytes, binds to terminally galactosylated glycans on glycoproteins, i.e. glycans that have lost sialic acid, and facilitates their endocytosis (Ashwell and Harford, 1982, Schwartz, 1984 and, Fukuda *et al*., 1989).
The second mechanism of clearance is *via* recognition by mannose-binding lectins. These are either membrane-associated to macrophage and hepatic endothelial cells and facilitate endocytosis (Taylor and Drickamer, 1993) or, in the case of collectins and ficolins, free in sera where they activate complement in response to targeted aberrant glycosylation (Ikeda *et al*., 1987). In both instances, the mannose-binding lectin binds to glycoconjugates with terminal GlcNAc and Man residues. This serves as an important innate immune response against pathogenic organisms, such as the parasitic yeast *Candida albicans*, bearing abnormal terminal monosaccharides (Ezekowitz *et al*., 1990).

1.1.4 Mammalian glycosylation and sialic acids

An important aspect of mammalian glycosylation is the capping of complex-type glycans with sialic acid. Sialic acids are a particularly diverse family of monosaccharides, comprising a common backbone with various functional groups and the ability to form multiple different glycosidic linkages. Sialic acids warrant a more in-depth discussion because of their physiological importance and the pharmacokinetic benefits that sialic acids bestow upon biotherapeutic proteins.

First discovered on salivary mucins (Greek *Sialos*), sialic acids are a group of nine-carbon backbone monosaccharides (fig. 1.6) predominantly found at the terminus of N- and O-glycans, glycosphingolipids and some GPI-anchors, on proteins and cell surfaces.
The sialic acids are diverse. As well as having α/β anomericity, the fifth carbon (C5) group can be N-acetylated (N-acetylneuraminic acid, Neu5Ac) or hydroxylated (2-keto-3-deoxynononic acid, Kdn). The C5 N-acetyl group can also be hydroxylated to yield N-glycolylneuraminic acid (Neu5Gc), a monosaccharide that humans have lost the ability to

biosynthesise. Additionally, the C5 amino group may not be acetylated, which gives neuraminic acid (Neu).

Based on these four core sialic acids further diversification occurs. Each of the backbone carbon hydroxyl-groups and the C1 carboxyl group may be modified, while there are multiple possibilities for glycosidic linkages. Nonetheless, the most common sialic acid is Neu5Ac. The most common linkage formed by this sialic acid involves its α-2 position in linkage with the C3 or C6 position of galactose (Varki *et al*., 2017).

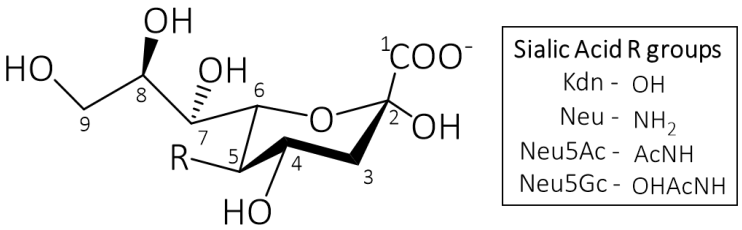


Figure 1.6: Haworth projection of the molecular architecture common to all sialic acids. This structure is depicted in α-anomeric configuration, where the C2 hydroxyl group sites below (axial) the plane of the heterocyclic pyranose ring structure. In the β-anomer the C2 hydroxyl group is equatorial. The R groups listed form the four core sialic acids of which all other sialic acids are derivatives.

High levels of sialic acids are present on glycoproteins, in the glycocalyx and on matrix-associated proteins such as mucins. The concentration of sialic acids within the glycocalyx of a single B-cell has been estimated to be over 100 mM (Collins *et al*., 2004).
The negative charge on sialic acids sets this monosaccharide apart from the other, neutral, monosaccharides which make up complex-type N-linked glycans. This negative charge is an important contributor to the biophysical landscape of biological systems, for example the negative charge on the surface of human erythrocytes prevents unwanted cell-cell interactions in blood circulation and sialylation enhances mucin viscosity (Raju and Davidson, 1994).
Thus, the density of sialic acids has a significant biophysical effect on the molecules and cells to which the sialic acids are conjugated and is an important factor in the circulatory life-time of glycoproteins (Fukuda *et al*., 1989 and, Yang *et al*., 1992), immunogenicity (Fenouillet *et al*., 1986), receptor interaction (Sumer-Bayraktar *et al*., 2011) and, stabilisation and solubility (Dissing-Olesen *et al*., 2008).

1.2 Complement

1.2.1 Three complement activation pathways converge on the formation of C3b

Complement is one of the oldest evolutionary components of the immune system, bridging the gap between innate and adaptive immunity (Dunkelberger and Song, 2010). Complement is, therefore, one of the first lines of defence against pathogens.

Composed of over 40 proteins, present in the blood or on membrane surfaces, activation of complement causes a sequential enzymatic activation cascade that leads to the generation of proinflammatory anaphylatoxins, opsonisation of surfaces (i.e. tagging them for phagocytosis) and cell lysis by assembly of the pore-forming membrane-attack complex (Dunkelberger and Song, 2010).

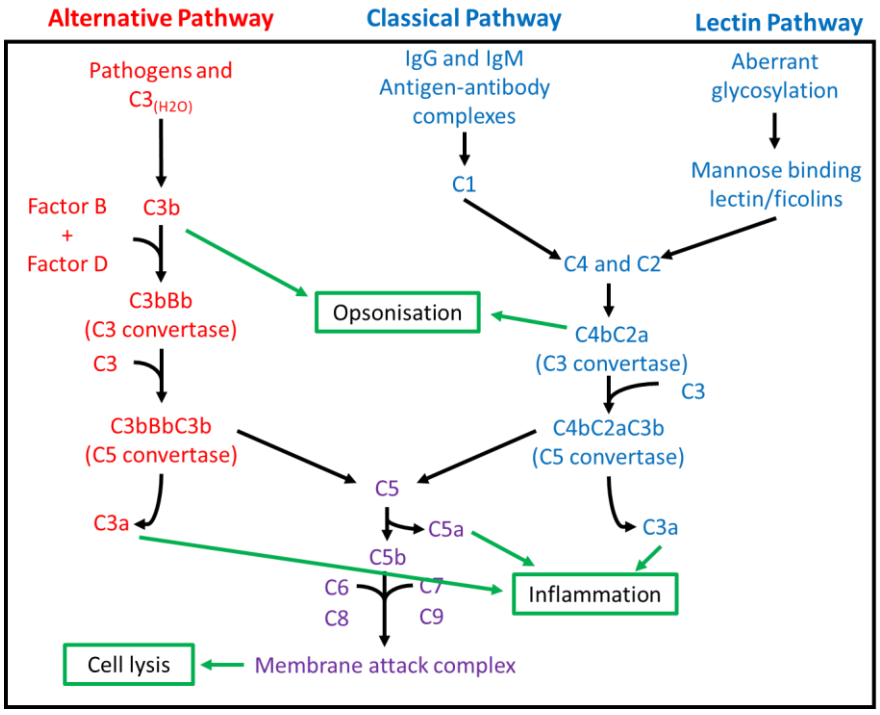


Figure 1.7: Overview of the complement process. The three complement-activation pathways converge on the formation of the opsonin C3b which, when in complex with Bb, forms a powerful self-amplification loop to generate more molecules of C3b. Opsonisation of surfaces by C3b and downstream complement cascades facilitated by the formation of C3 convertase complexes, C3bBb and C4bC2a, lead to phagocytosis, inflammation and, eventually, cell lysis. Complement pathways of activation are coloured red (alternative pathway) and blue (classical and lectin pathways), the downstream complement pathway is coloured purple and the downstream effects of complement activation are boxed in green.

There are three pathways of activation of complement (fig. 1.7). All three pathways involve formation of a complex that cleaves the key complement protein C3 to C3b, as described

below. The classical pathway (CP) and the lectin pathway (LP), are activated by pathogen- or danger-recognition molecules.

The CP is activated in response to immunoglobulin (Ig) G and IgM binding to pathogens or non-self antigens. In this process, the multimeric complement C1 complex binds, *via* C1 component C1q, to the Fc portion of IgG clusters or IgM bound to antigen, which causes a conformational change and activation of the serine protease C1r. C1r cleaves and activates another serine protease component of C1, C1s. Activated C1s catalyses the hydrolysis of the fluid-phase protein molecules C4 and C2 to form the complex C4bC2a. This complex is called the CP and LP C3-convertase as it cleaves the abundant fluid-phase protein C3 to form C3a, an anaphylatoxin, and C3b.

The LP leads to formation of the same C3-convertase i.e. C4bC2a. It does so in response to glycosylation patterns present on yeast, bacteria, parasites and viruses. These glycosylation patterns are recognised by the mannose-binding lectin and ficolins which then trigger mannose-binding lectin-associated protein (MASP) 2 (Wallis *et al*., 2007) to cleave C4 and C2, resulting, as in the CP, in C3-covertase (C4b2a) complex formation and thus production of C3a and C3b.

The third pathway, the alternative pathway (AP), differs from the lectin and classical pathway in that it does not require pattern recognition molecules for activation. In a process known as “tick over” C3 is randomly hydrolysed to form $C3_{(H2O)}$ (Isenman *et al*., 1981).

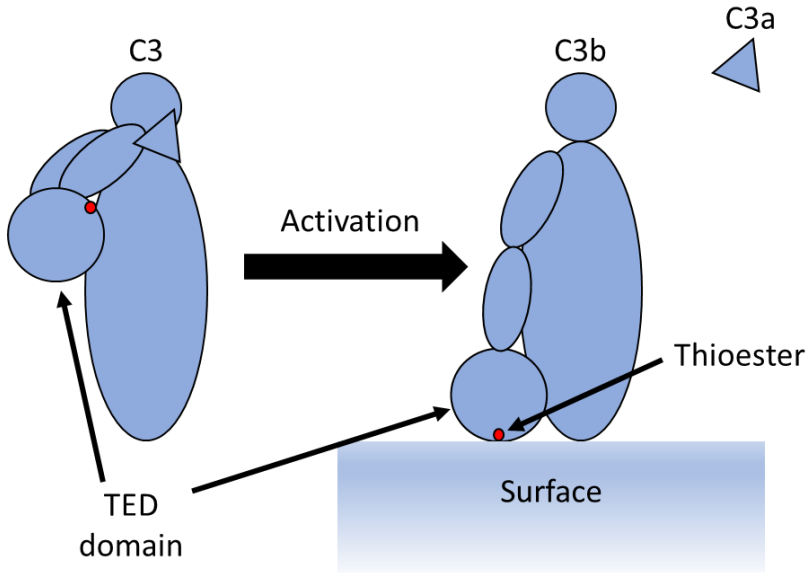


Figure 1.8: Schematic diagram of the conversion of C3 to C3b. Proteolytic cleavage of C3a from C3b results in a repositioning of the TED domain of C3b, thus exposing the previously buried reactive thioester for attachment of C3b to surfaces.

$C3_{(H2O)}$ undergoes a structural rearrangement (Chen *et al*., 2016) whereby the TED domain rotates and migrates to reveal a binding site for factor B. In this context factor B is a substrate for factor D, which cleaves away the Ba fragment leaving C3bBb. The remaining Bb fragment has protease activity for C3 and this complex ($C3_{(H2O)}$Bb) is called the initiating C3 convertase. It generates C3b (177 kDa) and the anaphylatoxin, C3a. C3b has a similar conformation and shape to $C3_{(H2O)}$ but formation of C3b exposes a previously buried reactive thioester (fig. 1.8) (Janssen *et al*., 2006). Nucleophilic attack of the newly exposed thioester by surface associated hydroxyl groups allows attachment of C3b, to these surfaces (Sahu, *et al*., 1994 and, Nagar *et al*., 1998). C3b, like $C3_{(H2O)}$ is a platform for factor B cleavage and the product, C3bBb is called the AP C3 convertase. This stokes an amplification process in which, unless regulated, a positive-feedback loop produces a large number of C3b molecules very rapidly. These, produced in their millions, can rapidly coat any surfaces that have nucleophilic groups, tagging these for phagocytosis.

Thus, all three pathways converge on formation of C3b. Regulation of C3b production, and in particular, the positive feedback loop that generates large amounts of C3b, is critical on self surfaces.

1.2.2 Factor H deactivates complement in fluid phase and on self-surfaces

The AP is always "on" (*i.e.* in tick over mode) and C3b, that can participate in a positive feedback loop, binds to nearly any surface - including self surfaces. Clearly there is a constant danger of C3b overproduction and collateral damage to host tissue. To ensure this process is kept under tight control, multiple membrane-associated and fluid-phase regulators of complement exist to maintain the homeostatic balance between fast, efficient pathogen destruction and the prevention of autologous complement-mediated attack.

The soluble glycoprotein factor H (FH) is a key regulator of complement activation. FH is encoded by the *CFH1* gene located in the Regulator of Complement Activation gene cluster on the long arm of chromosome 1, 1q32 (Weis, *et al*., 1987 and, Rodríquez De Córdoba, *et al*., 1999). FH is a large 1,213-amino acid, 155-kDa glycoprotein composed of a single polypeptide chain folded into 20 complement control protein (CCP) domains arranged like "beads on a string" (Ripoche *et al*., 1988 and, Vik *et al*., 1988). The structure of FH CCPs has

been extensively characterised (Perkins *et al*., 1988, Barlow *et al*., 1991, Barlow *et al*., 1992, Barlow *et al*., 1993, Makou *et al*., 2012 and, Morgan *et al*., 2012), each CCP forms a globular structure containing two pairs of disulphide bonds, thus full-length FH contains 40 disulphide bonds.

FH has eight sites of N-linked glycosylation; CCPs 9, 12, 13, 14, 17 and 18 contain one glycosylation site each and the two remaining sites are both located on CCP 15 (fig. 1.9). Fenaille *et al*. (2007*a*) have shown that the glycans on human FH are exclusively complex-type, and the majority of them feature identical diantennary disialylated structures, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$, collectively accounting for 67.4% of total glycan content. Monosialylated diantennary (9.2%), $Sia_1Gal_2GlcNAc_2Man_3GlcNAc_2$, and trisialylated triantennary (3.9%), $Sia_3Gal_3GlcNAc_3Man_3GlcNAc_2$, structures are also present. These three structures are additionally present in a fucosylated form (8.9%, 2.0% and 4%, respectively). However, 4.6% of the glycan content of FH was uncharacterised.

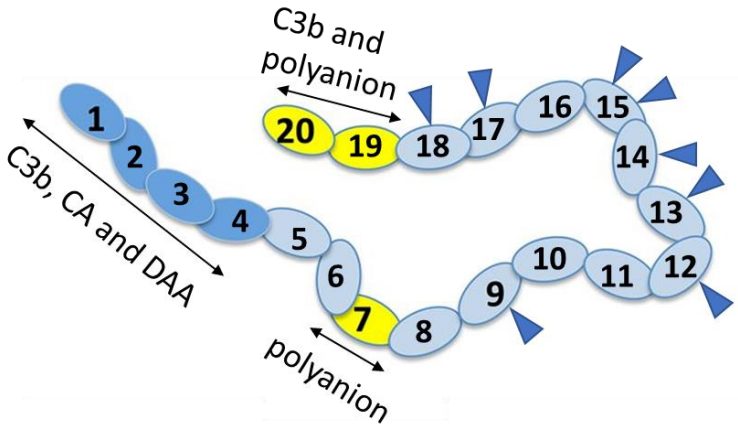


Figure 1.9: Schematic of FH. CCPs are labelled 1 – 20 from N-terminus to C-terminus. CCPs 1-4 (dark blue) are responsible for C3b binding, cofactor activity (CA) and decay acceleration activity (DAA). CCPs 19-20 are also capable of binding to C3b and, in addition, polyanions. Polyanion binding CCPs (coloured yellow) are responsible for discriminating self from non-self surfaces. Glycosylation sites are marked by dark blue triangles. The image was taken and adapted with permission from Herbert *et al*. (2015).

FH is expressed predominantly in the liver, but also locally in the eye (Skerka *et al*., 2007 and, Scholl *et al*., 2008), circulates in blood plasma at a concentration of 200 – 300 milligrams per litre (Hakobyan, *et al*., 2008 and, Sofat *et al*., 2013) and regulates complement at the level of C3b (Weiler *et al*., 1976) both in fluid-phase and on self-surfaces. FH regulates complement in three ways. First, FH competes with FB for binding to C3b, preventing the formation of the potent C3bBb, C3-convertase complex, thus attenuating C3b self-amplification. Second, FH can bind to the labile C3bBb complex and accelerate the rate of dissociation of this complex by a factor of about 50. This is FH's decay-acceleration activity (DAA). Thirdly, FH is a cofactor for the serine protease factor I

(FI) for cleavage of C3b; this is FH's cofactor activity (CA). Binding of FH to C3b enables the recruitment of FI which catalyses the first step in the degradation of C3b, conversion to iC3b. Continual degradation steps eventually leave just the TED domain of C3b (called C3d) on self-surfaces.

The N-terminal CCPs 1-4 of FH have been shown to bind to C3b (fig. 1.10) and are primarily responsible for CA and DAA of FH (Gordon *et al*., 1995, Kuhn and Zipfel, 1996, Hocking *et al*., 2008, Schmidt *et al*., 2008 and Pechtl, *et al*., 2011). The crystal structure of N-terminal CCPs 1-4 in complex with C3b shows that CCPs 1-4 bind C3b along the length of C3b at a site overlapping with the Bb interaction site on C3b and CCPs 1-2 directly occlude the Bb binding site on C3b (fig. 1.10 C). This suggests that DAA of FH is mediated exclusively by these CCPs via a combination of competition between, and electrostatic repulsion of, Bb in the C3 convertase complex (Wu *et al*., 2009). Crystal structures of C3b in complex with FI and "mini-FH", composed of CCPs 1-4 and CCPs 19-20 joined by a short linker, show that FI binds to FH across CCPs 2-3 (fig. 1.10 B) and re-orientates the C-terminal domain of C3b to allow FI access to the first, of a potential three, scissile bonds (Xue *et al*., 2017). Thus, CA of FH facilitates the docking of FI to C3b and results in the disruption of the FB binding site *via* cleavage of two to three scissile bonds in C3b.

The C-terminal CCPs 19-20 bind tightly to C3b and C3d but do not participate directly in CA and DAA (Ferreira *et al*., 2006, Schmidt *et al*., 2008, Ferreira *et al*., 2009 and Lehtinen *et al*., 2009). CCPs 19-20 have been shown to interact with C3b in the TED domain (fig. 1.10 A) and this binding site is located near to, but distinct from, the interaction site in the TED domain of the N-terminal CCP 4 (Morgan *et al*., 2011 and, Kajander *et al*., 2011). Mutations in CCPs 19-20 do not affect FH-mediated complement regulation in the fluid phase, but these mutations hinder FH-mediated complement regulation on self-surfaces (Ferreira *et al*., 2006). Thus, key to the function of FH is the ability of the molecule to recognise self from non-self. This allows FH to attenuate complement activation on self-surfaces (not on foreign surfaces) and this is crucial for surfaces that do not already possess membrane-associated complement regulators.

Self-surfaces are decorated in negatively charged polyanions, e.g. glycosaminoglycans (GAGs) and sialic acids, and FH has been shown to interact with these structures via at least two sites located on CCP 7 (Blackmore *et al*., 1996 and Giannakis *et al*., 2003) and CCP 20

(Blackmore *et al*., 1998). Other putative binding sites have been reported (Pangburn *et al*., 1991; Ormsby *et al*., 2006), however, these may not be of physiological relevance (Schmidt *et al*., 2008). Polyanions may engage CCPs 7 and 19-20 by electrostatic interactions. Blaum *et al*., (2015), using a crystal structure of FH CCPs 19-20 in complex with C3d (TED domain of C3b) and an α(2,3)-Neu5Ac linked trisaccharide, showed that the carboxyl-group of Neu5Ac formed a salt bridge with the FH arginine 1215 side chain.

Although the complete structure of FH has not yet been characterised, piecing together collections of small angle X-ray scattering (SAXS), analytical ultracentrifugation and nuclear magnetic resonance data of stretches of CCP modules and crystal structures of CCPs 1-4 and 19-20 in complex with C3b indicted that FH appears to adopt a bent back structure when in complex with C3b (Aslam and Perkins, 2001 and, Morgan *et al*., 2011), enabling

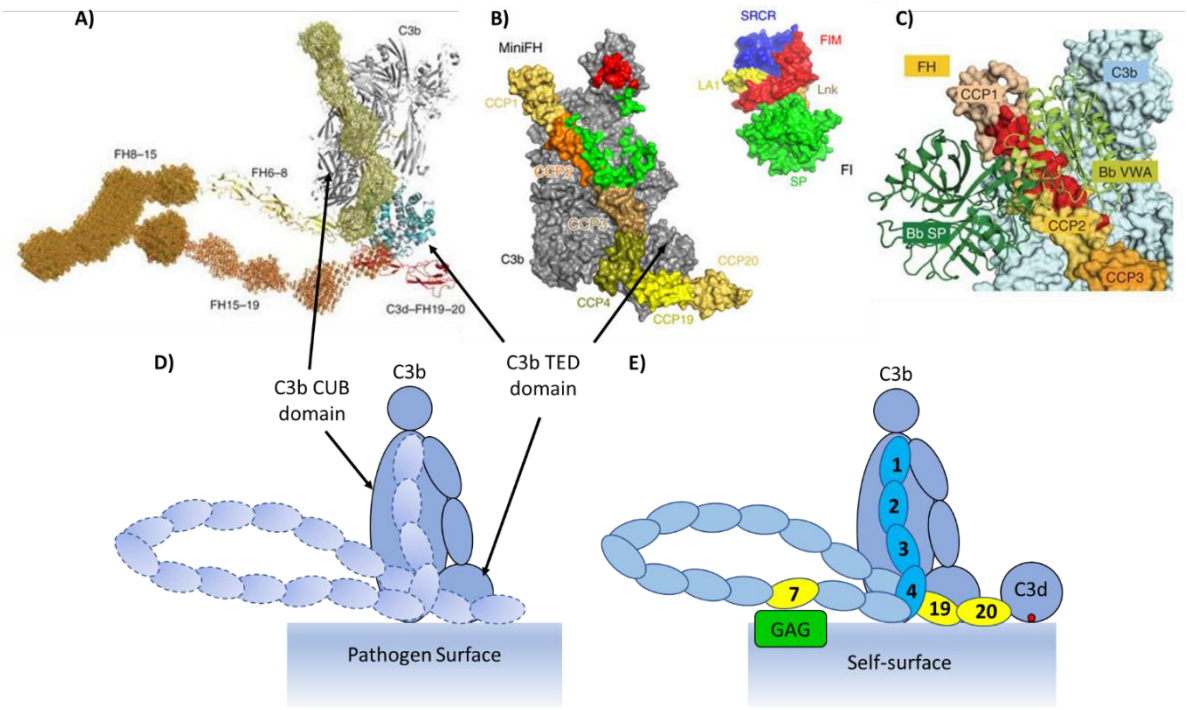


Figure 1.10: Potential model for the engagement of FH, FI and Bb with C3b. A) Superimposed X-ray crystallography-derived structures of C3b:FH1-4 and C3b:FH19-20, with SAXS envelope for CCPs 1-5 overlaid and juxtaposed with NMR derived structures of CCPs 6-8 and SAXS derived envelopes of CCPs 8-15 and 15-19. Image adapted with permission from Morgan *et al*., 2011. B) Contact regions of FI with C3b in complex with FI, figure taken and adapted with permission from Xue, *et al*. (2017). C) Overlay of the FH and Bb binding sites on C3b, the red region indicates where FH and Bb overlap, figure taken and adapted with permission from Wu *et al*. (2009). D) Schematic diagram of C3b on a pathogenic, non-self surface. FH does not bind to C3b on non-self surfaces that lack polyanions such as GAGs, the dotted outline represents where FH would bind to C3b. E) Schematic diagram of the interaction of FH with C3b on a self-surface. On self-surfaces polyanionic structures mark these surfaces as self and stabilise FH interaction with C3b enabling complement regulation. On self-surfaces already protected from complement activation by FH binding, FH may remain anchored to C3d via CCPs 19-20 and search out other, nearby molecules of C3b for deactivation by CA and DAA mediated via N-terminal CCPs 1-4.

simultaneous interaction of the N-terminal CCPs 1-4 and C-terminal CCPs 19-20 (fig. 1.10 A).

In addition, FH has been shown to have enhanced affinity for C3b and C3d and enhanced DAA in the presence of a bacterial FH-binding protein PspCN (Herbert, *et al*., 2015). Cross-linking mass spectrometry data show that PspCN cross-links with FH at CCPs 9-10 and that, in the absence of PspCN, FH preferentially adopts two conformations: in the first conformation CCPs 4 and 5-7 form cross-links with CCPs 19-20, however, in the second conformation these CCPs do not form cross-links and the cross-links that do form occur between immediately adjacent CCPs, e.g. CCP 3 with 4, 4 with 5, 5 with 6, etc (Herbert *et al*., 2015). This indicates that FH can have either a compact conformation or a more extended conformation. However, in the presence of PspCN, the conformational preference of FH is shifted toward the first, more compact conformation. This, together with the enhanced C3b and C3d affinity and DAA of FH in the presence of PspCN, indicates that the compact conformation of FH is more active. This led Herbert *et al*., (2015) to reason that circulating FH *in vivo* also adopts compact and elongated conformations and that the active compact conformation is stabilised in the presence of self-surfaces by simultaneous interaction of CCPs 1-4 and 19-20 with C3b and CCPs 7 and 20 with self-surface polyanionic markers (fig 1.11).

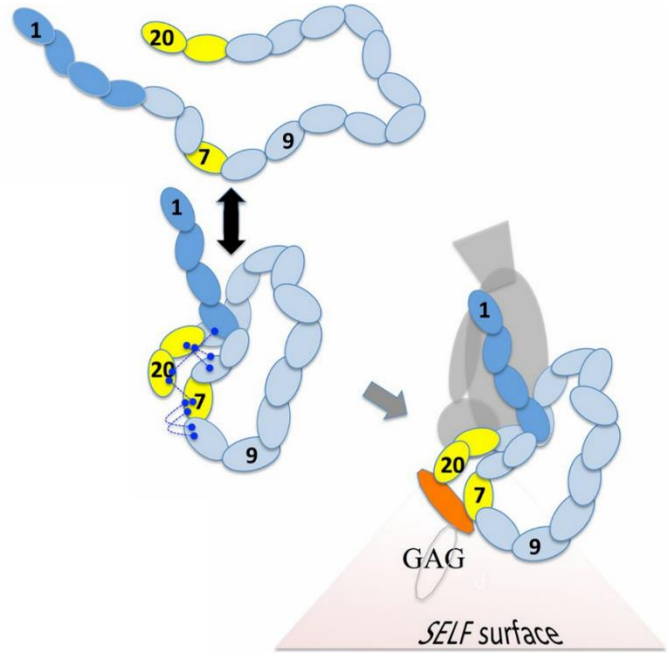


Figure 1.11: Schematic diagram showing the interconversion of FH from the latent, extended conformation (upper) to the compact, more active conformation (middle). Cross-linked sites in the compact structure of FH are marked by blue lines. CCPs 19-20 are shown to cross-link with CCPs 4 and 6-7 holding the two C3b binding domains in close proximity to one another. On self surfaces (bottom right) Herbert *et al*. (2017) reason that CCPs 7 and 20 interaction with self-surface polyanionic markers (labelled here as GAG) stabilise the compact, more active conformation of FH facilitating complement regulation on self-surfaces. The image was taken and adapted with permission from Herbert *et al*. (2017).

1.2.3 Diseases associated with factor H

Given the central role of FH in regulating complement activation, it is unsurprising that mutations and single nucleotide polymorphisms (SNPs) in FH lead to autoimmune diseases. Mutations that result in impaired ability to secrete FH into the circulation have been described in humans and animals (Ault et al., 1997, Sánchez-Corral *et al*., 2000, Hegasy et al., 2003). Most cases of the rare kidney disorder, membranoproliferative glomerulonephritis-type II (MPGNII) or "dense deposit disease" are caused by a lack of FH in circulation. This leads to uncontrolled fluid-phase complement activation resulting in depletion of plasma C3 and other complement components (Thompson and Winterborn, 1981 and, Ault et al., 1997) and deposition of C3b on the basement membrane of the glomerulus causing wall thickening and fibrosis (Licht *et al*., 2007). In support of a link between FH deficiency and MPGNII, mice deficient in *cfh* (*cfh*$^{-/-}$) have uncontrolled C3 activation and spontaneously develop MPGNII. Subsequent deletion of the mouse gene encoding FB abolishes the *cfh*$^{-/-}$ phenotype (Pickering *et al*., 2002). Development of MPGNII in early life has severe repercussions; children have a poor prognosis and often develop late-stage renal disease. The median survival time of children with MPGNII is less than nine years (Schwertz *et al*., 1996).

FH is also linked to another rare kidney disorder, the heritable disease atypical haemolytic uraemic syndrome (aHUS). aHUS is characterised by haemolytic anaemia, thrombocytopaenia and acute renal failure (Ferreira *et al*., 2006). The recessive form of the disease is strongly associated with mutations clustering in the C-terminal region of FH, particularly in CCP 20 (Rodríguez De Córdoba *et al*., 2004). Functional analysis of common mutations associated with aHUS in the C-terminal domains of FH show an impaired ability to bind to C3b and polyanionic self-markers (Pangburn, 2002, Ferreira *et al*., 2009 and, Lehtinen *et al*., 2009). In addition, transgenic mice modified to express a version of FH lacking the C-terminal five CCPs spontaneously developed aHUS but were still able to regulate complement in plasma (Pickering *et al*., 2007). This led to the premise that aHUS-causing FH mutants are able to regulate complement in fluid phase, preserving levels of C3 in plasma, but C3b amplification is poorly regulated on self-surfaces due to a reduced ability of the mutant FH to bind either or both of polyanions and self-surface associated C3b (or C3b fragments).

Age-related macular degeneration (AMD) is the leading cause of blindness in the developed world (Klein *et al*., 2002). Early-stage AMD predominantly affects rod-mediated dark adaptation (Owsley *et al*., 2015), but in its later stages AMD is characterised by a central scotoma and loss of vision (Toomey *et al*., 2017). A hallmark of AMD is the presence of numerous soft drusen, yellow deposits containing oxidised proteins and lipids as well as FH. The "dry" form of AMD features retinal pigment epithelial (RPE) atrophy (Toomey *et al*., 2017). The "wet" form is neovascular and vision loss is caused by the formation of abnormal blood vessels of choriocapillaries in the Bruch's membrane of the retina (Toomey *et al*., 2017).

FH has been strongly linked with the onset of AMD. Four separate studies found a tyrosine-to-histidine allelic variant at amino acid 402 that in homozygotes increases the risk of developing AMD over sevenfold. The Y402H variant is the cause of ~43% of cases in older adults with AMD (Edwards *et al*., 2005, Hageman *et al*., 2005, Haines *et al*., 2005 and, Klein *et al*., 2005). The Y402H polymorphism occurs in the polyanion-binding CCP 7 of FH. The Y402H variant was thought to have a reduced affinity for self-surface polyanions, reducing the efficiency of FH recruitment to the RPE and other layers in the Bruch's membrane. Studies on truncated recombinant fragments of FH, CCPs 6-8, support this hypothesis and show that the H402 variant has a reduced ability to bind to the polyanion heparin *in vitro* and *ex vivo* (Clark *et al*. 2006, Clark *et al*., 2010 and, Clark *et al*., 2013). However, these results have not been validated with full-length FH (Ormsby *et al*., 2008 and, Kelly *et al*., 2010).

1.2.4 Potential biotherapeutics for the treatment of factor H-linked disease

Molecular replacement strategies offer a possible route to treat diseases caused by a deficiency of active FH. In support of this, plasma infusions containing functional FH have been shown to be effective in restoring partial complement regulation in mice (Paixão-Cavalcante *et al*., 2009 and, Fakhouri *et al*., 2010), pigs (Høgåsen *et al*., 1995) and humans (Appel *et al*., 2005, Licht *et al*., 2005 and, Licht *et al*., 2006). However, the need for high circulating concentrations of FH, and its relatively short half-life of approximately six days (Licht *et al*., 2005), mean that regular plasma infusions would be required. In addition, the risk of infection associated with plasma transfusion means there is a strong case for using recombinant FH for therapeutic treatment.

"Mini-FH" molecules composed of the functional N- and C-terminal domains of full-length FH have shown therapeutic promise in *in vitro* and *ex vivo* assays (Hebecker *et al*., 2013, Schmidt *et al*., 2013 and, Nichols *et al*., 2015). However, of the three "mini-FH" molecules described to date, just one (Nichols *et al*., 2015) has been assessed *in vivo*. Nichols *et al*. (2015) describe the recombinant expression, in a Chinese hamster ovary (CHO) cell line, of a "mini-FH" with a molecular weight of 59 kDa and consisting of the N-terminal CCPs 1-5 and C-terminal CCPs 18-20 of FH joined together by a short, twelve glycine polypeptide linker. This protein was thus designed to be significantly smaller than FH (and easier to produce) but to retain its key regulatory and C3b-binding domains. "Mini-FH" had comparable CA and DAA compared to full-length FH, *in vitro* and *ex vivo*, however, *in vivo* "mini-FH" had a reduced effectiveness at restoring complement regulation in *cfh*$^{-/-}$ mice and it was completely cleared from circulation in just six hours. Conversely, in "mini-FH"-treated animals, C3 deposition in the glomerulus was reduced, suggestive of protection from complement on self-surfaces, however, the effect of self-surface protection by "mini-FH" in comparison with full-length FH was not tested over longer periods
The short half-life of "mini-FH" and the inability to regulate complement in fluid phase prevents use of "mini-FH" as an effective replacement therapy for diseases caused by FH deficiency.

Alternative approaches to the development of FH-replacement therapies have focused on recombinant production of full-length FH (Sharma and Pangburn, 1994, Sánchez-Corral *et al*., 2002, Schmidt *et al*., 2011, Büttner-Mainik *et al*., 2011 and, Michelfelder *et al*., 2017). However, with protein yields of 20 micrograms per gram of dry-cell weight in the moss *Physcomitrella patens* (Büttner-Mainik *et al*., 2011), 1 milligram per litre of *P. patens* cell culture (Michelfelder *et al*., 2017), 2 milligrams per litre of COS7 cell culture (Sánchez-Corral *et al*., 2002) and 5 milligrams per litre of Sf9 cell culture (Sharma and Pangburn, 1994) only Schmidt *et al*. (2011) report yields of recombinant FH (>10 milligrams per litre *P. pastoris* cell culture) high enough to have potential for production on the scale required for therapeutics.

Importantly, the glycosylation profile of mammalian cell- derived recombinant FH has not been characterised (Sánchez-Corral *et al*., 2002). Other cells used for recombinant FH production cannot glycosylate with mammalian complex-type glycosylation (Sharma and

Pangburn, 1994, Sánchez-Corral *et al*., 2002, Schmidt *et al*., 2011, and, Büttner-Mainik *et al*., 2011).

Therefore, there remains a pressing need to produce therapeutically useful quantities of recombinant FH with human glycosylation. This calls for a comparison of the expression systems available for production of N-linked glycoproteins, with a specific focus on FH and the need for high yields.

1.3 Glycoprotein-production systems

Agalsidase alpha and beta, epoetin alpha and beta, follitropin alpha and beta, lutropin alpha, transforming growth factor-β1, antithrombin, thyrotropin alpha, lenograstim, sargramostim, interleukin-3, prourokinase, lymphotoxin, coagulation factor VIIa, VIII and IX are just some of the therapeutic proteins whose stability and efficacy have been improved by glycosylation (Solá and Griebenow, 2010).

Below, various glycoprotein-production systems are explored with an emphasis on the type of glycan produced and consideration of any additional glycoengineering requirements in the context of human recombinant factor H production.

1.3.1 Bacteria, Plants, and Insects

Escherichia coli offers several advantages for heterologous expression of recombinant proteins. It is cheap, grows readily on a variety of media and lends itself to genetic manipulate. Indeed, the first US Food and Drug Administration (FDA) approved recombinant biotherapeutic, human insulin, was recombinantly produced in *E. coli*. But, unlike FH, insulin is not a glycoprotein. The lack of N-linked glycosylation machinery in *E. coli* and poor disulphide bond formation are major constraints for the use of *E. coli* as an expression system for recombinant glycoproteins.

The discovery of a bacterial homologue of a subunit of OST, called pglB, in *Campylobacter jejuni* (Szymanski *et al*., 1999, Szymanski *et al*., 2002 and, Nita-Lazar *et al*., 2005) and the successful recombinant expression of this protein in *E. coli* (Feldman *et al*., 2005) raised hopes for an approach to humanise glycoprotein production based on glycoengineered *E.*

coli. To date, however, efforts to produce fully humanised glycosylation (Valderrama-Rincon *et al*., 2012) in *E. coli*, and to improve glycosylation site efficiency, (Ding *et al*., 2017) have met with limited success. In addition, although disulphide bond-forming strains of *E. coli* have been developed, such as SHuffle (Lobstein *et al*., 2012) and Origami, the inefficiency of disulphide bond formation in these strains coupled with the high disulphide bond requirement of FH, which contains forty disulphide bonds, severely limits the use of *E. coli* for production of recombinant human FH with humanised glycosylation.

Advances have been achieved in glycoengineering of insect cells to produce humanised glycosylation (Jarvis and Finn, 1996, Hollister *et al*., 1998) with terminal sialylation (Jarvis *et al*., 2001) and exclusion of the highly immunogenic α(1,3)-fucose from glycans (Mabashi-Asazuma *et al*., 2013). However, insect cells are slow growing and often produce low protein yields, which limits the applicability of this system for production of therapeutically useful quantities of FH, especially considering the high concentration of FH in blood plasma.

Like insect cells, plant cells offer a non-microbial recombinant glycoprotein production option. The moss *Physcomitrella patens* is attractive in this respect as it is well characterised, it can be cultured in simple media and is fully differentiated and, therefore, genetically stable (Decker *et al*., 2014). Plants can glycosylate proteins with mammalian-like diantennary terminally digalactosylated complex-type glycans, $Gal_2GlcNAc_2Man_3GlcNAc_2$, but the galactose forms a β(1,3)-glycosidic linkage with GlcNAc, which differs from the human Gal-β(1,4)-GlcNAc glycosidic linkage. Like insects, plants incorporate the immunogenic α(1,3)-fucose as well as the immunogenic β(1,2)-xylose monosaccharide and efforts have focused on removal of these antigenic epitopes in glycoengineered *Arabidopsis thaliana* (Strasser *et al*., 2004) and moss (Koprivova *et al*., 2004).

In support of the potential of glycoengineered moss strains for biotherapeutic glycoprotein production Michelfelder *et al*. (2017) successfully expressed full-length human FH in *P. patens* glycoengineered to supress fucosyl- and xylosylation and to produce and attach to proteins the diantennary GlcNAc-terminal glycan, $GlcNAc_2Man_3GlcNAc_2$. However, a multi-step purification procedure, lack of protein purity, relatively low yields (1 milligram per litre cell culture), and the likelihood of accelerated clearance from blood circulation due to the

presence of GlcNAc terminal glycans, limits the biotherapeutic applicability of this source of recombinant human FH.

1.3.2 Mammalian cells

Mammalian expression systems are the dominant method for production of recombinant therapeutic glycoproteins. Of these CHO cells are the most frequently utilised. For example, in 2015 and 2016 over half of the FDA-approved new biologics were CHO cell-derived recombinant proteins (Lalonde and Durocher, 2017).

The major advantage of using mammalian expression systems for recombinant glycoprotein production is the commonality in mammalian and human glycosylation – both are capable of producing complex-type, terminally sialylated glycans. However, mammalian systems produce the immunogenic glycan epitopes Neu5Gc and galactose-α(1,3)-galactose (Bosques *et al*., 2010 and, Ghaderi *et al*., 2010), and CHO cells specifically are incapable of generating α(2,6)-sialylation and α(1,3)-fucosylation (Patnaik and Stanley, 2006). These factors result in glycoengineering, either *in vivo* or *in vitro* downstream in the manufacture process, being a key requirement for optimising the pharmacokinetic properties of glycoprotein therapeutics expressed in CHO cells. In addition to this, CHO cells, being immortalised cells, have low long-term genetic stability often resulting in impaired process yields (Bailey *et al*., 2012).

Human expression systems are increasingly used to produce therapeutic glycoproteins. Since they are also immortalised cell lines they face similar limitations in terms of genetic stability as other mammalian systems. In addition to this, human cell lines, when cultured on animal-derived media, can incorporate Neu5Gc on glycoproteins (Bardor *et al*., 2004) and the cell surface which, when used for transplantation procedures, represented the first instance of a “xeno-auto-antigen” (Padler-Karavani and Varki, 2011).

1.3.3 Yeast

FH circulates in blood plasma at 200 – 300 µg/mL and, therefore, high yields of recombinant protein are likely required to meet the needs of therapeutic use in humans. Matthews *et al*. (2017) estimate that, for the same number of cells, CHO cells would occupy 75% of a bioreactors volume whereas microbial cells would occupy just 10% of the

bioreactor volume. In addition to this, microbrial eukaryotes have faster growth times leading to more rapid strain engineering as well as higher volumetric productivity (Matthews *et al*., 2017).

Yeast offers a compromise between the human-like glycosylation of mammalian expression systems and the fast growth rates and ease of genetic manipulation of prokaryotic systems. To add to this FH has been expressed in the methylotrophic yeast *P. pastoris,* to therapeutically useful 10s-of-milligram quantities in the Barlow group (Schmidt *et al*., 2011). In addition, fully humanised glycosylation of both N- and O-linked glycosylation in this species has been reported (Hamilton *et al*., 2006 and, Hamilton *et al*., 2013).

Considering this, the yeast *P. pastoris* offers the potential to produce recombinant human FH with humanised glycosylation whilst also meeting the high demands on yield of therapeutic use in large numbers of human patients.

1.4 *Pichia pastoris* and glycoengineering

Pichia pastoris has grown in popularity for heterologous protein production and is used widely in the biotechnology and biopharmaceutical industries. An advantage of *P. pastoris* is its ability to grow to high cell densities, over 130 grams of dry cell weight per litre has been achieved in continuous culture (Cereghino and Cregg, 2000), on simple chemically defined media. Moreover, its doubling time is short compared to mammalian cells. Importantly for our purposes, it can efficiently perform post-translational modifications such as N-linked glycosylation and disulphide bond formation.

A significant contributor to the usefulness of this strain is the tightly regulated and very strong inducible *AOX1* promoter (Cregg *et al*., 1985). *P. pastoris* is a methyltrophic yeast meaning that it is capable of growing on media containing methanol as the sole carbon source. The enzyme responsible for initiating methanol metabolism is alcohol oxidase (AOX), which has been shown to constitute 33% of the total soluble protein content of cells cultured in methanol (Couderc and Baratti, 1980). There are two genes that encode AOX in *P. pastoris*, *AOX1* and *AOX2*. These differ in strength of expression. *AOX1* expresses more mRNA than *AOX2* (Cregg *et al*., 1998).

The *AOX1* promoter (*AOX1*p) was first used for recombinant protein expression by Tschopp *et al*., (1987) and has become the most commonly used promoter for expression of recombinant proteins in *P. pastoris*. The promoter is strongly repressed in the presence of glucose, glycerol and ethanol (Inan and Meagher, 2001), but fully active in the presence of methanol when these other carbon sources have been depleted. This allows for a two-stage cell culture. In the first stage, growth is carried out on non-methanol media which allows cells to reach high cell density. In the second stage, non-methanol carbon sources are depleted and replaced with methanol allowing for high-level expression of heterologous protein without compromising growth of the cell culture. As a little-used alternative to AOX1p, the glycercaldehyde-3-phosphate promoter (GAPp) is a constitutive promoter that can achieve high-level expression comparable to the inducible AOX1p (Waterham *et al*., 1997).

These two promoters have enabled the establishment of *P. pastoris* as a powerful recombinant protein expression platform. For example, Werten *et al*. (1999) reported the production of 14.8 grams per litre of cell culture of mouse gelatin under the control of the AOX1p and Mallem *et al*. (2014) report the production of 10 grams per litre of cell culture of human serum albumin.

The Barlow group has capitalised on the advantageous protein-production characteristics of *P. pastoris* in order to achieve high yields of recombinant, disulphide-rich CCP modular segments of the protein FH (Schmidt *et al*., 2008, Schmidt *et al*., 2010, Makou *et al*., 2012) and full-length FH to 10s of milligram quantities (Schmidt *et al*., 2011, Herbert *et al*., 2015 and, Kerr *et al*., 2017).
In these studies (Schmidt *et al*., 2011, Herbert *et al*., 2015 and, Kerr *et al*., 2017), the *CFH* gene was cloned into the *P. pastoris* expression plasmid pPICZα downstream of the AOX1p, to enable tightly controlled inducible expression, and placed in frame with a gene sequence encoding the *S. cerevisiae* α-secretion signal, for secretion of FH into the culture media. The plasmid was linearised at a site in the AOX1p by restriction endonuclease digestion and introduced into *P. pastoris* by electroporation. Linearisation of the plasmid facilitated single-site homologous recombination between the AOX1p sequence in the plasmid and the homologous AOX1p site in the genome of *P. pastoris* thus, generating a genome integrated expression plasmid – this is standard procedure for heterologous protein expression in *P. pastoris*. All strains of *P. pastoris* used in the current study were

constructed, previously, to express recombinant human FH in this way. The focus of the current study was to engineer the glycosylation pathway of *P. pastoris* that could already express recombinant FH.

1.4.1 Glycoengineering of *P. pastoris*: galactose-terminal glycans

Initial attempts to glycoengineer yeast started in *S. cerevisiae* with the discovery of the mannosyl transferase-encoding gene *OCH1*. This gene was found to be responsible for catalysing the formation of the first step in the hypermannosylation process of yeast glycan synthesis. Knockout of *och1* prevented further glycosylation of the $Man_8GlcNAc_2$ glycan that leaves the ER (Nakayama *et al*., 1992) (fig. 1.12).

Chiba *et al*. (1998) developed this further, by transforming *S. cerevisiae Δoch1* and *Δmnn1*, a gene responsible for outer glycan mannosylation (Shindo *et al*., 1993), with an α(1,2)-mannosidase I from *Aspergillus saitoi* fused with the ER-retention signal peptide sequence HDEL (Pelham, 1988) at the C-terminus. Expression of ManI in the double-knockout strain of *S. cerevisiae* produced the high-mannose glycan $Man_5GlcNAc_2$, as intended. However, the efficiency was low – approximately 20% of the heterogenous glycoprofile of recombinant carboxypeptidase Y produced in this strain was $Man_5GlcNAc_2$ glycans.

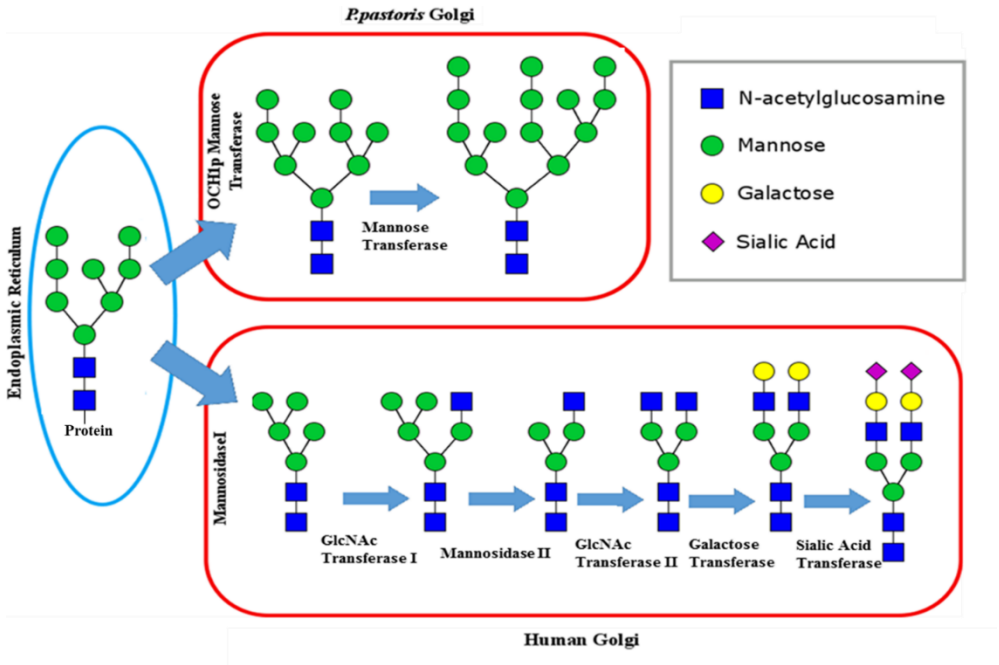


Figure 1.12: Comparison of the glycosylation pathway in yeast and humans. The initial stages of glycosylation are conserved in the ER lumen where a pre-synthesised high-mannose glycan is transferred onto an asparagine in the sequence Asn-X-Ser/Thr (where X is any amino acid except proline) of the nascent polypeptide. The glycan that leaves the ER consists of eight mannose units and two N-acetylglucosamine units ($Man_8GlcNAc_2$). In the golgi of yeast the gene product of *OCH1*, a mannosyltransferase, catalyses the transfer of a single mannose onto the eight-mannose glycan to form $Man_9GlcNAc_2$. This is the substrate for further mannosylation. In humans, however, the eight-mannose glycan that leaves the ER is cut back by mannosidase I to a five-mannose glycan, $Man_5GlcNAc_2$, and this is then extended by the sequential action of multiple monosaccharide transferases to produce complex-type glycans terminating in sialic acid, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$. For simplicity only diantennary complex-type glycans are shown here, however, one to two additional antennae can be generated by the action of N-acetylglucosamine transferases III and IV and core fucosylation is also possible.

Callewaert *et al*. (2001) identified that the glycoprofile of proteins expressed in *P. pastoris* contain $Man_8GlcNAc_2$ glycans and therefore hypothesised that the *OCH1* homologue in *P. pastoris* is less active than *OCH1* in *S. cerevisiae*. In addition, Callewaert *et al*. (2001) also noted that, unlike with *S. cerevisiae*, *P. pastoris* does not produce the hyper-immunogenic outer chain terminal α(1,3)-mannose. Callewaert *et al*. (2001) identified that these factors would make *P. pastoris* a more suitable host for glycoengineering attempts than *S. cerevisiae*. Callewaert *et al*. (2001) expressed HDEL-tagged ManI from *Trichoderma reesei* in *P. pastoris* and found variable results when co-expressed with recombinant glycoproteins. The glycoprofile of *trans*-sialidase co-expressed in the *P. pastoris* ManI strain produced glycans predominantly of the intended five-mannose type, $Man_5GlcNAc_2$ (fig. 1.12). However, co-expression of influenza A haemagglutinin produced a glycoprofile containing $Man_9GlcNAc_2$ as the smallest glycan.

Choi *et al*. (2003) were able to greatly improve the efficiency of production of the target five-mannose, $Man_5GlcNAc_2$ glycan, the chassis for hybrid- and complex-type glycan

formation, by utilising the *Δoch1* deletion together with expression of the *Caenorhabditis elegans* ManI enzyme. This strain was able to produce the $Man_5GlcNAc_2$ structure with high efficiency, >75% (Choi *et al*., 2003). Choi *et al*. (2003) then took this a step further. They expressed *H. sapiens* GnTI, fused with a yeast golgi localisation domain, together with a UDP-GlcNAc transporter from *Kluyveromyces lactis*, which transports the GnTI substrate, UDP-GlcNAc, across the golgi membrane from the cytoplasm in which it is synthesised, to generate $GlcNAc_1Man_5GlcNAc_2$. Although yields were low, this was the first-time hybrid-type glycosylation was reported in yeasts.

Using a combinatorial library approach, Hamilton *et al*. (2003) were the first to report the production of a complex-type glycan in yeast. Building upon the strain engineered by Choi *et al*. (2003), Hamilton *et al*. (2003) sequentially introduced *Drosophila melanogaster* ManII, responsible for converting $GlcNAc_1Man_5GlcNAc_2$ to $GlcNAc_1Man_3GlcNAc_2$, followed by *Rattus norvegicus* GnTII which converted the hybrid-type glycan $GlcNAc_1Man_3GlcNAc_2$ into the complex-type glycan $GlcNAc_2Man_3GlcNAc_2$. In both instances, the catalytic domains of the two introduced enzymes ManII and GnTII were fused to *S. cerevisiae* Mnn2 golgi transmembrane localisation domains.

In a parallel approach, Vervecken *et al*. (2004) developed a strain of *P. pastoris* expressing a hybrid-type glycan terminating in galactose, $Gal_1GlcNAc_1Man_5GlcNAc_2$. The key features of this study were two-fold. Firstly, Vervecken *et al*. (2004) were able to create the *Δoch1* strain by simple single homologous recombination between a truncated *och1* gene on an expression plasmid and the endogenous *OCH1* site in the genome. This achievement simplified the otherwise laborious task of *Δoch1* deletion used previously (Nakayama *et al*., 1992, Nakanishi-Shindo *et al*., 1993 and, Choi *et al*., 2003) and allowed for the ManI gene to be integrated simultaneously with *Δoch1* deletion by including the ManI gene on the same plasmid as the truncated *och1* sequence. This represented the first instance of galactose incorporation into glycans biosynthesised *in vivo* in yeast.

In yet another parallel approach, Bobrowicz *et al*. (2004) were able to generate a complex-type glycan terminating in galactose, $Gal_2GlcNAc_2Man_3GlcNAc_2$, in *P. pastoris*. This study differed from previous studies in two ways. First, unlike previous studies that focused on engineering glycosylation in the golgi lumen, this study aimed to block the glycosylation pathway at an earlier stage, before *OCH1*, when the glycan is biosynthesised on the

dolichol linker. Bobrowicz *et al*. (2004) deleted *Δalg3* which prevented extension of the dolichol-linked $Man_5GlcNAc_2$-P-P-Dol glycan. This glycan could then be transferred onto the nascent glycoprotein in the ER. Second, building on the work of Choi *et al*. (2003) and Hamilton *et al*. (2003), Bobrowicz *et al*. (2004) were able to generate the galactose-terminal complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$, by expressing a fusion protein construct of the catalytic domains of *Schizosaccharomyces pombe* UDP-galactose 4-epimerase and *H. sapiens* β(1,4)-galactosyltransferase, in a process inspired by Chen *et al*. (2000), to the *S. cerevisiae* Mnn2 golgi membrane-localisation domain. This construct allows for the simultaneous production of a pool of UDP-galactose substrate, generated by epimerisation of UDP-galactose, and transfer of galactose from this substrate onto the nascent $GlcNAc_2Man_3GlcNAc_2$ glycan.

1.4.2 Glycoengineering of *P. pastoris*: sialic acid terminal glycans

The final, and arguably the most important, step in humanising the glycosylation pathway of *P. pastoris* was realised by Hamilton *et al*. (2006) who accomplished the impressive feat of introducing five exogenous genes involved in the biosynthesis of CMP-sialic acid, transport of CMP-sialic acid into the golgi and transfer of sialic acid onto the nascent glycan to form the fully humanised diantennary terminally sialylated glycan $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$ with sialic acid in an α(2,6) linkage (fig. 1.13). Hamilton *et al*. (2006) were also able to consolidate the sialic acid biosynthesis pathway onto a single expression plasmid and, by using a combinatorial library optimisation approach, were able to achieve 90.5% disialylated and 7.9% monosialylated diantennary complex-type glycans on recombinant erythropoietin.

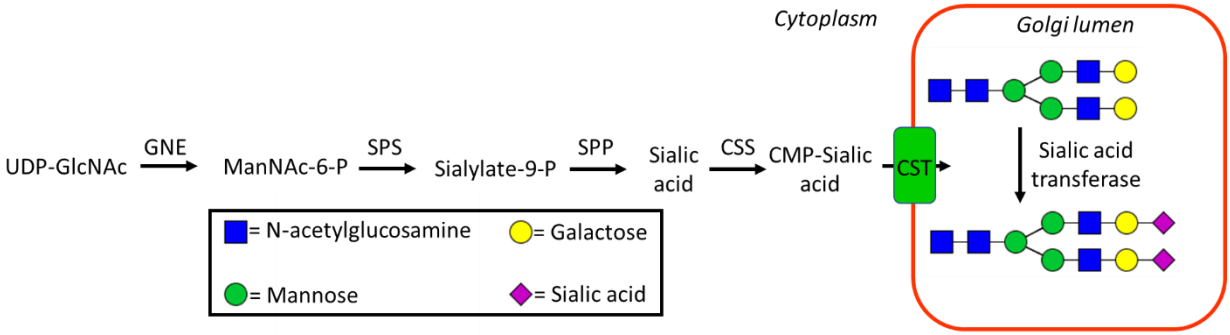


Figure 1.13: Schematic diagram of the steps required to sialylate galactose-terminal complex-type glycans in yeast golgi. The pathway is initiated by conversion of uridine diphosphate (UDP)-GlcNAc to N-acetylmannosamine-6-phosphate (ManNAc-6-P) by UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase (GNE). ManNAc-6-P is converted to sialylate-9-phosphate by N-acetylneuraminate-9-phosphate synthase (SPS) and then to sialic acid by sialylate-9-phosphate phosphatase (SPP). CMP-sialic acid synthase (CSS) catalyses the charging of sialic acid with cytidine monophosphate (CMP) the CMP-sialic acid product is transported into the golgi lumen from the cytoplasm by CMP-sialic acid transporter (CST). The nascent glycan terminal β(1,4)-galactose is sialylated by sialic acid transferase using CMP-sialic acid as substrate.

1.4.3 *P. pastoris* glycoengineering technology: GlycoSwitch vs GlycoFi

Jacobs *et al*. (2009) combined the previous works (Callewaert *et al*., 2001, Hamilton *et al*. 2003, Bobrowicz *et al*., 2004 and, Vervecken *et al*., 2004) to generate the licensable GlycoSwitch technology.

GlycoSwitch consists of a single starting strain of *P. pastoris*, called SuperMan 5, that has been engineered to contain *Δoch1* and *T. reesei* ManI, in a process similar to that described in Vervecken *et al*. (2004), as well as four vector plasmids encoding *H. sapiens* GnTI, *D. melanogaster* ManII, *R. norvegicus* GnTII and the *S. pombe* UDP-glucose 4-epimerase/ *H. sapiens* GalT fusion construct described by Bobrowicz *et al*. (2004).

Introduction of this technology into SuperMan 5 yields a humanised galactose terminal diantennary complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$. Jacobs *et al*. (2009) report that the doubling time of the glycoengineered strain, also expressing recombinant mouse interleukin-10, is almost doubled compared to a wild-type strain, 2.4 hours compared to 4.6 hours, respectively. Although this is a significant increase in doubling time and, therefore, could negatively affect protein production yields, it is still less than the 20 – 24 hour doubling time of CHO cell lines expressing recombinant proteins (Sunstrom *et al*., 2000). Thus, demonstrating that glycoengineered *P. pastoris* still retains an advantageous growth rate compared to CHO cells. Therefore, given that *P. pastoris* is a well-established

system used for high-level production of full-length human recombinant FH (Schmidt *et al*., 2011, Herbert *et al*., 2015 and Kerr *et al*., 2017), it seems that the GlycoSwitch technology offers an attractive route for the production of potentially therapeutic human recombinant FH with humanised glycosylation.

The *in vivo* sialylation technology developed by Hamilton *et al*. (2006) as part of GlycoFi Inc. was sold to Merck-Serono in 2008 (Roumeliotis, 2008). This technology relies upon the GlycoSwitch technology to achieve terminal galactosylated complex-type glycans *in vivo* and then builds upon this glycan species by the addition of sialic acid. However, Merck-Serono's ownership of this technology, and their restricted offerings of licenses for its use, has hampered the academic research community in applying this technology for the production of recombinant glycoproteins with humanised glycosylation in *P. pastoris*. This does not preclude the replication of this technology *in vivo* for research purposes. An alternative is to purchase a license for GlycoSwitch to humanise the glycosylation as far as terminal galactose *in vivo* followed by enzyme-catalysed sialylation of recombinant humanised glycoproteins *in vitro*.

The current study focused on implementing the GlycoSwitch technology to develop a galactose-terminal glycan *in vivo*. Once this was achieved the next step was an attempt to implement *in vivo* sialylation, modelling the method on the GlycoFi technology. Ancillary to this, *in vitro* sialylation of purified recombinant FH carrying terminally galactosylated glycans was performed with a recombinant sialyltransferase.

1.5 Glycobiology toolkit

Glycans are a diverse group of carbohydrate molecules. As yet, no single method exists for complete characterisation of glycans and therefore a repertoire of techniques must be employed in order to acquire information on the type of glycosylation (N- or O-linked, high-mannose, hybrid- or complex-type), the monosaccharide composition, the nature of the glycosidic linkages, etc. Detailed below are the key techniques used in this study for the analysis and characterisation of glycans.

1.5.1 Endo- and exoglycosidases

Glycosidases are a group of enzymes that catalyse the hydrolysis of O-glycosidic bonds. This group can be divided into two categories: 1) glycosidases that cleave glycosidic bonds at the terminus of a polysaccharide chain – exoglycosidases – and 2) those that cleave glycosidic bonds from within a polysaccharide chain – endoglycosidases.

The enzymes most frequently used as tools in glycobiology, Endoglycosidase H (endo H) and Peptide-N-glycosidase F (PNGase F), cleave different types of bond and have different substrate specificities. Endo H is an endoglycosidase that cleaves the glycosidic linkage between the two GlcNAc of the chitobiose core of N-linked glycans (fig. 1.14). In contrast, PNGase F cleaves the glycopeptide linkage between the reducing terminal chitobiose core GlcNAc and the N-linked asparagine residue on the polypeptide chain; in other words, PNGase F is an amidase but, like endo H, it liberates N-linked glycans from glycoproteins (fig. 1.13).

PNGase F can cleave all major types of N-linked glycan (Tretter *et al*., 1991), except for those with an α-1,3-fucose linked to the chitobiose core.
Conversely, Endo H has a narrower substrate specificity in that it can cleave only high-mannose and some hybrid-type glycans from protein conjugates (Maley *et al*., 1989) (fig. 1.14).

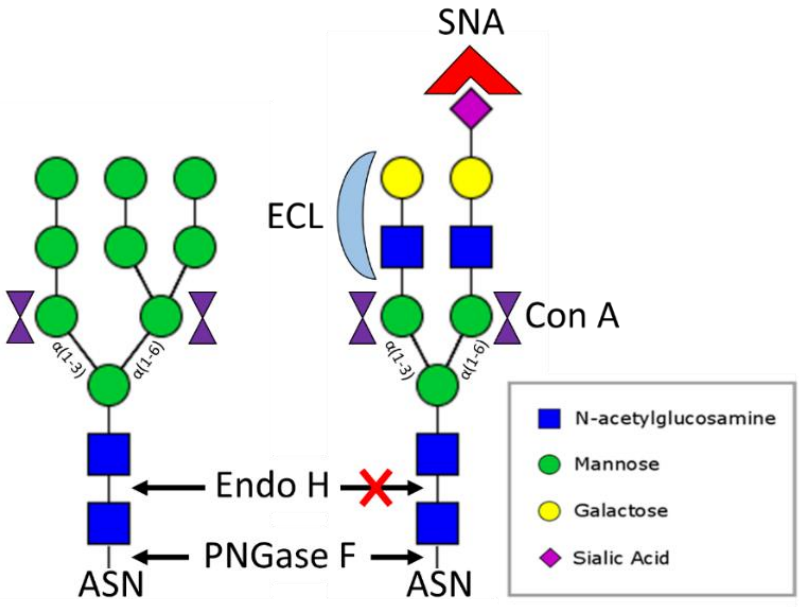


Figure 1.14: Schematic diagram showing the substrate specificities of the endoglycosidases Endoglycosidase H (endo H) and Peptide-N-glycosidase F (PNGase F) toward high-mannose (left) and complex-type (right) glycans. Also shown are the glycan-binding specificities of the lectins *Erythrina cristagalli* (ECL, blue crescent), *Sambucus nigra* (SNA, red indented triangle) and, Concanavalin A (Con A, purple triangles joined point to point). ASN – asparagine residue linked to N-linked glycan at the reducing terminus. The α-linked mannose of the trimannosyl-core are labelled (however, high-mannose glycans have many more).

1.5.2 Lectins – carbohydrate binding proteins

Lectins are a group of carbohydrate-binding proteins that exhibit extraordinary specificity for their intended target. For example, *Sambucus nigra* lectin (SNA) will bind preferentially to N-acetylneuraminic acid (Neu5Ac) in an α(2,6)-glycosidic linkage with galactose (Gal) compared to Neu5Ac α(2,3)-Gal. SNA can weakly bind to free Gal but not to free Neu5Ac (Shibuya *et al*., 1987), indicating that the glycosidic linkage is an essential part of SNA recognition.

In this thesis, three lectins were used for western blot and ELISA to detect specific glycan moieties. The lectins used were (fig. 1.14):

(i) SNA (see above).

(ii) *Erythrina cristagalli* lectin (ECL) – binds strongly to the disaccharide N-acetyllactosamine (LacNAc) (Iglesias *et al*., 1982) (providing it is not sialylated) that is a key component of the antennae of most complex mammalian glycans (Debray *et al*., 1986), therefore ECL binds epitopes on the terminus of the glycan.

(iii) Concanavalin A (ConA) – binds to α-mannose (as well as α-glucose) (Goldstein *et al*., 1973) when it is situated within the trimannosyl-core common to all N-linked glycans. Additionally, Goldstein *et al*., (1973) found that ConA can bind to its epitope regardless of whether it is at the glycan terminus or buried within the glycan chain.

Throughout this study, ConA was used to detect for the presence of a glycan, irrespective of the glycan type (e.g. high-mannose, complex-type, etc). SNA was used to detect for the presence of sialic acid (specifically Neu5Ac). Finally, ECL was used to detect the presence of LacNAc. In general, lectins have served as an important part of the glycan-analysis toolkit to evaluate the glycosylation of human plasma-purified and recombinant glycoforms of FH.

1.5.3 BlotGlyco glycan purification and MALDI-TOF mass spectrometry

A significant obstacle in whole glycan characterisation is that a lengthy purification processes must be employed to achieve the clean, highly pure glycan samples required for glycan analysis (Pilobello and Mahal, 2007). Sialic acids present an additional limitation in whole-glycan characterisation due to the presence of the C1 COO^- group. This group contributes to the inherent instability of the sialic acid α-glycosidic linkage, often resulting in loss of sialic acid during the harsh ionisation conditions used in matrix-assisted laser-desorption ionisation (MALDI) and, even, during electrospray ionisation (ESI). In addition to this, both positive and negative ions form, resulting in splitting of mass signals. Moreover, partial alkali salt formation can interfere with the mass spectrum profile, often leading to complex spectra that are difficult to decipher (Powell and Harvey, 1996). The negatively charged sialic acids also interfere with analysis by high performance liquid chromatography (HPLC). The distinct chemistry of sialic acids, compared to the other neutral monosaccharides that make up complex-type glycans, often results in the requirement for multiple HPLC steps.

BlotGlyco is a hydrazide-functionalised solid-support resin technology platform that overcomes these limitations in whole-glycan characterisation (fig. 1.15). BlotGlyco covalently fixes glycans with a free reducing terminus to the hydrazide-support resin by facilitating, under mild acidic conditions, the formation of a Schiff base hydrozone bond between the carbonyl oxygen of the reducing terminal monosaccharide and the hydrazide group (fig. 1.15 B). This process is highly specific for glycans and, because the reducing terminus is common to all free glycans, is non-discriminatory (Furukawa *et al*., 2008). Glycans can then be simultaneously released and tagged by transamination, substituting the resin for an appropriate tag via an oxime or hydrozone bond.

The procedure to purify and tag glycans for analysis takes six hours, compared to 2-3 days for standard procedures, and in addition, the functionalised resin has a very high hydrazide density of ~460 µmol/mL (Furukawa *et al*., 2008). Therefore, whole-glycome analysis can be achieved rapidly with high sensitivity.

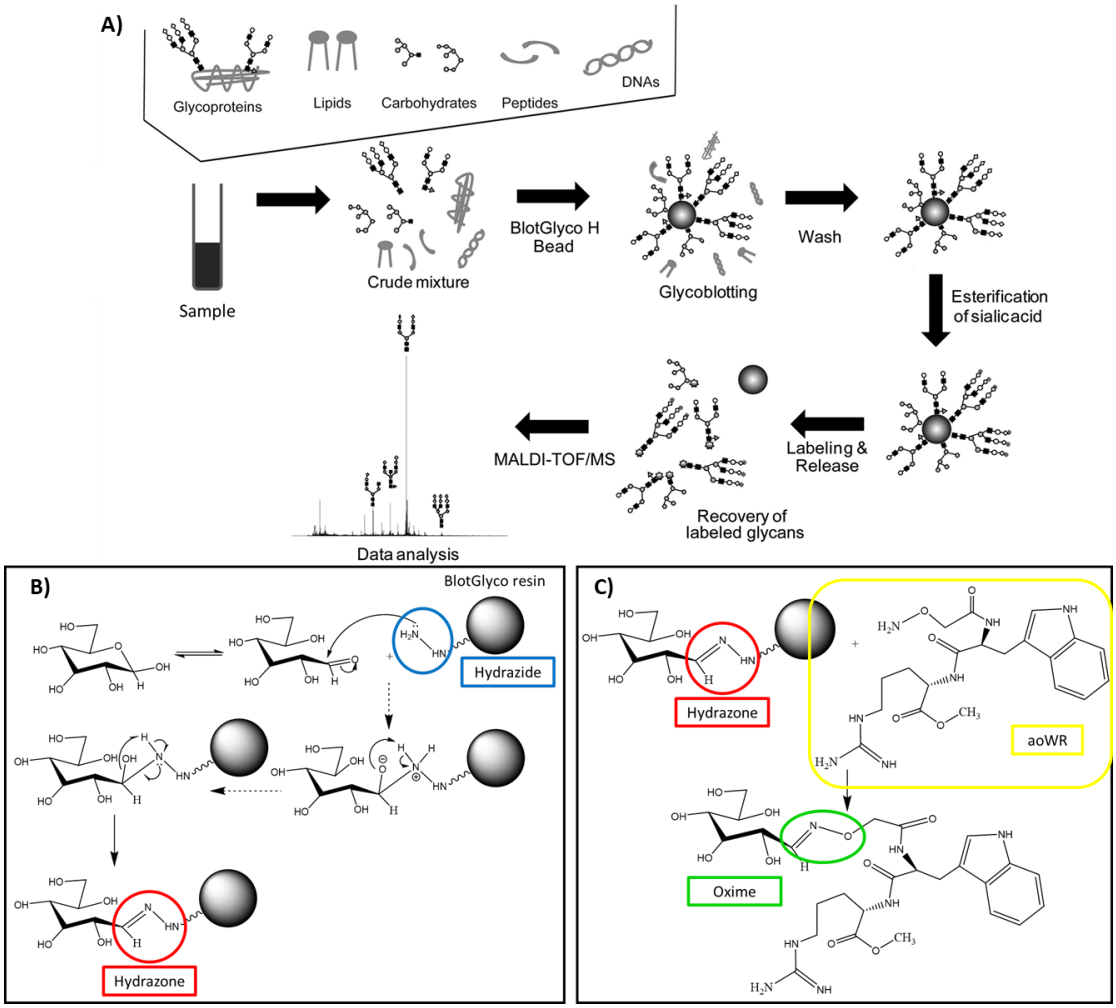


Figure 1.15: A) Schematic diagram of the steps involved in BlotGlyco glycan purification. The sample can be pure protein or a crude mixture e.g. cell culture media. The sample is enzymatically deglycosylated before applying to the BlotGlyco resin. Glycans are selectively extracted from the sample by covalent linkage to the resin by hydrozone bond formation in mild acidic conditions. The glycan-associated resin is then subjected to multiple, exhaustive wash steps to remove non-covalently associated impurities. Methyl esterification is carried out in solid phase to neutralise the COO^- group on sialylated glycans and the glycan is then released by hydrozone-to-hydrozone or hydrozone-to-oxime transamination substitution. The released, tagged glycans are now pure enough for analysis by a method of choice. Image taken and adapted from Terashima *et al*. (2014). B) Reaction mechanism of the covalent attachment of carbohydrates to BlotGlyco resin by the formation of a hydrazone bond. C) Glycans are released from BlotGlyco beads by a hydrozone-to-oxime transamination substitution reaction in the presence of aoWR tag. AoWR tagged glycans are then washed off the beads and collect for analysis by MALDI-TOF MS. Glucose is shown here as an example, but any glycan with a free reducing terminus, and thus a hemiacetal/carbonyl group, can be covalently linked to the resin.

In addition to this, BlotGlyco utilises a solid-phase esterification procedure for methyl esterification of sialic acid COO^-, forming the neutral group $COOCH_3$. This process eliminates the need for use of toxic, halogenated solvents to achieve the same end, neutralises the positively charged sialic acids, and reduces the lability of the sialic acid glycosidic linkage (Miura *et al*., 2007) thus enabling analysis of sialylated glycans by MALDI time of flight (TOF) mass spectrometry and single step HPLC analysis.

BlotGlyco was used throughout this study to purify whole glycans enzymatically cleaved from both purified samples of different glycoforms of FH and in the purification of glycans from crude *P. pastoris* culture media. Shinohara *et al*. (2004) found that guanidino and hydrophobic functional groups on tags used to derivatise glycans greatly improved the ionisation potential. Building upon this Uematsu *et al*. (2005) developed the dipeptide tag aminooxy-tryptophanylarginine (aoWR). This tag was used to release BlotGlyco captured glycans for efficient, highly sensitive qualification of glycan mass and structure by MALDI-TOF mass spectrometry.

1.6 inABLE: Multipart DNA assembly technique

As well as the more traditional DNA assembly techniques, e.g. cloning, a multipart DNA assembly technique called inABLE was used throughout this present work to assemble individual DNA sequences into a single contiguous sequence. A detailed description of the inABLE procedure is given below, as well as a more complex "nested" inABLE assembly approach.

1.6.1 The inABLE procedure

inABLE is a proprietary 'onepot' DNA assembly technology (Che *et al*., 2010). inABLE allows for the highly specific assembly of multiple DNA 'parts' to create a single, much larger DNA sequence. This process is useful in joining upwards of two sequences for a range of purposes, such as, inserting a gene sequence into an expression vector or creating multi-gene operons on a single DNA plasmid.

inABLE works by taking each part and dividing it into three double-stranded segments. These are, in order of 5' to 3', i) Linker Oligonucleotide (LO), ii) Part Oligonucleotide (PO) and iii) Truncated Part (TP). All LOs have the same overhang triplet sequence at the 5'-end of each sense-strand and the complementary triplet sequence is on the 5'-end of the TP anti-sense strand. This would allow for any TP to be ligated to any LO.
The assembly specificity is provided by the triplet sequence on the 5'-end of the TP sense strand and its complement on the 5'-end of the anti-sense strand of the TP associated PO. Thus, only specific TPs can be associated with specific POs.

Each associated PO and LO have long, (~16 bp), complementary overhangs that facilitate the association of two or more different parts by complementary base pairing followed by ligation.

As an example (fig. 1.16), a two-part assembly consisting of a gene – part A – and a backbone vector plasmid – part B – would be assembled such that the PO of part A (PO_A) is attached to the 5'-end and the LO of part B (LO_B) attached to the 3'-end of TP_A. For part B, PO_B is attached at the 5'-end and LO_A attached to the 3'-end, respectively. These two ligations are carried out independently.

The sequences PO_B – TP_B – LO_A and PO_A – TP_A – LO_B are then incubated together resulting in ligation of the two parts by complementary base pairing between the overhangs of PO_A with LO_A and PO_B with LO_B, respectively.

The long sequence complementarity between associated PO and LO (e.g. PO_A and LO_A) offers exquisite specificity in part-part association, whilst the paired triplet sequence on LOs and TPs allows for flexibility in part-part associations (e.g. TP_A can be associated with either LO_B or LO_C or LO_D, etc). Thus, this allows for the controlled assembly of multiple parts in multiple different orders.

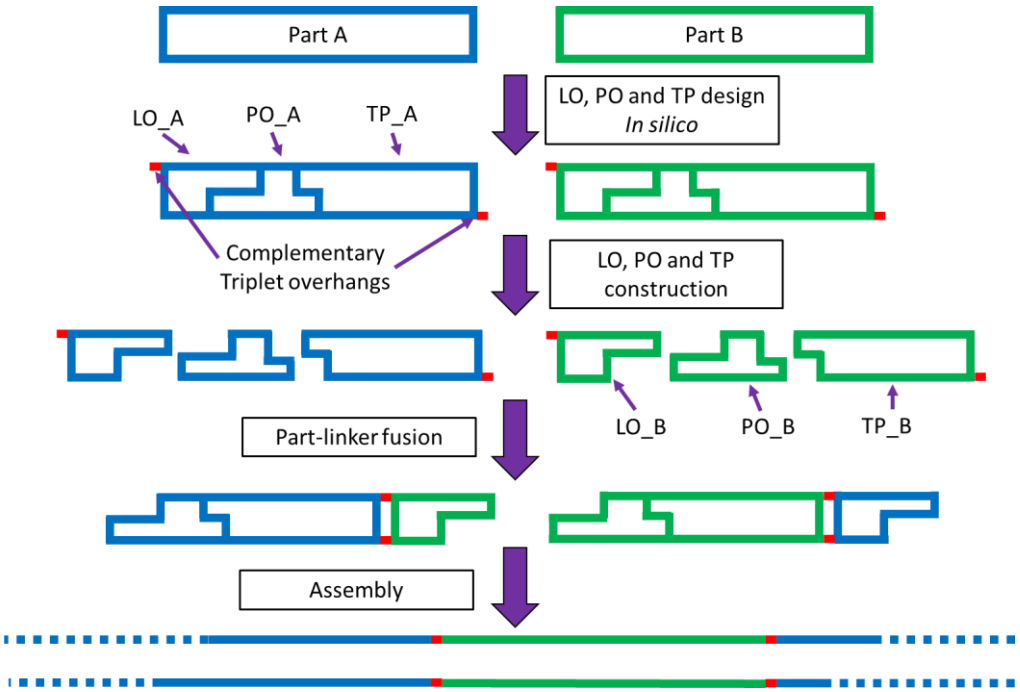


Figure 1.16: Schematic diagram of a two-part inABLE assembly. First, gene sequences A and B are segmented *in silico* to give the LO, PO and TP sequences. LO and PO sequences are assembled from single-stranded oligonucleotide sequences annealed together and phosphorylated and the TP is constructed either by DNA synthesis or PCR. Part-linker fusion reactions are set up to fuse PO_A to TP_A to LO_B and PO_B to TP_B to LO_A. The complementary triplet overhangs (red lines) are identical for all DNA parts and allow for any TP to be joined to any LO. Part-linker fusion products are placed in a single-pot reaction. Complementary base pairing of LO_A to PO_A and LO_B to PO_B results in the joining together of Part A and B as a single contiguous sequence.

In the original part, placed upstream and downstream of the triplet sequences – either by inclusion of the sequence in PCR primers or when the part is synthesised – is the same type II S restriction endonuclease (RE) recognition sequence, such that there are two copies of this sequence flanking the part in opposite orientation - or mirror images - of one another. Type IIS RE are a group of RE which cleave at a location a defined distance from the DNA recognition sequence.

Each TP, when generated by synthesis or PCR, can be cloned into a backbone plasmid to generate a library of TPs. The library of TPs are in a readily useable format for direct ligation to the desired PO and LO segments by type IIS RE-mediated excision.

1.6.2 The inABLE procedure: a “nested” approach

A limitation of the inABLE technology is that assembly efficiency decreases as the number of parts increases. DNA assemblies of no more than 6-7 parts are feasible by the standard inABLE methodology. However, by using a 2-step “nested” approach the assembly efficiency limitations are overcome, which allows for many more parts to be used in the process.

To use a nine-part assembly as an example (fig. 1.17), in the first step of the “nested” approach the nine parts are divided into three groups of three. Each group (comprising parts A1-3, B1-3 and C1-3) is assembled independently by the inABLE procedure to form three “nested” parts (A, B and C). Crucial to this first step is that the PO (PO_x and its complimentary LO_x), associated with parts A1, B1 and C1, contains a type IIS restriction site. Crucially a fourth part, a plasmid backbone for selection and amplification in *E. coli*, is introduce that has in its LO the same type IIS-RE site in reverse orientation to the one in the “nested” LO of A1, B1 and C1. In the second step of a nested assembly, digestion of “nested” parts A, B and C with the type II S RE removes the “nested” LO_x and PO_x and creates a three-base pair overhang (triplet sequence) at the 5’ and 3’ ends of “nested” parts A, B and C. The resulting truncated “nested” parts A, B and C can now be ligated, using DNA ligase, to the true LO and PO associated with part A1, B1, and C1, respectively. These parts can then be assembled into one single contiguous DNA sequence.

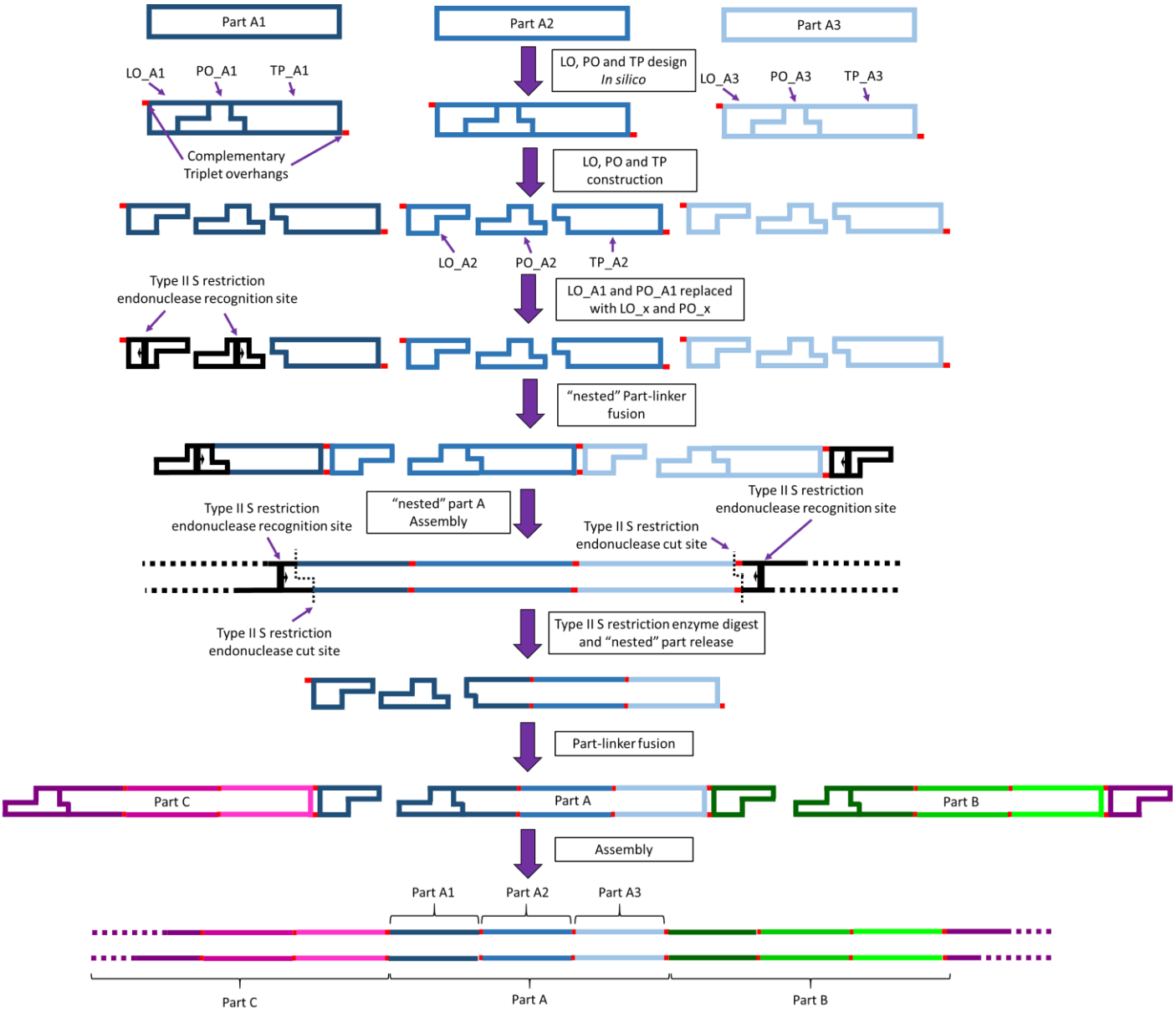


Figure 1.17: Schematic diagram of a nine part "nested" assembly. The nine parts are divided in to three groups (A1-3, B1-3 and C1-3) and assembled into three "nested" parts (A, B and C) in a process identical to that set out in figure 1.15 (just assembly of nested part A is shown here), except that the LOs and POs associated with part A1, B1 and C1 is swapped for the "nested" LO_x and PO_x. LO_x and PO_x contain type II S restriction endonuclease (RE) sites. After assembly of each "nested" part (A, B and C), the nested parts can be digested with the type II S RE which create overhangs for the attachment to the original PO and LO associated with parts A1, B1 or C1, in a part-linker fusion reaction. This enables association of "nested" part A to B to C as one contiguous sequence.

1.7 Aims

The overall goal of the current study was to produce therapeutically useful quantities of recombinant human FH with humanised glycosylation in *P. pastoris*. In endeavouring to achieve this principle aim the following objectives were set:

(i) The glycosylation profile of human FH must be characterised to verify, or otherwise, the sole, detailed study in the literature (Fenaille *et al*., 2007*a*) and to determine the target glycan for glycoengineering.

(ii) The GlycoSwitch technology should be introduced, sequentially, into *P. pastoris* and a strain expressing recombinant FH with human-like complex-type diantennary digalactosylated glycans ($Gal_2GlcNAc_2Man_3GlcNAc_2$) selected by lectin western blot and the glycoprofile validated by MALDI-TOF mass spectrometry.

(iii) The GlycoSwitch plasmid technology, currently consisting of four genes on four plasmids, should be consolidated onto a single expression plasmid, using inABLE "nested" assembly approach, in order to make available antibiotic selectable markers for *in vivo* sialylation.

(iv) Time limitations prevented attempts at *in vivo* sialylation of glycans on FH. However, recombinant FH expressed in glycoengineered *P. pastoris* was purified and sialylated *in vitro* using the bacterial α(2,6)-sialyltransferase, recombinantly expressed in *E.coli.*

(v) The functional activity of terminally galactosylated and terminally sialylated recombinant FH is to be compared to the well characterised human plasma-derived FH and deglycosylated recombinant FH. Analysis will first be conducted *in vitro* by Biacore-based assays, comparing binding affinity of the different glycoforms to C3b and C3d and measuring the ability to accelerate the decay of the C3 convertase complex C3bBb, and a gel based cofactor activity assay. *In vivo* analysis will be carried out in wild-type mice, to assess half-life in blood circulation, and in $cfh^{-/-}$ mice to assess ability to recover plasma C3 levels.

Chapter 2

Investigating the role of glycosylation on the function of human factor H

2.1 Overview

This chapter used the substrate recognition properties of the endoglycosidases and lectins referenced in Chapter 1 section 1.5.1 and MALDI-TOF mass spectrometry to confirm the findings of Fenaille *et al*. (2007*a*) in terms of the whole-protein glycoprofile of human FH. Further, and as discussed in Chapter 1 section 1.2.2, FH may interact with its own sialic acids possibly permitting tighter regulation of FH binding to self-surfaces. Therefore, the aim of the current work is to explore the role FH sialic acids have in the function of FH.

2.2 Materials and Methods

2.2.1 Purification of human factor H from blood plasma

2.2.1.1 PspCN affinity chromatography

Non-sterile, citrated human blood plasma (TCS Biosciences, PR100-1L) was clotted by the addition of 1 M $CaCl_2$ in a 1:20 ratio ($CaCl_2$ solution to plasma) and incubation at 37 °C for 90 minutes. Coagulated proteins were removed by centrifugation, in a Fiberlite F12-6X500 LEX fixed angle rotor (ThermoFisher Scientific) in a Sorval Lynx4000 centrifuge (ThermoFisher Scientific), at 6000 g for 30 minutes at 4 °C. The supernatant was mixed 1:1 with buffer (Phosphate buffer saline (PBS), containing 300 mM NaCl and 20 mM EDTA) before filter sterilisation through 0.22-µm filters (Nalgene Rapid-flow PES filter units, FIL8416).

Purification was carried out on a Tricorn 5/100 column (GE Healthcare Life Sciences), prepared in-house by Dr. Andrew Herbert, containing UltraLink Iodoacetyl resin (ThermoFisher Scientific, 53155) modified with PspCN. The column was equilibrated in 20 mM potassium phosphate buffer, pH 7.0; approximately 400 mL diluted human blood plasma, free of clotting factors, was loaded at 5 mL/min onto the PspCN-UltraLink column. The column was washed with 7 column volumes (CV) buffer A1 (20 mM potassium phosphate buffer, pH 7.0), a 2 CV linear gradient to 50% buffer B (20 mM potassium phosphate buffer, 1 M NaCl, pH 7.0) and a 5 CV equilibration into 50% buffer B and 50%

buffer A2 (0.1 M glycine, pH 2.5). Fractions were eluted in a 4 CV gradient of 50% - 100% buffer A2. 1.5 mL fractions were collected and to each fraction was added a one-tenth volume of 1 M Tris buffer, pH 8.0. This process was repeated for a total of four runs.

Aliquots of 2 µL of (i) blood plasma after removal of clotting factors, (ii) column flow through ten-times concentrated by centrifugation of 500 µL in 10 kDa molecular weight cut-off Pierce Protein Concentrators (PES, 0.5 mL, ThermoFisher Scientific, 88513) at 10,000 rpm for 5 minutes (Centrifuge 5415R, Eppendorf), and (iii) selected fractions (400 – 445 mL (wash fraction) and 510 mL – 541 mL (D11-F8) – elution volume) (fig. 2.1), were each mixed with 14 µL SDS-PAGE loading buffer (in duplicate, with and without reducing agent). These samples were loaded onto polyacrylamide gels (NuPAGE 4-12% bis-tris, ThermoFisher Scientific, NP03222). The gels were run (XCell *Sure*Lock Mini-Cell, MOPS buffer, 200 V, 1 hour) and stained with Coomassie Brilliant Blue G250.

2.2.1.2 Cation exchange chromatography

Based on the chromatogram and the appearance of the gels, appropriate fractions were pooled, placed in 10 kDa molecular weight cut-off dialysis tubing (SnakeSkin, ThermoFisher Scientific, 88245) and dialysed overnight, at 4 °C, against cation-exchange buffer A (20 mM potassium phosphate, pH 7.0), with buffer exchange after 2 hours. Cation-exchange chromatography was carried out on a tricorn 5/100 column (GE Healthcare Life Sciences, 28406410) pre-packed with Source 15S cation-exchange resin (GE Healthcare Life Sciences, 17094401), the column was equilibrated in 3 CV buffer A, sample was loaded at 6 mL/min and the column was washed with 3 CV buffer A before eluting by a 16 CV linear gradient from 0 – 50% buffer B (20 mM potassium phosphate, 1 M NaCl, pH 7.0) and a final high salt wash over 1CV from 50% - 100% buffer B. 1 mL Fractions were collected and analysed by SDS-PAGE (fig. 2.3 and 2.4) as before. Appropriate fractions (C9 – D9, 158.8 mL to 170.9 mL elution volume) were pooled and placed in 10 kDa molecular weight cut-off dialysis tubing (SnakeSkin, ThermoFisher Scientific, 88245) and was dialysed against phosphate buffered saline, pH 7.4, at 4 °C, overnight, with buffer exchange after 2 hours. Protein was dispensed into 250 µL aliquots and stored frozen at -80 °C.

2.2.2 Human factor H glycan characterisation: Endoglycosidase sensitivity

2.2.2.1 SDS-PAGE and Lectin western blot analysis

For SDS-PAGE and lectin-based western blot analysis, Fetuin (New England Biolabs, P6042S), Ribonuclease B (RNase B) (New England Biolabs, P7817S) and plasma-derived human (h)FH (prepared in-house) were either left untreated or were deglycosylated with endoglycosidase H_f (Endo H_f, a maltose-binding protein-tagged version of endoglycosidase H, New England Biolabs, P0703S) or Peptide-N-glycosidase F (New England Biolabs, P0704S).

For Endo H_f treatment, 20 µg protein was made up to 10 µL with 1 µL glycoprotein-denaturing buffer (0.5% (w/v) SDS, 40 mM DTT) and double distilled (dd) H_2O, and denatured by heating to 100 °C for 10 minutes. After cooling, the reaction was made up to 20 µL with ddH_2O, 2 µL glycobuffer 3 (50 mM sodium acetate, pH 6.0) and 1 µL (1kU) Endo H_f. For PNGase F treatment, 20 µg protein was made up to 10 µL with 1 µL glycoprotein-denaturing buffer ddH_2O, and denatured by heating to 100 °C for 10 minutes. After cooling, the reaction was made up to 20 µL with ddH_2O, 2 µL glycobuffer 2 (50 mM sodium phosphate, pH 7.5), 2 µL 10% (v/v) NP40 and 1 µL (500U) PNGase F. For the untreated samples, 20 µg protein was made up to 20 µL as with the PNGase F reaction, but 1µL ddH_2O was added instead of PNGase F.
After mixing, both treated and untreated samples were incubated overnight (16 hours) at 37 °C. Then 500 ng (for SDS-PAGE) or 200 ng (for lectin-based western blot) protein was loaded onto gels (Bolt 4 – 12% bis-tris plus gels, ThermoFisher Scientific, NW04120BOX), and run (Invitrogen Mini Gel Tank, MES buffer, 200 V, 35 minutes). One gel was stained using Instant*Blue* (Expedeon, ISB1-1L) Coomassie stain whilst the rest were subjected to three lectin-based western blots to assess binding by the core N-linked glycan binding lectin ConA, the sialic acid binding lectin SNA and the galactose binding lectin ECL.

All gels for lectin western blotting are run with three hFH derived glycosylation standards. These standards are: i) untreated hFH, terminal sialic acid allows for binding by SNA and is therefore a positive control for SNA and ConA ii) hFH desialylated with α(2-3,6,8)-neuraminidase (New England Biolabs), exposure of galactose allows for binding by ECL and

is therefore a positive control for ECL and ConA; and iii) hFH deglycosylated with PNGase F (New England Biolabs), acts as a negative control for all three lectin westerns.
Lectin western blot standards were prepared as follows. For the untreated hFH standard, 10 µg plasma purified hFH was diluted in ddH_2O to a concentration of 91 ng/µL. For the desialylated hFH standard, a solution containing 10 µg plasma purified hFH, 10 µL glycobuffer 1 (50 mM sodium acetate, 5mM $CaCl_2$, pH 5.5), 500 U α(2-3,6,8)-neuraminidase was made up to 110 µL. For the deglycosylated standard, 10 µg plasma purified hFH was mixed with 10 µL denaturing buffer and made up to 50 µL with ddH_2O. This solution was briefly mixed before incubating at 100 °C for 10 mins. The solution was cooled on ice and centrifuged briefly before 10 µL each of glycobuffer 2 and 10% NP-40 was added together with 5 kU PNGase F. The solution was made up to 110 µL with ddH_2O. All standards were incubated at 37 °C for 16 hours.

After gel electrophoresis of samples for lectin western analysis, polyacrylamide gels were placed in a western blot transfer stack (iBlot 2 Transfer Stacks, PVDF, mini, ThermoFisher Scientific, IB24002) and transferred to the iBlot 2 Dry Blotting System (ThermoFisher Scientific). Protein was transferred onto the polyvinylidene fluoride (PVDF) membrane by electrophoresis at 25 V for 7 minutes. Membranes were removed and placed in blocking buffer (1 x Carbo-Free Blocking Solution, Vector Laboratories, SP-5040), a buffer free of carbohydrates that could otherwise cause interference with lectin-glycan binding, and incubated at room temperature (20 – 25 °C) for 30 minutes and 40 rpm rocking. For all three gels blocking buffer was poured off and replaced by a 20 mL solution of 0.5 µg/mL lectin in blocking buffer. The first PVDF membrane was placed in diluted biotinylated lectin ConA (Vector Biolaboratories, B-1005) solution, the second in biotinylated lectin SNA (Vector Biolaboratories, B-1305) solution and the third in biotinylated lectin ECL (Vector Biolaboratories, B-1145) solution. PVDF membranes were incubated in lectin solution at room temperature (20 – 25 °C) for 30 minutes and 40 rpm rocking. Lectin solutions were poured off and the membranes were washed thrice in TPBS buffer (1% (v/v) tween-20, 10 mM phosphate buffer, 137 mM NaCl, 2.7 mM KCl, pH 7.4). After washing, membranes were placed in a 20 mL solution of Streptavidin-Peroxidase polymer (Ultrasensitive, Sigma-Aldrich, S2438-250UG) diluted 1:20,000 in blocking buffer and incubated at room temperature (20 – 25 °C) for 30 minutes and 40 rpm rocking. Diluted Streptavidin-Peroxidase polymer solution was poured off and membranes were washed thrice as before. Bands were developed by incubating membranes in a 10 mL volume of metal

enhanced DAB (ThermoFisher Scientific, 34065) diluted 1:10 with stable peroxide buffer (ThermoFisher Scientific) until bands begin to appear, at room temperature (20 – 25 °C) and 40 rpm rocking. The reaction was stopped by dilution with ddH_2O, before pouring off diluted DAB solution.

2.2.2.2 Whole protein MALDI-TOF mass spectrometry

Whole protein MALDI—TOF mass spectrometry analysis was carried out on hFH, fetuin and RNase B and samples were either untreated or deglycosylated with either PNGase F and Endo H_f. 2.5 mU PNGase F or 5 mU Endo H_f was added to 48.4 pmol of protein sample. Each enzyme-treated sample, together with a sample consisting of 48.4 pmol untreated protein, was made up to 7.5 µL with 60% acetonitrile. Samples were then mixed and incubate at 37 °C for one hour. A 0.5 µL sample was taken and mixed 1:1 with 10 mg/mL sinapinic acid (in 50% v/v acetonitrile, 0.1% v/v TFA, in H_2O).
Proteins were analysed by MALDI-TOF mass spectrometry by mixing the protein solution 1:1 with 10 mg/mL sinapinic acid (Sigma-aldrich, D7927-1G) in 1:1 acetonitrile:HPLC grade H_2O (Fisher Scientific) and 0.1% (v/v) trifluoroacetic acid (Sigma-aldrich) before spotting 2 µL onto a MALDI-TOF plate and allowing to air dry. Samples were subjected to MALDI-TOF mass spectrometry analysis using an Ultraflex II TOF-TOF mass spectrometer (Bruker Daltonics) controlled by the FlexControl 2.0 software package. Spectra were obtained in linear positive ion mode using an acceleration voltage of 25 kV. All spectra were the result of signal averaging of 2,000 laser shots.

2.2.3 Human factor H glycan characterisation: Monosaccharide composition

2.2.3.1 SDS-PAGE and Lectin western blot analysis

For desialylation of the native protein, human plasma - purified FH was dialysed against 1x glycobuffer 1. α(2-3,6,8)-neuraminidase (New England Biolabs, P0720) was combined with dialysed plasma-derived pure FH in a ratio of 6.8 kU neuraminidase to 1 mg FH. This desialylation reaction was allowed to proceed for 24 hours at 37 °C.

For analysis, 500 ng (for SDS-PAGE) or 200 ng (for lectin-based western blot) were loaded onto either 3-8% tris-acetate gel (for SDS-PAGE, Thermofisher, NuPAGE, EA0375) or 4-12%

bis-tris gel (for lectin western blot). The 4-12% bis-tris gels were run, protein western transfer and stained as before. The molecular weight ladder used for 3-8% tris-acetate gel was HiMark pre-stained (Thermofisher, LC5699) and the gel was run at 150 V for 60 minutes and stained with Instant*Blue*.

2.2.3.2 Glycan release, purification and analysis by MALDI-TOF mass spectrometry

Glycans were released for purification and analysis by MALDI-TOF mass spectrometry as follows. 20 µg hFH was denatured with 1 µL denaturing buffer and heating to 100 °C for 10 minutes. After cooling the solution was made up to 25 µL with 2.5 µL glycobuffer 2, 2.5 µL 10% NP-40 and dH_2O before adding 550 U PNGase F, mixing and incubating for approximately 16 hours at 37 °C. hFH desialylated as above was also deglycosylated in the same way. Released glycans were purified using BlotGlyco spin columns (Sumitomo Bakelite, BS-45403Z). Crude sample containing released glycans was loaded onto a single BlotGlyco spin column, containing hydrazide functionalised resin, before adding 180 µL 2% (v/v) acetic acid in acetonitrile and incubating at 80 °C for one hour. Resin conjugated glycans were washed extensively with 200 µL each of 2 M guanidine HCL twice, ddH_2O twice and 1% (v/v) triethylamine in methanol twice. Residual hydrazides were blocked by incubating the glycan conjugated resin in 100 µL 10% (v/v) acetic anhydride in methanol at room temperature for 30 minutes. The glycan conjugated resin was washed again with 200 µL each of 10 mM HCl twice, methanol twice and DMSO twice. For hFH samples the sialic acids were “protected” by methyl esterification as follows, the glycan conjugated resin was incubated in 100 µL solution of 500 mM 3-methyl-1-(*p*-tolyl)triazene in DMSO at 60 °C for 1 hour. Followed by washing with 200 µL each of methanol twice and ddH_2O twice. Enzymatically desialylated hFH was not subjected to the methyl esterification procedure. Glycans were released by the addition to the glycan conjugated resin of 20 µL of the dipeptide labelling reagent N-α-aminooxyacetyltryptophanylarginine methyl ester (aoWR) and 180 µL 2% (v/v) acetic acid in acetonitrile followed by incubation at 60 °C for 1 hour. Released and aoWR tagged glycans were collected by addition of 50 µL 10 mM HCl and centrifugation of the spin columns (2000 g, 10 seconds). The glycan solution was diluted 1:20 with acetonitrile and applied to BlotGlyco clean-up columns to remove excess aoWR label reagent. Cleaned glycans were eluted from the BlotGlyco clean-up column by the addition of 50 µL ddH_2O and centrifugation (2000 g, 10 seconds).

Glycans were analysed by MALDI-TOF mass spectrometry by mixing the glycan solution 1:3 with 10 mg/mL super DHB (Sigma-aldrich, 50862-1G-F) in 3:7 acetonitrile and HPLC grade H_2O (Fisher Scientific) before spotting 2 µL onto a MALDI-TOF plate and allowing to air dry. Samples were subjected to MALDI-TOF mass spectrometry analysis using an Ultraflex II TOF-TOF mass spectrometer (Bruker Daltonics) controlled by the FlexControl 2.0 software package. Spectra were obtained in positive ion mode with reflectron using an acceleration voltage of 25 kV, a reflector voltage of 22.6 kV and a pulsed ion extraction of 160 ns. All spectra were the result of signal averaging of 10,000 laser shots. Peaks were picked using GlycoWorkBench 2 software.

2.2.4 Human factor H glycan characterisation: Role of sialic acids in the function of factor H

2.2.4.1 Purification of desialylated human factor H

Cation-exchange chromatography of desialylated human plasma FH was carried out on a Mono S 5/50 GL column (GE Healthcare Life Sciences, 17516801), conditions as before, and SDS-PAGE gels run. 1 µg of each fraction was loaded onto the gel, except fraction E6 where 10 µL of fraction was loaded, in duplicate under reducing and non-reducing conditions.

2.2.4.2 C3b and C3d binding assays

C3b and C3d binding assays were performed using a Biacore T200 instrument at 25 °C. 150 and 300 RUs C3b and 50 RUs C3d were attached by standard amine coupling to different flow cells of a C1 sensor-chip (GE Healthcare Life Sciences), the other available flow cell was left blank for background subtraction. 2-fold dilution series of each glycoform of FH (hFH, desialylated hFH and deglycosylated recombinant (r) FH) was performed in 10 mM HEPES, 150 mM NaCl, 0.05% (v/v) surfactant P20 (HBS-P^+ buffer, GE Healthcare Life Sciences). Glycoform dilutions were injected for 150 s at a flow rate of 30 µL/min followed by a dissociation time of 300 s. For each dilution series, a negative control sample containing PBS diluted in HBS-P^+ buffer was also injected. The sensor-chip was regenerated by three injections of 1 M NaCl for 30 s at a flow-rate of 30 µL/min. Data were analysed using the Biacore Evaluation software and a 1:1 steady state binding model.

2.2.4.3 SPR-based decay acceleration assay

Decay-accelerating activity was measured in a surface plasmon resonance (SPR)-based assay performed using a Biacore T200 instrument at 25 °C. 1000 RUs of C3b was attached to a single flow-cell of Biacore CM5-sensor chip (GE Healthcare Life Sciences) using standard amine coupling. A second flow-cell was left blank for background subtraction. C3bBb was assembled on the chip by a 180 s 10 µL/min injection of a solution containing 500 nM Factor B (CompTech, A135) and 50 nM Factor D (CompTech, A136). The formation of the C3 convertase complex C3bBb was monitored (~150 RUs), and then allowed to decay naturally, during a dissociation phase lasting 240 s, to observe intrinsic convertase decay. Glycoforms of FH, or PBS, were injected for 180 s at a flow-rate of 10 µL/min and a further dissociation phase was monitored. The sensor-chip was regenerated between cycles by a single 30 s injection of 0.1 M hFH followed by three 30 s injections of 1 M NaCl at a flow rate of 10 µL/min.

2.2.4.4 Cofactor assay

The activity of FH as a cofactor for C3b cleavage catalysed by FI was measured by detection of C3b 63 and 39 kDa α'-chain proteolytic products following SDS-PAGE. The non-cleaved β-chain served as an internal control. The 16-µl reaction mixture contained 1.7 µM C3b, 14 nM factor I (FI) (CompTech, A138), and a range of concentrations (80, 40, 20, 10 nM) of each glycoform of FH. The reaction mixture was incubated for 30 min (37°C), then stopped by addition of SDS-PAGE loading buffer. The mixture was boiled for 5 minutes and subjected to polyacrylamide gel electrophoresis, as before. Densitometry values were obtained using ImageJ software.

2.3 Results

2.3.1 Purification of human factor H from blood plasma

Because of its role as the principal soluble regulator of the complement system, FH is a target for microbial hijack (Zipfel *et al*., 2002). By anchoring FH on their surfaces, bacteria attempt to protect themselves from C3b deposition and amplification that would otherwise lead to clearance or cell killing. The N-terminal region of the choline-binding protein (PspCN), from the nasopharyngeal bacterium *Streptococcus pneumoniae*, binds FH with

picomolar affinity (Herbert *et al*., 2015). This high binding affinity can be exploited to capture FH in blood plasma.
PspCN, when attached to a supporting resin, was used to capture FH from citrated human blood plasma in the first of a two-step chromatographic purification process. (fig. 2.1 and 2.2). In the second “polishing” step, cation-exchange chromatography (fig. 2.3 and 2.4) was used to prepare homogenous material.

In the first purification step, approximately 400 mL diluted citrated human blood plasma with clotting factors removed (blood serum) was loaded onto a 7.5 mL PspCN-UltraLink iodoacetyl resin tricorn column. Protein was eluted and selected peak fractions were evaluated for FH content by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (fig. 2.1 and 2.2). This process was repeated 4 times until all 1.6 L diluted citrated human blood sera was subjected to the first step of the human FH purification process.

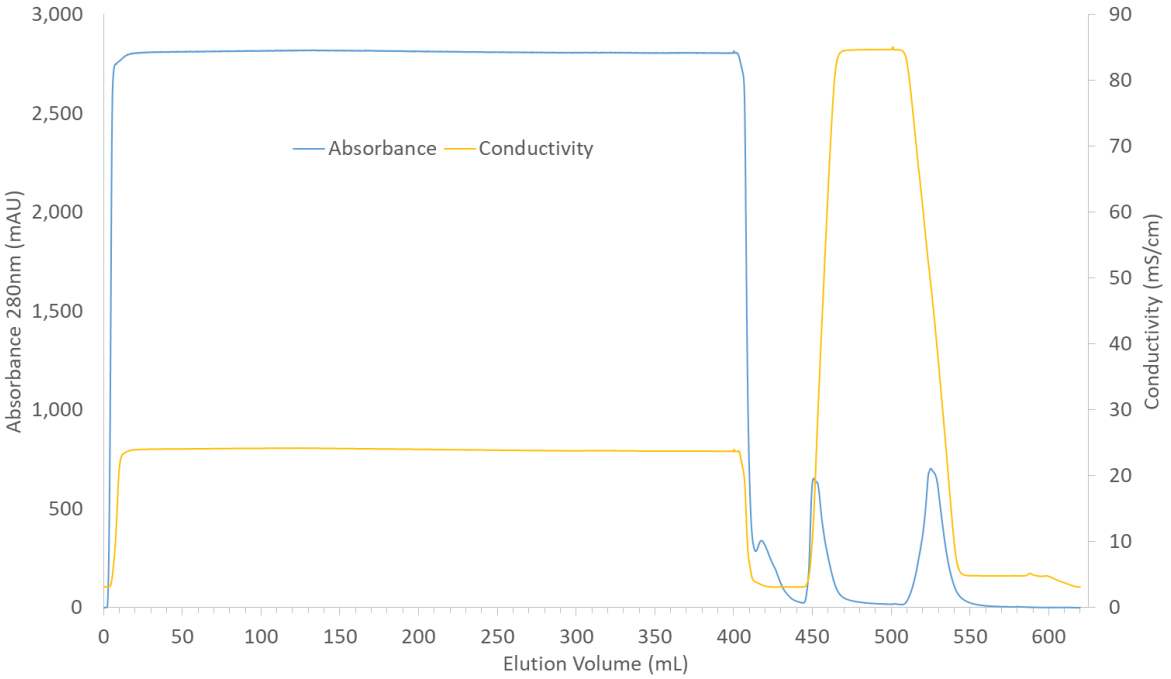


Figure 2.1: Chromatogram of the purification of FH from citrated human blood plasma by PspCN affinity chromatography. The peak just after 500 mL elution volume contains FH according to analysis using SDS-PAGE.

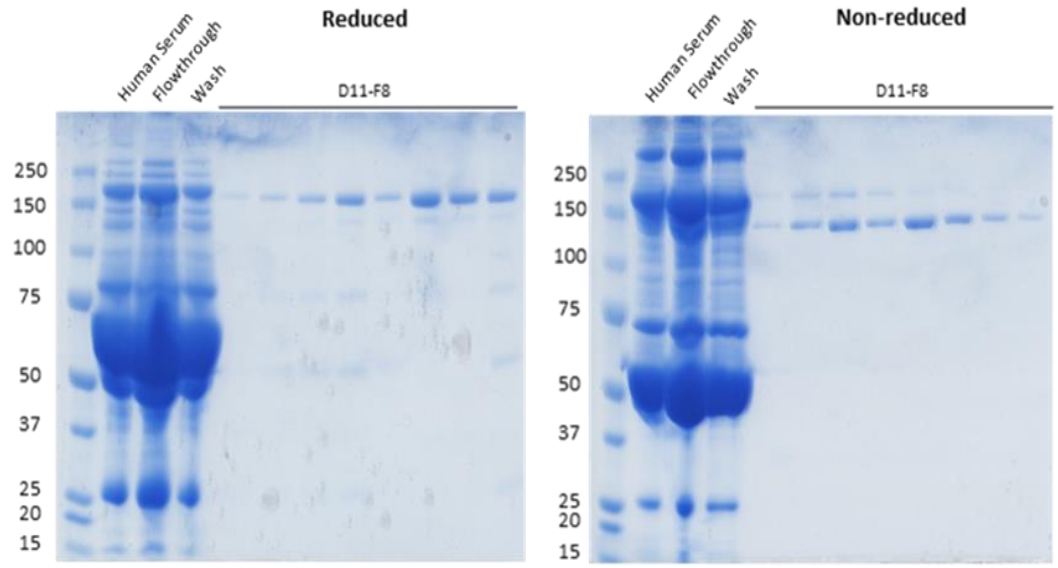


Figure 2.2: SDS-PAGE analysis of PspCN – affinity chromatography fractions corresponding to peaks. Flow-through and wash fractions were concentrated ten times prior to loading. Fractions were collected in 1.5 mL volumes. D11 – F8 corresponds to elution volume of 510 mL to 541 mL on chromatogram (fig. 2.1). Every third fraction from D11 – F8 was loaded onto the polyacrylamide gels. 2 µL human blood serum, flow-through, wash and collect fractions were added to 14 mL SDS-Buffer and all 16 µL was loaded onto polyacrylamide gels. The molecular weight of FH is 155 kDa and molecular weight markers measured in kDa.

The 40 disulphide bonds in FH cause the protein to have different electrophoretic mobilities by SDS-PAGE depending on whether its reduced or not. When not reduced, the disulphide bonds compact the polypeptide chain resulting in greater electrophoretic

mobility. When these bonds are reduced, the polypeptide chain is more open resulting in reduced electrophoretic mobility. This difference in mobility may be seen by comparing the two gels in Figure 2.2.

Based on inspection of the outcome of SDS-PAGE (fig 2.2), fractions E8 to F8 (523.2 – 541.3 mL), inclusive, were pooled together with a further 67 fractions from five repeated chromatographic runs in a 120 mL volume.
As can be seen from table 2.1, the protein concentration (as measured by Ultraviolet – visible light spectroscopy absorbance readings at 280 nm) is 0.446 mg/mL yielding 53.5 mg of protein. However, after dialysis of pooled fractions against low-salt buffer and filtration through 0.22µm filter the concentration dropped to 0.22 mg/mL whilst the volume remained approximately the same, yielding 26.4 mg of protein.
This difference could be due to human error in the measurement of either concentration or it could be that a small molecular weight contaminant that absorbed at 280nm was co-eluting with FH on the PspCN column and which was subsequently lost after dialysis.

400 mL diluted citrated human blood sera was loaded onto the PspCN column for each of the five repeated runs. Approximately the same volume was collected as flow-through and 45 mL collected as a wash fraction. Although the flow – through and wash fractions loaded onto the polyacrylamide gels (fig. 2.2) were ten-times concentrated, there is a significant amount of FH in both of these fractions.
This suggests that the column was overloaded which prevented complete (or near complete) depletion of FH from blood sera. Re-applying the flow-through and wash fractions to the PspCN column would have increased yields of FH.

Pooled fractions were dialysed against low – salt buffer before being subjected to a second purification step, cation – exchange chromatography (fig 2.3).

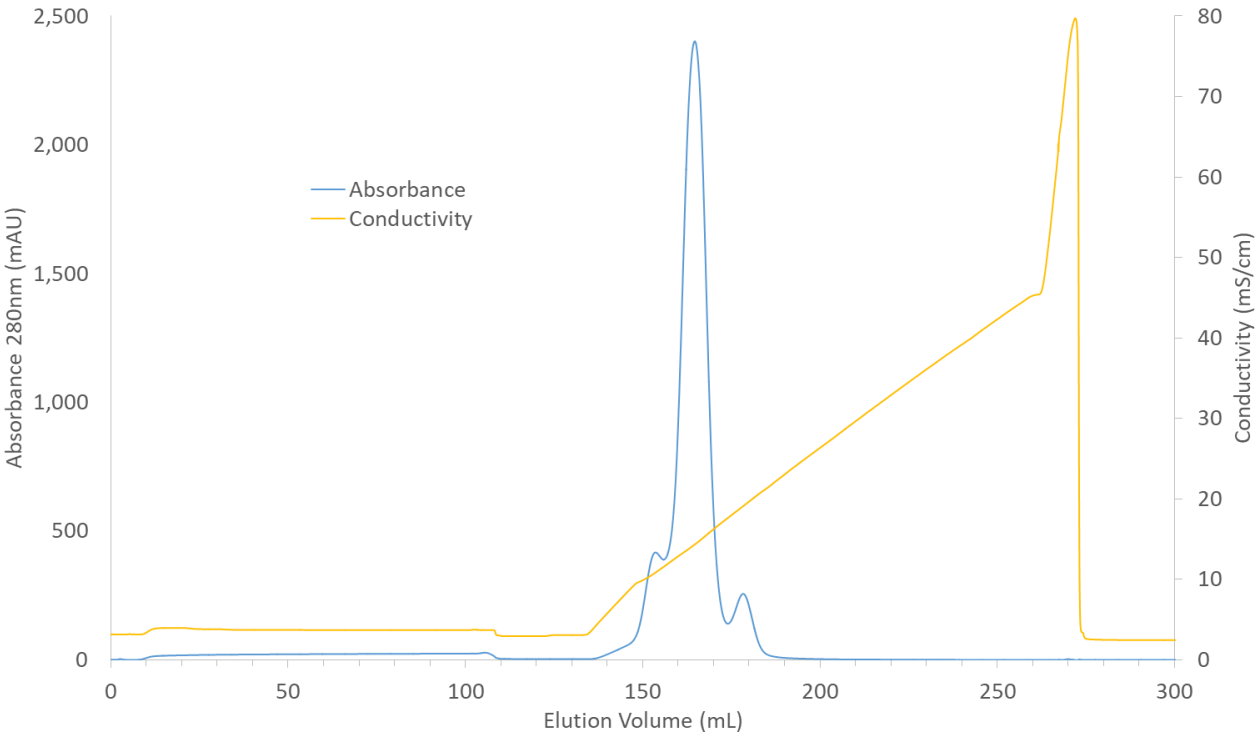


Figure 2.3: Chromatogram of the second and final step in the purification of FH from human blood plasma, obtained by cation exchange chromatography.

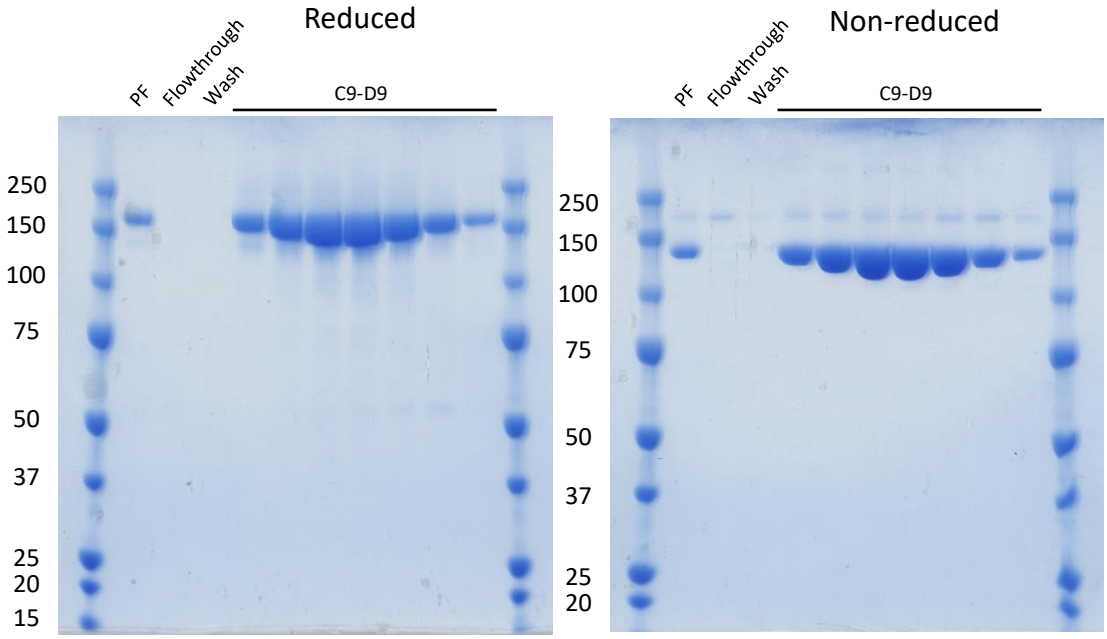


Figure 2.4: SDS-PAGE analysis of cation – exchange chromatography peak fractions. Fractions were collected in 1.5mL volumes. C9 – D9 corresponds to elution volume of 158.8 mL to 170.9 mL on chromatogram (fig. 2.3). Every second fraction from C11 – D9 was loaded onto the polyacrylamide gels. 2 µL PspCN chromatography pooled fraction (PF), flow-through, wash and collect fractions were added to 14 mL SDS-Buffer and all 16 µL was loaded onto polyacrylamide gels. The molecular weight of FH is 155 kDa and molecular weight markers measured in kDa.

After dialysis against cation-buffer A, all 120 mL PspCN purified protein was loaded onto a 5/100 tricorn column pre – packed with source 15S cation – exchange resin.

The chromatogram (fig. 2.4) shows protein eluting at 140 – 185 mL and at this elution volume there is a central peak (155 – 175 mL) flanked by two smaller peaks.
SDS-PAGE analysis of fractions C9 – D9 (158.8 mL – 170.9 mL elution volume) show that FH is contained within this large central peak. These thirteen fractions were pooled together and dialysed against phosphate – buffered saline and filtered through 0.2 µm filters. The UV-Vis absorbance readings for wavelengths 280, 260 and 320 nm were measured and protein concentrations predicted (see table 2.1).
26.4 mg PspCN – column purified protein was loaded onto the cation – exchange column which – after a purification run, dialysis and filtration – yielded of 21.7 mg of high purity, low DNA contamination (A260/280 nm ratio is approximately 0.6) FH.

			Absorbance			Concentration				Purity
Purification Step	**Extinction coefficient FH ($M^{-1}cm^{-1}$)**	**Molecular weight FH (kDa)**	**280nm**	**260nm**	**320nm**	**(µM)**	**(mg/mL)**	**Volume (mL)**	**Yield (mg)**	**Ratio A260/A280**
PspCN affinity chromatography	246,800	155	0.142	0.091	0.014	2.88	0.446	120	53.5	0.641
Dialyis and filtration			0.070	0.023	-0.044	1.42	0.220	120	26.4	0.330
Cation-exchange chromatography			0.587	0.356	0.006	11.89	1.840	12	22.1	0.606
Dialysis and filtration			0.575	0.317	0.009	11.65	1.810	12	21.7	0.551

Table 2.1: Table containing protein absorbance readings, concentrations, volumes, protein yields and purity at each stage in the purification of FH from human blood plasma.

The average circulatory concentration of FH in human blood sera is approximately 200 – 300 mg/L (Hakobyan *et al.,* 2008 and Sofat *et al*., 2013). After clotting factor removal but before dilution 1:1 in dilution buffer (see 2.1.1 materials and methods), human FH was purified from approximately 900 mL human blood sera.
In this sera there is expected to be 180 – 270 mg of FH. A yield of 21.7 mg FH (table 2.1) indicates that just one tenth of the available FH was purified from human blood sera.
This supports the finding that a significant amount of FH is eluting in the flow – through and wash fractions after the PspCN chromatography, as assessed by SDS-PAGE (fig. 2.2).
Yields of FH for PspCN – column purification could have been improved by reducing load volumes or increasing the capacity (resin bed volume) as well as reapplying flow – through and wash fractions, containing a significant amount of FH.
The cation – exchange chromatography purification step had the smallest loss of protein of the two purification steps, with 26.4 mg protein in a 120 mL volume applied to the column and 21.7 mg protein in a 12 mL volume after elution, dialysis against phosphate – buffered saline and filtration (table 2.1). There is no FH in the flow – through and wash fractions (fig. 2.4) indicating that the column was not overloaded.

Reducing the fractions sizes and/or reducing the steepness of the elution gradient could have increased the separation of the smaller peaks from the main central peak – potentially improving yields of FH slightly.

2.3.2 Human factor H glycan characterisation: Endoglycosidase sensitivity

As previously discussed in Chapter 1 section 1.5.1, endoglycosidases are a group of enzymes that cleave glycans attached to proteins within the polysaccharide chain. Due to their different substrate specificities the two enzymes, Endo H_f and PNGase F, can be used to broadly characterise the glycosylation-type of glycoproteins.

Fenaille *et al*. (2007*a*) showed that all of the glycans present on human FH are complex-type. If this is the case, human FH should be insensitive to glycan cleavage by the high-mannose and hybrid-type glycan-cleaving enzyme Endo H_f. In contrast, human FH should be sensitive to glycan cleavage by PNGase F which can cleave the same glycan-types as Endo H_f but, in addition, can also cleave complex-type glycans.

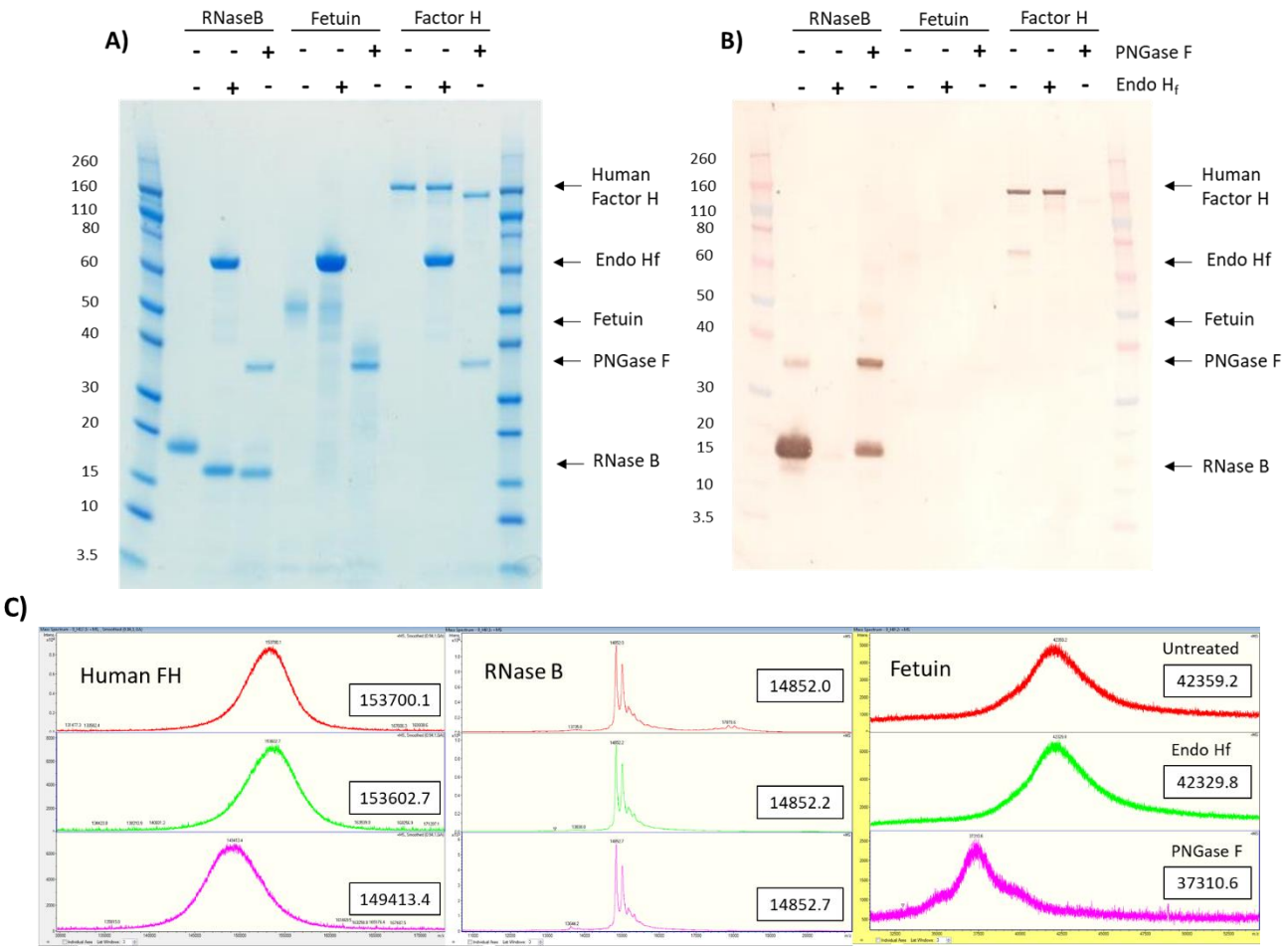


Figure 2.5: Different glycoproteins were treated with either PNGase F, Endo H_f or were untreated and analysed by A) SDS-PAGE, B) concanavalin A lectin western blot and C) whole protein MALDI-TOF mass spectrometry. Values in black boxes are peak m/z values. The molecular weight of glycosylated RNase B is 17 kDa, Fetuin 64 kDa and FH 155 kDa. Molecular weight markers measured in kDa.

Bovine pancreas ribonuclease B (RNase B) is a well-studied model mammalian glycoprotein with a single N-linked glycosylation site conjugated to a high-mannose glycan (Plummer and Hirs, 1964; Plummer *et al*., 1968 and, Liang *et al*., 1980). Calf serum fetuin is another well-studied mammalian glycoprotein that contains three N-linked, as well as three O-linked, glycans. In contrast to RNase B, the N-linked glycans conjugated to Fetuin are complex-type (Spiro and Bhoyroo, 1974; Baenziger and Fiete, 1979 and, Green *et al*, 1988). These proteins were used as standards in an experiment conducted to confirm the type of N-linked glycosylation present on human plasma-derived FH.

Analysis by molecular-weight shift on SDS-PAGE (fig 2.5 A) shows that human FH glycans can be removed by PNGase F, but not by Endo H_f. This seems to support the published findings that FH has complex-type glycans. However, the sensitivity limit of protein detection by Coomassie staining is about 30 ng per band. Therefore, it remains possible

that not all of the FH has been deglycosylated by PNGase F. MALDI-TOF mass spectrometry (MS) (fig. 2.5 C) is also consistent with glycan cleavage by PNGase F but cannot confirm complete deglycosylation due to the broadness of the mass peaks.

Concanavalin A (ConA)-based western blotting, however, is more sensitive than coomasie staining for detecting glycosylated proteins (compare the first RNase B lane in SDS-PAGE against the same lane in the ConA blot (fig. 2.5 A and B); a band at ~35kDa is visible in the ConA blot but not present in SDS-PAGE). The more sensitive ConA western blot thus further confirms that FH is exhaustively deglycosylated by PNGase F, but not by Endo H_f.

In contrast, RNase B appears not to be completely deglycosylated by PNGase F given that a faint band for glycosylated RNase B remains visible on SDS-PAGE, while a strong band at the same molecular weight is present in the lectin-based western blot. This could be because the glycosylation site on RNase B is partially obstructed, hindering cleavage by PNGase F that cleaves between Asn and the first GlcNAc. Endo H_f, which cleaves between the two GlcNAc residues in the chitobiose core, may be less susceptible to hindrance. However, in both cases RNase B was denatured prior to deglycosylation – it could be that RNase B was more thoroughly denatured in the Endo H_f reaction compared to the PNGase F reaction. This would account for the reduced deglycosylation efficiency in the PNGase F reaction. An alternative, but unlikely explanation is that, PNGase F displays a reduced activity towards glycans with an α(1,3)-fucose linked to the chitobiose core (Tretter *et al.,* 1991) and that these might be present in RNase B. Plummer and Hirs (1964) report the presence of fucose in the glycosylation of RNase C and D. However, Liang *et al*., (1980) and Prien *et al*, (2009) do not detect fucosylation of RNase B glycans.

In contrast to the SDS-PAGE and lectin western analysis, RNase B does not appear to be deglycosylated by either PNGase F or Endo H_f when assessed by whole protein MALDI-TOF MS (fig. 2.5 C).
The conditions used in this study for deglycosylation for MALDI-TOF MS were different to those used for SDS-PAGE and western blot analysis, however the same conditions for MALDI-TOF analysis were used for all three glycoproteins studied – and deglycosylation was shown to occur in at least the PNGase F treated FH and Fetuin samples.

Additionally, Fenaille *et al*., (2007*b*) were able to show, by MALDI-TOF MS, that the MW of RNase B shifts from 14.8 kDa (untreated) to 13.685 and 13.887 kDa when treated with PNGase F and Endo H, respectively.
Fenaille *et al*., (2007*b*) used 2', 4', 6' – trihydroxyacetophenone monohydrate (THAP) as a MALDI-TOF MS matrix, compared to the sinapinic acid used in this study. Perhaps the sinapinic acid favours ionisation of the glycosylated form of RNase B, however this is unlikely as the same effect is not see with Fetuin or FH.

Regardless, RNase B deglycosylation with PNGase F and Endo H_f must be repeated to include longer incubation times with both enzymes in preparations for western blot and MALDI-TOF analysis.

Fetuin has multiple N- and O- linked glycosylation sites and demonstrable glycan heterogeneity (Nwosu *et al*., 2011). This is supported in the current work by the smeared appearance of its bands in SDS -PAGE. Additionally, Fetuin migrates as though it had a lower molecular weight (MW) on SDS-PAGE than its actual MW (~50 kDa by SDS-PAGE vs its calculated MW of 64 kDa). Fetuin is a heavily sialylated glycoprotein (Spiro, 1960) with three sialic acid monomers per N-linked glycan (Baenziger and Fiete, 1978; Green *et al*, 1988) and at least some of the O-linked glycans terminating in one or more sialic acid monomers (Spiro and Bhoyroo, 1974; Windwarder and Altmann, 2014). The additional negative charge conferred by the extensive sialylation could enhance the electrophoretic mobility of fetuin. Fetuin N-linked glycans are insensitive to enzymatic removal by Endo H_f but are however, sensitive to removal by PNGase F, as assessed by molecular weight (MW) shift in SDS-PAGE and by whole protein MALDI-TOF MS analysis (fig. 2.5 A and C). This supports the literature findings that the N-linked glycans on Fetuin are complex-type.

Taken together, the insensitivity of FH glycans to cleavage by Endo H_f, but sensitivity to cleavage by PNGase F, indicate that FH glycans are complex-type mammalian glycans.

Finally, this study demonstrates that the endoglycosidase toolkit can be used to distinguish complex-type glycans from high-mannose glycans.

2.3.3 Human factor H glycan characterisation: Monosaccharide composition

Most mammalian complex-type glycans terminate in sialic acid and, indeed, Fenaille *et al*. (2007*a*) show that human plasma FH glycans are predominantly sialylated.
As established above, the current work set out to explore the role of sialic acids in the function of FH. First it was important to show that our methodology is able to confirm FH is predominantly sialylated and then show that these sialic acids have been removed under native, non-denaturing conditions.

Human plasma FH was treated, under native conditions (i.e. not under reducing conditions), with α(2-3,6,8)-neuraminidase, an enzyme able to cleave a range of sialic acid glycosidic linkages. The removal of sialic acid from native, fully folded human FH was assessed by molecular weight-shift on SDS-PAGE, the presence or absence of signals by lectin-based western blots and by MALDI-TOF MS of glycans after protein denaturation, enzymatic cleavage and purification.

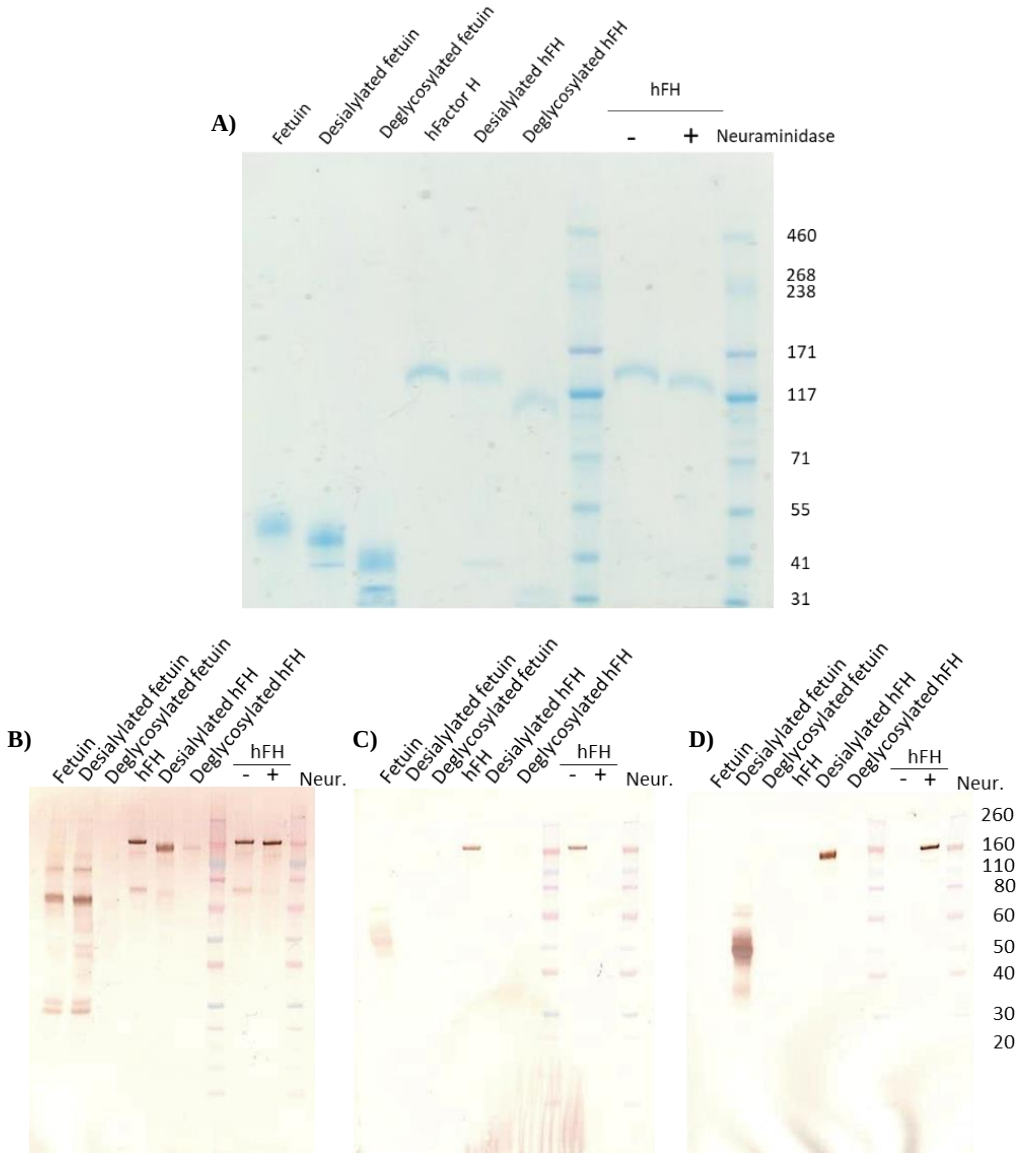


Figure 2.6: Human plasma-derived FH (hFH) treated with α(2-3,6,8)-neuraminidase (Neur.) under non-denaturing conditions and analysed for desialylation by A) SDS-PAGE, B), C) and D) lectin western blot. Concanavalin A (ConA) was used in B) to detect for glycans, *Sambucus nigra* lectin (SNA) was used in C) to detect for terminal sialic acid and *Erythrina cristagalli* lectin (ECL) was used in D) to detect for terminal galactose. Fetuin and human plasma standards were desialylated with Neur and deglycosylated with PNGase F under optimal, denatured conditions. MW markers measured in kDa.

Judged by SDS-PAGE (fig 3.6 A), a slight reduction in MW occurs when native, fully folded human FH is treated with the neuraminidase indicating sialic acid removal. In further support of sialic acid having been removed, no sialic acid was detected by *Sambucus nigra* lectin (SNA) following treatment of native FH with neuraminidase (fig. 2.6 C). Complementary to this, galactose is detected after native FH treatment with neuraminidase consistent with removal of sialic acid to expose galactose for binding by *Erythrina cristagalli* lectin (ECL) (fig. 2.6 D).

In further analysis, human FH and human FH treated with neuraminidase under native conditions was subsequently denatured and glycans enzymatically cleaved from FH by incubation with PNGase F. Free glycans were purified by “glycoblotting” to a hydrazide resin and released with a dipeptide tag at the reducing terminus (reviewed in chapter 1 section 1.5.2) and analysed by MALDI-TOF MS. This gave a more detailed picture of the glycans present on plasma FH (fig 2.7).

MALDI-TOF MS characterisation of glycans released from human plasma FH, using PNGase F, showed that, in agreement with the findings of Fenaille *et al*. (2007*a*), the predominant glycan present has a MW identical to what would be expected for a diantennary disialylated glycan, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$. The peak with the second highest intensity corresponds to the monosialylated diantennary glycan, $Sia_1Gal_2GlcNAc_2Man_3GlcNAc_2$. Some of the glycans on FH are fucosylated, as can be seen by peak masses of 2507.6 and 2813.9 (fig. 2.7 A) which correspond to fucosylated mono- and disialylated diantennary complex-type glycans, respectively. And peak masses of 2201.8 and 2567.6 (fig. 2.7 B) which correspond to fucosylated desialylated di- and triantennary complex-type glycans, respectively. Additionally, there are some peaks (e.g. peaks above m/z 3000 in figure 2.7 A) that cannot be assigned to known glycan structures.

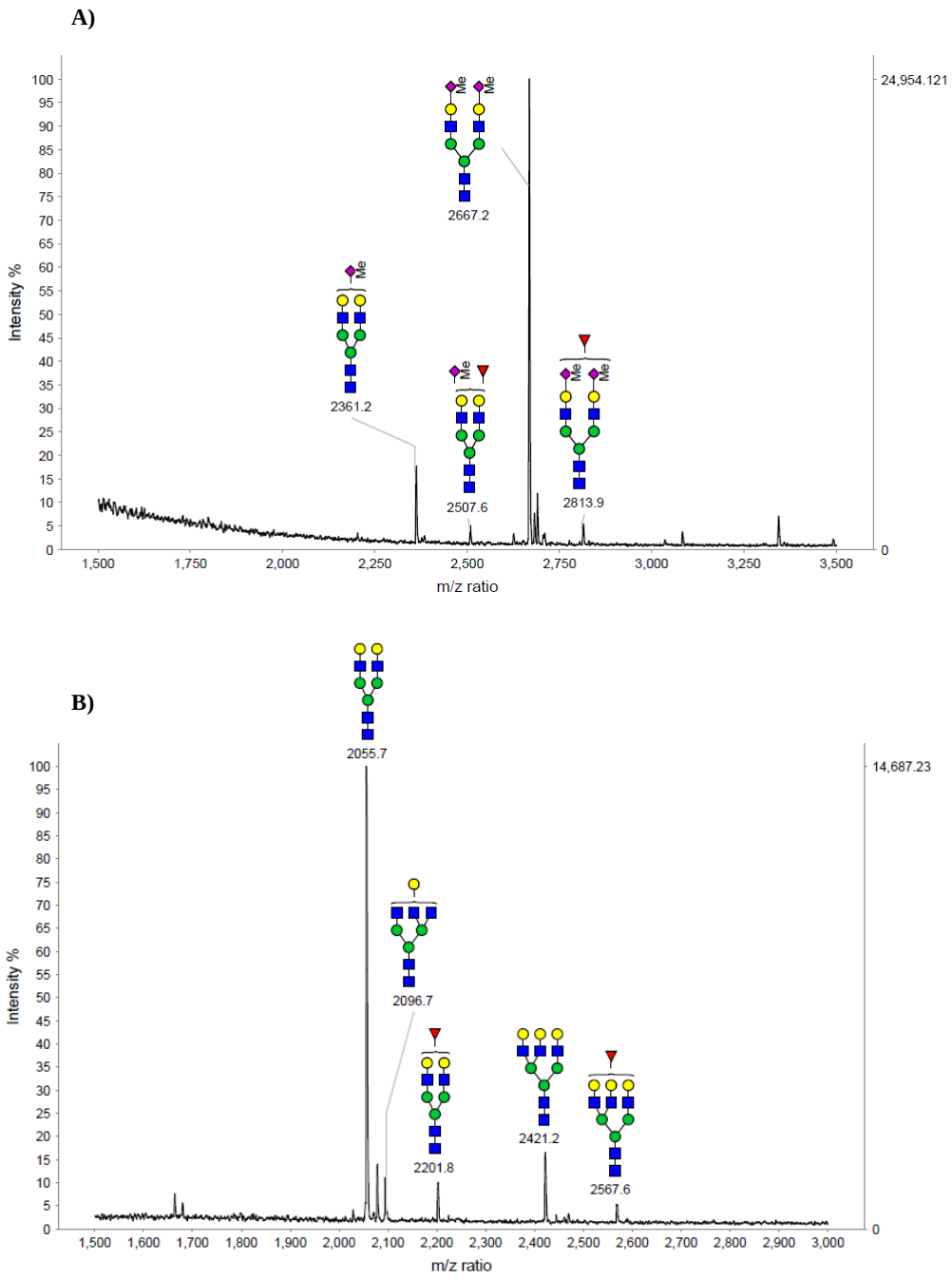


Figure 2.7: Glycans isolated and purified from **A)** human plasma FH and, **B)** human plasma FH after treatment with α2-3,6,8 neuraminidase under native, folded conditions.

Contrary to the findings of Fenaille *et al*. (2007*a*), no triantennary structures could be detected amongst the glycans cleaved from human plasma-derived FH (fig. 2.7 A).
But in contrast, desialylation of human plasma FH glycans yielded detectable quantities of triantennary glycans (fig 2.7 B) implying that triantennary species are present on human plasma-derived FH. It is possible that some of the unassigned peaks over m/z 3000 in figure 2.7 A are (mono-, di- and tri -) sialylated triantennary structures and that desialylation enriches these structures into a single homogenous non-sialylated triantennary peak in

figure 2.7 B. Alternatively, MALDI-TOF mass spectrometry was carried out in positive ion mode, the excessive negative charge conferred by sialic acids on trisialylated glycans may affect detection of these species in positive mode, desialylation of these species reduces the net negative charge facilitating detection in positive ion mode.
The lack of sialylated glycans assignable to any peaks in the mass spectrum shown in figure 2.7 B is consistent with near-complete prior desialylation of human plasma FH by α(2-3,6,8)-neuraminidase.

2.3.4 Production and purification of desialylated human plasma-derived factor H

In order to elucidate the role, if any, of sialic acids in the function of FH, desialylated plasma FH must be prepared in a highly purified form. This involves purifying it away from the contaminating α(2-3,6,8)-neuraminidase and any free sialic acids. To prepare pure desialylated FH, cation-exchange chromatography was performed (figs. 2.8 and 2.9). The ExPASy ProtParam tool predicted the pI of α(2-3,6,8)-neuraminidase to be 5.47. This is well below the neutral pH at which cation-exchange chromatography was carried out.
ExPASy ProtParam pI predicts the pI of FH as 6.12. At pH 7.0 FH the protein must have an overall net negative charge. However, FH is known to interact ionically with negatively charged sialic acids and sulphated glycosaminoglycans (GAGs) (Meri and Pangburn, 1990 and, Pangburn *et al*., 1991) at physiological pH. These sialic acid- and sulphated GAG-binding regions on FH are likely to be pockets of positive charge within an overall electronegative surface at physiological pH. This can be exploited for purification of FH by cation-exchange chromatography at pH 7 (Herbert *et al*., 2006; Schmidt *et al.,* 2008 and, Makou *et al*., 2012).

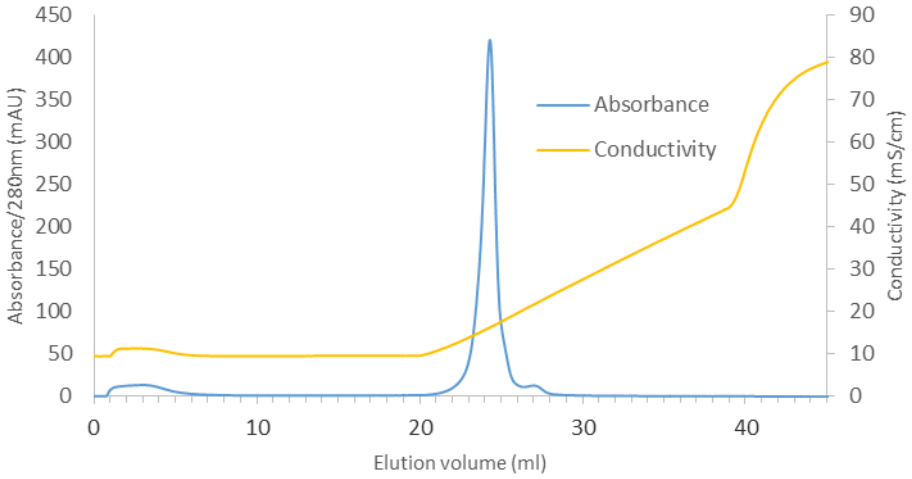


Figure 2.8: Purification of desialylated human plasma FH from α(2-3,6,8)-neuraminidase by cation exchange chromatography.

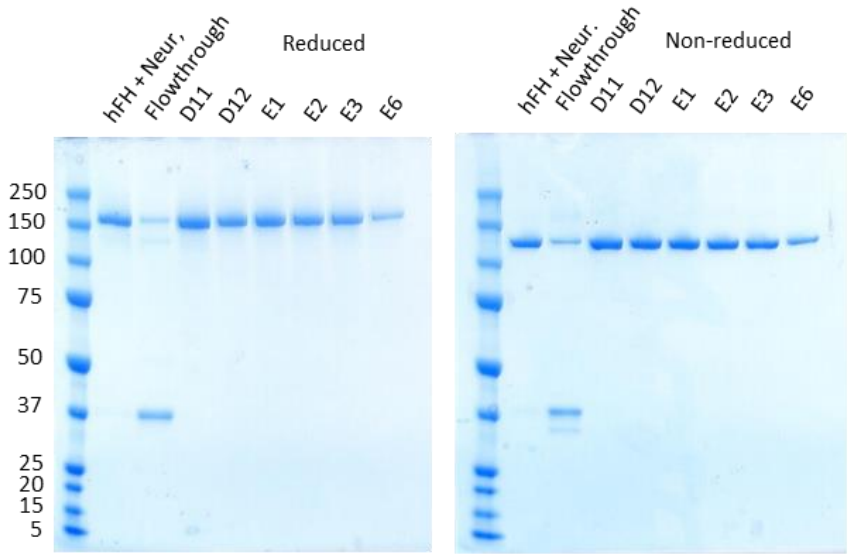


Figure 2.9: SDS-PAGE of peak fractions for the separation of desialylated human plasma-derived FH from α(2-3,6,8)-neuraminidase (Neur.) by cation-exchange chromatography on a Mono S 5/50 GL column. 1 µg of each fraction, except E6 where 10 µL was used, was made up to 16 µL in SDS loading buffer and all 6µL was loaded onto the 4-12% bis-tris polyacrylamide gel. 0.5 mL fractions were collected, D11 – E3 corresponds to an elution volume of 23.14 – 25.14 mL and E6 to 26.14 – 26.64 mL. The flow-through fraction was ten times concentrated in a pierce 10 kDa MW cut-off PES protein concentrator. The MW of α(2-3,6,8)-neuraminidase is 43 kDa. MW markers measured in kDa.

There is no detectable α(2-3,6,8)-neuraminidase in peak fractions D11-E3 (23.14 – 25.14 mL) or the peak shoulder fraction E6 (23.14 – 25.14 mL and 26.14 – 26.64 mL elution volume, respectively) (fig 2.9).

Approximately 3 mg of human FH desialylated, under native conditions, with α(2-3,6,8)-neuraminidase was loaded onto the cation-exchange column. 0.5 mL fractions D11 – E2 (23.14 – 24.64 mL) were shown, by SDS-PAGE analysis (fig. 2.9), to contain high purity FH and were measured, by absorbance at 280 nm wavelength, to have the highest protein concentration. These fractions were pooled and dialysed against phosphate buffered saline to yield approximately 1 milligram of purified desialylated human plasma FH.

2.3.5 Human factor H glycan characterisation: Role of sialic acids in the function of factor H

The effect of sialic acids on the function of FH was assessed, *in vitro*. First, surface plasmon resonance (SPR)-based assays were used to measure the affinity of FH for the protein ligands C3b and C3d (C3d is a proteolytic cleavage product of C3b corresponding to its thioester domain), and to assess the decay-acceleration activity of FH on the C3-convertase complex, C3bBb. Second, the ability of FH to recruit Factor I (FI) for cleavage of the α'-chain of C3b was assessed using an SDS-PAGE-based cofactor assay.

Deglycosylated rFH (made in *Pichia pastoris*) was used as a control in these experiments. This protein has been well characterised in regard to its C3b and C3d binding, its decay acceleration and its cofactor activity (Schmidt *et al*., 2011 and, Kerr *et al*., 2017).

The C3b/d-binding affinity measurements were conducted by immobilising, by amine coupling, 150 and 300 RUs C3b and 50 RUs C3d to three flow cells of a C1 sensor-chip. A fourth flow cell was used as a blank. Binding of each concentration of a serial dilution series (1 in 2) of three glycoforms of FH (sialylated (i.e. native, untreated) and enzymatically desialylated human plasma-derived FH (see section 2.2.4) and, enzymatically deglycosylated human rFH expressed in *Pichia pastoris*) was measured, in duplicate.

The interaction between the three different glycoforms of FH and C3b was investigated (fig. 2.10). Duplicates of each glycoform at each concentration were performed, but the second of each duplicate was omitted from the sensograms for clarity. The sensorgrams are reference subtracted and blank subtracted.

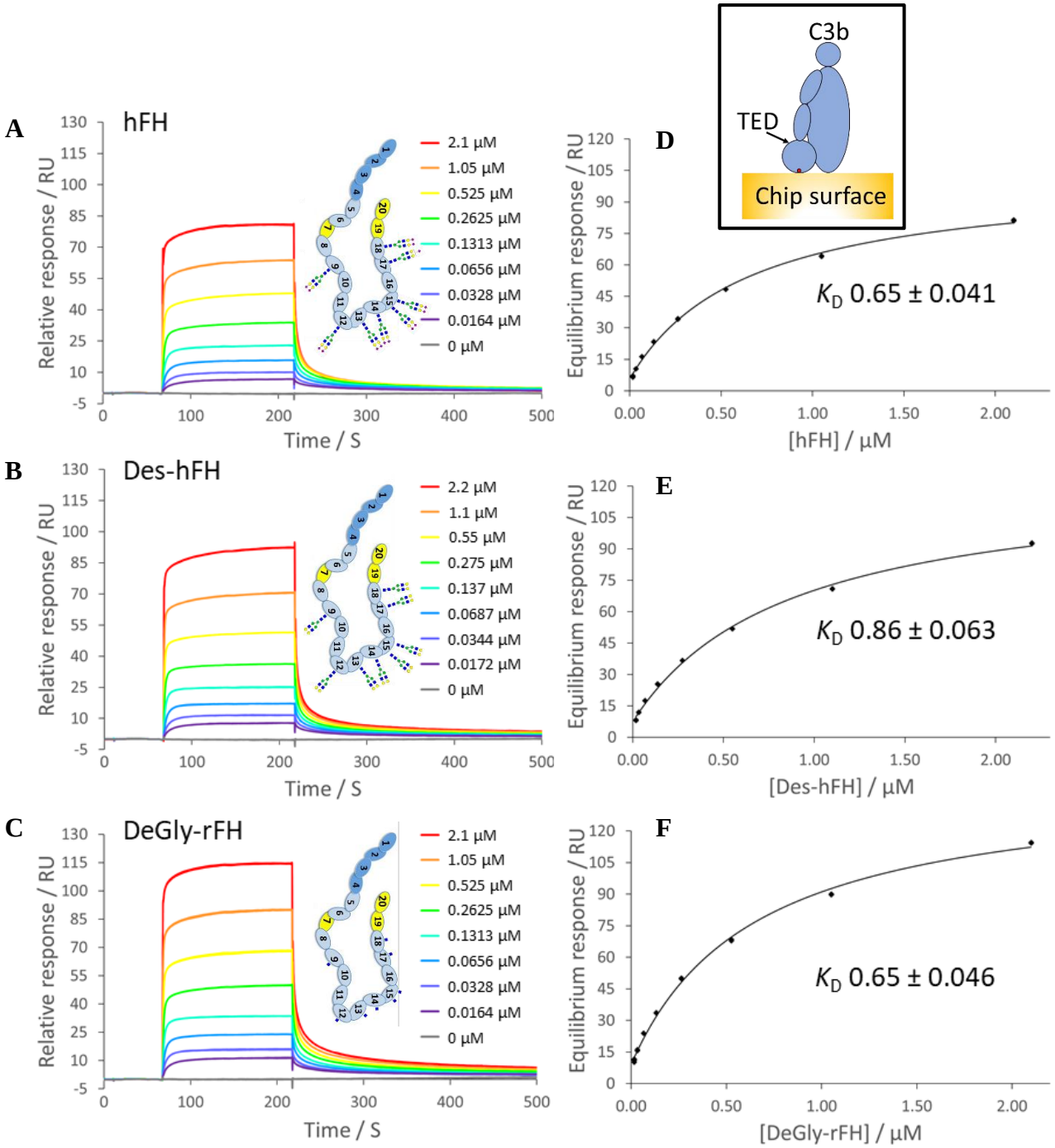


Figure 2.10: C3b binding assays. C3b was immobilised on a Biacore sensor chip surface (inset). Sensograms (A), B) and C)) and affinity curves (D), E) and F)) of plasma FH (A) and D)), desialylated plasma FH (B) and E)) and recombinant deglycosylated FH variant VYE (C) and F)) binding to 300 RUs of C3b. Sensograms were generated from duplicate injections of eight concentrations of each glycoform of FH and a blank for 150 s at 30 µL/min followed by dissociation for 360 s. Sensograms were reference and blank subtracted and 1:1 steady state kinetic model was fitted to sensograms to generate the affinity curves. Sia-FH is human plasma FH, DES-FH is desialylated human plasma FH and VYE-rFH is deglycosylated recombinant FH variant VYE.

A second C3b binding experiment, to 150 RUs C3b, was also conducted (sensograms and affinity data not shown) and K_D values are in table 2.2.

Next the interaction between the three different glycoforms of FH and C3d, a breakdown product of C3b that remains surface bound after degradation of C3b, was investigated (fig. 2.11). Duplicates of each glycoform at each concentration were performed, but the second of each duplicate was omitted from the sensorgrams for clarity. The sensorgrams are reference subtracted and blank subtracted.

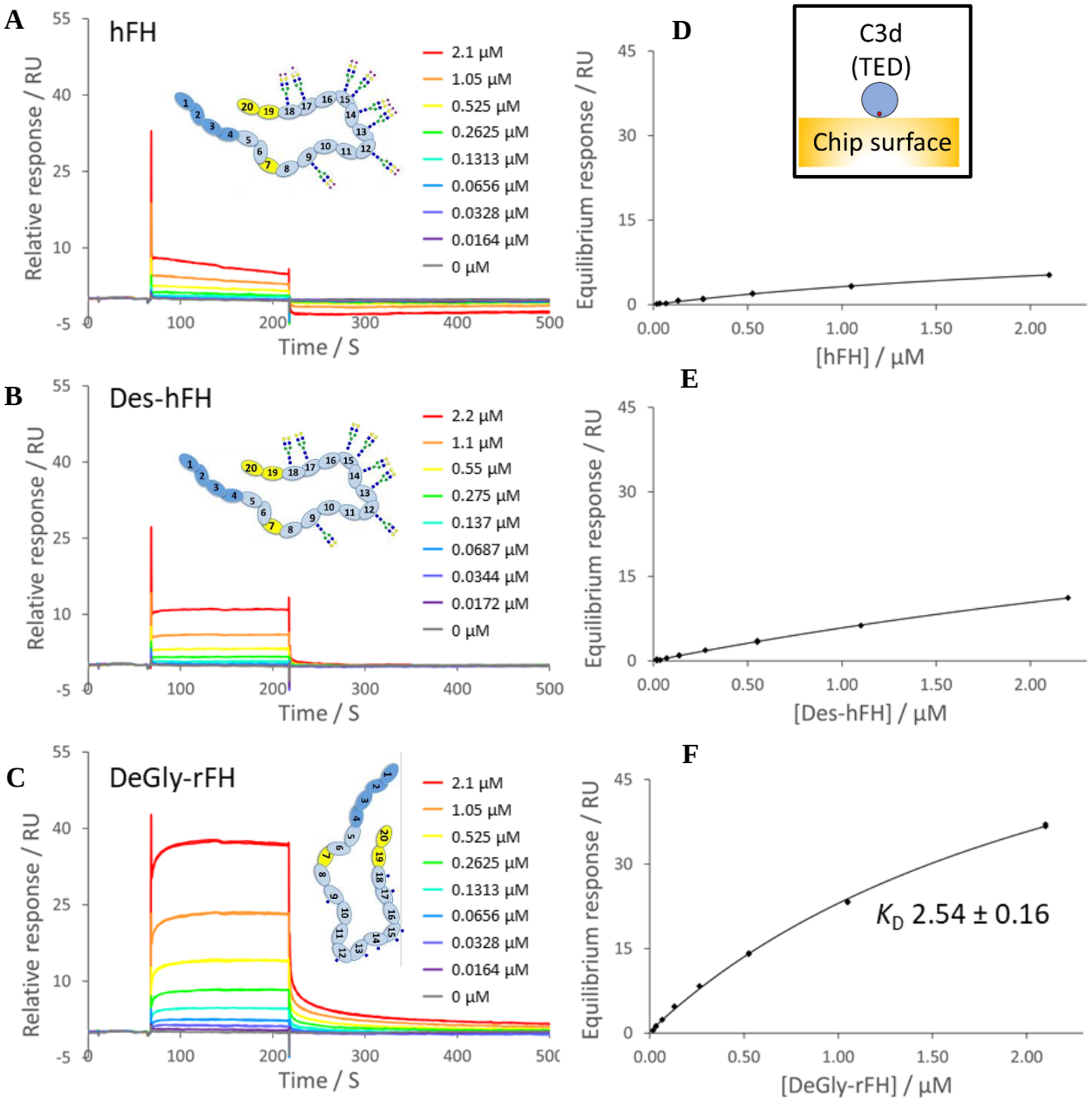


Figure 2.11: C3d binding assays. C3d was immobilised on a Biacore sensor chip surface (inset). Sensorgrams (A), B) and C)) and affinity curves (D), E) and F)) of plasma FH (A) and D)), desialylated plasma FH (B) and E)) and recombinant deglycosylated FH variant VYE (C) and F)) binding to 50 RUs of C3d. Sensograms were generated from duplicate injections of eight concentrations of each glycoform of FH and a blank 150 s at 30 µL/min followed by dissociation for 360 s. Sensograms were reference and blank subtracted and 1:1 steady state kinetic model was fitted to sensorgrams to generate the affinity curves. Sia-FH is human plasma FH, DES-FH is desialylated human plasma FH and VYE-rFH is deglycosylated recombinant FH variant VYE.

Duplicate concentrations of FH glycoforms were assessed for binding to C3b (at 150 RUs and 300 RUs C3b) and C3d (50 RUs) to obtain steady state sensorgrams. The steady state sensorgrams were used to calculate the affinity of the three different glycoforms of FH for C3b and C3d. The affinities were calculated by fitting the steady-state sensorgrams to a simple 1:1 interaction model. The calculated affinities (dissociation constant, K_D), theoretical maximum response (R_{max}) and chi-squared (χ^2) values are displayed in table 2.2.

	150RUs C3b					300RUs C3b					50RUs C3d				
Sample	K_D (μM)	SE K_D	R_{max} (RU)	χ^2 (RU^2)	Offset (RU)	K_D (μM)	SE K_D	R_{max} (RU)	χ^2 (RU^2)	Offset (RU)	K_D (μM)	SE K_D	R_{max} (RU)	χ^2 (RU^2)	Offset (RU)
Plasma FH	0.67	0.035	51.43	0.309	2.34	0.65	0.041	97.31	1.63	5.71					
Desialyl-FH	0.97	0.057	66.67	0.461	3.02	0.86	0.063	117.00	2.51	7.36					
Deglycolyl-rFH	0.72	0.048	76.46	1.040	3.70	0.65	0.046	135.00	4.08	9.27	2.54	0.16	80.46	0.144	0.27

Table 2.2: Summary of the parameters fitted for the SPR-based binding studies of three different glycoforms of FH to the ligands C3b and C3d. Binding to 150 RUs and 300 RUs of C3b and, 50 RUs of C3d was measured. The three glycoforms of FH are plasma derived (sialylated) human FH, enzymatically desialylated plasma derived human FH and enzymatically deglycosylated recombinant human FH. SE = standard error.

No parameters could be fitted for the very weak binding of plasma-derived FH and desialylated FH for C3d. Sialylated FH and deglycosylated rFH both have similar K_D values of about 0.7 μM for binding to C3b, whilst desialylated FH has a slightly higher K_D value of about 1 μM (Table 2.2). The slightly weaker affinity of desialylated FH, compared to sialylated FH, suggests that the negatively charged sialic acids on FH may play a small role in facilitating the interaction of FH with C3b, but these differences are small and it is hard to draw many conclusions regarding the relative affinities of these molecules. The similarity in K_D values for human plasma-derived (sialylated) and deglycosylated human rFH is surprising given that previous comparison of these two glycoforms in the literature (Schmidt *et al*., 2011 and, Kerr *et al*., 2017) consistently shows deglycosylated rFH having a lower K_D compared to human plasma-derived FH. For example, Schmidt *et al*. (2011) measured a K_D of human plasma-derived FH and deglycosylated rFH for C3b binding to be 2.9 and 1.4 μM, respectively. Similarly, Kerr *et al*. (2017) K_D values for C3b binding by human plasma-derived FH were measured to be 0.33 μM and two isoforms of deglycosylated rFH (I62 and V62) were measured to be 0.21 and 0.27 μM, respectively.

The presence in the FH N-glycans of sialic acids could influence interactions between FH and C3b in two ways. The sialic acids might interact directly with C3b, possibly by ionic interactions. Another possibility is that electrostatic repulsion between sialic acids (and other negatively charged sites on the molecule), or favourable intramolecular interactions between sialic acids and sialic acid-binding sites such as found in CCP 7 and CCP 20, limit

the conformational freedom of FH, potentially holding the molecular in a conformation primed for C3b-binding.

If the latter possibility were the case then it would be expected that removal of the entire glycan chain (including the sialic acids) from FH would have a similar effect to removing just the sialic acids – i.e. an improvement in affinity of FH for C3b. But, in the present study no marked improvement in C3b-binding affinity was observed.

Deglycosylated rFH and desialylated human plasma-derived FH both have higher R_{max} values than (sialylated) human plasma-derived FH (Table 2.2). The relative difference in R_{max} values of these glycoforms of FH is consistent across the two loadings of C3b. In addition, the χ^2 values are less than 10% of R_{max} indicating an acceptable fit of the data to a 1:1 steady state kinetics.

Neither sialylated or desialylated plasma FH bind C3d strongly enough to give a reliable K_D. Conversely, deglycosylated recombinant FH does bind C3d relatively well, with a R_{max} of 80.46 and a calculated K_D of 2.5 μM. This is in good agreement with findings in the literature, for example Kerr *et al*. (2017) also reported binding of human plasma-derived FH to C3d to be too weak to measure, and that two isoforms of deglycosylated rFH (I62 and V62) bind C3d appreciably (K_D values of 1.8 and 2.1 μM, respectively).

Fitting of the sensograms curves allows estimation of R_{max} values. This is a measure of the maximum amount of FH resident on the chip surface per unit area when all binding sites are occupied – approximately 1 RU corresponds to 1 pg mm^{-2} (Stenberg *et al*, 1991). In table 2.2 it can be seen that sialylated human plasma FH has the lowest R_{max} value, deglycosylated recombinant FH has the highest and desialylated human plasma FH has an intermediate R_{max} and that this relationship is consistent across the two loadings of C3b (150 and 300 RUs). The predicted value of R_{max} is dependent on multiple considerations: i) the number of ligand molecules (C3b/d) per unit area immobilised on the chip surface, ii) the MW of the analyte (FH) and the iii) stoichiometry of the complex. Of these three considerations, the amount of C3b/d loaded is consistent for all glycoforms and the different MWs of the each glycoform of FH (the analyte) was taken into account when fitting the 1:1 steady state kinetic model. Therefore, the differences in R_{max} values indicate that these differences are due to stoichiometry and show that more desialylated human

plasma FH is binding to the C3b-decorated chip surface, compared to sialylated human plasma FH, despite the sialylated material having a higher affinity. In addition, more deglycosylated rFH is binding to the chip, then either sialylated or desialylated human plasma FH.

As discussed in Chapter 1 section 1.2.2, FH is capable of binding to two primary sites on C3b (Schmidt *et al*., 2008): i) the N-terminal CCPs 1-4 bind the length of C3b and ii) the C-terminal CCP 20 binds in the TED domain of C3b (which becomes C3d after proteolytic degradation of C3b).

Additionally, studies where FH was assessed for binding to C3b and C3d, by SPR, in the presence and absence of PspCN showed a marked increase in affinity (K_D) and R_{max} when FH was in the presence of PspCN (Herbert *et al*., 2015 and Kerr *et al*., 2017). This suggests PspCN activates FH facilitating tighter binding and binding to a greater proportion of C3b molecules – possibly by permitting initial engagement with C3b at either of the two C3b primary binding sites rather than at just one site (followed by a further binding event at the other binding site once FH is engaged with C3b).

C3b is presumed to be in different orientations on the chip surface because the amine which tethers a particular C3b molecule to the chip surface is random.

Therefore, the two primary binding sites on C3b will be differentially exposed. It is conceivable that a glycoform of FH that is able to bind to either or both primary binding sites might find more molecules of C3b it can interact with compared to a glycoform that can only interact with one site.

The R_{max} values indicate that deglycosylated rFH is capable of forming the most FH:C3b complexes and that desialylated FH could form more FH:C3b complexes than sialylated plasma-derived FH.

Removal of the glycans from FH may liberate the conformational freedom of the molecule, allowing for independent interaction at both primary C3b interaction sites. In addition, removal of sialic acids may confer a similar, but less pronounced, increase in conformational freedom, perhaps by reducing electrostatic repulsion between sialic acids or neighbouring glycans.

However, considering FH binds to both C3b and the surface to which C3b is attached it is not known what electrostatic effect sialic acid removal would have on the interaction of FH with C3b on a physiological surface.

Next, the ability of the three different glycoforms of FH to accelerate the decay of the bimolecular C3-convertase complex, C3bBb, was investigated using an SPR-based decay acceleration activity (DAA) assay. In these experiments, 1000 RUs of C3b was immobilised on the surface of a CM5 sensor-chip by standard amine coupling. Then the C3bBb, C3-convertase, complex was generated by the addition of complement factors B (FB) and D (FD). Formation of the C3-convertase complex was monitored (~ 150 RUs) and then allowed to decay naturally (in the absence of FB and FD) for a short time to observe intrinsic convertase decay, prior to addition, in duplicate, of the different glycoforms of FH to be tested. Plotting several such experiments on the same graph allowed a direct comparison of the dissociation rate of the C3-convertase complex on its own and in the presence of the different glycoforms of FH (fig. 2.12). Each glycoform DAA assay was carried out in duplicate. However, duplicate sensorgrams traces were omitted for clarity.

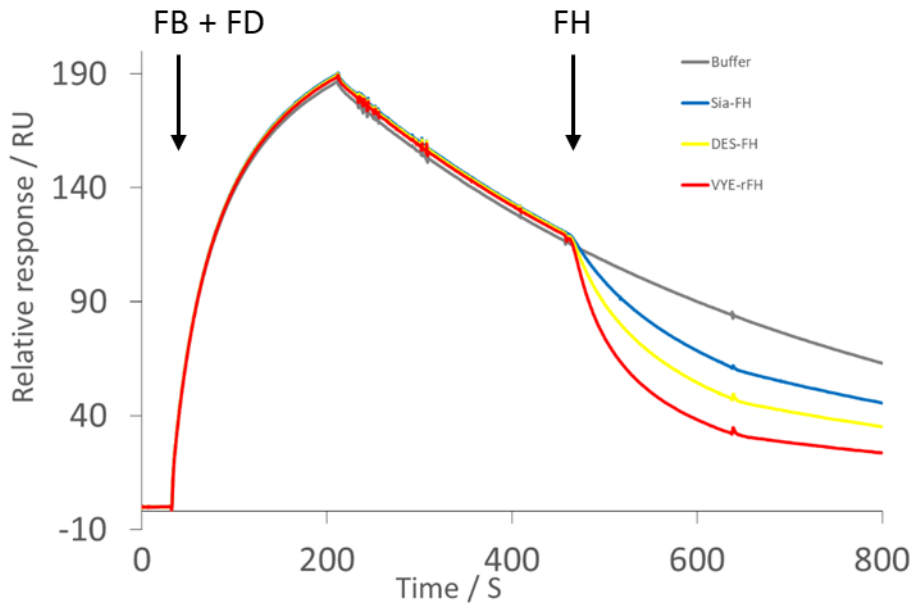


Figure 2.12: Sensorgram showing the results of a decay acceleration activity (DAA) assay of various glycoforms of FH on the C3-convertase complex, C3bBb. Approximately 1000 RUs C3b was amine coupled to the surface of a CM5 sensor-chip, C3bBb was assembled by an injection of 500 nM FB 50 nM FD lasting 180 s. After formation of ~150 RUs C3bB, the complex was allowed to naturally decay for 240 s followed by the addition of three glycoforms of FH, in duplicate, at 10 µL/min for 180 s. Duplicate traces have been omitted for clarity. Sia-FH is human plasma FH, DES-FH is desialylated human plasma FH and VYE-rFH is deglycosylated recombinant FH variant VYE.

C3b was amine coupled to the surface of a CM5 sensor-chip prior to the addition of FB and FD (marked on fig 2.12). A rising response occurs over the range of ~20 s – 200 s (fig. 2.12), which is consistent with the binding of Factor B to surface immobilised C3b, and the subsequent cleavage of Factor B by Factor D to form the C3-convertase complex, C3bBb. An

exponential decline in response units between ~200 s – 450 s represents the natural dissociation of the C3bBb complex. At approximately 460 s (marked on fig 2.12), the FH samples (20 nM) were injected, resulting in an accelerated decline in RU. This corresponds to the more rapid decay of the complex catalysed by the presence of FH – its DAA.

	Sample	Half-life (s)	Mean half-life (s)	SD	Change in half-life (approx.)
Pre-FH	Sia-FH 1	382.7	391.2	12.0	
	Sia-FH 2	399.6			
	DES-FH 1	381.8	387.7	8.3	
	DES-FH 2	393.6			
	VYE-rFH 1	378.8	383.1	6.1	
	VYE-rFH 2	387.4			
	Buffer 1	371.7	381.8	14.2	
	Buffer 2	391.8			
Post-FH	Sia-FH 1	246.5	258.9	17.5	
	Sia-FH 2	271.2			x1.5
	DES-FH 1	196.5	204.7	11.5	
	DES-FH 2	212.8			x2.0
	VYE-rFH 1	150.7	156.7	8.4	
	VYE-rFH 2	162.6			x2.5
	Buffer 1	387.4	405.6	25.7	
	Buffer 2	423.8			x1 .0

Table 2.3: Calculated half-life (s) for decay of the C3bBb complex before and after the addition of FH. The change in half-life was calculated by dividing the mean half-life after the addition of FH by the mean half-life before the addition of FH.

Duplicates of each glycoform were assessed for DAA of C3bBb and decay half-life of the C3bBb complex was calculated before and after the addition of each glycoform. Averages of each duplicate were taken to find the mean half-life for each glycoform (Table 2.3). Changes in half-life were calculated for each sample relative to the initial natural decay of the C3bBb complex. Deglycosylated rFH has the highest change with an approximately 2.5-fold reduction in half-life after the addition of deglycosylated rFH. Desialylated human plasma-derived FH has the next highest change, with an approximately 2-fold reduction in half-life after the addition of desialylated FH. Human plasma-derived (sialylated) FH has the lowest change in half-life, a half-life decay 1.5-fold greater than C3bBb decaying naturally. The data suggest that glycans and sialic acid may have a cumulative effect on reducing the DAA of FH on the C3-convertase complex C3bBb. Although duplicate samples were in good agreement for all glycoforms analysed, inferring a trend from just 6 data points is difficult. Therefore, the experiments need to be repeated to validate the findings.

Finally, the role of sialic acids in the ability of FH to act as a cofactor for Factor I (FI)-mediated proteolysis of C3b was investigated. Four different concentrations (80, 40, 20 and 10 nM) of each glycoform of FH were incubated with C3b (1.9 μM) and FI (1.6 nM), and the extent of cleavage of the α'-chain of C3b was assessed by SDS-PAGE (fig. 2.12 A). Each reaction was carried out as independent duplicates. Image J software densitometry function was used to measure density of each band. The β-chain was used as an internal standard against which band intensities for the α'-chain, 63 kDa and 39 kDa fragments were normalised against. The mean of normalised duplicate band intensities were plotted against concentration of FH (fig. 2.12 B C and D).

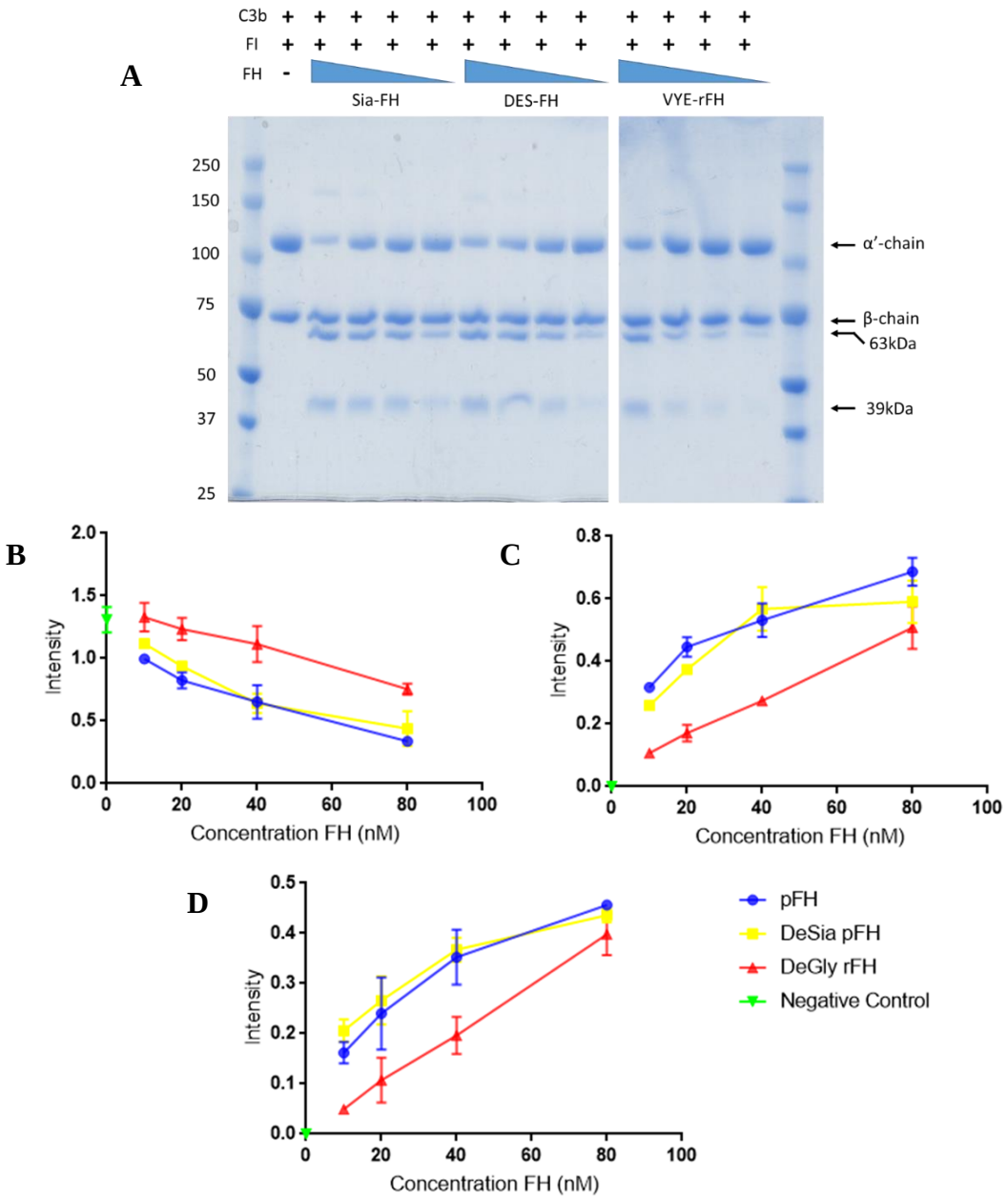


Figure 2.13: A) SDS-PAGE of the cofactor assay of three different glycoforms of FH incubated with C3b and Factor I. B) Band intensity of the α'-chain, C) band intensity of the 63 kDa chain, D) band intensity of the 39 kDa chain. Band intensities were measured using Image J software and normalised against the β-chain in the same lane as the measured band. Each sample was run in duplicate (duplicates not shown in A)). Error bars are standard deviation. Sia-FH is human plasma FH, DES-FH is desialylated human plasma FH, VYE-rFH is deglycosylated recombinant FH variant VYE and the negative control is C3b and FI incubated in the absence of FH. MW markers measured in kDa.

Inspection of the SDS-PAGE gels (fig 2.13 A) shows that, as expected, in the presence of FH, FI cleaves the α'-chain of C3b to form two products of molecular weight 63 kDa and 39 kDa whilst leaving the β-chain of C3b intact. In the absence of FH the α'-chain is not cleaved by FI. Densitometry performed on the gels (fig 2.13 A, B, C and D) indicates that sialylated (i.e. plasma-derived FH) and enzymatically desialylated plasma FH appear to facilitate the

cleavage of the α'-chain to similar extents, while less α'-chain cleavage was obtained when using equivalent concentrations of deglycosylated recombinant FH.

These findings contrast with the C3b/d-binding and DAA assays. They suggest the absence of sialic acids makes no difference to the cofactor activity of FH. These data further suggest that, although not essential for cofactor activity, the presence of the majority of the glycan chain enhances the cofactor activity of FH.

2.4 Discussion

Results in this chapter confirmed findings in the literature (Fenaille *et al*, 2007*a*), that the N-linked glycosylation on plasma FH is moderately heterogeneous – comprising mainly diantennary mono- and disialylated species as well as triantennary species and, in addition to being non-fucosylated, these species can also be fucosylated. Nonetheless, all the glycan species detected are complex-type and the majority are diantennary disialylated glycans.

However, further work needs to be done to understand the glycosylation of FH. There is currently no information on the types of glycosidic linkage between monosaccharides. For example, in humans, sialic acid can exist in an α(2,3)- and an α(2,6)- linkage with galactose. It is not known whether either or both of these sialic acid linkages are present on FH and what effect having either or both has on the function of FH.
The presence of fucose on FH could also be investigated further. It was previously shown that the absence of fucose on the chitobiose core of N-linked antibody glycans improves antibody binding to lymphocyte receptors, and therefore, improves the triggering of antibody-dependent cytotoxicity (ADCC) (Shields *et al*., 2002). In contrast, other monosaccharide constituents, such as galactose, were shown to have no effect on enhancing ADCC (Shinkawa *et al*., 2003). Fenaille *et al* (2007*a*) showed that fucosylation is not specific to a particular glycan at a particular glycosylation site, but that fucosylation is present at low levels on glycans at all eight of the glycosylation sites on FH. This would suggest that fucosylation is an optional extra in the glycosylation of the protein. But is it a by-product of errors in the glycosylation pathway or does having low levels of fucosylated FH confer an advantage in some way? For example, FH was shown to bind human L-selectin, a member of the sialic acid binding class of lectins, and that this interaction is inhibited in the presence of fucoidan (Malhotra *et al*, 1999), a polysaccharide composed of

sulphated fucose. However, Malhotra *et al*. (1999) showed that the strength of interaction was reduced upon desialylation of human FH and, therefore, the interaction is unlikely to be fucose dependent. The fact is that there is a sub-population of fucosylated FH molecules and it is attractive to speculate that fucosylation has an intended purpose *e.g.* enrichment of FH at a particular surface via binding to a fucose receptor or modulation of activity in a similar way to non-fucosylated versus fucosylated antibodies.

FH is expressed in a range of tissue-types including: i) hepatocytes (Morris *et al.,* 1982 and Schwaeble *et al*., 1987), thought to be the major source of plasma FH, ii) blood monocytes (Whaley, 1980), iii) skin fibroblasts (Katz and Strunk, 1988), iv) platelets (Devine and Rosse, 1987), v) endothelial cells (Brooimans *et al*., 1990) and vi) retinal pigment epithelial cells (Chen *et al*., 2007). This, therefore, raises the question: is the glycosylation profile of FH expressed in different tissues the same or different? For example, Ceciliani *et al*. (2007) found that α_1-Acid glycoprotein (AGP) has differential glycosylation depending on the cell type in which it is expressed. More diantennary and fucosylated structures were present on somatic cell expressed AGP compared to mammary gland expressed AGP.
If the glycosylation of FH expressed in different tissues is different, what effect does this have on the function of FH? For example, glycodelin (Gd) is a glycoprotein with four reported glycoforms (Gd –A, -F, -C and –S) which have identical protein cores yet markedly different glycosylation patterns. Yet, Gd-A and Gd-F inhibit proliferation, supress interleukin-2 secretion and induce cell death of Jurkat cells and blood mononuclear cells whilst Gd-C and Gd-S exhibit no immunosuppressive effects (Lee *et al*., 2009). In addition to this, Gd-S has been shown to inhibit sperm-zona pellucida binding whilst Gd-A does not inhibit this process (Morris *et al*., 1996). The only difference in these glycoforms is the glycosylation profile, therefore the difference in function of these glycoforms is likely due to differential glycosylation.

A comparison of the functions of native (sialylated) plasma-derived FH and enzymatically desialylated plasma-derived FH suggested that removal of sialic acids from plasma FH reduces the binding affinity of FH for C3b. But, adding deglycosylated rFH in to the comparison suggested that complete removal of the glycan chains from FH abrogates this reduction in C3b affinity. However, the literature consistently shows that deglycosylated rFH has a higher affinity for C3b than plasma-derived FH (Schmidt *et al*., 2011 and, Kerr *et*

al., 2017) and therefore the C3b binding assay conducted here must be repeated for validity.

The C3b binding regions on FH, N-terminal CCPs 1-4 and C-terminal CCPs 19-20, were shown to bind to C3b on a CM5 chip and C1 chip with K_Ds of 14 μM and 3.5 μM and, 10 μM and 4.5 μM, respectively (Schmidt *et al*., 2008). These values are well above the K_Ds reported here and elsewhere in the literature for full length FH (Schmidt *et al*., 2011 and Kerr *et al*., 2017) but they indicate that the C-terminal region of FH binds with tighter affinity than the N-terminal region of FH.

Additionally, the bacterial FH binding protein PspCN has been shown to significantly improve affinity of FH for C3b and to confer affinity of FH for C3d (the TED domain and CCP 19-20 binding region of C3b) (Herbert *et al*., 2015 and, Kerr *et al*. 2017). Cross-linking mass spectrometry analysis of FH in the presence and absence of PspCN suggests that a conformational change in FH occurs and that this conformational change is attributable to the improvement of affinity of FH for C3b and C3d (Herbert *et al*., 2015) *i.e.* PspCN bound to FH induces a conformational change in FH to expose the previously hidden C-terminal C3b (TED domain) binding site on FH which, as alluded to above, has a higher affinity for C3b than the N-terminal C3b binding site.

It is significant that deglycosylated rFH binds to C3d here and in the literature (Schmidt *et al*., 2011 and, Kerr *et al*. 2017) whilst glycosylated human plasma-derived FH does not. In addition, it has been shown here (although this needs repeating for validity) that desialylated human FH may bind to C3d. Therefore, this would suggest that sialylated human FH is held in a conformation that prevents exposure of the CCP 19-20 C3b binding site. Desialylation and, to a greater extent, deglycosylation make the CCP 19-20 binding site available for C3b binding in the TED domain thus improving affinity for C3b and C3d. In addition, this also accounts for the greater R_{max} values for deglycosylated rFH and desialylated FH compared to human plasma-derived (sialylated) FH. Exposure of the CCP 19-20 site on FH by desialylation and deglycosylation allows for binding at two sites on C3b (fig. 2.14), therefore a greater proportion of randomly orientated C3b molecules is available for binding.

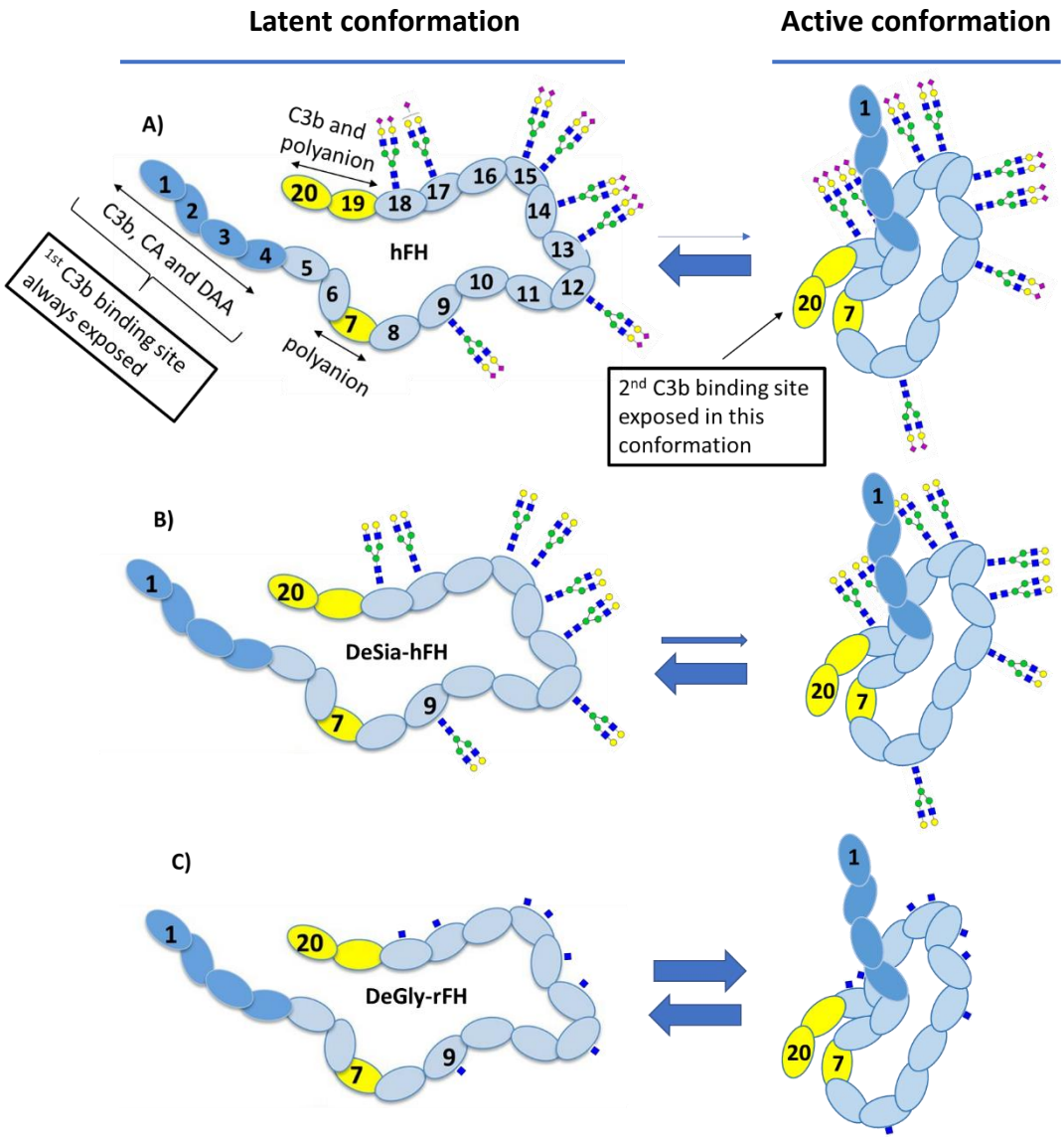


Figure 2.14: Schematic diagram of the propensity for different glycoforms to form a more open, active conformation by exposing the second C3b binding site (CCPs 19-20) for binding with C3b. A) in the absence of self-surface polyanions and GAGs, fluid phase FH is predominantly in the latent conformation. In this conformation, the N-terminal CCPs 1-4 are available to bind fluid phase C3b, albeit with low affinity relative to the second C3b binding site in CCPs 19-20, and thus regulate complement in fluid phase. In the presence of self-surface polyanions and C3b on self-surfaces, FH is pulled into a more active conformation in which the second, and relatively high affinity, C3b binding site in CCPs 19-20 is exposed for binding to C3b. Thus, FH with sialylated glycans (i.e. the natural glycosylation state of FH) circulates predominantly in the latent conformation and only converts to the more active conformation in the presence of self-surface polyanions and self-surface associated C3b. The negatively charged sialylated glycans on hFH may provide steric hinderance and charge repulsion, thus preventing latent hFH from adopting the active conformation in the absence of self-surface polyanions and self-surface associated C3b. B) In the absence of self-surface polyanions, like in the SPR C3b binding studies carried out here, removal of sialic acids (DeSia-hFH) may increase the propensity with which FH can adopt the active conformation by reducing charge repulsion. C) Removal of glycans altogether appears to drastically enhance the ability of FH to form the active conformation in the absence of self-surface polyanions. This could be due to removal of the steric clashes and charged repulsion effects conferred by glycans that would otherwise hinder the adoption of the active conformation.

DAA assays show that deglycosylated rFH has the highest increase (2.5-fold reduction in half-life) in decay acceleration of the C3bBb complex whilst human plasma-derived (sialylated) has the lowest (1.5-fold reduction in half-life) and desialylated human plasma-derived FH is intermediate (2.0-fold reduction in half-life). However, this ties in with the idea set out above, that desialylation and deglycosylation may expose the CCP 19-20 C3b binding site on FH. The C3b molecules on the chip surface are randomly orientated in the DAA assay. If desialylated and deglycosylated FH are able to bind to a second site on C3b, a greater proportion of C3bBb complexes are available for desialylated and deglycosylated FH mediated decay, compared to human plasma derived (sialylated) FH. This would result in a higher change in response and, therefore, a higher apparent change in half-life.

It was shown in this study that removal of sialic acid had no effect on the cofactor activity of FH. But, removal of total glycan appears to significantly reduce cofactor activity. In contrast to the C3b/d and DAA assays the cofactor assay was carried out in solution. Therefore, the limitations conferred by differential orientation of C3b on a sensor chip surface do not apply in solution cofactor assays – *i.e.* the two primary FH binding sites on C3b are both available for binding, thus a glycoform of FH able to bind both sites is able to do so. Therefore, this could account for the lack of difference seen in cofactor activities of human plasma-derived FH and desialylated human plasma derived FH. However, deglycosylated rFH shows a reduced cofactor activity compared to the glycosylated human plasma-derived FH samples.

Crystal structures of FI in complex with C3b and "miniFH", a truncated version of FH consisting of the C3b binding N-terminal CCPs 1-4 and the C-terminal CCPs 19-20 joined by a linker of 12 glycine residues (Schmidt *et al*., 2013), show that in this complex FI contacts FH in CCPs 2 and 3 (Xue *et al*., 2017).

Human FH has 8 N-linked glycosylation sites (Fenaille *et al*., 2007a) two of which are located near to the TED domain binding region of FH, on CCPs 17 and 18. A Third glycosylation site is on CCP 9 near to CCPs 6-8, a region on FH implicated in binding to C3b at a third site (Schmidt *et al*., 2008). Additionally, the 5 remaining glycosylation sites are located on CCPs 12-15 – a potential "hinge" region of FH (Schmidt *et al*., 2008). To add to this, Aslam and Perkins (2001) were able to show that FH has a bent back structure enabling the C3b binding regions of CCP 1-4 and 19-20 to come into close proximity and that the bent back structure is potentially facilitated by the long linkers located between CCPs in central region of FH (i.e, CCPs 11-14). The reduced cofactor activity of

deglycosylated FH, compared to glycosylated human plasma derived FH (both sialylated and desialylated) could indicate that the glycosylation may play a role in cofactor activity of FH. For example, could the glycans located near the C3b binding sites on FH facilitate docking of FI in an optimal orientation or could the glycans located near the hinge region of FH stabilise a structure of FH conducive for optimal cofactor activity?

Further work needs to be done on investigating glycosylation and, by extension, the role of sialic acids in the function of FH. In the work included here sialylated and enzymatically desialylated human FH was compared with deglycosylated rFH. Future work will aim to compare the human glycoforms with enzymatically deglycosylated human FH – to ensure a comparison of "like with like". As well as this, further replicates would confirm the validity of this work. However, this work also raises more questions which would be worthy of answers. For example, in a physiological setting, considering FH is able to interact with sialic acids and polyanions and that these structures enhance FH interaction with C3b (Meri and Pangburn, 1990), how does differential FH glycosylation effect function on a physiological surface and *in vivo*? And finally, does glycosylation differ between FH expressed in different tissues and if so, what effect does this have on the function of FH?

Chapter 3

Glycoengineering *P. pastoris* for production of recombinant human factor H carrying humanised, complex-type glycans

3.1 Overview

In Chapter 2 it was shown that the predominant glycan species on hFH is N-linked diantennary disialylated complex-type, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$. This glycoform became the target for glycoengineering work on *P. pastoris*.

As discussed in Chapter 1 section 1.4, the GlycoSwitch glycoengineered strain, SuperMan 5, had been modified by Dr Heather Kerr (in the Barlow group) to co-express *CFH* and a chaperone protein. For the current work, the resulting strain is designated SuperMan 5 *CFH*.

The *CFH* gene had been cloned into the plasmid pPICZα (fig. 3.1) downstream of the alcohol oxidase 1 (AOX1) promoter and *S. cerevisiae* prepro-α-secretion signal thus enabling methanol inducible expression and secretion of rFH.

Unpublished data showed that co-expression of *CFH* with the chaperone gene resulted in a 10-fold increase in protein yield of rFH compared to the 10s of milligram yields reported in the Barlow lab by Schmidt *et al*. (2011) in a *P. pastoris* strain without the chaperone.

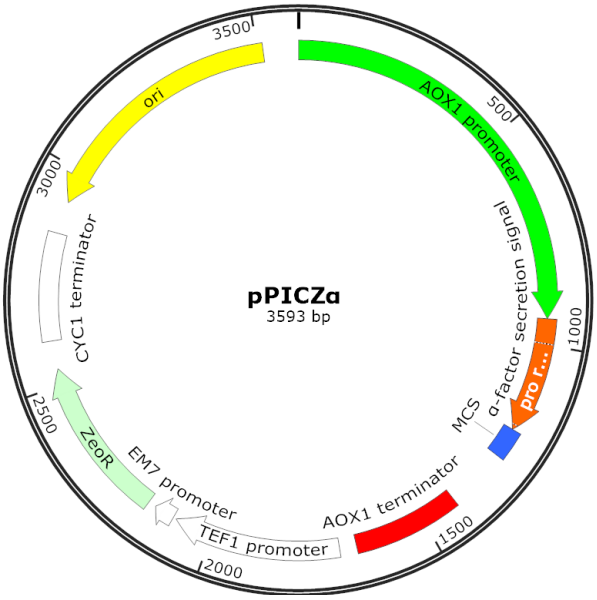


Figure 3.1: Vector plasmid map of plasmid pPICZα. The *CFH* gene was cloned into the multiple cloning site (MCS) by Dr. Heather Kerr. This placed *CFH* under the control of the methanol inducible AOX1 promoter (green arrow) and downstream and in frame of a sequence encoding the *S. cerevisiae* prepro-α-secretion signal (orange arrow). Also encoded in this plasmid is a gene conferring resistance to the antibiotic zeocin (faint blue arrow), for selection purposes, and an origin of replication (yellow arrow) for plasmid amplification in *E. coli*.

This chapter documents the process of using the GlycoSwitch technology in order to modify the glycan biosynthetic pathway to produce human galactose-terminated diantennary complex-type glycans, $Gal_2GlcNAc_2Man_3GlcNAc_2$. This glycan is two sialic acid residues short of the fully humanised complex-type N-linked glycans that predominant on human plasma FH.

3.2 Materials and Methods

3.2.1 Large-scale fermentation and purification of recombinant factor H expressed in strain SuperMan 5

3.2.1.1 Large-scale fermentation

A single colony was picked from YPD agar and used to inoculate a starter culture of 10 mL BMGY in 50 mL conical tube (Falcon). The starter culture was incubated for approximately 16 hours at 30 °C, 250 rpm shaking and used to inoculate two lots of 1.5 L BMGY media in 5 L baffled shake flasks. The cultures were left to incubate for approximately 72 hours, 30 °C at 200 rpm. The shake flask cultures were used to inoculate a 10 L BioFlow 4500 (New Brunswick Scientific) fermenter. Initial culture growth occurred by glycerol-fed batch fermentation, expression of rFH was induced by switching to a methanol/glycerol continuous feed for four days.

3.2.1.2 Purification of recombinant factor H from large scale P. pastoris *fermentation: cation exchange chromatography*

Culture media was centrifuged initially at low speed, the collected supernatant (8 – 9 L) was then clarified by a faster centrifugation (6,000 g, 30 minutes). The clarified supernatant was collected and the pellet was discarded. Protease inhibitors EDTA (1% (v/v) of a 500 mM stock) and PMSF (0.5% (v/v) from 100 mM stock in isopropanol) were added followed by dilution of supernatant with ddH_2O (1:5). At 4°C, dilute supernatant was flowed over a cation-exchange column (50 mL Toyopearl SP 650 M, Sigma-aldrich, XK 26 column, GE Healthcare Life Sciences) using an AKTA Pure sample pump at 30 mL/min.
After sample loading, the column was washed with 2 CV buffer A (20 mM potassium phosphate buffer, pH 6.0) and protein was eluted over a gradient of 0-100% buffer B (20 mM potassium phosphate buffer, 0.5 M sodium chloride, pH 6.0). 2 mL fractions of wash and elution steps were collected in 15 mL conical tubess. 2 µL each of peak, wash and flow-

through fractions, together with recombinant FH positive control, were mixed with 14 µL reducing SDS loading buffer, loaded onto a 4-12% bis-tris polyacrylamide gel and run (200 V, 1 hour) before staining with Coomassie Brilliant Blue G250.

3.2.1.3 Purification of recombinant factor H from large scale P. pastoris *fermentation: anion exchange chromatography*

Appropriate fractions were pooled, placed in 10 kDa molecular weight cut off dialysis tubing (SnakeSkin, ThermoFisher Scientific) and dialysed overnight at 4 °C into anion exchange chromatography buffer A (20 mM glycine, 80 mM NaCl, pH 9.5), with a buffer exchange after 2 hours dialysis. Anion exchange chromatography was carried out using a 6 mL Resource Q column (GE Healthcare Life Sciences). At a flow-rate of 6 mL/min, the column was equilibrated with 6 CV of buffer A, protein sample was loaded onto the column and washed with 5 CV buffer A. Protein was eluted in a linear gradient from 0 - 50 % buffer B (20 mM glycine, 1M NaCl, pH 9.5). 2 mL fractions were collected. 2 µL each of peak, wash and flow-through fractions, together with recombinant FH positive control, were mixed with 14 µL reducing SDS loading buffer, loaded onto a 4 - 12% bis-tris polyacrylamide gel and run (200 V, 1 hour) before staining with Coomassie Brilliant Blue G250. Appropriate fractions were pooled, placed in 10 kDa molecular weight cut off dialysis tubing and dialysed against phosphate buffered saline, pH 7.4, at 4 °C for approximately 16 hours, with a buffer exchange after 2 hours dialysis.

3.2.1 Transformation of P. pastoris *strain SuperMan 5 with GlycoSwitch plasmids*

3.2.1.1 Plasmid amplification in E. coli

In separate transformations, 1 µL of each GlycoSwitch plasmid (Biogrammatics) (fig. 3.6) was used to transform electrocompetent *E. coli* strain TOP10 (One Shot TOP10 Electrocomp *E. coli*, ThermoFisher Scientific, C404050) by electroporation (1.7 kV, 25 µF). Transformed cells were recovered in 1 mL SOC media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) for 1 hour, 37 °C and 250 rpm shaking. Recovered cells were spread plated onto LB agar (1% (w/v) tryptone, 0.5% yeast extract, 170 mM NaCl, 1.5% (w/v) agar, pH 7.5) containing the appropriate selection antibiotic (pGlycoSwitch-GnTI – no selection, pGlycoSwitch-GalT 50 µg/mL nourseothricin (Carbosynth), pGlycoSwitch-ManII 50 µg/mL kanamycin sulphate (ThermoFisher Scientific)

and pGlycoSwitch-GnTII ampicillin 50 µg/mL (ThermoFisher Scientific)) and incubated for approximately 16 hours. Clones were picked and used to inoculate 100 mL LB media, containing the appropriate antibiotic, in 500 mL baffled shake flasks. Cultures were left to incubate for approximately 16 hours at 37 °C, 250 rpm shaking. Cultures were centrifuged (3,500 rpm, 20 minutes) and supernatant was discarded. Plasmids were isolated from the pelleted cells by Midi prep (Qiagen) standard procedure (Qiagen plasmid purification handbook, www.qiagen.com/handbooks).

3.2.1.2 Plasmid linearisation by restriction endonuclease digestion and linear plasmid concentration by isopropanol precipitation

Isolated plasmids were linearised using either the restriction endonuclease AvrII (New England Biolabs, R0174S) (pGlycoSwitch-GnTI, -GalT and -ManII) or SapI (New England Biolabs, R0569) (pGlycoSwitch-GnTII). Briefly, >10 µg plasmid was mixed with 50 µL cutsmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/mL bovine serum albumin, pH7.9, New England Biolabs), either 50 U AvrII or 100 U SapI, and made up to 500 µL with ddH_2O before incubating for 16 hours at 37 °C. Linearisation was assessed by 1% (w/v) agarose DNA electrophoresis of digested and undigested plasmid (data not shown).

Linearised plasmids were concentrated using isopropanol precipitation to yield >2 µg linear plasmid. Briefly, sodium acetate was added to linearised DNA to yield a final concentration of 0.3 M and pH 5.2. To this, room-temperature isopropanol was added at 0.7 of the volume of linear DNA and the solution was mixed by pipetting before centrifugation (15,000 g, 30 minutes, 4 °C). The supernatant was removed and the pellet washed with 1.5 mL 70% (v/v) ethanol in distilled water followed by centrifugation (15,000 g, 15 minutes, 4 °C). The supernatant was removed and the pellet air dried for 10-20 minutes. The DNA was dissolved in ddH_2O, in a water bath (40-50°C), prior to transformation.

3.2.1.3 Transformation of P. pastoris *strain SuperMan 5* CFH *with GlycoSwitch plasmids pGlycoSwitch-GnTI and pGlycoSwitch-GalT*

P. pastoris SuperMan 5 *CFH* was made competent according to the methodology of Lin-Cereghino *et al*. (2005). Briefly, *P. pastoris* strain SuperMan 5 *CFH* was cultured in a 50 mL

volume of buffered complex media containing glycerol (BMGY) (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base with ammonium sulphate and without amino acids, $4x10^{-5}$% (w/v) biotin, 1% (v/v) glycerol, 100 mM potassium phosphate buffer, pH 6.0) to an optical density at 600 nm wavelength (OD_{600}) of between 0.8 – 1.5. Cells were centrifuged (3000 rpm, 5 minutes), the supernatant discarded and the pellet suspended in 9 mL BEDS solution (3% (v/v) ethylene glycol, 5% (v/v) DMSO, 1 M sorbitol, 10 mM bicine-NaOH, pH 8.3) and 1 mL 1M DTT, followed by incubating at 30 °C, 250 rpm for 5 minutes. Cells were centrifuged as before, supernatant discarded and the pellet was suspended in 1 mL BEDS solution.

Linearised plasmids pGlycoSwitch-GnTI and -GalT were mixed, by pipetting, in equimolar quantities before co-transformation by electroporation (1.5 kV, 25 µF) into *P. pastoris* strain SuperMan 5. Transformed cells were placed into 14 mL round bottom tubes with 1 mL recovery media (1:1 mixture of BMGY and 2 M sorbitol) before incubating at 30 °C, 250 rpm for 2 – 3 hours. After transformation, cells were plated onto yeast peptone dextrose sorbitol (YPDS) (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose, 1 M sorbitol, 2% (w/v) agar) media, prepared according to the method outlined in the "EasySelect *Pichia* Expression Kit" user manual (Invitrogen, manual part number 25-0172, pp 54-58), including the antibiotics 300 µg/mL Zeocin (Invitrogen) and 100 µg/mL nourseothricin sulphate. Plates were incubated at 30°C for 10 days. Clones were picked and assessed for the presence of both pGlycoSwitch plasmids (-GnTI and -GalT) and assessed for the presence of the target hybrid, galactose terminal glycan catalysed by the *GnTI* and *GalT* gene products. A strain identified to produce this hybrid glycan was called HyGal *CFH*. HyGal *CFH* was made competent by the same procedure outlined above. Colony PCR and hybrid glycan analysis methods are detailed below.

3.2.1.4 Transformation of P. pastoris *strain HyGal* CFH *with GlycoSwitch plasmids pGlycoSwitch-ManII and pGlycoSwitch-GnTII*

The same procedure, detailed in section 3.2.1.3, was used for the co-transformation of linearised plasmids pGlycoSwitch-ManII and -GnTII into HyGal *CFH*, except that transformed cells were spread plated onto YPDS agar containing the antibiotic 350 µg/mL G418 (ThermoFisher Scientific, 10131035) and 150 µg/mL hygromycin B (Sigma) as well as Zeocin and nourseothricin as stated previously. A clone identified to contain all four

plasmids, by colony PCR, and shown to biosynthesise the target complex type galactose terminal glycan was called SuperGal *CFH*.

3.2.2 Transformed clone selection by colony polymerase chain reaction (PCR)

Two sets of colony PCR were carried out. The first set of colony PCR was used to identify clones of SuperMan 5 *CFH*, transformed as detailed in section 3.2.1.3, containing pGlycoSwitch-GnTI and -GalT. The second set of colony PCRs was used to identify clones of HyGal *CFH*, transformed as detailed in section 3.2.1.4, containing all four GlycoSwitch plasmids (-GnTI, -GalT, -ManII and -GnTII). The colony PCR procedure was as follows. Colonies of *P. pastoris* were picked from a fresh transformation plate and were used to inoculate 2 mL YPD (YPDS without sorbitol) media containing appropriate antibiotic. Inoculated media was incubated at 30 °C, 250 rpm for approximately 16 hours. 1mL of each culture was centrifuged (14,000 rpm, 5 minutes) and the supernatant was discarded. The remaining 1 mL culture was stored at 4 °C. 5 µL of each cell pellet was suspended in 10 µL SPZ buffer (75 U zymolase (Zymo Research), 1.2 M sorbitol, 81 mM sodium phosphate buffer, pH 7.5) and incubated at 37 °C for 1 hour. Zymolase was deactivated by heating to 95 °C for 5 minutes. Samples were diluted with 90 µL ddH_2O and then centrifuged (14,000 rpm for 1 minute). The supernatant, containing crude genomic DNA, was transferred to a fresh 1.5 mL tube (Eppendorf) prior to PCR. 2 µL crude genomic DNA was mixed with 5 µL One*Taq* Quick-Load Master mix (New England Biolabs) and 1.5 µL each of 10 µM forward primer and reverse primers (primers used for each colony PCR given below). PCR reactions were placed in a thermocycler (Eppendorf Mastercycler nexus X2) and PCR was carried out with the following parameters: the annealing temperature for all PCR reactions was 55 °C and the extension time used was 60 s per 1 kbp, rounded up to the nearest minute.
The primers used and the expected PCR product sizes for colony PCR of GlycoSwitch plasmids were:

pGlycoSwitch-GnTI
3.i) For (5082): 5’-GAGGAAGCGGCCGCGCTCTTCGCAGGTGCGGTTATTGTTC-3’
3.ii) Rev (4048): 5’-CGATGATGGGGAATAGCTCAGCCGAGGGC-3’
PCR product size: 364 base pairs (bp)

pGlycoSwitch-GalT
3.iii) For (5088): 5’-GAGGAATCTAGAGCTCTTCGTAGTTTTGATATTGTGCGGGC-3’

3.iv) Rev (4511): 5’- CAACGTCCCCTGCACGTCTAGGGGTGACC-3’
PCR product size: 1024 (bp)

pGlycoSwitch-ManII
3.v) For (5684): 5’-GGACACAGCTCTATTCACCCACATGATGC-3’
3.vi) Rev (5734): 5-TTAGATCTCGAGGCTCTTCGGGCTCAGCTTGAGTGACTGC-3’
PCR product size: 2376 (bp)

pGlycoSwitch-GnTII
3.vii) For (5680): 5’-ACTTTTCCTGGAAGAAGACCACTACTTAGC-3’
3.viii) Rev (5721): 5’-GAGGAAGCGGCCGCGCTCTTCGTCTCCCCACCCTCCATTTTTC-3’
PCR product size: 534 (bp)

3.2.3 Test expression and selection of glycoengineered strains HyGal CFH *and SuperGal* CFH

3.2.3.1 Shake-flask growth and expression

Clones identified, by colony PCR, to contain pGlycoSwitch plasmids GnTI, ManII (for HyGal *CFH*), GnTII and Gal (for SuperGal *CFH*) were cultured in shake flasks and induced to express rFH. Colonies were picked and used to inoculate 5 mL BMGY media, containing appropriate antibiotics, and incubate overnight at 30 °C, 250 rpm shaking. Cultures were grown until an OD_{600} of between 0.8 – 1.5 was reached. 50 µL cell culture was used to inoculate 25 mL fresh BMGY media in 100 mL baffled shake flask, these were incubated at 30 °C, 250 rpm for approximately 60 hours. Glycerol stocks were made (200 µL 50 % (v/v) glycerol and 800 µL cell culture) and stored frozen at -80 °C. Cell cultures were transferred to 50 mL conical tubes and centrifuged (1,600 rpm, 20 mins, 4 °C). The supernatants were discarded and the pellet suspended in 5 mL induction media BMMY (same as BMGY except glycerol replaced with 0.5% (v/v) methanol). Cultures were left to incubate in BMMY for approximately 96 hours at 16 °C, 250 rpm shaking. Cultures were fed with 50 µL (1% culture volume) methanol 24, 48, 72 and 80 hours post-induction. Cultures were transferred to fresh 50 mL conical tubes and centrifuged (1,600 rpm, 20 minutes, 4 °C). Supernatant was transferred to fresh 50 mL conical tubes and 1% (v/v) 500 mM EDTA and 0.5% 100 mM PMSF was added. A 1 mL aliquot of each culture was taken for analysis, detailed below in section 3.2.3.2, the remainder was stored frozen at -20 °C.

3.2.3.2 SDS-PAGE and Lectin Western blot

SDS-PAGE and ECL lectin-based western blot were carried out as detailed in Chapter 2 section 2.2.2.1. Specifically, for analysis of the HyGal *CFH* colonies, crude culture media, 10 µL of supernatant was mixed with 40 µL reducing and non-reducing SDS-loading buffer, and 10 µL of this was loaded onto 4-12% bis-tris polyacrylamide gels. For analysis of SuperGal *CFH* colonies crude culture media, 25 µL of supernatant was mixed with 25 µL of reducing and non-reducing SDS-loading buffer and 20 µL of this mix was loaded onto 4-12% bis-tris polyacrylamide gels. The reason the volumes used for SDS-PAGE and lectin-based western blot are different for HyGal *CFH* and SuperGal *CFH* is that the volumes used for HyGal *CFH* did not yield strong bands in either SDS-PAGE or lectin-based western blots, therefore, for SuperGal *CFH* analysis higher volumes of crude supernatant were used.
Aliquots from the SuperGal colony culture media (after spinning out cells) were run on polyacrylamide gels before and after treatment with endo H_f. For endo H_f treatment, endo H_f (1 MU/mL) was added to crude culture supernatant at a volumetric ratio of 1:150 and incubated overnight at 37 °C

3.2.4 Large-scale fermentation and purification of recombinant factor H expressed in glycoengineered strain SuperGal CFH

Large scale fermentation of glycoengineered strain SuperGal *CFH* was carried out as in section 3.2.1.1 and purification by cation exchange chromatography as in section 3.2.1.2, except 10 mL fractions were collected, and anion exchange chromatography as in section 3.2.1.3. However, 1 µL was mixed with reducing and non-reducing SDS loading buffer before application to gels for SDS-PAGE analysis of cation and anion exchange chromatography eluted fractions (figs. 3.15 and 3.16).

3.2.5 Removal of endogenous yeast-type glycans from recombinant factor H expressed in glycoengineered P. pastoris *strain SuperGal* CFH *catalysed by endoglycosidase* H_f

3.2.5.1 Endoglycosidase H_f catalysed yeast glycan removal

Recombinant FH expressed and purified in glycoengineered strain SuperGal *CFH*, see section 3.2.4, is glycosylated with yeast glycans as well as the target diantennary,

diagalactosylated complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$. Yeast glycans were enzymatically removed using endo H_f, which has specificity for yeast and hybrid glycans, but cannot cleave complex-type glycans. 100 kU of endo H_f (New England Biolabs) was mixed with approximately 70 mg of purified recombinant FH expressed in glycoengineered strain SuperGal *CFH* and incubated at 37 °C for 2 hours.

3.2.5.2 Purification of endoglycosidase H_f from recombinant factor H

Endo H_f is tagged with maltose binding protein (MBP) which allows for efficient removal of endo H_f by MBP-tagged affinity chromatography. A 1 mL prepacked MBPTrap column (GE Healthcare Life Sciences) was equilibrated in 10 CV buffer A (PBS, pH 7.4). The recombinant FH endo H_f mix, in PBS, pH 7.4, was applied directly to the column at a flow-rate of 1 mL/min. Flow-through was collected in 50 mL conical tubes. The column was washed with 10 CV buffer A and protein was eluted by washing the column in 5 CV of buffer B (10 mM maltose in PBS, pH 7.4). 2 µL each of flow-through, wash and peak fractions, together with purified recombinant FH before endo H_f treatment and pure endo H_f, were mixed with 14 µL reducing and non-reducing SDS-loading buffer and applied to 4-12% bis-tris polyacrylamide gels and run (200 V, 1 hour). The flow-through and wash fractions containing recombinant FH (fig. 3.17) were pooled, placed in 10 kDa molecular weight cut off dialysis tubing and dialysed against PBS, pH 7.4.

3.2.6 Glycan release, purification and analysis by MALDI-TOF mass spectrometry of crude culture media and purified recombinant factor H

Glycan release, purification and MALDI-TOF mass spectrometry analysis were the same as in chapter 2 section 2.2.3.2. However, for HyGal *CFH* and SuperGal *CFH* crude culture media analysis, 1.5 mL culture media was concentrated in a freeze dryer and the dried contents was suspended in 20 µL ddH_2O before PNGase F deglycosylation and BlotGlyco glycan purification (figs. 3.9 and 3.13).

3.3 Results

3.3.1 Characterising the extracellular glycome of *P. pastoris* strains SuperMan 5 *CFH* and KM71H *CFH*

Before attempts at glycoengineering using the GlycoSwitch technology, it was first important to assess whether the SuperMan 5 strain is capable of producing proteins decorated with five-mannose glycan, $Man_5GlcNAc_2$, since this lays the foundation for further efforts to humanise the glycosylation pathway in *P. pastoris*.

To do this, recombinant human FH (rFH) was expressed in two strains of *P. pastoris* that were expected to produce two different patterns of glycosylation. The strain KM71H *CFH* has no modifications to its glycosylation pathway, and therefore should produce what we can term wild – type glycosylation. The other strain, SuperMan 5 *CFH*, as outlined in Chapter 1 section 1.4.1, had its glycosylation pathway modified to produce a shortened five mannose yeast-type glycan, $Man_5GlcNAc_2$. A clone of each strain, taken from a freshly streaked YPD agar plate, was cultured in baffled shake flasks and induced to express rFH by the method outlined in section 3.2.3.1. Culture media was deglycosylated with PNGase F, glycans purified by BlotGlyco glycan purification and analysed by MALDI-TOF mass spectrometry as outlined in section 3.2.6 (fig 3.2).

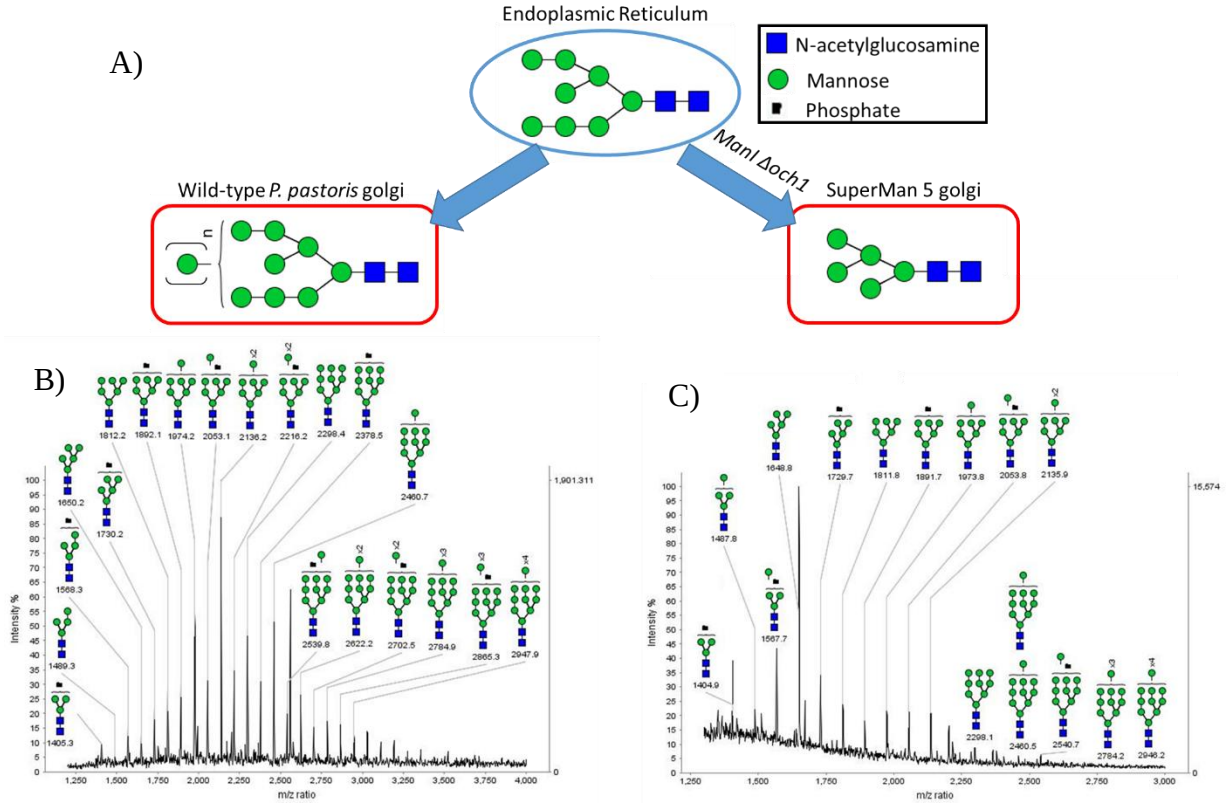


Figure 3.2: A) Schematic diagram of glycan processing in wild-type *P. pastoris* strain KM71H (left) and *P. pastoris* strain SuperMan5 (right) that has been glycoengineered to produce a five-mannose glycan by knock-out of the *och1* gene and expression of the exogenous *ManI* gene from *Trichoderma reesei*. MADLI-TOF mass spectrums of glycans released and purified from crude supernatant after rFH expression in the *P. pastoris* strains KM71H *CFH* (B) and SuperMan5 *CFH* (C). The glycan structures used here to annotate m/z peaks are archetypal isoforms thus, the structures presented here do not represent all possible isoforms for each m/z value.

The mass spectra (fig 3.2) represent the glycans in the extracellular environment that are sensitive to cleavage by PNGase F and, therefore, these mass spectra do not represent the glycans on exclusively rFH but they represent the entire N-linked extracellular glycome. The mass spectra show that there is substantial heterogeneity in the glycosylation of both strains. Additionally, both strains appear to incorporate phosphomannose into the glycosylation chain. The presence of phosphomannose in protein glycans is an important signal for targeting proteins to the lysosome (Sleat *et al*, 2006). However, why phosphomannosylation occurs in the extracellular environment is unknown, it could be the remnants of cell lysis during culturing.

KM71H *CFH* shows the greater heterogeneity and its glycans tend to be hypermannosylated – structures containing upwards of 13 mannoses are present – with no particular species of glycan dominating. In comparison, the glycan mass spectrum for SuperMan 5 *CFH* has shifted to the left, relative to the KM71H *CFH* glycan mass spectrum. The glycans are, in general, substantially smaller, with a 10-mannose containing glycan being the largest detectable species in SuperMan 5 (the position of the 12- and 13-mannose glycan species are included in the SuperMan 5 *CFH* spectrum to allow comparison with KM71H *CFH* however, these species are not detectable, above background levels, in the SuperMan 5 *CFH* spectra).

A five-mannose glycan species is indeed present in the SuperMan 5 *CFH* strain and, in contrast to KM71H *CFH*, this appears to be the most abundant species.

These experiments confirmed that that the SuperMan 5 *CFH* strain produces the five-mannose glycan as expected. However, SuperMan 5 *CFH* also produced many other species that could not be developed into human glycans. Since the mass spectra in figure 3.2 were generated from crude supernatant, they represent the glycoprofile of all glycoconjugates in secreted proteins that are releasable by PNGase F. It was decided to purify rFH expressed in Superman 5 to analyse the glycoprofile specific to rFH.

3.3.2 SuperMan 5 *CFH* expression and purification of recombinant factor H

By methods outlined in section 3.2.1, a 10-litre fermentation of SuperMan 5 *CFH* was carried out and rFH was purified in a two-step purification process, cation exchange chromatography (fig. 3.3) followed by anion exchange (fig. 3.4) chromatography.

The cation exchange chromatography procedure exploits the basic, polyanion interacting CCPs (7 and 20) on FH that, at the slightly acidic pH of *P. pastoris* culture media, are positively charged. 10 L of culture supernatant, clarified by centrifugation and diluted 1:5 in ddH_2O, was applied to a column containing 50 mL strong cation exchange resin (Toyopearl SP 650 M, XK 26 column) (fig. 3.3 A). 2 mL fractions were collected and select fractions were analysed for purity by SDS-PAGE (fig. 3.3 B).

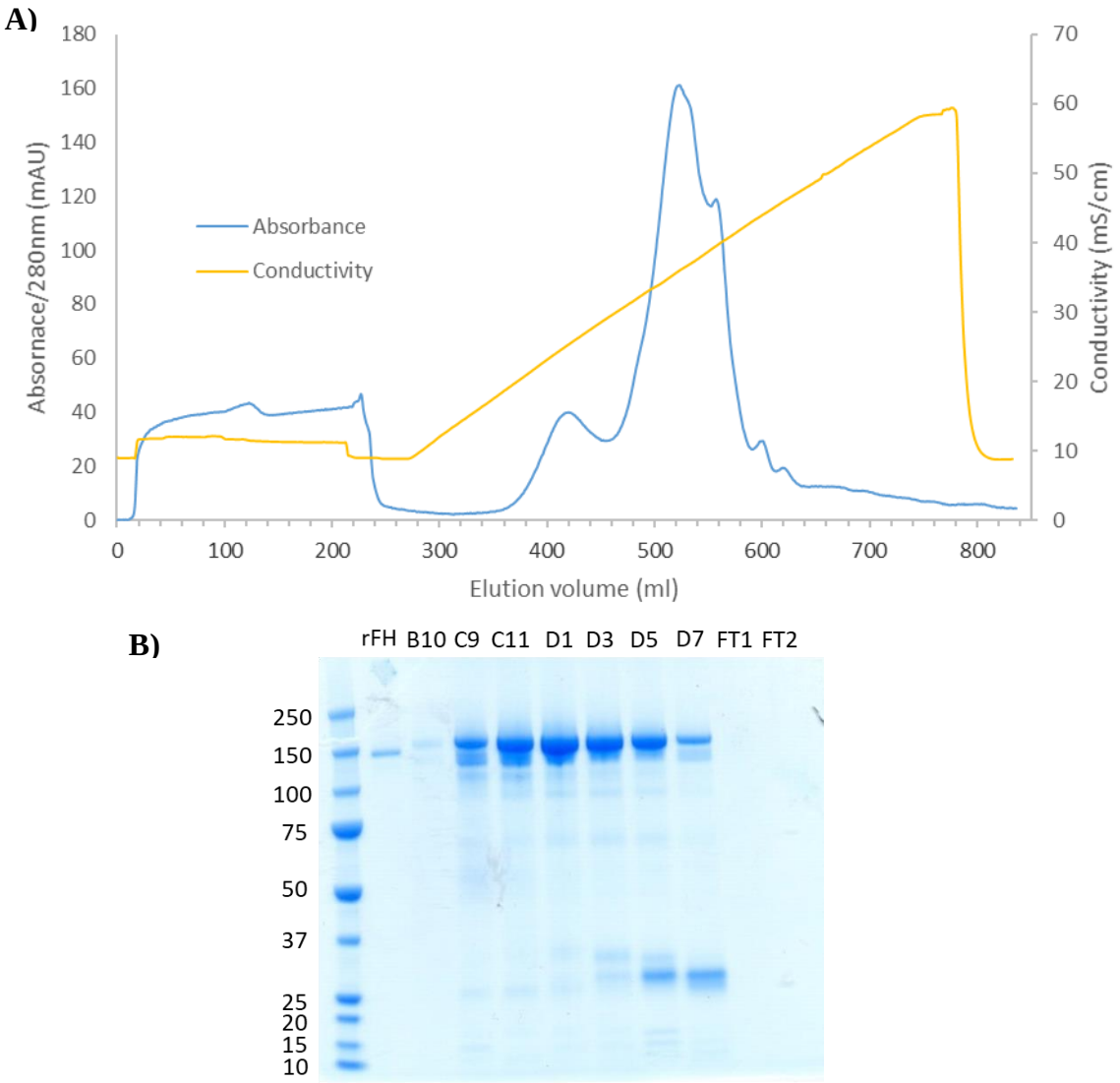


Figure 3.3: A) Cation-exchange chromatography purification from crude supernatant of rFH expressed in *P. pastoris* strain SuperMan 5. B) SDS-PAGE gel of rFH positive control, wash B10 (124.5 mL), peak fractions C9-D7 (146.5 – 166.5 mL) and flow-through (FT) fractions. 2 µL of each fraction was mixed with 14 µL reducing SDS loading buffer. MW of markers measured in kDa.

Peak fractions (as shown in figure 3.3) were run on SDS-PAGE to assess purity. Fractions C11-D3 (elution volume 150.5 – 158.5 mL) were considered to have the highest concentration of rFH and relatively low levels of impurities.

After pooling the appropriate fractions and dialysis against 20 mM glycine, 80 mM NaCl, pH 9.5, rFH was subjected to a second purification step by anion-exchange chromatography (fig. 3.4) as outlined in section 3.2.1.3. Approximately 25 mL of sample, purified by cation exchange chromatography and dialysed as described above, was applied to a column containing 6 mL strong anion exchange resin (Source 15Q, Resource Q column) (fig. 3.4 A). 2 mL fractions were collected and analysed for purity by SDS-PAGE (fig. 3.4 B).

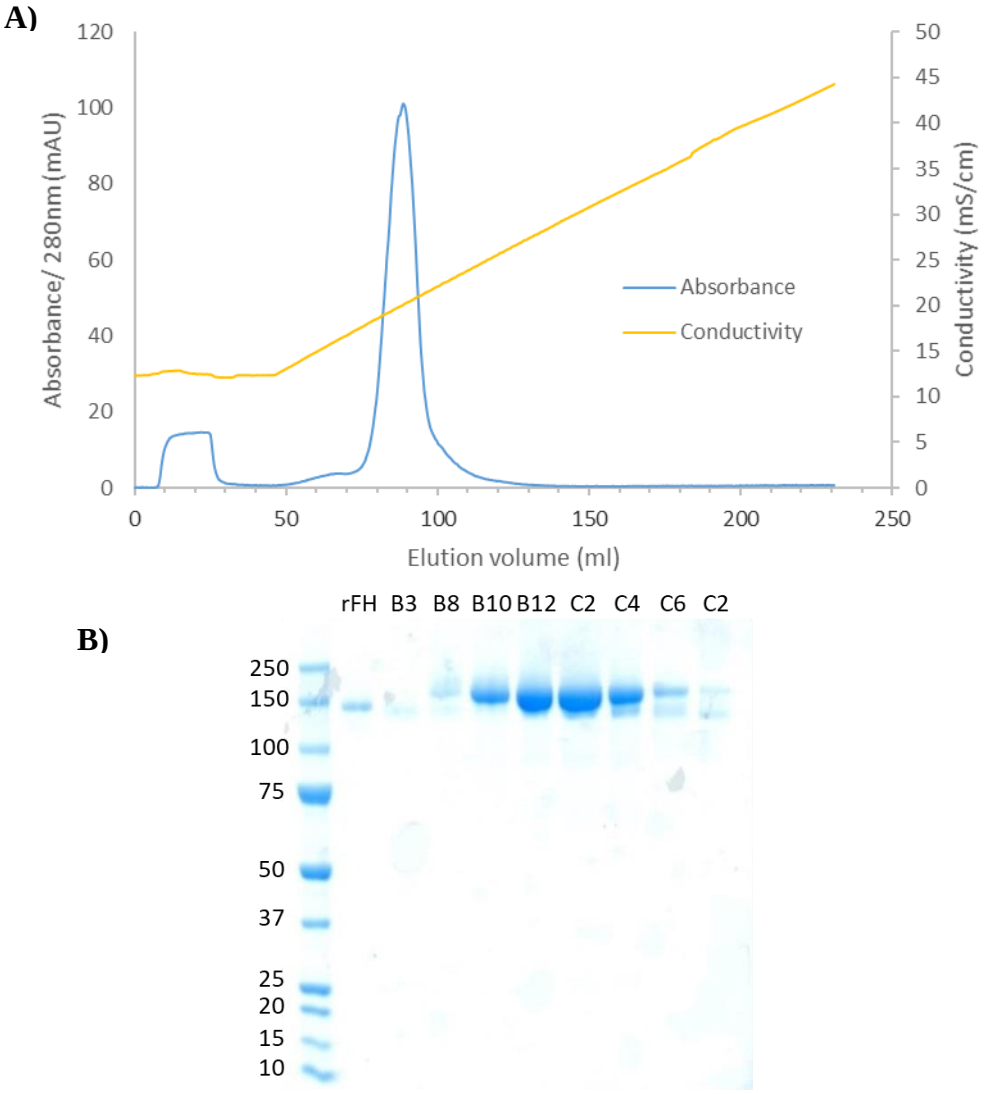


Figure 3.4: A) Anion exchange chromatography purification of rFH expressed in *P. pastoris* strain SuperMan 5. B) Reduced SDS-PAGE gel of rFH positive control, wash B3 (65.4 mL) and peak fractions B8-C2 (75.4 – 87.4 mL). 2 µL of each fraction was mixed with 14 µL reducing SDS loading buffer. MW of markers measured in kDa.

rFH eluted in a single, broad peak (fig 3.4 A). Fractions B10-C2 (79.4 – 87.4 mL) (fig 3.3 B) were pooled and dialysed against PBS, pH7.4. The concentration was measured by UV-VIS spectrophotometry to be 2.2 µM, purified rFH was aliquoted into 250 µL aliquots and stored frozen at -80 °C. Fraction C4, although having a high concentration of rFH, was not

pooled because it had a relatively high concentration of a cleavage product (a band just below the rFH main band at ~150 kDa) that presumably arises from the action of a *P. pastoris* protease.

3.3.3 Characterising the glycosylation of purified recombinant factor H expressed in SuperMan 5 *CFH*

After purification, the N-glycans were released from a 20 µg sample of rFH expressed in SuperMan 5 by incubating at 37 °C overnight with 0.5 U PNGase F and analysed by MALDI-TOF mass spectrometry by methods outlined in section 3.2.6 (fig. 3.5).

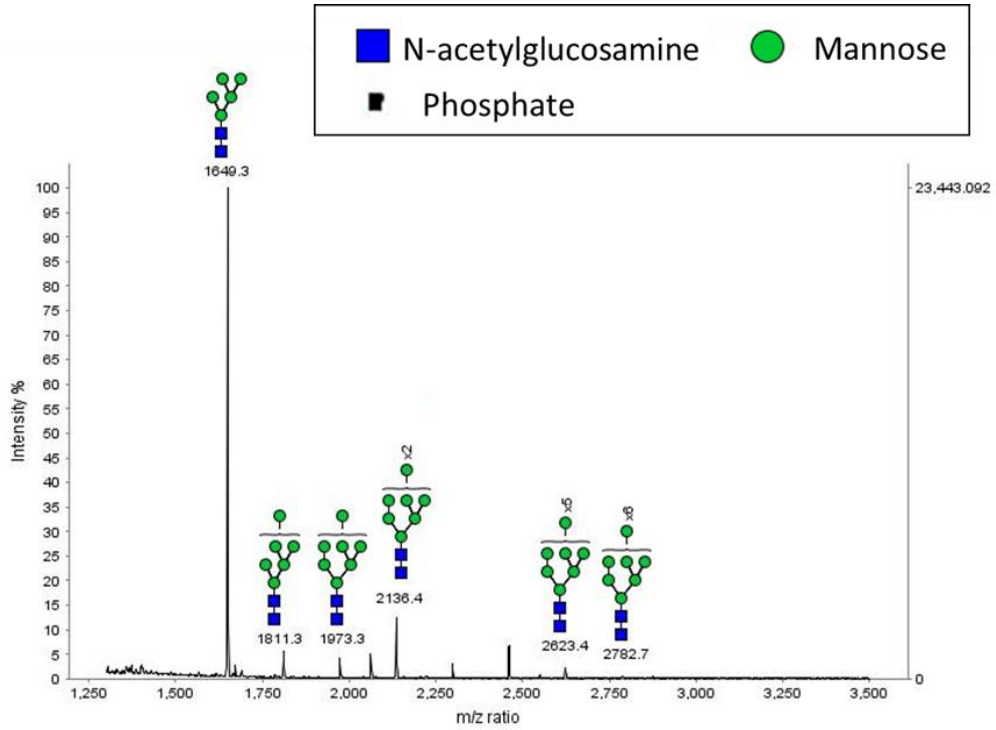


Figure 3.5: MADLI-TOF mass spectrum of glycans released and purified from purified rFH expressed in *P. pastoris* strain SuperMan 5.

The glycoprofile is dominated by five-mannose species, $Man_5GlcNAc_2$, but shows a degree of heterogeneity, with some glycans having 12 mannose residues, $Man_{12}GlcNAc_2$. This species is larger than the largest species detected in crude supernatant of SuperMan 5 (fig 3.1), however, this difference could be due to the purity of the starting material and therefore produces less background allowing for detection of less abundant glycan species. In contrast to the glycoprofile of SuperMan 5 crude supernatant, there are no phosphomannose glycan species present on rFH expressed in this strain.

The prominence of the five-mannose species, $Man_5GlcNAc_2$, suggest that the majority of glycans on rFH expressed in SuperMan 5 have the potential to be humanised.

3.3.4 Glycoengineering *P. pastoris* to produce hybrid-type, galactose-terminated glycans

The GlycoSwitch vector technology (fig 3.6) was then introduced into *P. pastoris* strain SuperMan5 to humanise the glycosylation of *P. pastoris*. This was achieved by a two-step process. In the first-step, the target was to engineer an intermediate, hybrid-type, galactose terminal glycan, $Gal_1GlcNAc_1Man_5GlcNAc_2$. To do this the GlycoSwitch vectors containing genes encoding GnTI and GalT (in fig. 3.6 A and B) were introduced together in a co-transformation procedure. Clones were screened by a combination of colony PCR, SDS-PAGE and ECL lectin-based western blot to detect for rFH carrying the hybrid-type glycan. In the second-step, a clone identified to biosynthesise the intermediate, hybrid-type glycan was co-transformed with the final two GlycoSwitch plasmids (fig. 3.6 C and D). These two plasmids encode ManII and GnTII which catalyse the conversion of the intermediate, hybrid-type glycan, $Gal_1GlcNAc_1Man_5GlcNAc_2$, into a diantennary, digalactosylated complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$.
The results for the production of the hybrid-type glycan are described in this section, and the results for production of the humanised, galactose terminal glycan are described in the next section, 3.3.5.

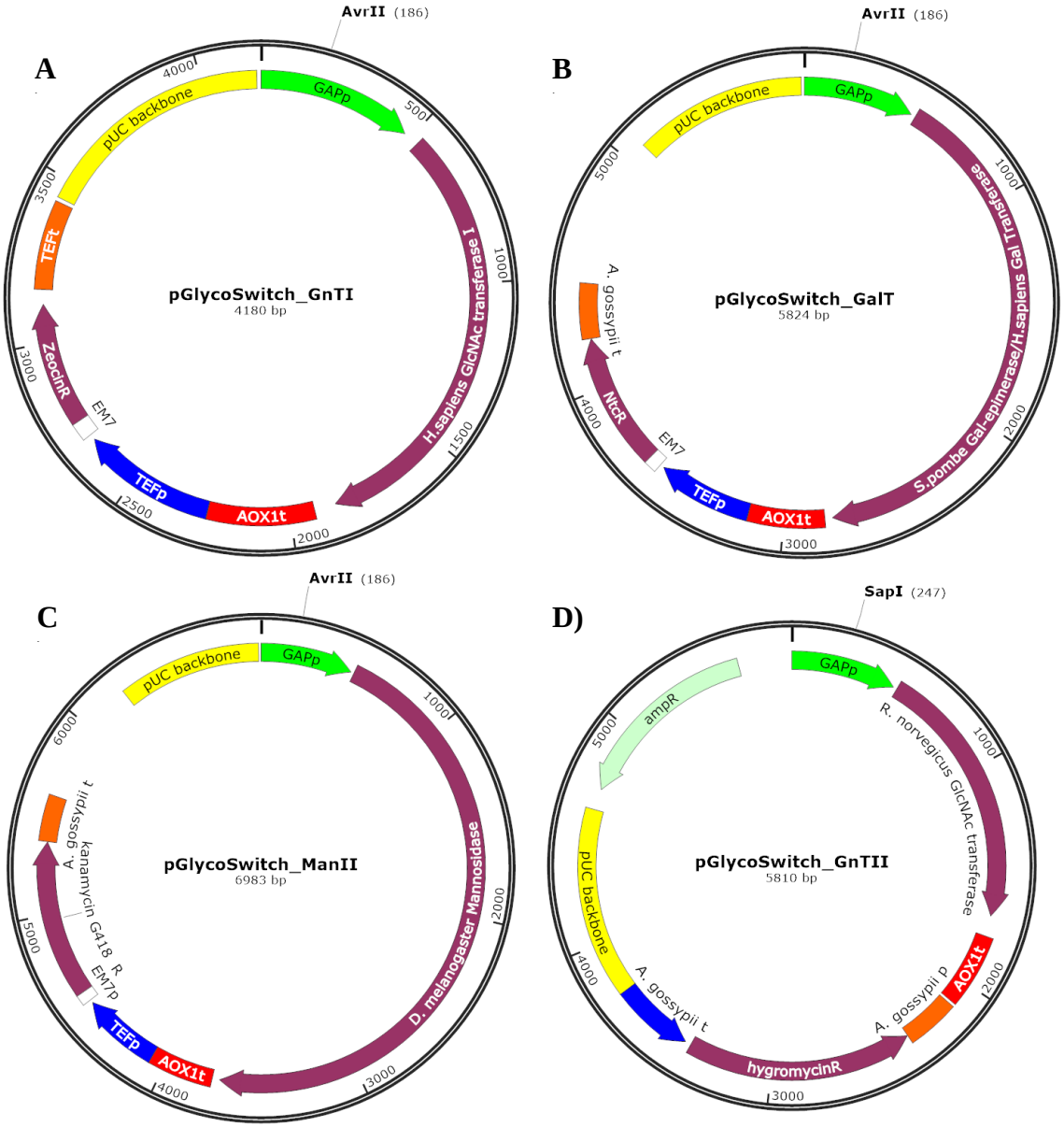


Figure 3.6: GlycoSwitch plasmid maps. Each plasmid was linearised in the GAP promoter prior to transformation into *P. pastoris*. Marked on each plasmid is the restriction endonuclease used to linearise the plasmid. These plasmids were introduced pairwise in a two-step procedure. First, pGlycoSwitch-GnTI (A) and -GalT (B) were introduced. Clones that were identified to produce a galactose terminal hybrid glycan were then made competent and transformed with the second pair of GlycoSwitch plasmids, pGlycoSwitch-ManII (C) and -GnTII (D).

Initially, Jacobs *et al*. (2009) recommend that each plasmid is introduced sequentially. In this way, selection for the target glycan could be done using lectin western blot, before introducing the next plasmid in the sequence.

However, a caveat had to be applied to this process. Both candidates for the first plasmid to be introduced, i.e. those encoding N-acetylglucosamine transferase I (GnTI) (fig. 3.6 A) or

mannosidase II (ManII) (fig. 3.6 C), fail to produce products ($GlcNAc_1Man_5GlcNAc_2$ or $Man_3GlcNAc_2$, respectively) that are distinguishable from the parental 5-mannose glycan, $Man_5GlcNAc_2$. Therefore, the lectin-based western detection method would not be able to report on success or otherwise of this first step.

Therefore, an alternative method was devised in which the first GlycoSwitch plasmid, encoding N-acetylglucosamine transferase I, was introduce at the same time into *P. pastoris* SuperMan5 together with the second GlycoSwitch plasmid (co-introduced), encoding galactosyltransferase (fig. 3.6 A and B). Thus, the resulting glycan structure generated from the sequential action of N-acetylglucosamine transferase I and galactosyltransferase would incorporate galactose, a monosaccharide not naturally present in yeast glycans. This structure is a hybrid glycan intermediate between a yeast-type and human complex-type glycan, $Gal_1GlcNAc_1Man_5GlcNAc_2$. ECL, a lectin that requires galactose as a core competent for binding, could then be used in a western blot style assay to detect for clones of *P. pastoris* that had taken up and were expressing the glycosylation genes on the co-introduced plasmids.

The second step in the process was to co-introduce the plasmids pGlycoSwitch-ManII and pGlycoSwitch-GnTII (fig. 3.6 C and D). Expression of ManII should cut back the two mannose residues linked to the α1,6-mannose of the trimannosyl-core of the hybrid-type glycan generated in the first step, converting $Gal_1GlcNAc_2Man_5GlcNAc_2$ to $Gal_1GlcNAc_1Man_3GlcNAc_2$. Not only does the enzymatic action of ManII generate the substrate for attachment of GlcNAc by GnTII, it also renders the glycan insensitive to cleavage by the endoglycosidase Endo H_f. This final point allowed for the selection of clones expressing the diantennary galactose-terminated complex-type glycan from the mono-antennary galactose terminal hybrid-type glycan using the lectin ECL – i.e. to distinguish $Gal_1GlcNAc_1Man_5GlcNAc_2$ generated in the first step from $Gal_2GlcNAc_2Man_3GlcNAc_2$ generated in the second step.

Although *P. pastoris* autonomous replication sequences do exist (Cregg *et al*, 1985) their use is not routine due to stability issues. Therefore, the most common method for stable transformation of *P. pastoris* is genomic integration. This is done by linearising the plasmid (by restriction endonuclease digestion) in a stretch of DNA that is homologous to a region in the genome of *P. pastoris*. Genomic integration can then occur by single homologous

recombination using the cells endogenous recombination machinery. This additional need for genomic integration causes the success of transformation in *P. pastoris* to be low relative to *E. coli*.

The strain SuperMan 5 was duly co-transformed with equivalent numbers of moles of the enzymatically linearised GlycoSwitch plasmids pGlycoSwitch-GnTI and pGlycoSwitch-GalT (fig. 3.6 A and B). Because two plasmids were to be co-transformed, it was especially important to assess colonies, post-transformation, for the presence of both GlycoSwitch plasmids. This was done by colony PCR (fig 3.7).

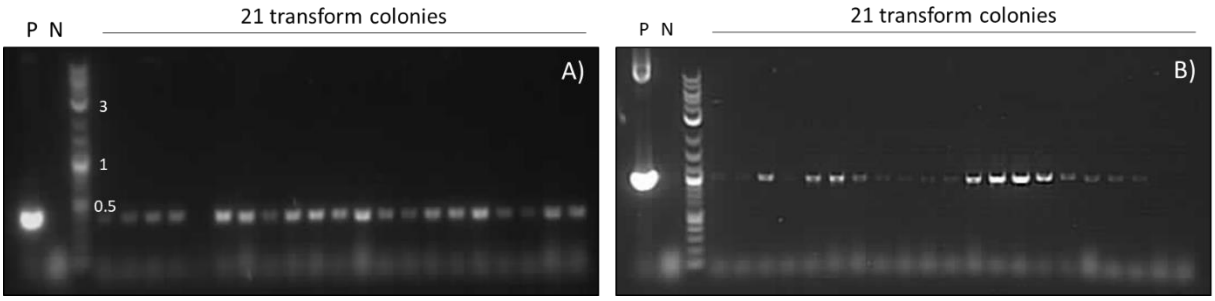


Figure 3.7: Agarose electrophoresis gel images of colony PCR carried out using two sets of primers, one set was used to amplify a region in the GnTI gene (A) and the other set was used to amplify a region in the GalT gene (B). P is positive control – pGlycoSwitch-GnTI for A and pGlycoSwitch-GalT for B. N is negative control, this was the parental strain SuperMan 5 without either plasmid introduced. Molecular weight markers measured in kilobase pairs (kbp). Expected band sizes for PCR of GnTI is 364 bp and GalT is 1024 bp.

Colony PCR analysis show that of the twenty-one colonies analysed only one does not contain pGlycoSwitch-GnTI (colony 5), and two do not contain pGlycoSwitch-GalT (colonies 20 and 21) – this represents a transformation efficiency of ((18/21)*100=) 86%.

Nine colonies, that had been shown to contain both pGlycoSwitch-GnTI and GlycoSwitch–GalT, were cultured in 25 mL BMGY growth media in baffled shake flasks for 60 hours and induced to express rFH, by transfer of each culture to 5 mL BMMY induction media for 96 hours, as outlined in section 3.2.3.1. After 96 hours, the OD_{600} had reached between 95 - 120, and culture supernatants were harvested, and they were then subjected to SDS-PAGE and ECL lectin-based western analysis to detect for production of rFH and galactose-containing hybrid-type glycans, respectively (fig. 3.8).

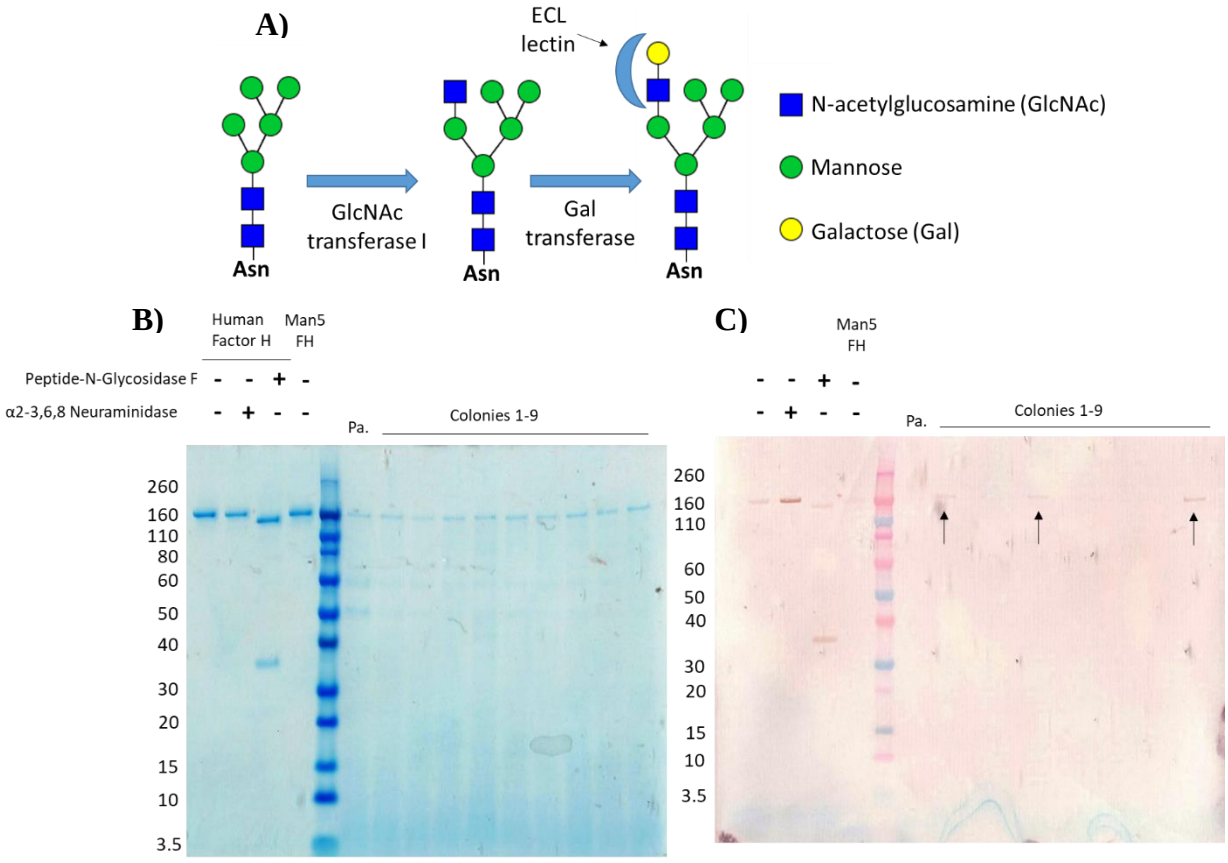


Figure 3.8: A) Schematic diagram of the reaction catalysed by the genes encoded by the co-introduced plasmids pGlycoSwitch-GnTI and pGlycoSwitch-GalT. GnTI (GlcNAc transferase) converts a five mannose glycan into a hybdrid glycan with a single GlcNAc terminal antenna. GalT (Gal transferase) then catalyses the addition of galatose. The resulting LacNAc antennae (Gal-GlcNAc) can be bound by ECL lectin in a western blot assay.SDS-PAGE (B) and ECL lectin-based western blot for N-acetyllactosamine (C) analysis of the expression of rFH in *P. pastoris* strain SuperMan 5 transformed with two GlycoSwitch plasmids pGlycoSwitch-GnTI and pGlycoSwitch-GalT. Left of the markers are different glycoforms of FH – in order from left to right, sialylated human FH, desialylated human FH, deglycosylated human FH and purified rFH expressed in SuperMan 5. Pa. is crude supernatant from SuperMan5 induced to express rFH. Black arrows on ECL lectin western blot indicate lanes that contain bands. Molecular weight markers are measured in kilodaltons (kDa). The molecular weight of α(2-3,6,8)-neuraminidase is 43 kDa and PNGase F is 36 kDa.

The lectin-based western blot (fig. 3.8 C) unexpectedly detected weak bands in the three negative controls, namely non-enzymatically treated human FH (lane 2), Endo H_f-deglycosylated human FH (lane 3), and SuperMan 5-produced (untreated) rFH (lane 4). Critically, though, the signal for N-acetyllactosamine in the sialidase-treated human FH sample is substantially stronger than that for the untreated FH sample. These results suggest that, despite some background signals, this method can clearly distinguish between glycoproteins with and without N-acetyllactosamine-terminated glycans.

All nine GlycoSwitch-engineered colonies produced detectable quantities of rFH (fig 3.8 B). However, only in colonies 2, 4 and 9 did the rFH band give a positive signal according to the ECL lectin-based N-acetyllactosamine analysis (fig 3.8 C). It seems therefore that only the

cells in these three colonies have co-expressed the *GnT1* and *GalT* genes. The glycoprofile of supernatant from one of these – colony 9 – was analysed by MALDI-TOF mass spectrometry (Figure 3.9 B).

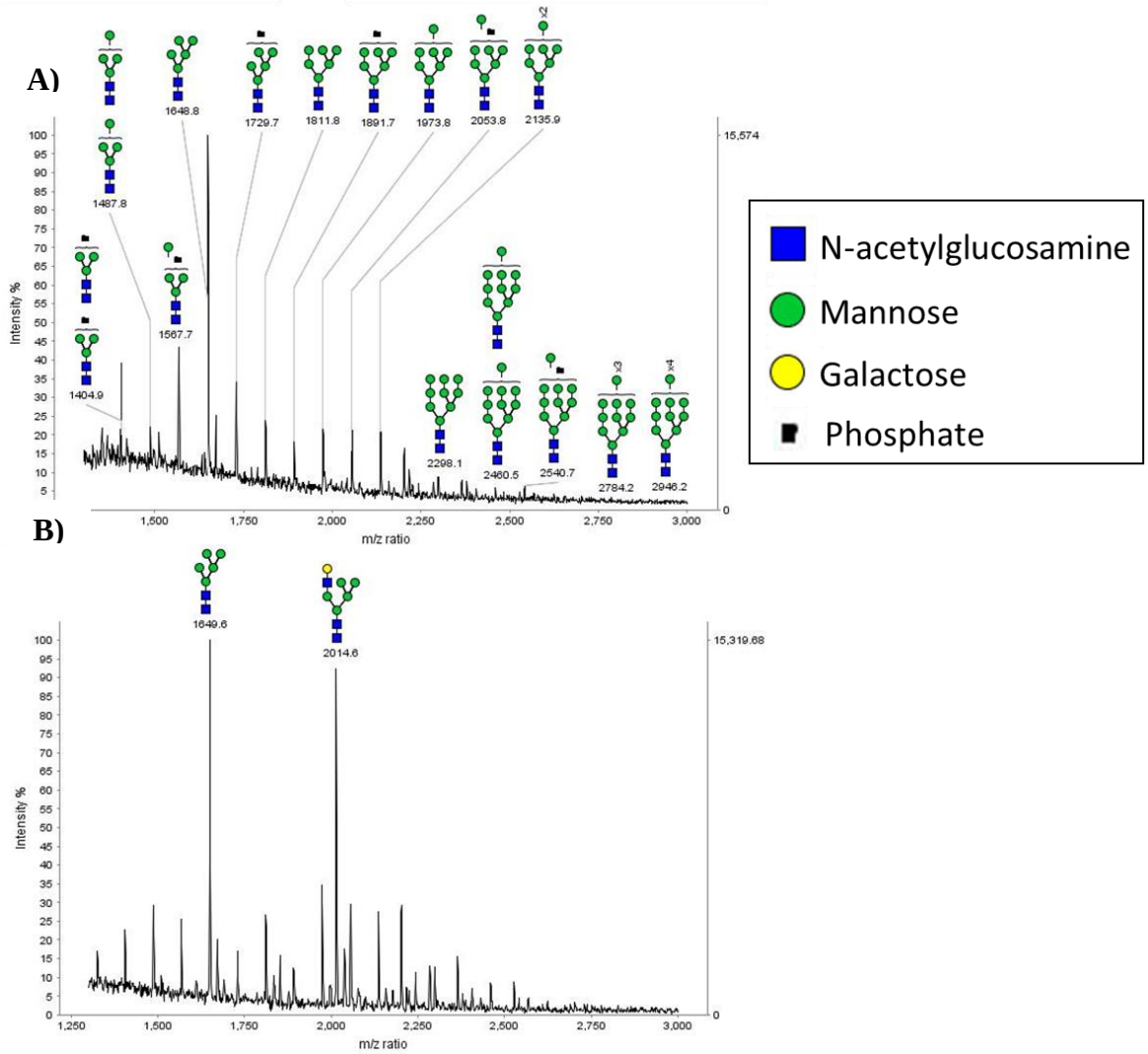


Figure 3.9: Glycoprofile based on MALDI-TOF mass spectrometry of crude supernatant from the parental strain of SuperMan 5 (A) and the glycoprofile based on MALDI-TOF crude supernatant from colony 9 after transformation of the parental strain SuperMan 5 with GlycoSwitch plasmids encoding GnTI and GalT (B).

A peak corresponding to the mass (2,014 Da) of the hybrid galactose-terminated glycan is present in the mass spectrum of colony 9 following transformation with pGlycoSwitch-GnTI and -GalT (fig. 3.9 B). This peak is not present in SuperMan 5 crude supernatant before transformation with these plasmids (fig 3.9 A). Along with the western blot, these results confirm that colony 9 has been successfully genetically modified to biosynthesise hybrid-type, galactose-terminated glycans, $Gal_1GlcNAc_2Man_5GlcNAc_2$ – this strain was named HyGal (for Hybrid-glycan, galactose terminal). Glycoanalysis of the supernatants from colonies 1 and 4, which also gave signals in ECL lectin western analysis, likewise featured a

peak at 2,014 Da, but in this case, the relative peak intensities were much lower (data not shown).

While very encouraging, these results also reveal some minor heterogeneity in the glycoprofiles. In addition to the target, hybrid glycan there are also high-mannose species. The most abundant of these high-mannose species is the five-mannose glycan, $Man_5GlcNAc_2$.

3.3.5 Further glycoengineering of *P. pastoris* to produce complex-type, galactose terminal glycans

HyGal colony 9 was used as the platform for further modifications to the glycosylation biosynthetic pathway of *P. pastoris*. The final two GlycoSwitch plasmids (fig. 3.6 C and D) encoding ManII and GnTII were introduced, together, into HyGal by transformation. Expression of these two enzymes, ManII and GnTII, should build a second N-acetyllactosamine antennae on the hybrid-type glycan, $Gal_1GlcNAc_1Man_5GlcNAc_2$, facilitating the conversion to a galactose terminal, diantennary complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$.

All four of the GlycoSwitch plasmids (fig. 3.6) contain the constitutive GAP promoter (GAPp) and all four GlycoSwitch plasmids were linearised in the GAPp for targeted genomic integration at the *P. pastoris* endogenous GAPp site (fig. 3.10). Therefore, for each plasmid-genome integration event, an exogenous GAPp sequence will be introduced, resulting in multiple homologous repeat sequences. Spontaneous homologous recombination could result in loss of one or more pGlycoSwitch plasmids previously integrated into the genome.

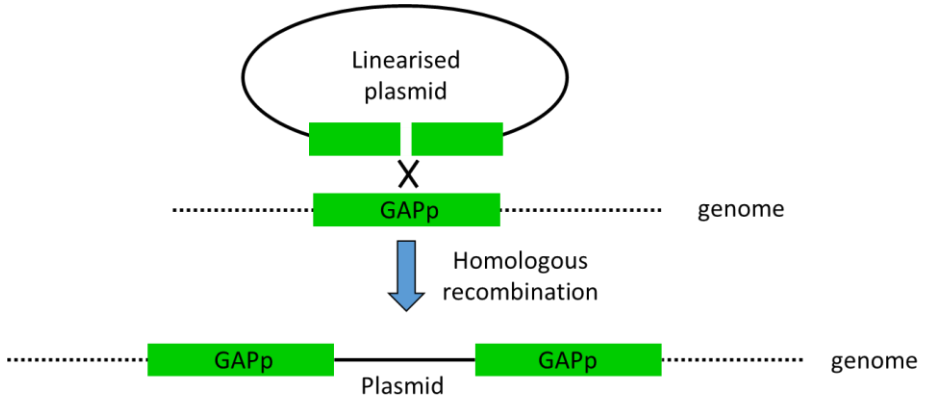


Figure 3.10: Schematic diagram of GlycoSwitch plasmid integration *via* homologous recombination between the exogenous GAPp in the plasmid and the endogenous GAPp in the *P. pastoris* genome. All four GlycoSwitch plasmids (encoding GnTI, GalT, ManII and GnTII) were linearised in the exogenous GAPp and directed for integration at the endogenous, genomic GAPp. Thus, for each plasmid, an additional GAPp sequence is introduced into the genome. The HyGal strain will have three GAPp sequences, the original endogenous GAPp and the two GAPp contained in pGlycoSwitch-GnTI and -GalT. SuperGal, which is transformed with all four GlycoSwitch plasmids, will have five GAPp sequences, the endogenous GAPp as well as the exogenous GAPp sequences contained in pGlycoSwitch-GntI, -GalT, -GnTII and -ManII.

Consequently, it was necessary to not only detect for colonies which had taken up the new plasmids (pGlycoSwitch-ManII and -GnTII), but also to detect for the two GlycoSwitch plasmids which had previously been introduced (pGlycoSwitch-GnTI and -GalT) (fig. 3.11).

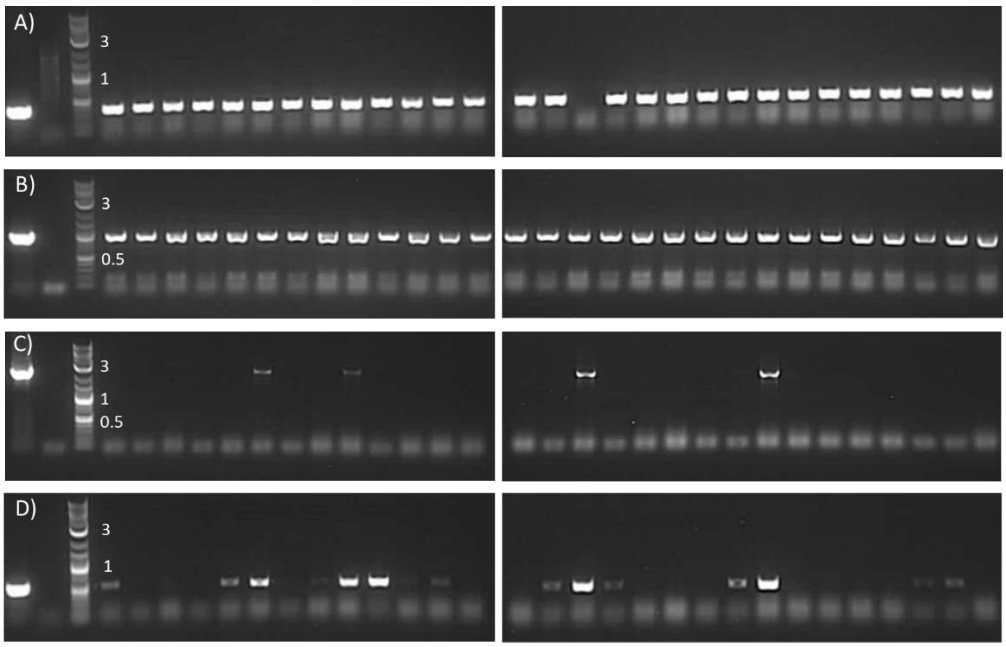


Figure 3.11: Agarose electrophoresis gel images of colony PCRs carried out using four sets of primers, the first two are targeted toward amplifying parts of coding region of *GnTI* (A) and *GalT* (B), which had previously been integrated into the *P. pastoris* genome to generate the strain HyGal. The other two sets of primers are targeted toward amplifying parts of the coding regions of the genes *ManII* (C) and *GnTII* (D) on the newly introduced set of GlycoSwitch vectors. P is positive control – pGlycoSwitch-GnTI for A, -GalT for B, -ManII for C and –GnTII for D. N is negative control, this was the parental strain SuperMan 5 without any of the above four GlycoSwitch plasmids introduced. Molecular weight markers measured in kilobase pairs (kbp). Expected band sizes for PCR of GnTI is 364bp, GalT is 1024bp, ManII is 2376bp and GnTII is 534bp.

Thus, ensuring that the colonies carried forward for rFH expression contain all 4 GlycoSwitch plasmids.

Colony PCRs of HyGal co-transformed with pGlycoSwitch-ManII and – GnTII show that, of twenty-nine colonies analysed, only one has lost the pGlycoSwitch-GalT plasmid. Just five colonies contained the introduced plasmid pGlycoSwitch-ManII and fifteen contained the other introduced plasmid pGlycoSwitch-GnTII. Of these fifteen, only four colonies contain all four plasmids *i.e.* the previously introduced plasmids pGlycoSwitch-GnTI and -GalT (responsible for producing the hybrid glycan) plus the two newly co-introduced plasmids.

The four colonies which contain all four GlycoSwitch vectors were cultured in 25 mL BMGY growth media in baffled shake flasks for 60 hours and induced to express rFH, by transfer of each culture to 5 mL BMMY induction media for 96 hours, as outlined in section 3.2.3.1. After 96 hours, the OD_{600} had reached between 90 - 115, and culture supernatants were harvested, some of the culture supernatant from each of the four colonies was treated with the endoglycosidase Endo H_f and all culture supernatant was then subjected to SDS-PAGE and ECL lectin-based western analysis to detect for production of rFH and galactose-containing complex-type glycans, respectively (fig. 3.12). As outlined Chapter 1 section 1.5.1, Endo H_f has substrate specificity for high-mannose glycans and, importantly, hybrid-type glycans. But crucially, it cannot cleave complex-type glycans at an appreciable rate (fig. 3.12 B). Therefore, our rFH, which has been modified to include complex-type glycosylation, should not be deglycosylated by endo H_f. Thus, we expect (i) that after Endo H_f treatment there will be no molecular weight shift by SDS-PAGE and (ii) that there will be no loss in signal according to ECL lectin-based western blot.

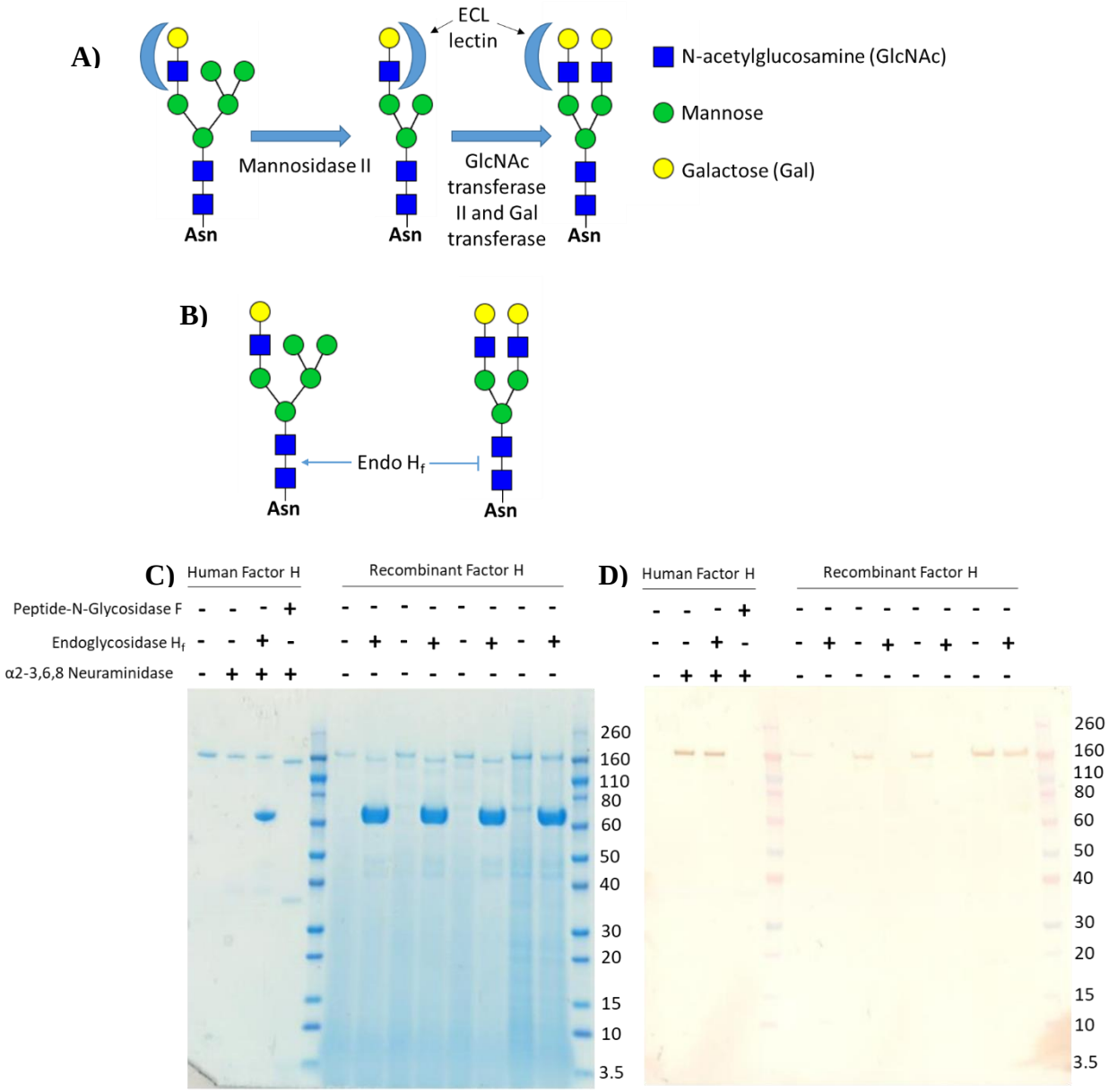


Figure 3.12: A) Schematic diagram of the reaction scheme catalysed by the co-introduced plasmids pGlycoSwitch-ManII and pGlycoSwitch-GnTII when transformed into *P.pastoris* strain HyGal. The gene encoded by pGlycoSwitch ManII, mannosidase II, cuts off two mannose from a hybrid glycan. GlcNAc transferase II, encoded by pGlycoSwitch-GnTII, then catalyses the addition of N-acetylglucosamine which can then be galactosylated by galactose transferase. B) Schematic diagram showing that Endo H_f can cleave the parental hybdrid glycan but not the engineered complex-type glycan. SDS-PAGE (C) and ECL lectin western blot (D) analysis of the expression of rFH in *P. pastoris* strain HyGal *CFH* transformed with pGlycoSwitch-ManII and –GnTII. Left of the markers are different glycoforms of FH – in order from left to right, sialylated human FH, desialylated human FH, desialylated human FH treated with Endo H_f and desialylated human FH treated with PNGase F. Culture supernatant were either untreated or treated with Endo H_f prior to loading onto SDS-PAGE gel. Molecular weight markers are measured in kDas. The molecular weight of α(2-3,6,8)-neuraminidase is 43kDa, Endo H_f is 70kDa and PNGase F is 36kDa.

Looking at the controls in figure 3.12 C and D, there is no molecular weight shift or loss of signal for neuraminidase-treated human FH after treatment with Endo H_f, as assessed by SDS-PAGE and ECL lectin-based western blot, respectively. This confirms that Endo H_f cannot cleave the complex-type glycans present on human FH. However, when treated with PNGase F there is a molecular-weight shift according to SDS-PAGE and a loss of signal in the ECL lectin-based western analysis. These results confirm that complex-type glycans are cleavable by PNGase F but that complex-glycans are not substrates of Endo H_f.

Looking at the four strains of HyGal *CFH* transformed with pGlycoSwitch-ManII and pGlycoSwitch -GnTII, three show a reduction in molecular weight, by SDS-PAGE, and a loss of signal by ECL lectin-based western blot for N-acetyllactosamine. This suggests that, unfortunately, the glycosylation biosynthetic pathway has not been modified beyond the hybrid-type, galactose terminal glycan, $Gal_1GlcNAc_1Man_3GlcNAc_2$. The fourth colony, however, shows no loss of molecular weight according to SDS-PAGE, and no loss of signal in the western blot. Thus, in this one colony was the glycosylation pathway of HyGal *CFH* modified to go beyond the hybrid-type glycan. This strain is named SuperGal *CFH* from here on.

To investigate the glycoprofile of SuperGal *CFH* colony 4 further, crude supernatant was enzymatically deglycosylated using PNGase F, the glycans purified from crude supernatant using BlotGlyco purification and analysed by MALDI-TOF mass spectrometry (fig. 3.13), as detailed in section 3.2.6.

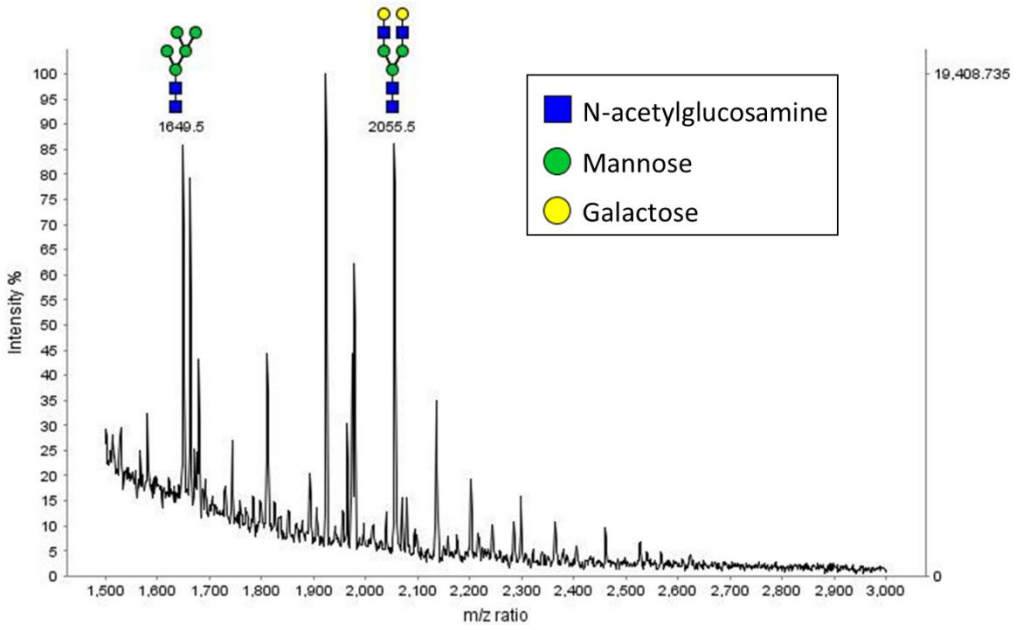


Figure 3.13: Glycoprofile MALDI-TOF mass spectrum of crude supernatant from the glycoengineered *P. pastoris* strain SuperGal *CFH* colony 4.

The mass spectrum for SuperGal *CFH* shows a peak corresponding to the mass, 2055.5 Da, of diantennary digalactosylated complex-type glycans, $Gal_2GlcNAc_2Man_3GlcNAc_2$. This, together with Endo H_f insensitivity, assessed by SDS-PAGE and western blot, suggest that the glycosylation pathway of *P. pastoris* has indeed been modified to biosynthesise humanised, complex-type glycosylation. However, as was observed with the HyGal *CFH* strain, there is some glycan heterogeneity present; the five-mannose glycan is probably as abundant as the complex-type, diantennary digalactosylated glycan.

3.3.6 Scaled-up production of recombinant factor H with humanised glycosylation in *P. pastoris* glycoengineering strain SuperGal *CFH*

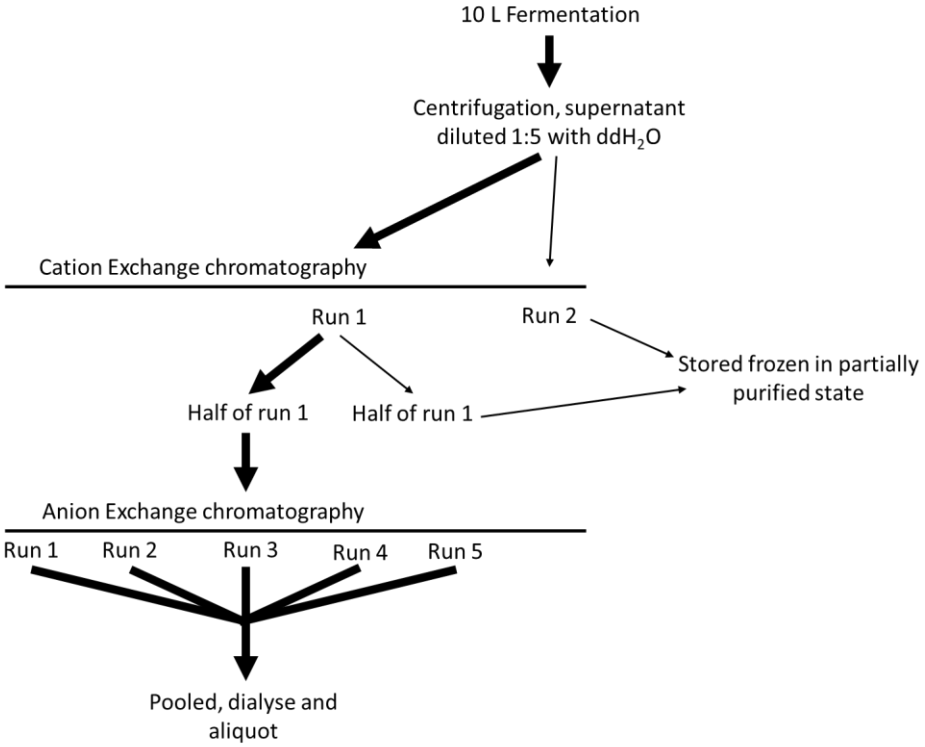


Figure 3.14: Flow diagram of the purification process of recombinant FH expressed in glycoengineered strain SuperGal *CFH*. After 10 L fermentation of SuperGal *CFH*, the supernatant was harvested by centrifugation (6,000 g for 30 minutes). The supernatant was diluted 1 in 5 with distilled water. In cation exchange run 1, approximately half of the diluted culture supernatant was applied to a cation exchange chromatography column and eluted. In cation exchange chromatography run 2, the second half of the diluted culture supernatant was applied to the same cation exchange column and eluted. Eluted fractions from cation exchange chromatography run 1 and run 2 were assessed for purity by SDS-PAGE, appropriate fractions were pooled and dialysed against buffer for a second purification step, anion exchange chromatography. Half of cation exchange chromatography run 1 and all of run 2 pooled and dialysed fractions were stored frozen. The other half of cation exchange chromatography run 1 was subjected to a further purification step, it was divided into five lots and each lot was applied separately to, and eluted from, an anion exchange chromatography column. The eluted fractions from the five anion exchange chromatography runs were assessed for purity by SDS-PAGE and appropriate fractions were pooled, dialysed against PBS, pH7.4, aliquoted and stored frozen.

A 10 L batch fermentation of SuperGal *CFH* was carried out (fig. 3.14). The recombinant glycoprotein-producing strain was cultured, in two 5 L baffled shake flasks containing 1.5 L BMGY growth media, for 72 hours at 30°C. The shake flask cultures were then used to inoculate a 10 L BioFlow 4500 fermenter. The culture was grown by glycerol-fed batch fermentation and expression of recombinant FH was induced by switching to a methanol/glycerol continuous feed for four days. 8 L of culture supernatant was harvested by centrifugation (6,000 g, 30 minutes), diluted 1 in 5 with double distilled water and recombinant FH was purified from the dilute supernatant according to the same procedure used to purify recombinant FH from the *P. pastoris* strain SuperMan 5, *i.e.* using cation-exchange chromatography followed by anion-exchange chromatography (figs 3.15 and 3.16).

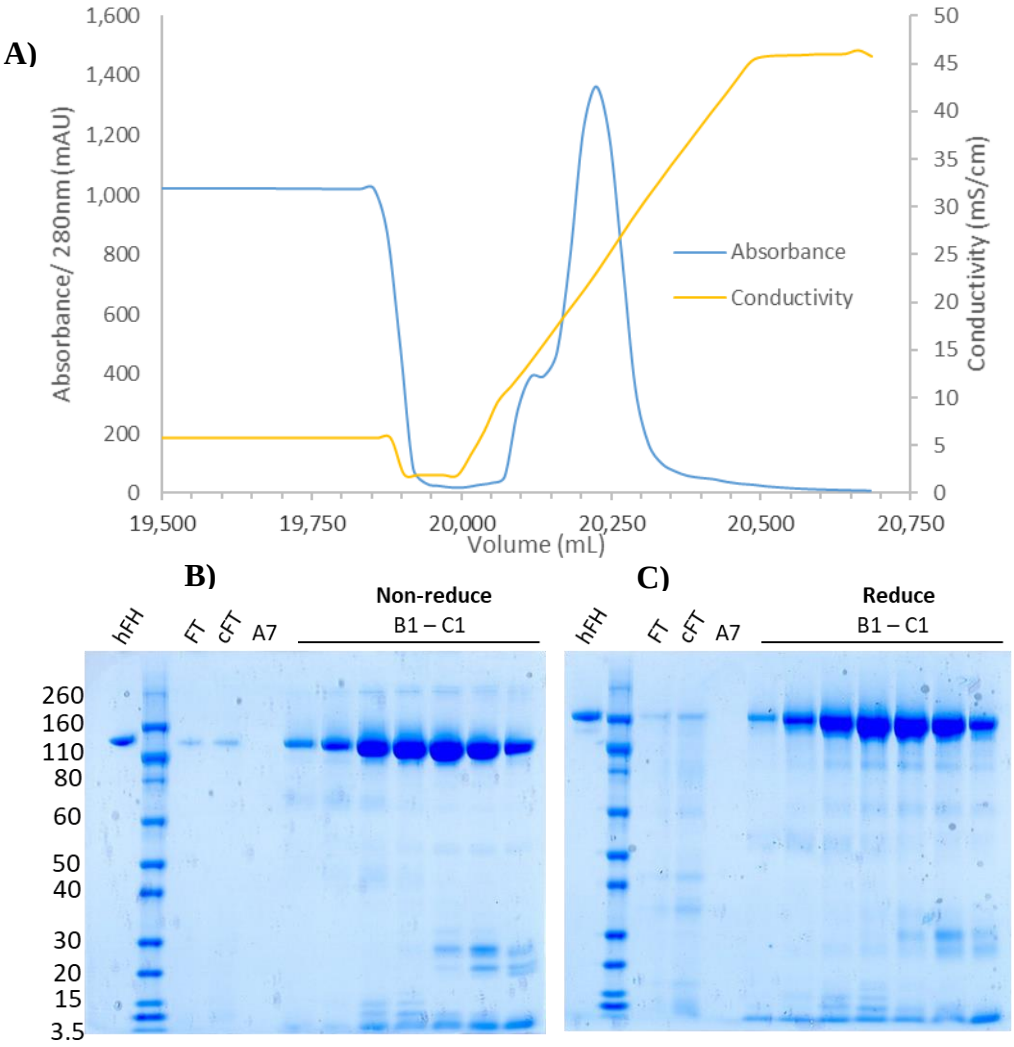


Figure 3.15: Cation-exchange chromatogram (A) and peak fractions run either non-reduced (B) or reduced (C) on SDS-PAGE gel. hFH is human FH, FT is flow through, cFT is 10x concentrated flow through, A7 (20,034.6 mL) is a wash fraction and B1-C1 (20,118.6 – 20,286.6 mL) are cation-exchange chromatography peak fractions. Molecular weight markers are measured in kDa. Every second fraction from B1 is loaded onto the gel.

Due to the large volume of diluted culture supernatant, two successive cation-exchange runs were carried out. Figure 3.15 shows the first of these two runs. Less culture supernatant was used in the second run (this accounts for the lower protein concentration in run 2, see table 3.1).

Recombinant FH eluted as a single peak at an elution volume of approximately 20.25 L, with a shoulder on the peak at elution volume approximately 20.1 L (fig 3.15 A). Avoiding the peak shoulder as much as possible, peak fractions B1-C1 (20,118.6 – 20,286.6 mL) were taken and run on SDS-PAGE to assess purity (fig 3.15 B and C).

Fractions B3-B11 (20,146.6 – 20,258.6 mL), for both runs, were considered to have the highest concentration of rFH relative to impurities. So, these fractions were pooled,

separately, and dialysed against anion-exchange running buffer (pH 9.5) and the concentration and nucleic acid contamination was assessed by absorbance at 280 nm and the ratio of 260 nm to 280 nm (table 3.1).

Half of the pooled and dialysed eluent from cation exchange chromatography run 1 and all of the pooled, dialysed eluent from cation exchange chromatography run 2 was stored frozen at - 80°C whilst the other half of run 1 pooled and dialysed eluent was further purified by anion-exchange chromatography (fig. 3.16). Five separate anion-exchange runs were carried out. In the first run, 6 mL of cation exchange purified protein was loaded onto a 6 mL Resource Q strong anion exchanger chromatography column. Protein was eluted in a linear gradient of 0 – 0.5 M NaCl and 1.5 mL fractions were collected and assessed for purity by SDS-PAGE (fig. 3.16). For the remaining four anion exchange chromatography runs 12 mL of cation exchange purified protein was loaded onto the Resource Q column (chromatograms and SDS-PAGE gels not shown for anion exchange runs 2-5). Thus, a total volume of 55 mL eluent from cation exchange chromatography run 1 was subjected to further purification by anion exchange chromatography.

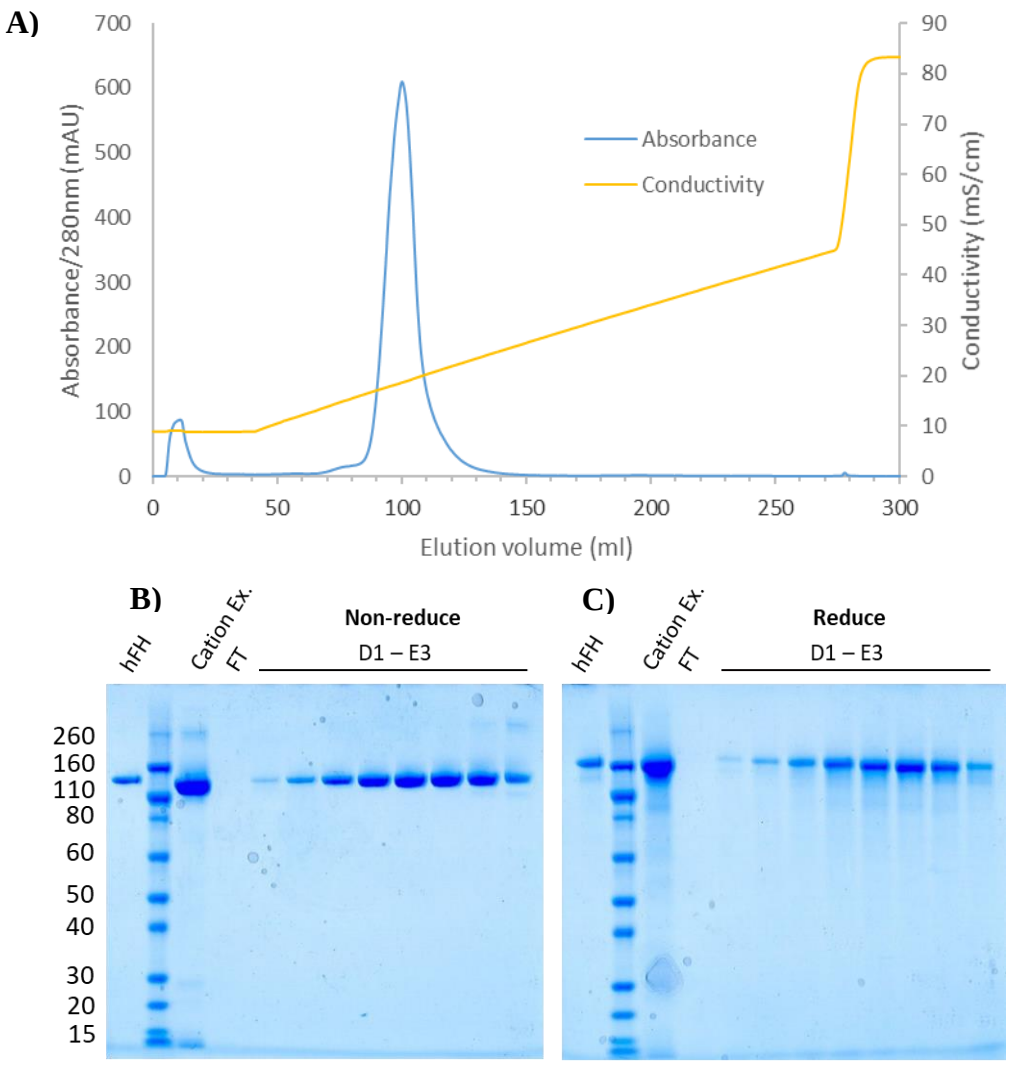


Figure 3.16: Anion exchange **c**hromatogram (A) and peak fractions run non-reduced (B) and reduced (C) on SDS-PAGE gel. hFH is human FH, Cation Ex. is pooled fractions from cation exchange chromatography run, FT is flow through and D1-E3 (85.4 – 106.4 mL) are anion exchange chromatography peak fractions. 1.5 mL fractions were collected. 1 µL of every second fraction from D1 was mixed with 13 µL LDS loading buffer and loaded onto the gel. Molecular weight markers are measured in kDa.

Recombinant FH eluted in a single peak (fig. 3.16) at approximately 100 mL elution volume; note, there is a slight shoulder at approximately 75 mL elution volume (fig 3.16 A). Avoiding the slight shoulder, peak fractions D1 – E3 (85.4 – 106.4 mL) were run on non-reducing (fig 3.16 B) and reducing (fig 3.16 C) SDS-PAGE to assess for purity. Fractions D5-E1 (91.4 – 103.4 mL) for run 1 (fig 3.16) and D3-E3 of runs 2-5 (data not shown) were considered to have the highest concentration and purity of rFH, these were pooled together and dialysed against PBS (pH 7.4).

Step:	Cation exchange chromatography				Absorbance							
Run	Pooled Fractions (elution volume ml)	Molecular weight (kDa)	Extinction coefficient ($M^{-1}cm^{-1}$)	Pathlength (cm)	A320	A260	A280	Concentration (μM)	Concentration (mg/mL)	A260/A280	Volume (mL)	Protein yeild (mg)
1	B3 - B11 (20,146.6 – 20,258.6)	150	246,800	1	0.64	6.45	9.07	34.16	5.12	0.71	~110	560
2				0.2	0.04	0.76	0.96	18.64	2.79	0.79		310
Step:	Anion exchange chromatography											
Run												
1	D5 - E1 (85.4 – 106.4 mL)	150	246,800	0.2	0.01	0.31	0.49	9.89	1.48	0.62	12	17.8
2	D3 - E3				0.00	0.42	0.72	14.61	2.19	0.58	18	39.4
3					0.01	0.43	0.73	14.57	2.18	0.59		39.3
4					0.00	0.43	0.72	14.40	2.16	0.60		38.9
5					0.01	0.41	0.67	13.37	2.01	0.61		36.1
Step:	Pooled anion run 1-5 and after dialysis against PBS, pH7.4	150	246,800	0.2	0.026	0.039	0.674	3.240	1.969	0.577	84	165.4

Table 3.1: Absorbance readings at 280, 260 and 320 nm of fractions from cation exchange chromatography runs 1 and 2 and anion-exchange chromatography runs 1-5 and runs 1-5 after pooling and dialysis against PBS. The absorbance reading at 280 nm was used to estimate protein concentration and the ratio of absorbance at 260 nm to 280 nm was used to assess for nucleic acid impurities.

Table 3.1 shows that after cation-exchange chromatography the protein yield in run 1 eluent is 560 mg. 55 mL (55 mL*5.12 mg/mL = 281.6 mg) of this was subjected to anion exchange chromatography and that after pooling peak fractions from all five anion-exchange chromatography runs and dialysis against PBS (bottom row table 3.1) the yield of rFH is 165.4 mg. Hence ((165.4 mg / 281.6 mg)*100 =) 58.7% of protein from cation-exchange chromatography was recovered after anion-exchange chromatography.

Protein from half of run 1 and all of run 2 of the cation-exchange chromatography was not further purified by anion-exchange chromatography and was instead stored frozen at -80 °C for further purification later when stocks of high purity SuperGal *CFH* expressed recombinant FH are depleted. The protein content of the remaining pooled fractions from cation exchange chromatography run 1 was approximately 280 mg and after run 2 was approximately 310 mg. Assuming a similar yield would be obtained for a subsequent anion-exchange step, it would produce 165.4 mg and 182.0 mg of purified rFH, respectively. Thus, it is fair to claim a total yield of (165.4 mg + 165.4 mg + 182.0 =) 512.8 mg purified rFH from this 10-L fermentation or about ~50 milligram per litre cell culture. This yield compares favourably with attempts by other groups to produce glycoengineered rFH, for example recently Michelfelder *et al*. (2017) report 1 milligram per litre cell culture of recombinant human FH with terminal GlcNac diantennary glycans, $GlcNAc_2Man_3GlcNAc_2$, expressed in a glycoengineered strain of moss.

Additionally, there was a substantial amount of rFH in the flow through (~40 L) after cation-exchange chromatography (fig 3.15 B and C) – suggesting that the capacity of the column was exceeded. Purification using a higher volume resin, or re-applying the flow through, could have increased yields of FH quite substantially.

3.3.7 Further glycoengineering: Removing high-mannose glycans

Figure 3.13 indicates that the glycoprofile of SuperGal *CFH* retains some heterogeneity and that the di-galactosyl species, $Gal_2GlcNAc_2Man_3GlcNAc_2$, is only about as abundant as the five-mannose species, $Man_5GlcNAc_2$, with other less abundant species corresponding to higher-mannose, yeast-type glycans. The ultimate aim of this work was to compare the function of recombinant human FH carrying human-like glycans on the one hand, with FH purified from blood plasma, on the other. The contaminating presence of five-mannose and high-mannose glycans would make this comparison difficult to interpret. It could have a particularly negative impact on *in vivo* studies in mouse models as mannose-rich glycans are highly immunogenic.

It was therefore necessary to remove the mannose-rich glycans from the sample of purified rFH that had been produced in the *P. pastoris* strain SuperGal. This was attempted by first incubating a batch of this rFH (33 mL at 2 mg/mL) with MBP-tagged Endo H_f (100 kU at 37 °C for 2 hours) to release mannose-rich glycans. Subsequently, mannose-rich glycans and MBP-tagged Endo H_f were removed by MBP-affinity chromatography (fig. 3.17). Glycosylation before and after this Endo H_f treatment step was compared to evaluate success.

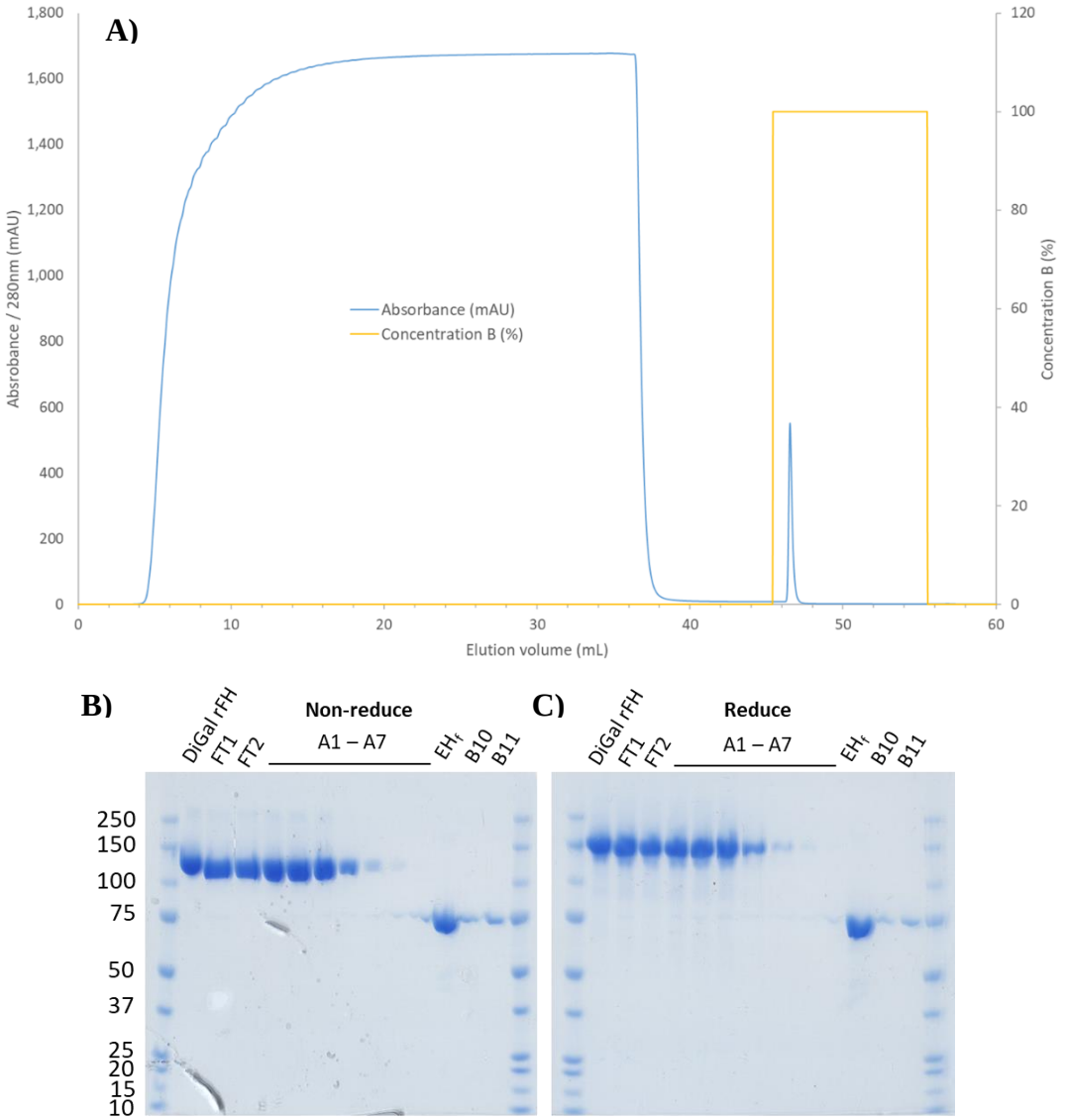


Figure 3.17: MBP-tag affinity chromatogram (A) and peak fractions run non-reduced (B) and reduced (C) on SDS-PAGE gel. DiGal is purified rFH expressed in *P. pastoris* strain SuperGal *CFH* but not treated with Endo H_f, FT is flow through (elution volume 4 – 35 mL), A1-A7 (35.8 – 38.8 mL) are wash fractions, EH_f is pure Endo H_f and B10 (45.9 – 46.4 mL) and B11 (46.4 – 46.9 mL) are peak fractions. FT1 and FT2 were collected as 15 mL volumes and wash and peak fractions were collected as 0.5 mL volumes. Molecular weight markers are measured in kDa.

The MBP-tag affinity chromatogram (fig 3.17 A) shows a large absorbance at 280 nm during sample application (flow through) and washing, with a small, sharp peak absorbance upon washing with 10 mM maltose (for elution). The SDS-PAGE gels (fig 3.17 B and C) show that the bulk of the protein content of the flow through and wash fractions A1-A6 (elution volume 35.8 – 38.3 mL) is rFH, and the small elution peak consists solely of Endo H_f. However, SDS-PAGE also show that there is a small amount of Endo H_f contaminating all fractions. FT1 and 2 had the lowest contaminating level of Endo H_f so were pooled, dialysed against PBS, pH 7.4, the concentration of protein assessed by UV-VIS spectrometry (table

3.2) and stored for future use. The remaining FH containing fractions were discarded because they had relatively high Endo H_f contamination and were only 0.5 mL in volume.

Sample	Molecular weight (kDa)	Extinction coefficient ($M^{-1}cm^{-1}$)	Pathlength (cm)	A320	A260	A280	Concentration (µM)	Concentration (mg/mL)	A260/A280	Volume (mL)	Protein yeild (mg)
SuperGal rFH	150	246,800	0.2	0.015	0.407	0.686	3.355	2.039	0.593	33	67.3
SuperGal rFH and Endo H_f				0.017	0.409	0.683	3.33	2.023	0.599	33	66.7
Pooled and dialysed FT1 and FT2				0.014	0.329	0.546	2.66	1.616	0.603	34	54.9

Table 3.2: Absorbance readings at 280, 260 and 320 nm of rFH expressed in SuperGal *CFH* before and after endo H_f treatment and after application to MBPTrap affinity column. The absorbance reading at 280 nm was used to estimate protein concentration and the ratio of absorbance at 260nm to 280nm was used to assess nucleic acid impurities.

Next the extent to which high-mannose (and hybrid) glycans had been removed from rFH expressed in SuperGal was assessed by glycan MALDI-TOF mass spectrometry (fig 3.18).

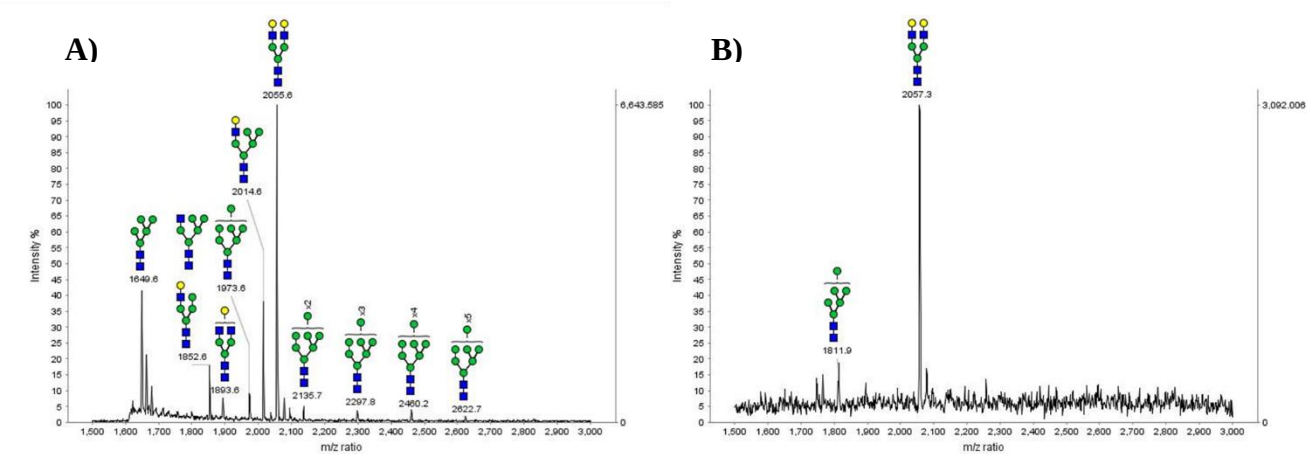


Figure 3.18: Glycan MALDI-TOF mass spectra for glycans cleaved and purified from rFH expressed in *P. pastoris* strain SuperGal *CFH* before (A) and after (B) treatment with the endoglycosidase Endo H_f.

Before treatment with Endo H_f the glycoprofile of purified rFH expressed in SuperGal *CFH* (fig 3.18 A) contains high-mannose and hybrid-type glycans as well as the target glycoengineered human complex-type diantennary digalactosylated glycan. After treatment with Endo Hf (fig 3.18 B) the glycoprofile is fairly homogenous, consisting almost exclusively of the diantennary digalactosylated complex-type glycan. There is also a small peak potentially corresponding to a six-mannose glycan – alternatively, this peak could be an artefact of the high background in the mass spectrum (compare backgrounds of fig 3.15 B with low background of fig 3.15 A).

3.4 Discussion

By implementing a two-stage approach, and after first confirming that the SuperMan5 *CFH* does indeed express recombinant FH carrying five mannose N-linked glycans, this chapter has outlined the methodology used to humanise the glycan biosynthesis pathway of *P. pastoris* using GlycoSwitch technology. In the first stage, the glycosylation biosynthetic pathway of *P. pastoris* was modified to produce a hybrid-type, galactose terminal glycan, $Gal_1GlcNAc_1Man_5GlcNAc_2$. Building upon this, in the second stage, the hybrid-type glycan was converted to a diantennary galactose terminal complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$.

However, the glycosylation profile of SuperMan 5 *CFH* was heterogenous and produced high-mannose glycans that could not be humanised. This heterogeneity carried through to the glycoengineered strain HyGal *CFH* and SuperGal *CFH*.

Multiple studies in the literature report expression of glycoproteins in glycoengineered *P. pastoris* (Hamilton *et al*., 2006, Li *et al*., 2006, Jacobs *et al*., 2009, Ha *et al*., 2011, Jiang *et al*., 2011, Gomathinayagam *et al*., 2012, Cukan *et al*., 2012, Gong *et al*., 2013 and Gomathinayagam *et al*., 2015). However, just some of these report heterogeneity in the glycosylation profile of recombinant glycoproteins expressed in glycoengineered *P. pastoris* (Jacobs *et al*., 2009 and Ha *et al*., 2011). The others report homogenous glycosylation profiles composed exclusively of complex-type glycans (Hamilton *et al*., 2006, Li *et al*., 2006, Jiang *et al*., 2011, Gomathinayagam *et al*., 2012, Cukan *et al*., 2012, Gong *et al*., 2013 and Gomathinayagam *et al*., 2015).

First introduced in the last paragraph of Chapter 1 section 1.4.1, GlycoFi Inc were the company that first developed *in vivo* sialylation technology in *P. pastoris* (Hamilton *et al.,* 2006) and were sold to Merck-Serono in 2008 (Roumeliotis, 2006).

A common thread that links the literature that does not report glycosylation heterogeneity is that they are all in association with GlycoFi Inc. The GlycoFi Inc. glycoengineered *P. pastoris* platform differs in two key ways from the GlycoSwitch *P. pastoris* platform used in this study. Firstly, the GlycoFi Inc. platform is capable of *in vivo* sialylation. Secondly, the GlycoFi Inc. platform has incorporated more deletions of endogenous *P. pastoris* glycan modification enzymes. In addition to the *Δoch1*, common to both the GlycoSwitch and GlycoFi Inc. technology, GlycoFi Inc have also deleted the phosphomannosyltransferase gene *PNO1* and the β-mannosyltransferase 2 gene *BMT2*. Both of these enzymes are responsible for catalysing the transfer of substrate monosaccharides onto the core N-linked glycan in yeast (Miura *et al*., 2004 and, Hopkins *et al*., 2011) and, therefore, deletion of

these two genes, together with *Δoch1*, appears to facilitate a strain of glycoengineered *P. pastoris* with a very homogenous glycosylation profile.

The work in this study has, therefore, identified a limitation to the GlycoSwitch technology, revealing a significant contamination of high-mannose glycans on proteins produced by the engineered strain. The mannose-binding lectin pathway of the complement system in mice may lead to rapid clearance of rFH bearing high-mannose N-linked glycans, therefore Endo H_f enzymatic removal of high-mannose glycans was essential for further *in vivo* studies.

Regardless of the limitations of the GlycoSwitch technology, the work in this study satisfies the two principle aims set out in chapter 1 section 1.7. Firstly, recombinant FH has been expressed in *P. pastoris* with humanised, complex-type diantennary digalactosylated glycans. Secondly, therapeutically useful quantities of recombinant FH were met by achieving production yields of 50 milligrams per litre cell culture.

Chapter 4

In vitro sialylation of human recombinant factor H expressed in the glycoengineered *P. pastoris* strain SuperGal

4.1 Overview

4.1.1 Selecting a sialyltransferase

The glycosylation of hFH and the need for sialylation of therapeutic glycoproteins in general were discussed in the Introduction (sections 1.1.3, 1.1.4 and 1.2.2). It was shown (Chapter 2) that plasma-purified hFH is sialylated but, the nature of the glycosidic bond that links sialic acid to galactose is unknown. Up to 20 sialyltransferases exist in the golgi (Büll *et al*., 2016). These attach sialic acid to galactose *via* two different glycosidic linkages, α(2,3) and α(2,6). Sialic acid linkages are known to influence the function of proteins to which they are conjugated, predominantly *via* interaction with immunomodulatory sialic acid-binding proteins including siglecs (Büll *et al*., 2016), and other lectins. For example, the human siglec, sialoadhesin, preferentially interacts with α(2,3)-linked sialic acid over α(2,6)-linked sialic acid (Hartnell *et al*., 2001). Conversely, α(2,6)-linked sialic acid, but not α(2,3)-linked sialic acid, protects T-cells from galectin 1-induced cell death (Toscano *et al*., 2007). These considerations made it difficult to choose the appropriate sialyltransferase for sialylation of SuperGal-expressed rFH. The key criteria that needed to be fulfilled when selecting a suitable sialyltransferase were that the enzyme must be sialylated with Neu5Ac and not the immunogenic mammalian Neu5Gc, the enzyme must also be expressed to high levels, active, soluble, stable and with a simple purification procedure.

The majority of circulating hFH is expressed in the liver (Skerka *et al*., 2007 and, Scholl *et al*., 2008). Within the liver, α(2,6)-sialyltransferase is the most highly expressed sialyltransferase. Although an α(2,3)-sialyltransferase is also highly expressed in the liver, α(2,6)-sialyltransferase was shown to have the higher activity (Paulson *et al*., 1989 and, Kitagawa and Paulson, 1994). Thus, it seems likely that the main sialic acid linkage on plasma derived hFH is α(2,6)-linked. Thus, it was decided to select a sialyltransferase that catalyses the formation of the α(2,6)-glycosidic linkage.

Mammals, and other eukaryotes, express a wide variety of sialyltransferase. However, recombinant production of these enzymes in mammalian cell lines is relatively slow, low yield, costly and often requires specialist equipment compared to an *E. coli* expression system. In contrast, bacterial expression is cheap, easy to do and fast.
However, recombinant eukaryotic α(2,6)-sialyltransferases often express poorly in bacterial systems, due to low stability and solubility resulting in low activity, and this necessitates protein engineering protocols to enhance solubility and protein folding (Ortiz-Soto and Seibel, 2016). Taking all these factors into account, to achieve the criteria set out here, it was ultimately decided that a bacterial α(2,6)- sialyltransferase was the best enzyme for this purpose; bacterial proteins are generally produced in higher yields from bacterial expression systems and have greater solubility, than eukaryotic proteins produced in bacterial expression systems.

Although there are many bacterial sialyltransferase, there are only four known bacterial α(2,6)-sialyltransferases: one from *Photobacterium damsela* (Yamamoto *et al*., 1998), two from *P. leiognathi* (Yamamato *et al*., 2007 and Mine *et al*., 2010) and one from *Photobacterium sp.* (Tsukamoto *et al*., 2008). All four enzymes belong to the same glycosyltransferase family GT80 in the Carbohydrate-Active Enzyme Database (CAZy), and have been produced recombinantly in *E. coli*, to high levels (Mine *et al*., 2010), with affinity tags for ease of purification (Ding *et al*., 2011 and Kang *et al*., 2015). However, in addition to α(2,6)-sialyltransferase activity, all four enzymes have been shown to act reversibly and hence to have inherent sialidase activity (Cheng *et al*., 2010; Mine *et al*., 2010 and Kang *et al*, 2015).

Interestingly, *Photobacterium sp.* α(2,6)-sialyltransferase (Ph-6ST) was reported to have significant sialidase activity when cultured at 30 °C for several hours (Kang *et al*, 2015) but not when cultured at either 30 °C (Tsukamoto *et al*., 2008) or 20 °C (Ding *et al*., 2011), for 3 minutes or 20 minutes, respectively. This indicates that sialidase activity may have a temperature- or incubation time- dependence. However, all reactions displayed a rapid initial α(2,6)-sialyltransferase activity.

Taking all of this into account, Ph-6ST was selected as the best candidate. The aim was to produce it recombinantly in *E. coli*, purify it and optimise conditions for its sialylation of rFH carrying galactose-terminal diantennary glycans, $Gal_2GlcNAc_2Man_3GlcNAc_2$.

4.2 Materials and Methods

4.2.1 InABLE plasmid assembly

4.2.1.1 Ph6-ST *PCR amplification*

All primers and oligonucleotide sequences were ordered from Sigma Genosys.
The *Ph-6ST* gene sequence was synthesised and codon optimised for expression in *E. coli* by Invitrogen GeneArt gene synthesis service (DNA sequence in appendix).
The *Ph-6ST* was amplified by PCR in a reaction containing 2 mM deoxynucleotide trisphosphate mix (New England Biolabs), 0.5 µM forward and reverse primers (see below), 1x Q5 reaction buffer (New England Biolabs) and 0.5 U Q5 High-fidelity DNA polymerase (New England Biolabs). The cycling conditions were: denaturation at 98 °C for 30 seconds, 30 cycles of: 98 °C for 10 seconds, 55 °C for 50 seconds and 72 °C for 45 seconds, followed by a final extension phase at 72 °C for 5 minutes. PCR reactions were carried out in a Mastercycler nexus X2 (Eppendorf).

Ph-6ST PCR primers:
4.i) Forward: 5' – gaggaagcggccgcgctcttcgcatcaacaaaaccatcatcgac – 3'
4.ii) Reverse: 5' – ttagatggatccgctcttcgggctcagtggtggtggtggtggtgtgcccaaaacaggacg – 3'

The PCR product was separated by 1.0% (w/v) agarose gel electrophoresis (100 V, 45 minutes) and yielded a PCR product of 1,462 bp. The PCR product was gel extracted using Zymoclean gel DNA recovery kit (Zymo Research, D4002). The excised gel band was weighed and placed in three volumes of agarose-dissolving buffer (Zymo Research) per milligram of gel and the agarose was left to dissolve by heating to 55 °C for 15 minutes. Dissolved agarose was transferred to Zymo-Spin columns (Zymo Research) and centrifuged (15,000 g for 60 seconds). An aliquot of 200 µL DNA wash buffer (Zymo Research) was added and the column was centrifuged (15,000 g for 30 seconds). The extracted DNA was eluted by adding 10 µL ddH_2O to the column and centrifuging (15,000 g for 30 seconds).

The concentration of the gel-extracted PCR product and a plasmid containing the assembly vector back-bone part, P59, was measured using fluorimetry (Qubit, ThermoFisher Scientific). A sample of 1 µL of gel-purified DNA was added to 199 µL Qubit reagent (Qubit, ThermoFisher Scientific) and 10 µL Qubit double-stranded (ds) DNA broad-range standard 1

(Tris-EDTA (TE) buffer, Qubit, ThermoFisher Scientific) and standard 2 (100 ng/µL DNA in TE buffer) were added, separately, to 190 µL Qubit reagent. All samples were mixed and left to incubate for two minutes at room temperature (20 – 25 °C) before measuring DNA concentrations. Concentration measurements of *Ph-6ST* PCR product and P59 yielded 52.6 ng/µL and 116 ng/µL, respectively. DNA molecular weights were calculated to be 1.59 MDa and 460 kDa, respectively, using online "Sequence Manipulation Suite: DNA Molecular Weight" software.

4.2.1.2 Linker and Part Oligo phosphorylation and annealing

In matched pairs, single-stranded oligonucleotides were subjected to a phosphorylation and annealing reaction to yield LO_730, LO_31, PO_730 and PO_31 (sequences listed below). Each phosphorylation and annealing reaction consisted of 1 µM of each matched pair of single-stranded oligonucleotides, 10 mU T4 polynucleotide kinase (New England Biolabs), 1 mM adenosine triphosphate (ATP), 5 µM dithiothreitol (DTT), 5 µL Polyethylene glycol 8,000 and 5 µL T4 polynucleotide kinase buffer (100mM $MgCl_2$, 700 mM Tris-HCl, 50 mM DTT, pH 7.6, New England Biolabs) and made up to 50 µL with ddH_2O. Phosphorylation and annealing reactions were placed in a thermocycler (Mastercycler nexus X2, Eppendorf) and the reaction was performed according to the following schedule (Table 4.1):

Temperature (°C)	Time (s)
37	1800
65	1200
90	300
85	60
80	60
75	60
70	60
65	60
60	30
55	30
50	30
45	30
40	30
35	30
30	30
25	30
20	180

Table 4.1: Thermocycler conditions for the phosphorylation and annealing of matched pair single-stranded oligonucleotides.

Single-stranded oligonucleotide sequences:

4.iii) LOlong_730: 5’ – gccatgtgtaataatagcgaagaaaatacccaga – 3’

4.iv) LOshort_730: 5’ – gctattattacacat – 3’

4.v) POlong_730: 5’ – atgtcgttcttgatgatgctctgggtattttcttc – 3’

4.vi) POshort_730: 5’ – gcatcatcaagaacga – 3’

4.iii) LOlong_31: 5’ – gcctgctaacaaagcccgaaaggaagctgagttg – 3’

4.iv) LOshort_31: 5’ – cgggctttgttagca – 3’

4.v) POlong_31: 5’ – ctcagcggtggcagcagccaactcagcttccttt – 3’

4.vi) POshort_31: 5’ – gctgctgccaccgct – 3’

4.2.1.3 Part-linker fusion reactions

Two separate part-linker fusion reactions were set-up, containing PCR product, part oligo (PO)_730 and linker oligo (LO)_31 in one reaction and, the plasmid containing P59, PO_31 and LO_730 in the other. The RE *Ear* I was used to liberate *tPh-6ST* and P59 in these two separate reactions. Standard part-linker fusion reactions conditions are as follows (final concentrations are given): a reaction mix consisting of 20 nM of truncated part (TP), 0.2 µM each of the LO and PO to be associated with the TP, 5 µL NEBuffer 4 (final concentrations of 5 mM potassium acetate, 2 mM Tris-acetate, 1 mM magnesium acetate, 0.1 mM DTT, pH 7.9, New England Biolabs), 1 mM ATP, 0.5 U T4 DNA ligase (New England Biolabs), 12.5 mU type II S restriction endonuclease (e.g. either *Ear* I or *Sap* I, New England Biolabs) and made up to 50 µL with ddH_2O. Part-linker reactions were placed in a thermocycler and subjected to the following temperature cycling protocol (Table 4.2):

Temperature (°C)	Time (minutes)
37	90
16	30
37	30
16	15
37	15
16	15
37	15
16	10
37	15
16	10
37	60
16	30
65	20

Table 4.2: Thermocycler conditions for the part-linker fusion reactions

The part-linker fusion products, PO_730 – *tPh-6ST* – LO31 and PO_31 – P59 – LO_730, were separated by 1% (w/v) agarose gel electrophoresis (100 V, 45 minutes) (fig. 4.1 B) and gel extracted using Zymoclean gel-DNA recovery kit (Zymo Research). Subsequent measurement of DNA concentration by fluorimetry (Invitrogen, Qubit) gave 8.12 ng/µL and 16 ng/µL, respectively (see section 4.2.1.1 for experimental details). The DNA molecular weight was calculated using the part-linker fusion DNA sequences and the online "Sequence Manipulation Suite: DNA Molecular Weight" software.

4.2.1.4 inABLE assembly of Ph-6ST *in a plasmid vector backbone for expression in* E. coli

The inABLE DNA assembly reaction was carried out by combining the two part-linker fusion products in the proportions outlined in table 4.3 and incubating at 37 °C for 30 minutes.

Reagents	Volume (µl)
PO_730 - *tPh-6ST* - LO_31	7.8
PO_31 - P59 - LO_730	1.2
Buffer 2 (10x)	1.2
PEG 8000 (50% [w/v])	1.2
EarI	1
Total volume	125

Table 4.3: inABLE assembly of part-linker fusion products to yield an *E. coli* expression vector plasmid containing a gene for expression of Ph-6ST with a C-terminal His_6-tag

A sample of 1 µL of assembled plasmid was used to transform electrocompetent *E. coli* strain TOP10 (One Shot TOP10 Electrocomp *E. coli*, ThermoFisher Scientific, C404050) by electroporation (1.7 kV, 25 µF). Transformed cells were recovered in 1 mL SOC media for one hour at 37 °C with 250-rpm shaking. Recovered cells were plated out onto LB agar containing 50 µg/mL kanamycin sulphate (ThermoFisher Scientific) and incubated at 37°C for approximately 16 hours.

Transformation colonies, containing the assembled vector plasmid, were selected by colony PCR followed by 1.0% (w/v) agarose electrophoreses (100 V, 45 minutes) (data not shown). The colony PCR procedure was as follows: colonies were picked from the LB agar transformation plate and suspended in 10 µL sterile ddH_2O. A 2 µL sample of suspended colony was added to a mix containing 5 µL One*Taq* 2X Master mix (New England Biolabs)

and 1.5 µL each of 10 µM stock of forward and reverse primers (see below for sequences). Colony PCR reactions were placed in a thermocycler and cycling conditions were as follows: DNA was denatured by heating to 94 °C for 30 seconds, then 30 temperature cycles of 94 °C for 30 seconds, 55 °C for 60 seconds and 68 °C for two minutes, were performed followed by a final extension at 68 °C for five minutes. The PCR reaction yielded a product of 2,276 bp. The primers used were: forward primer 5.iii) LOlong_730 (sequence given above) and the reverse primer:

4.vii) Reverse: 5' – ctccttcattacagaaacgg – 3'

DNA, from clones identified by colony PCR, to contain the assembled plasmid, was prepared using a QIAprep Spin Miniprep plasmid-preparation kit (Qiagen). Colonies were used to inoculate 5 mL LB media, containing appropriate antibiotic, in 50 mL conical tubes (Falcon) and left to incubate for approximately 16 hours at 37 °C, with 250 rpm shaking. Cultures were centrifuged (5,000 g, 20 minutes), supernatant discarded and pellet resuspended in 250 µL buffer P1 (Qiagen) and transferred to a fresh 1.5-mL microcentrifuge tube (Eppendorf). A 250 µL aliquot of buffer P2 (Qiagen) was added before mixing and incubating for five minutes at room temperature (20 – 25 °C). Then, 350 µL buffer N3 (Qiagen) was added, the contents of the tube were agitated and then centrifuged (15,000 g for 10 minutes). The supernatant was applied to a QIAprep spin column and centrifuged (15,000 g, 60 seconds). A 0.5-mL aliquot of buffer PB (Qiagen) was added and the spin column was centrifuged (15,000 g, 60 seconds). Subsequently, 0.75 mL of buffer PE (Qiagen) was added and the spin column was centrifuged (15,000 g, 60 seconds). Finally, 50 µL ddH_2O was added and the spin column was centrifuged (15,000 g, 60 seconds) to collected purified plasmid.

Restriction mapping of the purified plasmid was carried out to assess whether the plasmid was assembled correctly and with parts in the correct orientation. The restriction map was generated by restriction endonuclease digestion using the enzymes *Age* I and *Sph* I, and NEBuffer 1.1 (New England Biolabs). An aliquot containing 1 µg plasmid was incubated, at 37 °C for one hour, with 0.5 µL each of *Age* I and *Sph* I (New England Biolabs), 2 µL NEBuffer 1.1 in a reaction volume made up to 20 µL with ddH_2O. Correctly assembled plasmids were identified by bands at ~5300 bp and ~1300 bp (data not shown) by 1% (w/v) agarose electrophoresis (100 V, 45 minutes).

4.2.2 Expression in E. coli *strain BL21 (DE3)*

4.2.2.1 Transformation and selection

A clone identified, by colony PCR and restriction mapping, to contain the correctly assembled plasmid was used to inoculate 100 mL LB media, containing 50 µg/mL kanamycin sulphate, in baffled shake flasks. The culture was left to incubate for approximately 16 hours at 37 °C, with 250-rpm shaking. The culture was centrifuged (5,000 g, 20 minutes) and supernatant was discarded. The plasmid was isolated from the pelleted cells by Midi prep (Qiagen) standard procedure (Qiagen plasmid purification handbook, www.qiagen.com/handbooks). A 2-µL sample of pEcPh-6ST was mixed, by gentle flicking, with 50 µL *E. coli* BL21 (DE3) electro-competant cells (New England Biolabs) and transformation was carried out (1.7 kV, 25 µF). Transformed cells were transferred to 1 mL SOC media and incubated at 3.7 °C, with 250-rpm shaking for 1.5 hours. A sample of 50 µL of the recovered cells was spread plated onto LB agar containing 50 µg/mL kanamycin sulphate, and plates were then incubated for approximately 16 hours at 37 °C.

Colonies containing pEcPh-6ST were identified by the same colony PCR used in section 4.2.1.1 (fig. 4.3). PCR products were run on 4200 TapeStation 4200 (Agilent) using D5000 DNA analysis screen tape. D5000 sample buffer (Agilent formulation) was equilibrated at room temperature (20 – 25 °C) for 30 minutes. A 1-µL D5000 ladder (Agilent formulation) and 1 µL of each PCR product were added, separately, to 10 µL D5000 sample buffer mixed and spun (3,000 g, one minute). Samples were loaded into the 4200 TapeStation instrument, and the run was started using Agilent TapeStation Analysis Software.

4.2.2.2 Test expression of α(2,6)-sialyltransferase in E. coli

A single colony, identified by colony PCR and restriction mapping to contain pEcPh-6ST, was used to inoculated 10 mL LB media containing 50 µg/mL kanamycin sulphate. This seed culture was incubated for ~16 hours at 37 °C and 250-rpm shaking. A glycerol stock was made by taking 800 µL and mixing with 200 µL 50% (v/v) glycerol before storing frozen at -80 °C. The culture was back-diluted to an OD_{600} of 0.1 into twelve 10-mL volumes of LB media and 50 µg/mL kanamycin sulphate. Cultures were incubated at 37 °C, and 250-rpm shaking until an OD_{600} of between 0.4 – 0.8 was reached. Expression of Ph-6ST was then

induced by the addition of 1 M isopropylthio-β-galactoside (IPTG) (ThermoFisher Scientific) to a final concentration of either 0.1 mM, 0.3 mM and 0.9 mM IPTG – yielding four cultures at each concentration of IPTG. One culture of each concentration of IPTG was incubated, with 250-rpm shaking, at either 10 °C, 20 °C, 30 °C or 37 °C (Table 4.2) for a total of 20 hours. Samples of 600 µL were taken prior to induction, at four hours, and then at 20 hours post-induction. Aliquots of 100 µL were used to measure OD_{600} at each time point. The remaining 500 µL of each sample was placed in 1.5 mL micro-centrifuge tubes (Eppendorf) and centrifuged (16,000 g, 10 minutes). The supernatant was discarded and the pellet was used for expression analysis.

4.2.2.3 Analysis of test expression: cell lysis, SDS-PAGE and western blot

Cell lysis buffer was made up containing 124 µg/mL lysozyme (from chicken egg white, Sigma-aldrich) and >35 U of benzonase nuclease (EMD millipore) in BugBuster Protein Extraction Reagent (Novagen). Cell lysis buffer was added to cell pellets at a ratio of 32.2 µL per 1.0 unit OD_{600} (measured in section 4.2.2.2). Cells were incubated at room temperature (20 – 25 °C), with 40 rpm-rocking for 15 minutes followed by centrifugation (16,000 g, 20 minutes, 4 °C). Supernatant was transferred into fresh 1.5-mL microcentrifuge tubes (Eppendorf) and the pellet was suspended in ddH_2O at a ratio of 32.2 µL per 1.0 unit OD_{600} (measured in section 4.2.2.2). An aliquot of 10 µL was mixed with 40 µL SDS-PAGE loading buffer and 10 µL of that was loaded onto 4-12% bis-tris plus polyacrylamide gels for electrophoresis (Invitrogen Mini Gel Tank, MES buffer, 200 V, 35 minutes). Gels were either stained with Coomassie Instant*Blue* or subjected to western blot. Western blot was used to detect the presence of the His_6-tag on the bacterial α(2,6)-sialyltransferase. The western blot procedure was as follows: the polyacrylamide gel was placed in a western blot-transfer stack and transferred onto PVDF membrane as described in chapter 2 section 2.2.2.1. The membrane was placed in 50 mL blocking buffer (2.25% (w/v) milk powder (Sainsburys), 1% (v/v) tween-20, 10 mM phosphate buffer, 137 mM NaCl, 2.7 mM KCl, pH 7.4) and incubated at room temperature (20 – 25 °C), with 40 rpm rocking, for 1 hour. Blocking buffer was poured off and replaced with 20 mL of mouse monoclonal anti-His (poly) peroxidase conjugate antibody (Sigma-aldrich, A7058) diluted 1:2,000 in blocking buffer. After incubation at room temperature (20 – 25 °C), and 40-rpm rocking for one hour, the antibody solution was discarded and the membrane was washed thrice with wash buffer (TPBS, blocking buffer without milk powder), incubated at room temperature (20 – 25 °C)

with 40-rpm rocking for five minutes. Wash buffer was poured off and bands were developed by incubating membranes in 10 mL of metal-enhanced DAB (ThermoFisher Scientific, 34065) diluted 1:10 with stable peroxide buffer (ThermoFisher Scientific) until bands begin to appear, at room temperature (20 – 25 °C) and 40-rpm rocking. The reaction was stopped by dilution with ddH_2O, before pouring off diluted DAB solution.

4.2.2.4 Large-scale expression and nickel (II)-immobilised metal ion-affinity chromatography

A 400-mL culture of *E. coli* containing pEcPh-6ST was incubated and expression of *Ph-6ST* induced with a final concentration of 0.9 mM IPTG and incubation at 20 °C, with 250-rpm shaking (optimal expression conditions determined in section 4.2.2.3). Cells were lysed according to the conditions in section 4.2.2.3, and the soluble protein fraction was subjected to nickel (II)-immobilised metal ion-affinity chromatography using a 1-mL HisTrap HP column (GE Healthcare Life Sciences, 29-0510-21). Prior to use, buffer A (0.5 M sodium chloride, 50 mM tris-HCl, pH 7.5) and buffer B (250 mM imidazole, 0.5 M NaCl, 50 mM tris-HCl, pH 7.5) were filtered (0.2 µm filter) and degassed. The HisTrap column was equilibrated with 10 CV of 2% buffer B. The sample was then applied at a rate of 1 mL per minute. After sample loading, the column was washed with 8 CV 2% buffer B followed by washing with 20% buffer B. Protein was then eluted over a linear gradient of 20% - 100% buffer B. The protein concentration of select fractions, was estimated by A_{280} and selected fractions analysed by SDS-PAGE on 4 – 12% bis-tris polyacrylamide gels and western blot (10 µL fraction added to 40 µL SDS-loading buffer, 10 µL loaded, 200 V, 35 minutes; western blot procedure outlined in section 4.2.2.3). Fractions determined, by SDS-PAGE, to have a high yield and purity of Ph-6ST were pooled, placed in 3.5 kDa molecular weight-cut off dialysis tubing (SnakeSkin, ThermoFisher Scientific) and dialysed against 2 L of 200 mM Tris-HCl, 50 mM NaCl, pH 8.0 for approximately 16 hours, with a buffer exchange after four hours, at 4 °C. Dialysed protein samples were aliquoted in 250 µL volumes and stored, frozen at -20 °C.

4.2.2.5 Testing the activity of purified Ph-6ST

To test the activity of purified Ph-6ST, a 10 µL reactions was set up containing 200 mM tris-HCL, pH 8.0, 20 mM $MgCl_2$, 500 µM CMP-N-acetylneuraminic acid sodium salt (Carbosynth, MC04391) (CMP-sialic acid), 3.125 µM rFH expressed in SuperGal and 0.3 µM purified Ph-

6ST. Three control reactions were also set-up but omitting either CMP-sialic acid, rFH or Ph-6ST. Reactions were left to proceed over night at 20 °C. Sialylation reactions were subjected to SNA and ECL lectin-based western blot analysis as detailed in Chapter 2 section 2.2.2.1.

4.2.3 Sialylation of recombinant factor H expressed in glycoengineered P. pastoris *strain SuperGal*

4.2.3.1 Optimisation of sialylation reaction conditions

Final concentrations of reagents for sialyltransferase reactions were as follows: 200 mM Trizma base, pH 8.0, 20 mM magnesium chloride, 500 µM CMP-N-acetylneuraminic acid sodium salt (Carbosynth, MC04391), 470 µg/mL rFH expressed in SuperGal and carrying $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycans, 17 µg/mL Ph-6ST, made up to a defined volume with ddH_2O. In the first round of optimisation, a single master stock was made up and separated into three equal volumes in 1.5-mL microfuge tubes and incubated at 10 °C, 20 °C or 30 °C. Samples were taken one hour, four hours and 20 hours after the start of the reaction. In the second round, all samples were incubated at 10 °C for 20 hours. A sample of 70 µM CMP-N-acetylneuraminic acid sodium salt or ~10 U alkaline phosphatase (calf intestine, Sigma-Aldrich) or both were added at 30 minutes, two hours and four hours after the start of the reaction.

4.2.3.2 Lectin-based ELISA to detect sialylation of rFH expressed in SuperGal

Plasma-derived hFH standards for lectin ELISA were prepared under native conditions as follows, 1.3 mg/mL plasma-purified hFH was combined, to a final concentration of 91 ng/µL, with 5 µL 10x Glycobuffer 1 (New England Biolabs) and either 5 µL α(2-3,6,8)-neuraminidase (New England Biolabs) for the desialylated standard, or ddH_2O for the sialylated standard. Reactions were made up to 55 µL with ddH_2O and left to incubate for approximately 16 hours at 37 °C. An 11 µL aliquot of each standard was combined with 39 µL SDS loading buffer. To assess whether the standards were sialylated or desialylated, 200 ng of α(2-3,6,8)-neuraminidase-treated and untreated FH standards were loaded onto three 4-12% bis-tris plus polyacrylamide gels (ThermoFisher Scientific) and electrophoresis was carried out (200 V, 35 minutes). One gel was stained with Coomassie Instant*Blue*. The

other two gels were placed in western blot-transfer stacks and subjected to the lectin-based western blot procedure described in Section 2.2.2.1.

Lectin-based ELISA, to detect the extent of sialylation, were carried out as follows. SUMO-tagged PspCN (prepared in house by Dr Andrew Herbert) was mixed to a final concentration of 5 µg/mL in capture buffer (0.1 M sodium carbonate, pH 9.5), then 100 µL was added to wells in two 96-well plates (Falcon, 353072). One of these plates was designated for SNA lectin sialic acid detection, the other plate was designated for ECL lectin LacNAc detection. The two plates were incubated with gentle shaking at 40 rpm for 15 minutes at room temperature (20 – 25 °C) followed by incubation for approximately 16 hours at 4 °C. SUMO-PspCN in capture buffer was removed and 300 µL 1x carbo-free blocking buffer (Vector Laboratories, SP5040) was added before incubation with gentle shaking at 40 rpm for 30 minutes at room temperature. The sialylated and desialylated hFH standards, prepared as described above, were diluted to 10 µg/mL in 1x carbo-free blocking buffer. Aliquots of 100 µL of 10 µg/mL standard were loaded into wells A1 and B1 of each plate. Blocking solution was removed and sialylated standard was added to the SNA lectin-designated plate, while desialylated standard was added to the ECL lectin-designated plate. A 1:2 serial dilution of each standard was created (in duplicate) from lanes A1-11 and from lanes B1-11. Aliquots of 100 µL of 1:1000 and 1:2000 dilutions (for time and temperature studies) or 1:1000 and 1:1500 dilutions (for CMP-sialic acid and alkaline phosphatase studies) from each reaction condition were loaded into selected wells. Controls and samples were incubated with gentle shaking at 40 rpm for one hour at room temperature.

Subsequently, standards and samples were removed and the wells washed by the addition and removal of 300 µL wash buffer (0.05% (v/v) Tween-20 in PBS, pH 7.4). Washing was repeated five times. Separately, biotinylated SNA and biotinylated ECL lectins (Vector Laboratories, B-1305 and B-1145, respectively) were diluted to 10 µg/mL in carbo-free blocking buffer. An aliquot of 100 µL of diluted SNA was added to all wells in the SNA lectin-designated plate and the same was done, but with diluted ECL to the ECL lectin-designated plate. Plates were incubated with gentle shaking at 40 rpm for 1 hour at room temperature, then SNA and ECL solutions were removed and the plates were washed as previously. NeutrAvidin Protein, HRP (ThermoFisher Scientific, 34028) was mixed to a final concentration of 0.5 µg/mL in carbo-free blocking buffer and 100 µL was added to each well. The plates were incubated with gentle shaking at 40 rpm for one hour at room

temperature, then NeutrAvidin Protein, HRP solutions were removed and the wells were washed as previously. Subsequently, 100 µL of 1-Step Ultra TMB (3,3',5,5'-Tetramethylbenzidine)-ELISA Substrate Solution (ThermoFisher Scientific, 34028) was added to each well and incubated for 15 minutes, and reactions were stopped by the addition of 100 µL 2 M sulphuric acid. Absorbance values (650 nm) were measured using a SpectraMax M5 Multimode plate reader. Standards were plotted on a four-parameter logistic curve plotted using MyAssays.com software that was used to quantify amounts of terminal sialic acid/ galactose in sialylation assay samples.

4.2.3.3 Purification of sialylated rFH by cation-exchange chromatography

Recombinant FH that had been expressed in SuperGal and previously treated with Endo H_f to remove yeast glycans (see Chapter 3 section 3.3.7), was sialylated according to the optimal conditions determined as described above (incubation at 10 °C, for 20 hours with one, two or three additions of CMP-sialic acid at 30 minutes, two hours and four hours after the start of the reaction). Sialylated rFH was placed in 10-kDa molecular weight cut-off dialysis tubing (SnakeSkin, ThermoFisher Scientific, 88245) and dialysed overnight, at 4 °C, against cation-exchange buffer A (20 mM potassium phosphate, pH 7.0), with the buffer changed after two hours. Dialysed sialylated rFH was purified by cation-exchange chromatography using a 6-mL Resource S column (GE Healthcare Life Sciences). The column was equilibrated with three CV of buffer A, approximately 35 milligrams sialylated rFH in a 75 mL volume was loaded at 6 mL/min and the column was washed with three CV of buffer A before eluting with a 16-CV linear gradient from 0 – 50% buffer B (20 mM potassium phosphate, 1 M NaCl, pH 7.0), and a final high-salt wash with one CV from 50% - 100% buffer B. Fractions of 2 mL were collected. The protein concentration of selected fractions (flow through, first 80 mL of elution, and then every 2 mL from 150.5 to 182.5 mL) was estimated by A_{280} readings and then 2 µL of selected fractions were mixed with 14 µL of SDS-PAGE-loading buffer and subjected to PAGE (200 V, 1 hour) (fig. 4.9). Selected fractions (elution volume 158.5 – 174.5 mL) with high purity (fig. 4.5) and concentration of rFH were pooled, placed in 10-kDa molecular weight cut-off dialysis tubing (SnakeSkin, ThermoFisher Scientific) and dialysed against PBS, pH 7.4, for approximately 16 hours at 4 °C, with a buffer change after four hours.

4.2.3.4 Glycan release, purification and analysis by MALDI-TOF mass spectrometry

Glycans were released, purified and analysed by MALDI-TOF mass spectrometry as detailed in Chapter 2 section 2.2.3.2.

4.3 Results

4.3.1 Ph-6ST gene synthesis and inABLE plasmid preparation

The sequence for the gene encoding α(2,6)-sialyltransferase from *Photobacterium sp*. was obtained from the National Centre for Biotechnology Information (NCBI) (accession number: 2Z4T_A). A gene was designed with modifications in accordance with the method employed by Ding *et al*. (2011). Briefly, the gene was modified to yield a truncated, His-tagged fusion protein with amino acid deletions at the N- and C-terminus (Δ2-14 and Δ502-514, respectively). Removal of the N-terminal amino acids deletes a membrane-localisation domain, thus improving solubility and yield. Truncation of the C-terminal amino acids allowed for substitution with a polyhistidine (His_6)-tag for ease of purification whilst not affecting enzymatic activity (Ding *et al*., 2011). The truncated His_6-tagged gene was codon optimised for *E. coli* expression, synthesised and supplied by Invitrogen GeneArt gene synthesis. It is referred to as *Ph-6ST* and protein product Ph-6ST from here onwards.
The *Ph-6ST* gene sequence (codon-optimised sequence in appendix), including the 3' His_6-tag coding sequence, was evaluated using the ExPASy protparam tool (fig. 4.1).

A)

```
        10         20         30         40         50         60
MCNNSEENTQ SIIKNDINKT IIDEEYVNLE PINQSNISFT KHSWVQTCGT QQLLTEQNKE

        70         80         90        100        110        120
SISLSVVAPR LDDDEKYCFD FNGVSNKGEK YITKVTLNVV APSLEVYVDH ASLPTLQQLM

       130        140        150        160        170        180
DIIKSEEENP TAQRYIAWGR IVPTDEQMKE LNITSFALIN NHTPADLVQE IVKQAQTKHR

       190        200        210        220        230        240
LNVKLSSNTA HSFDNLVPIL KELNSFNNVT VTNIDLYDDG SAEYVNLYNW RDTLNKTDNL

       250        260        270        280        290        300
KIGKDYLEDV INGINEDTSN TGTSSVYNWQ KLYPANYHFL RKDYLTLEPS LHELRDYIGD

       310        320        330        340        350        360
SLKQMQWDGF KKFNSKQQEL FLSIVNFDKQ KLQNEYNSSN LPNFVFTGTT VWAGNHEREY

       370        380        390        400        410        420
YAKQQINVIN NAINESSPHY LGNSYDLFFK GHPGGGIINT LIMQNYPSMV DIPSKISFEV

       430        440        450        460        470        480
LMMTDMLPDA VAGIASSLYF TIPAEKIKFI VFTSTETITD RETALRSPLV QVMIKLGIVK

       490
EENVLFWAHH HHHH
```

B)

Number of amino acids	494
Molecular weight (Da)	56,571.62
Theoretical pI	5.23
Extinction Coefficient (M^{-1} cm^{-1})	69,790
Instability Index	31.85
Aliphatic Index	87.57
Grand average of hydropathy	-0.443

Figure 4.1: Amino acid sequences, predicted from *Ph-6ST* gene sequence (A) and protein paramaters (B) predicted by ExPASy protparam tool. Tryptophans are highlighted in the amino acid sequence.

The protparam-derived data indicates that the protein should be stable (from the instability and aliphatic indices) while the grand average of hydropathy score is slightly negative, indicating it is likely to be soluble. Overall, the sequence was judged to be suitable for production of a folded and well-behaved enzyme fit for the purpose of *in vitro* sialylation.

The gene sequence was amplified using primers designed by Ingenza's proprietary inABLE software. These primers were used to PCR amplify the *Ph-6ST* gene (fig 4.2 A) sequence to yield a truncated part (*tPh-6ST*) which was then inserted into the *E. coli* expression plasmid P59 – a plasmid made in house by inABLE-oriented modifications to the pET24 class of expression vectors – by restriction digestion (fig 4.2 B) and ligation according to inABLE plasmid-construction procedures.

This generated an *E. coli* expression vector (pEcPh-6ST) containing the *LacI* gene (and associated promoter), the *Ph-6ST* gene under the control of the T7-promoter, for isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible expression in conjunction with *LacI* (process outlined in pET system manual 11th edition), and with a C-terminal His_6-tag, and kanamycin

resistance-selectable marker for antibiotic-mediated selection of transformed clones (fig 4.2 C).

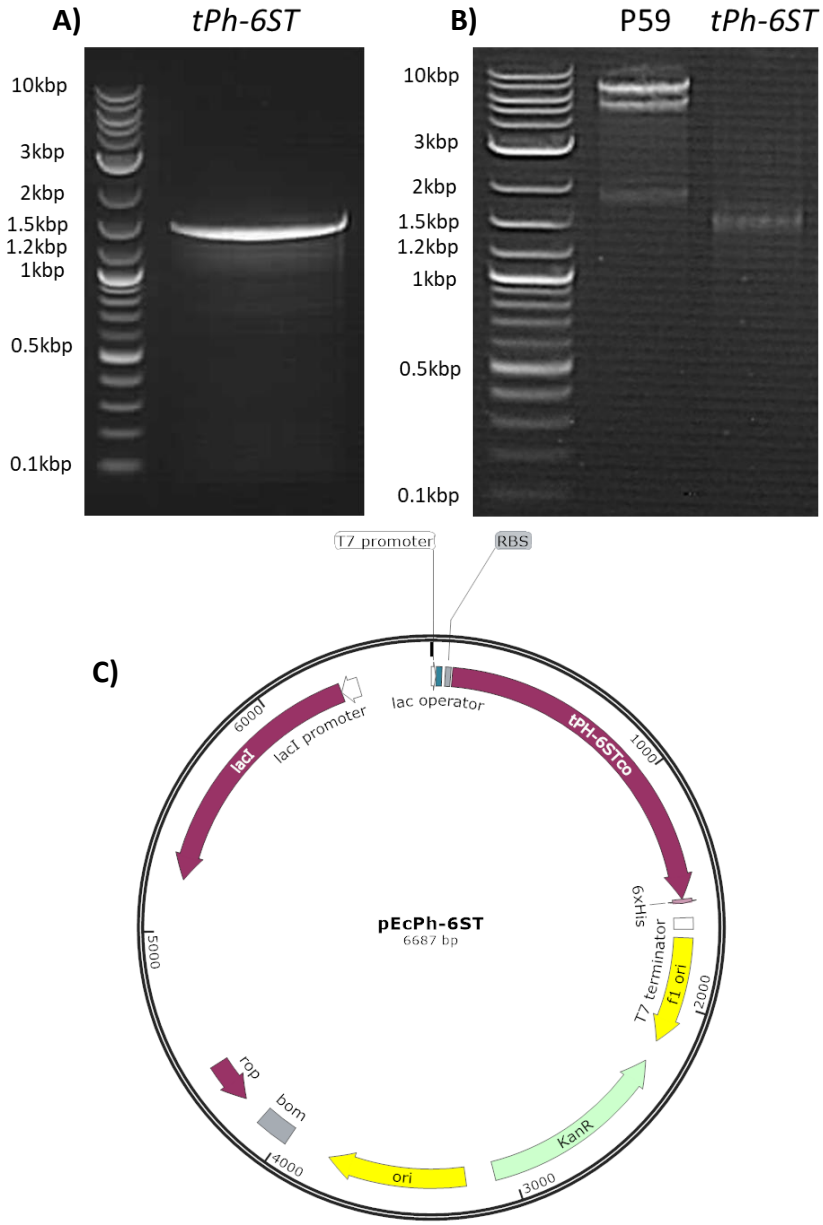


Figure 4.2: DNA agarose gel electrophoresis images of A) PCR amplification of *tPH-6ST* and B) restriction endonuclease digestion of the backbone plasmid p59 and PCR amplified *tPH-6ST*. Expected band sizes for *tPH-6ST* PCR product was ~1,462bp and, p59 and *tPH-6ST* after restriction digestion was 5150bp and 1438bp, respectively. And C), the plasmid map of the inABLE constructed vector plasmid pEcPH-6ST containing the C-terminally His_6-tagged gene *Ph-6ST*.

4.3.2 Transformation, selection and optimisation of expression conditions

After transformation, two colonies were selected for colony PCR analysis to identify whether they contain the introduced plasmid pEcPh-6ST (fig 4.3).

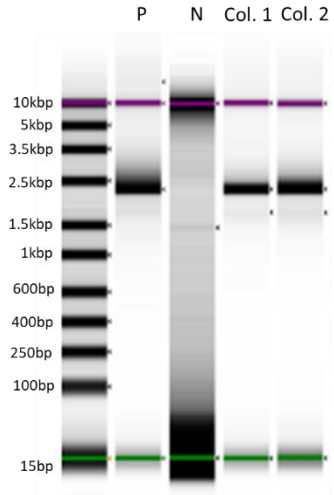


Figure 4.3: Tapestation gel electrophoresis virtual image of the colony PCR analysis of BL21 (DE3) *E. coli* strains transformed with expression plasmid pEcPh-6ST. The expected PCR-amplified product size is 2,276 bp. P is the positive control – the pEcPh-6ST plasmid. N is the negative control – a colony of parental BL21 (DE3) *E. coli* before transformation with pEcPh-6ST.

Both colonies analysed by colony PCR produced an amplification product of the expected size (2,276bp). Colony 1 was carried forward for test expression analysis.

Twelve replicate cultures of colony 1 were used to inoculate lysogeny broth (LB) media. After an initial growth phase at 37 °C, each culture was subjected to a different set of induction conditions (table 4.4).

Culture	Induction temperature (°C)	[IPTG] (mM)	Time post-induction (hrs)
1	10	0.1	4
			20
2		0.3	4
			20
3		0.9	4
			20
4	20	0.1	4
			20
5		0.3	4
			20
6		0.9	4
			20
7	30	0.1	4
			20
8		0.3	4
			20
9		0.9	4
			20
10	37	0.1	4
			20
11		0.3	4
			20
12		0.9	4
			20

Table 4.4: *Ph-6ST* expression induction condition for 12 cultures of colony 1 that had been confirmed to have been successfully transformed with plasmid pEcPh-6ST. Three cultures were cultured at one of four temperatures during expression induction, each of these three were induced with one of three IPTG concentrations, and samples of each culture were taken at four hours and 20 hours post-induction to assess for the presence and levels of the Ph-6ST.

Test expressions were carried out on colony 1 to test for the optimal temperature, incubation time and IPTG concentration for expression of *Ph6-ST*.

The OD_{600} was measured before induction of protein expression and at four hours and 20 hours post-induction (table 4.5).

Induction Temperature (°C)	[IPTG] (mM)	Pre-induction OD_{600}	4hr OD_{600}	20hr OD_{600}
10	0.1	0.54	0.73	1.04
	0.3	0.55	0.70	1.10
	0.9	0.54	0.68	1.07
20	0.1	0.72	2.06	6.89
	0.3	0.68	1.89	5.23
	0.9	0.64	1.81	5.65
30	0.1	0.66	3.55	7.00
	0.3	0.68	3.54	6.74
	0.9	0.63	3.30	6.45
37	0.1	0.60	2.65	3.70
	0.3	0.62	2.43	3.34
	0.9	0.68	2.41	3.08

Table 4.5: Table of the OD_{600} readings of 12 cultures pre-, four hours and 20 hours post-induction.

Of the temperatures explored, 30 °C was best while 10 °C was worst, for growth of *E. coli* cells that have been transformed with pEcPH-6ST (*i.e. E. coli Ph-6ST*). But, the data does not help to select optimal conditions for gene expression or protein production per cell.

To address this, OD_{600} readings were used to normalise cell density across the different cultures and then normalised protein expression levels were assessed by SDS-PAGE and western blot (fig 4.4).

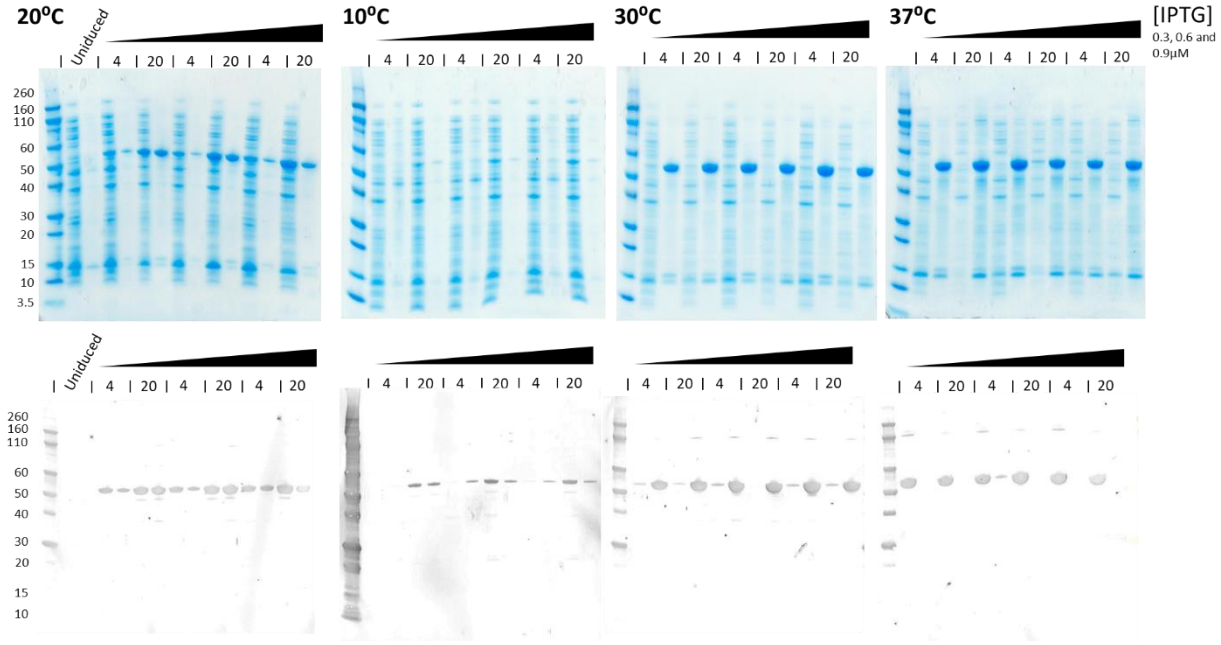


Figure 4.4: SDS-PAGE (top row) and anti-His-tag antibody western blot (bottom row) showing the results of test expression for *E. coli* BL21 (DE3) cells transformed with plasmid pEcPh-6ST containing *Ph-6ST*. The black triangles represent an increasing concentration gradient of IPTG (0.3 mM, 0.6 mM and 0.9 mM). For each time point, soluble (left) and inclusion body-containing samples (right) were run on SDS-PAGE – each time point is demarcated by short vertical black lines (top of SDS-PAGE images).

From the SDS-PAGE and anti-His-tag western blot images (fig. 4.4), expression appears tightly controlled since no Ph-6ST (~56.5 kDa) was detected in the non-induced culture. Post induction, Ph-6ST is expressed under all conditions. However, there is a tendency for the protein to form insoluble inclusion bodies, especially at higher incubation temperatures. For example, at 30 °C and 37 °C most of the protein appears to form inclusion bodies, while at 20 °C there is an approximately equal distribution of protein in the soluble and inclusion body samples. At 10 °C the protein seems more soluble – especially at higher IPTG concentrations. For this reason, inductions at 30 °C and 37 °C were deemed unsuitable for expression of Ph-6ST.

Even though there is a similar distribution of Ph-6ST in the soluble and inclusion body samples at 20 °C, incubation at 20 °C yields more total protein in the soluble samples compared to the soluble samples following incubation at 10 °C. Therefore, 20 °C was

judged to be the optimal incubation temperature for *Ph-6ST* expression and Ph-6ST production.

Additionally, slightly more protein was present in the soluble samples at 20 hours post-induction compared to four hours post-induction. Higher IPTG concentrations also yielded higher amounts of soluble protein. Therefore, the optimal conditions defined for expression of recombinant Ph-6ST were as follows: 20 °C, 0.9 mM IPTG and cells harvested at 20 hours post-induction.

4.3.3 Ph-6ST expression and purification

Next, *E. coli Ph-6ST* was cultured and *Ph-6ST* expressed according to the experimentally determined optimal conditions. Ph-6ST was then purified from 400 mL lysed *E. coli Ph-6ST* cell culture by nickel (II)-immobilised metal-affinity chromatography (fig 4.5).

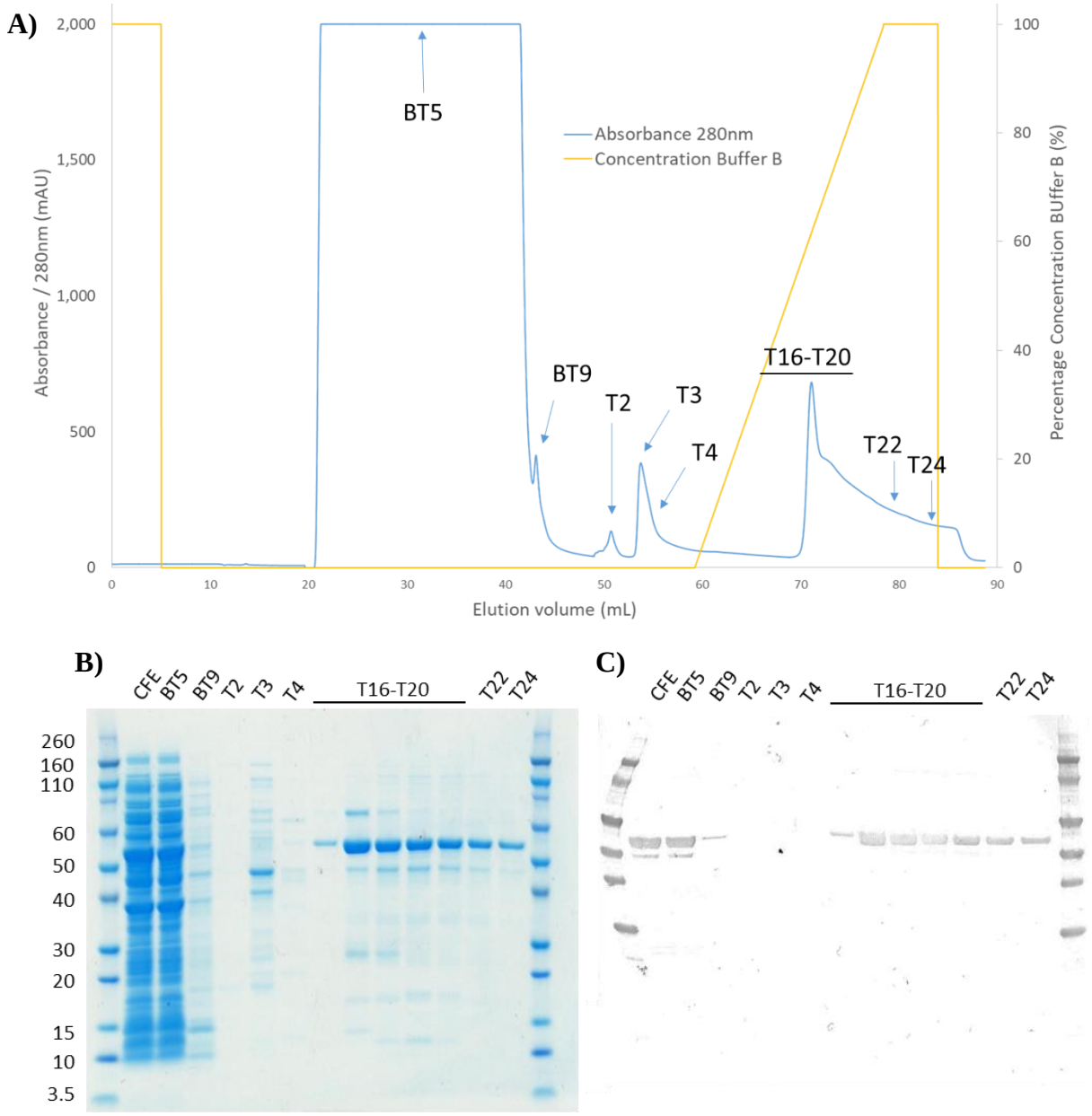


Figure 4.5: Chromatogram (A) of the nickel (II)-immobilised metal-affinity chromatography purification of His_6-tagged recombinant Ph-6ST from lysed *E. coli PH-6ST* cells. SDS-PAGE (B) and anti-His antibody western blot analysis (C) of selected fractions from the nickel (II)-affinity purification. CFE – is cell free extract (lysed *E. coli PH-6ST* cells prior to purification). BT5 and BT9 factions are from the flow-through; T2, T3 and T4 fractions are associated with peaks highlighted in the chromatogram (fig. 4.5 A) and T16-24 elution fractions.

The chromatogram (fig. 4.5 A) shows that the majority of protein in *E. coli Ph-6ST* cell lysate did not bind to the affinity resin. However, the SDS-PAGE and lectin western (fig. 4.5 B and C, respectively) show that a high proportion of Ph-6ST had eluted in the flow through – this suggests that the amount of Ph-6ST in the lysed *E. coli PH-6ST* cells exceeded the capacity of the 1-mL HisTrap column (the column capacity is reported to be 40 milligram His-tagged protein /mL resin, GE lifesciences). The chromatogram shows two peaks (T2, T3 and T4)

during the wash step of the chromatographic run but these two contain very little Ph-6ST protein according to western blot. The chromatogram also shows a peak with a shoulder during the elution gradient (T16 – 24) which, when analysed by SDS-PAGE and western blot, contains Ph-6ST as well as other contaminating proteins of varying molecular weights. Fractions T17-T24 were subsequently pooled, (along with a pool of equivalent fractions from a second pass of crude material through the column) and dialysed against Tris buffer. The absorbance at 280 nm, 260 nm and 320 nm was measured before and after dialysis (Table 4.6).

	Absorbance			Concentration				
Sample	280 nm	260 nm	320 nm	(μM)	(μg/mL)	A260/A280	Volume (mL)	Yeild (mg)
Crude cell lysate	16.30	27.67	2.63	233.14	13189.1	1.7	400	5276
Purified Ph-6ST	0.55	0.41	0.09	7.87	445	0.75	12	5.3

Table 4.6: Absorbance readings at 280 nm, 260 nm and 320 nm and calculated concentrations of pooled fractions from two nickel (II)-immobilised metal-affinity chromatography purification runs of Ph-6ST expressed in *E. coli Ph-6ST* strain.

The volume post-dialysis was approximately 12 mL and the amount of protein purified was extimated, using the ExPASy protparam derived molar extinction coefficient of 69,790 $M^{-1}cm^{-1}$ (fig. 4.1), to be ~5.3 mg protein. As the initial culture volume was 400 mL, this would have corresponded to a yield of ~13 mg of purified protein per litre of cell culture. However, from SDS-PAGE (fig 4.5 B) the sample contains other contaminating proteins at low levels. Additionally, the purity measurement (last column Table 4.6) is >0.6 (0.75) indicating that there is some nucleic acid contamination.

However, achieving high purity of Ph-6ST was not the primary objective – the purpose of the exercise was to produce enough functional enzyme to catalyse the sialylation of galactose terminal diantennary complex-glycans on rFH expressed in glycoengineered *P. pastoris* strain SuperGal. To test the activity of purified Ph-6ST, a 10 μL reaction was set up containing 200 mM tris-HCL, pH 8.0, 20 mM $MgCl_2$, 50 μM CMP-N-acetylneuraminic acid (CMP-sialic acid), 3.125 μM rFH expressed in SuperGal and 0.3 μM purified Ph-6ST. Three control reactions were also set-up but omitting either CMP-sialic acid, rFH or Ph-6ST. Reactions were left to proceed over night at 20 °C. Sialylation reactions were subjected to SNA and ECL lectin-based western blot analysis to detect for the presence of sialic acid (fig. 4.6).

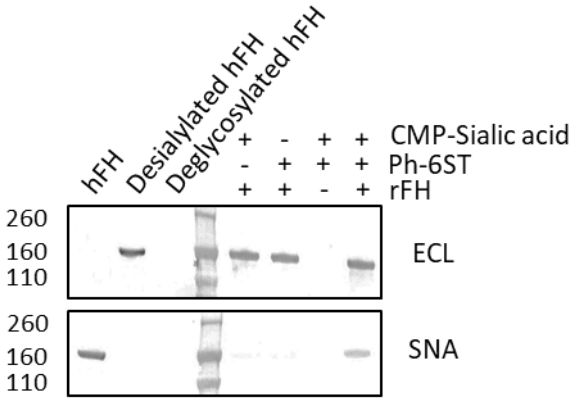


Figure 4.6: ECL and SNA lectin-based western blot analysis of SuperGal expressed rFH subjected to sialylation catalysed by purified Ph-6ST. Top is the ECL LacNAc binding lectin-based western blot. Bottom is the SNA sialic acid binding lectin-based western blot. To the left of the ladder are hFH controls. MW markers are measured in kDa.

A band present in the SNA lectin-based western for the sample that contains CMP-sialic acid, Ph-6ST and SuperGal-rFH (lane on far right of figure 4.6) indicates that Ph-6ST is active and is catalysing the sialylation of rFH. In the ECL lectin-based western, ECL cannot bind to glycans if sialic acid is present on the glycan. The presence of a strong band in the lane on the far right of the ECL lectin-based western blot indicates that there is still a substantial proportion of glycans on SuperGal-expressed rFH that have not been sialylated. Therefore, the next step was to optimise the reaction conditions for sialylation of rFH expressed in *P. pastoris* strain SuperGal.

4.3.4 Optimisation of sialylation of complex galactose terminal glycans by Ph-6ST

To assess sialylation of the galactose-terminating complex glycans carried by rFH, two lectin ELISA procedures were implemented. The first used SNA lectin to detect for the presence of sialic acid, while the second used ECL to detect for the presence of terminal N-acetyllactosamine. These two ELISAs together would show the relative proportion of sialylated to non-sialylated (terminal N-acetyllactosamine) glycans.

For standard curve production in the SNA ELISA, human plasma-purified FH was used. For the ECL ELISA standard curve, human plasma-purified FH that had been desialylated under native conditions (i.e. not denatured) was used. The extent of sialylation of both standards was confirmed by SDS-PAGE and lectin-based western blot (fig. 4.7).

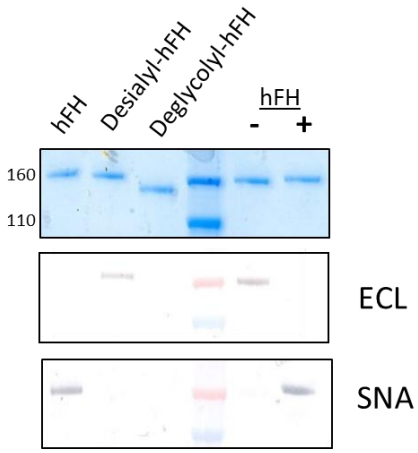


Figure 4.7: SDS-PAGE (top) and lectin-based western blots to detect for N-acetyllactosamie (middle) and sialic acid (bottom) of sialylated and enzymatically desialylated human plasma-purified FH (hFH). The standards to the left of the MW markers were prepared under optimal, denaturing conditions whilst test samples (standards for ELISA) were prepared under non-denaturing conditions. Molecular weight markers are measured in kDa. Desialyl-hFH is hFH desialylated with α(2-3,6,8)-neuraminidase. Deglycolyl-hFH is hFH deglycosylated with PNGase F. + and – indicated hFH treatment with or without α(2-3,6,8)-neuraminidase.

Inspection of the SDS-PAGE (fig 4.7) lanes to the left of the MW markers, reveals the expected small reduction in MW for desialylated hFH compared to sialylated. The ECL lectin-based western confirms that ECL lectin binds to desialylated hFH but not to sialylated hFH. Conversely, the SNA lectin-based western blot shows that SNA lectin binds to hFH but not to desialylated hFH. This confirms that hFH, to be used for SNA ELISA standard curve, is sialylated and that desialylated hFH, to be used for ECL ELISA standard curve, is largely desialylated.

Two separate previous studies explored conditions needed for *Photobacterium sp.* α(2,6)-sialyltranferase activity. Tsukamoto *et al*. (2008) studied full-length α(2,6)-sialyltransferase and a truncated version with deletion of the N-terminal amino acids 2-17 (Δ2-17α(2,6)-sialyltransferase). The other study (Ding *et al*. (2011)) used Ph-6ST (that is N- and C-terminally deleted *Photobacterium sp.* α(2,6)-sialyltransferase with a C-terminal His_6 tag), as used in the current study. In Tsukamoto *et al*. (2008), sialylation was carried out at between 25 - 30° C and reactions were allowed to proceed for up to 48 hrs. In Ding et al., sialylation reactions were carried out at 20 °C for 20 minutes. Neither study used glycoproteins as substrate and the substrates that were used were mono- and disaccharides of lactose or N-acetylgalactosamine conjugated to various tags. Sialylation was observed by HPLC analysis in both studies on all of the substrates tested. Due to this lack of previous reports for glycoproteins as substrates for this enzyme, it was necessary to optimise conditions for sialylation of galactose-terminal glycans on rFH, from scratch.

Importantly, because of inherent heterogeneity in the hFH glycoprofile, these ELISAs were not able to provide absolute quantitative information on sialic acid/galactose content for rFH. Instead, they gave an approximation of the relative amounts of sialic acid/galactose on Ph-6ST-treated rFH carrying humanised galactose-terminal glycans, under the different conditions tested. The temperatures and incubation times for the sialylation reactions were varied, and the extent of sialylation was measured using lectin ELISA (fig. 4.8).
For all sialylation reactions a ten-molar excess of CMP-sialic acid, relative to rFH, was used. The molar excess was based upon a prediction of the number of potential sialylation sites (LacNAc antennae) on the glycans carried by rFH. All eight N-linked glycosylation sites of rFH were assumed to be occupied by diantennary, digalactosylated, complex-type glycans. This would yield 16 LacNAc motifs (two per glycan) that could potentially be sialylated. Therefore, 3.125 µM rFH would yield (3.125 µM * 16 =) 50 µM of LacNAc for sialylation by Ph-6ST.

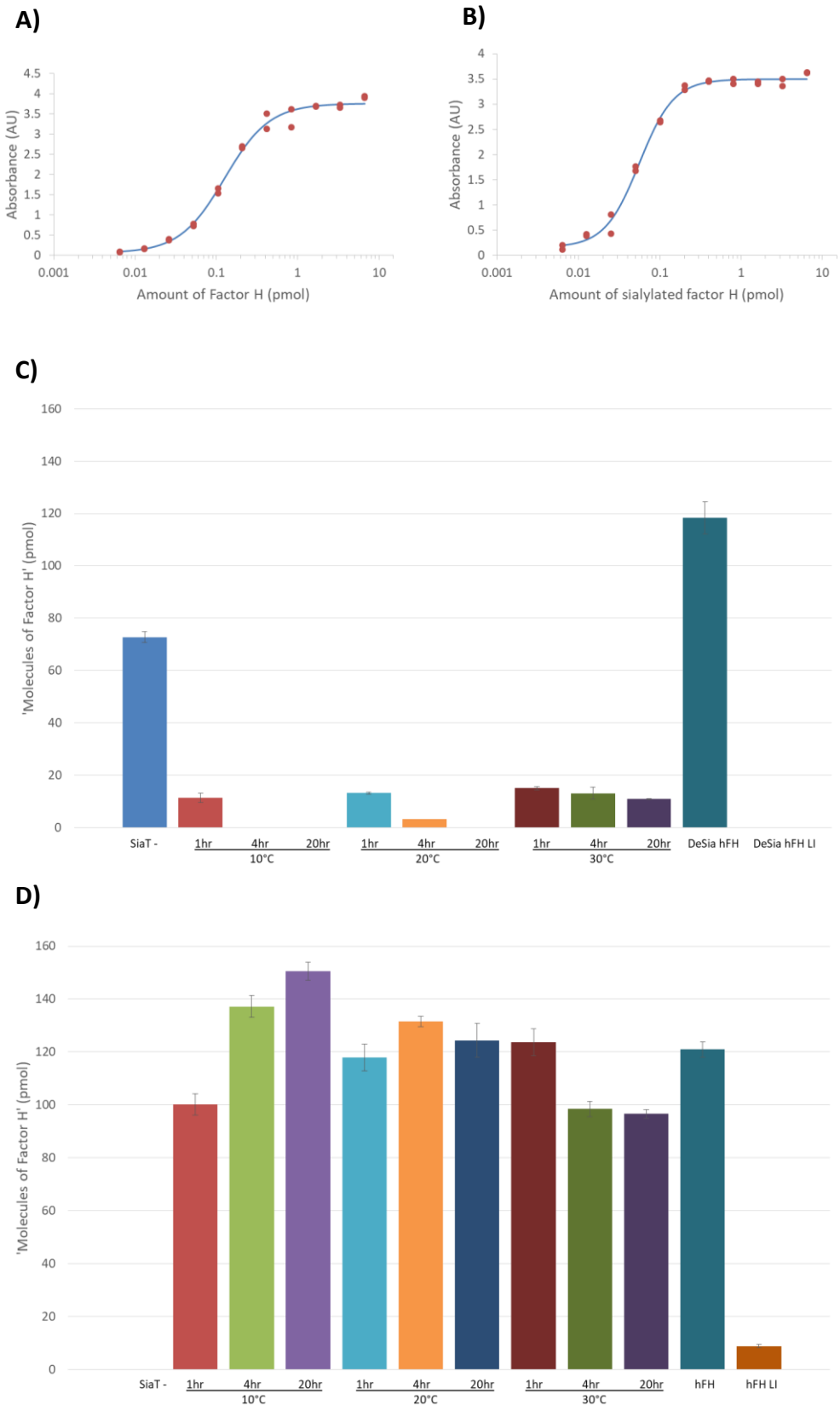


Figure 4.8: Standard curves for terminal N-acetyllactosamine (A) and for terminal sialic acid (B) detection by ECL lectin-based and SNA lectin-based ELISAs, respectively. Both standard curves were a 1:2 dilution series, starting at 10 µg/mL, of desialylated hFH and untreated hFH, respectively. Measurement of relative amounts of terminal N-acetyllactosamine (C) and terminal sialic acid (D), by ECL lectin-based or SNA lectin-based ELISA, on rFH treated with Ph-6ST at different incubation temperatures and reaction times. SiaT- is the reaction without Ph-6ST, hr is hours, hFH is human plasma-purified FH and hFH LI is hFH exposed to lectin incubated with the lectins inhibitory sugar (0.5 M galactose). DeSia hFH is enzymatically desialylated hFH.

The data for ECL ELISA (fig. 4.8 C) show that, as expected, in the absence of Ph-6ST galactose terminal rFH is not sialylated. The data also show that the positive control, desialylated hFH, is bound by ECL lectin and that, when ECL is incubated with an inhibiting sugar (0.5 M galactose), binding by ECL to desialylated hFH is abolished.
At all incubation times, at 30°C, there is still some glycan antennae that have not been sialylated. This was also the case for one hour and four hours incubation times at 20 °C and one hour incubationtion time at 10 °C. At four hours and 20 hours incubation times at 10 °C there is no detectable terminal N-acetyllactosamine on rFH, indicating that at an incubation temperature of 10 °C all available N-acetyllactosamine sites are sialylated by four hours.

Equally, the data from the SNA ELISA (fig. 4.8 D) confirmed that in the absence of Ph-6ST, rFH is not sialylated. The SNA lectin binds to the positive control hFH and, when SNA is incubated with an inhibitory sugar (0.5 M galactose), binding by SNA to hFH is almost completely abolished. The data show that at all incubation temperatures, rFH is sialylated after one hour and that incubation at 10 °C results in the highest level of sialylation. Sialylation continues to increase from four to 20 hours incubation at 10 °C. Conversely, after an initial increase in sialylation from one to 4 hours at 20°C, and after one hour at 30 °C, the level of sialylation decreases. This indicates that at the higher temperatures of 20 and 30 °C there is detectable sialidase activity. However, at 10 °C, there is no detectable sialidase activity and the forward, sialylation reaction is dominating. These data further indicate that the optimal temperature to maximise sialyltransferase activity is 10 °C; and that the optimal length of incubation, of these tested herein, is 20 hours.

Incomplete sialylation of rFH could lead to reduced circulatory half-life *in vivo* due to recognition and clearance by asialoglycoprotein receptors. Thus, the inherent sialidase activity of Ph-6ST could negatively affect *in vivo* half-life of rFH. Kang *et al*. (2015) investigated methods to reduce the sialidase activity of three α(2,6)-sialyltransferases, including Ph-6ST. Kang *et al*. (2015) employed two methods to bias the reaction in favour of the sialyltransferase reaction. In the first instance, Kang *et al*. (2015) hypothesised that as the sialyltransferase reaction proceeds the pool of CMP-sialic acid is depleted. Over time this then allows the sialidase reaction to dominate. Kang *et al*. (2015) showed that multiple additions of CMP-sialic acid, at various time points after the start of the reaction, helped improve the level of sialylated product.

For the second method, Cheng *et al*. (2010) show that the cytidine monophosphate (CMP) product of a sialyltransferase catalysed reaction, released when sialic acid is transferred from CMP-sialic acid to the exposed LacNAc motif at the glycan non-reducing terminus, enhances sialidase activity of the Ph-6ST related α(2,6)-sialyltransferase from *P. damsela*. Kang *et al*. (2015) tested the effect of different nucleotides on sialidase activity of three α(2,6)-sialyltransferases (including Ph-6ST). The authors found that CMP increased sialidase activity of all three α(2,6)-sialyltransferases, whilst CDP, CTP and cytidine slightly reduced sialidase activity. Therefore, Kang *et al*. (2015) hypothesised that the addition of alkaline phosphatase to the reaction would dephosphorylate CMP evolved in the sialylation step and thus, ameliorate sialidase activity. Kang *et al*. (2015) were subsequently able to show that alkaline phosphatase did indeed reduce sialidase activity of the three α(2,6)-sialyltransferases tested.

However, the temperature and time of incubation for sialyltransferase assays, conducted by Kang *et al*. (2015), were 30 °C for 30 minutes. These conditions are different from the optimal conditions for Ph-6ST determined in this study. Therefore, the effects of multiple additions of the substrate CMP-sialic and alkaline phosphatase, separately and together, throughout the reaction were explored and assessed by ECL and SNA ELISA. Three sets of nine sialyltransferase reactions were carried out with either one, two or three 'feeds' with either CMP-sialic acid (in addition to the CMP-sialic acid included in the reaction mixture at the start of the reaction) or alkaline phosphatase - on their own - or CMP-sialic acid plus alkaline phosphatase at 30 minutes, two hours and four hours from the start of each reaction. All reactions were carried out at 10 °C for 20 hours (fig 4.9).

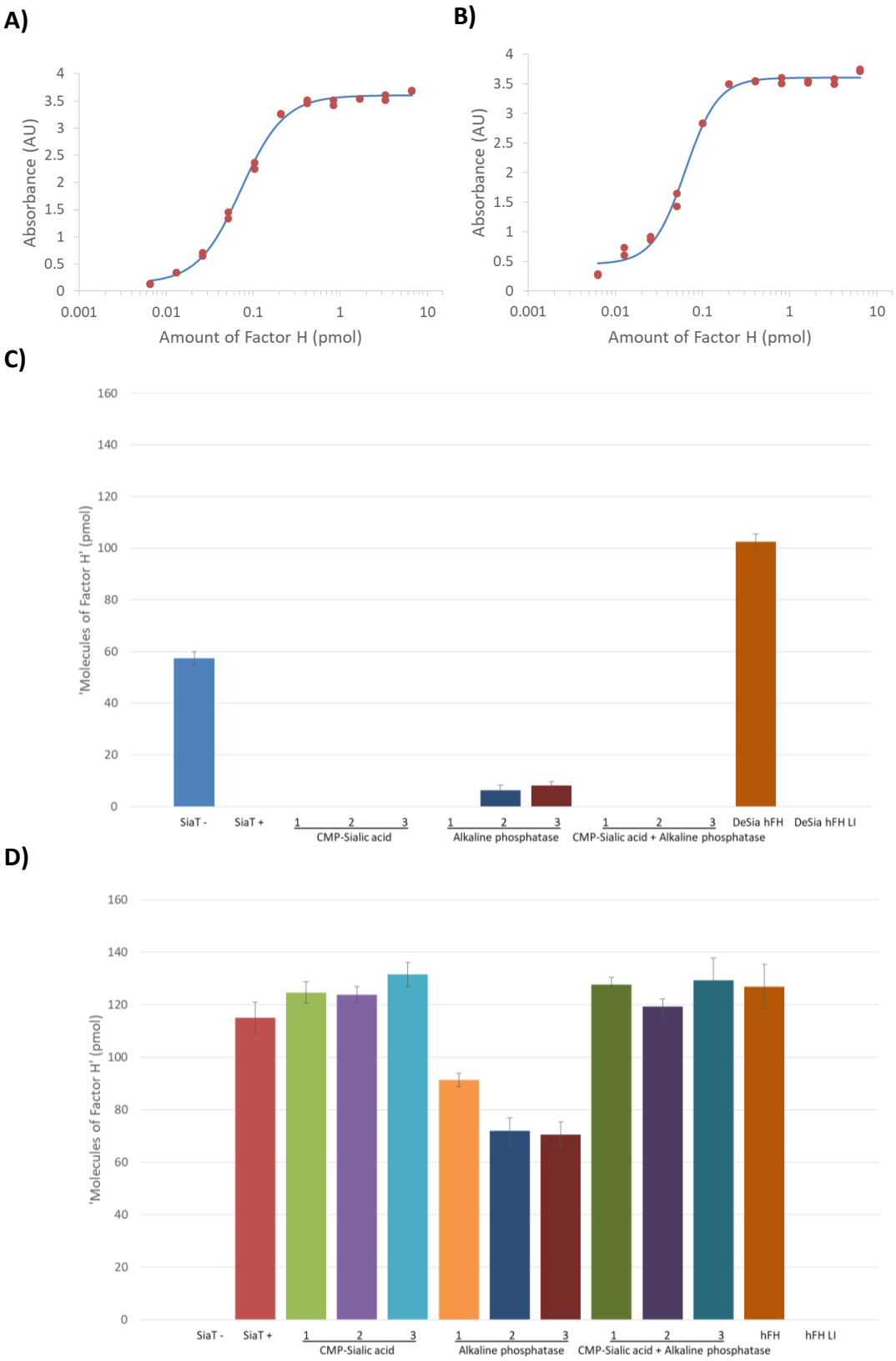


Figure 4.9: Standard curves for terminal N-acetyllactosamine (A) and for terminal sialic acid (B) detection by ECL and SNA lectin ELISA, respectively. Measurement of relative amounts of terminal N-acetyllactosamine (C) and terminal sialic acid (D), by ECL and SNA lectin ELISA, on rFH treated with Ph-6ST incubated at 10 °C for 20 hours. The numbers 1, 2 and 3 refer to the number of feeds with either CMP-sialic acid, alkaline phosphatase or both. Feeds were at 30 minutes, two hours and four hours from the start of the reaction. SiaT- is the reaction without Ph-6ST, SiaT+ is the reaction with Ph-6ST at 10 °C for 20 hours but with no additional feeds, hr is hours, hFH is human plasma-purified FH and hFH LI is hFH exposed to lectin incubated with the lectins-inhibitory sugar (0.5 M galactose). DeSia hFH is enzymatically desialylated hFH.

Alkaline phosphatase does not appear to enhance sialylation of rFH and instead multiple additions of alkaline phosphatase may actually enhance sialidase activity of Ph-6ST. There is no detectable N-acetyllactosamine after a single addition of alkaline phosphatase (fig. 4.9 C) but after two or three additions, the amount of N-acetyllactosamine detected on rFH increases slightly. Supporting this finding, the SNA-lectin ELISA (fig. 4.9 D) shows that sialylation is highest after a single addition of alkaline phosphatase, but the amount of sialic acid on rFH falls after a second and third addition of alkaline phosphatase.

Multiple additions of CMP-sialic acid improves the sialic acid content of rFH compared to no additions (SiaT+ reaction) (fig. 4.9 D). Three additions of CMP-sialic acid confers the greatest enhancement to sialylation of rFH by Ph-6ST. Addition of CMP-sialic acid and alkaline phosphatase together does not enhance sialylation of rFH beyond that seen with three additions of CMP-sialic acid alone. These data suggest that three 'feeds' of CMP-sialic acid, but not alkaline phosphatase, over the course of the reaction enhances sialylation of rFH.

Taken together, these studies suggested that suitable reactions conditions for Ph-6ST-catalysed sialylation of galactose-terminating diantennary complex glycans on rFH, containing 3.125 µM rFH (approximately 50 µM LacNAc), 500 µM CMP-sialic acid, 20 mM $MgCl_2$ in Tris buffer at pH 8.0, are as follows: incubation at 10 °C for 20 hours with up to three 'feeds' of ~1.6 nmol CMP-sialic acid (at 30 minutes, two hours and four hours after the start of the reaction).

4.3.5 Purification of Ph-6ST sialylated rFH

In Chapter 3 section 3.3.7, rFH expressed in glycoengineered *P. pastoris* strain SuperGal was purified, and yeast and hybrid-type glycans were enzymatically removed by treatment with Endo H_f. After Endo H_f treatment, rFH was purified and confirmed to carry exclusively diantennary digalactosylated complex-type glycans by MALDI-TOF mass spectrometry. Using the optimal conditions determined in this Chapter, this rFH was then sialylated and contaminants were purified away from sialylated rFH by cation-exchange chromatography using a Resource S column (fig. 4.10).

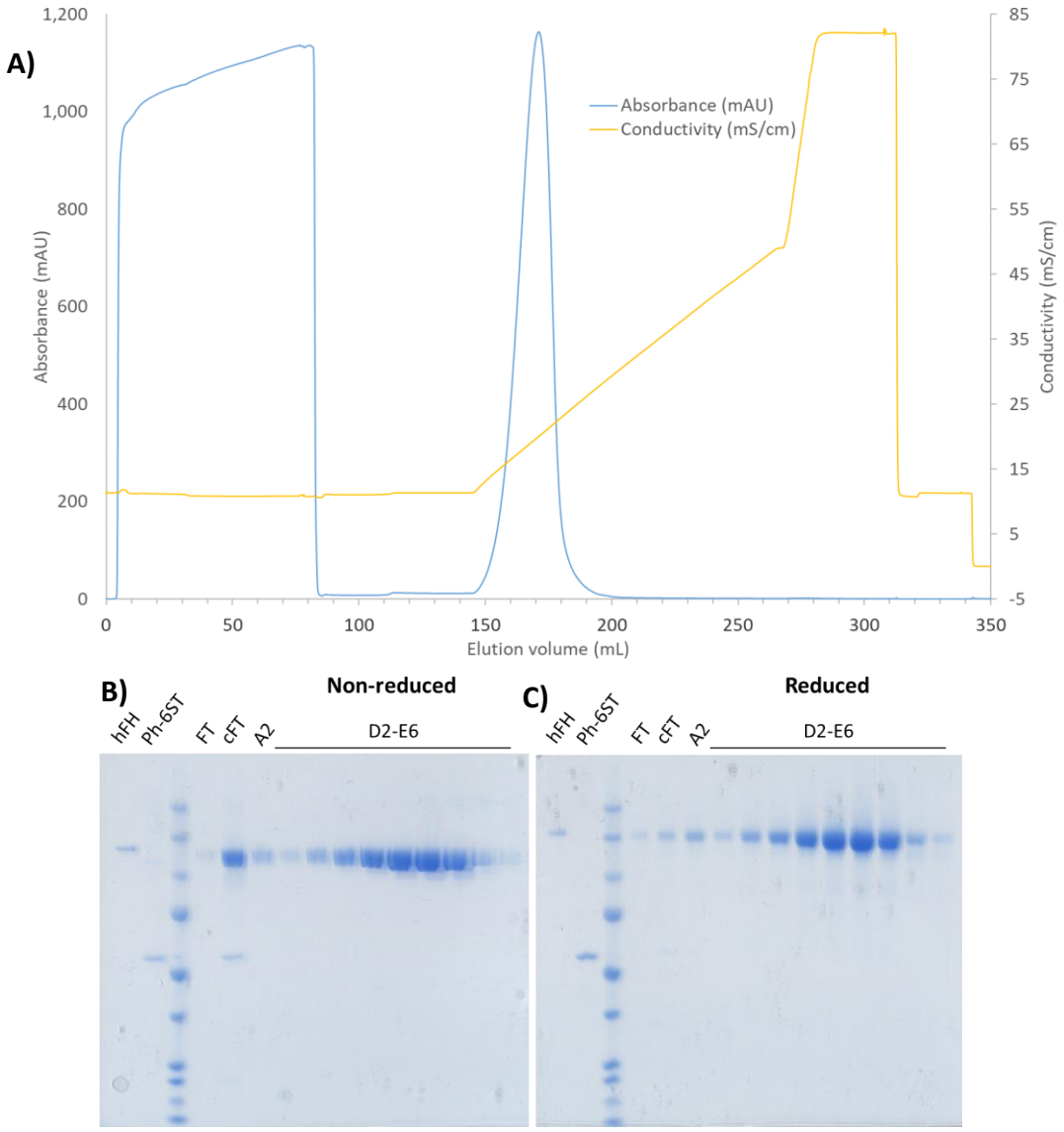


Figure 4.10: Cation-exchange chromatogram (A) of the purification of sialylated rFH expressed in glycoengineered *P. pastoris* strain SuperGal and after yeast glycan removal with Endo H_f. SDS-PAGE of human factor H (hFH), purified recombinant Ph-6ST expressed in *E. coli*, flow-through (FT, cFT is 10x concentrated flow through), wash fraction A2 (elution volume = 80 mL) and peak fractions D2-E6 (elution volumes 150.5 – 182.5 mL) (every second fraction run on the gel) under non-reducing (B) and reducing (C) conditions.

The chromatogram (fig. 4.10 A) for purification of sialylated rFH shows a single peak eluting early on in the elution phase (at an elution volume of 150-200 mL). The SDS-PAGE gels show a band corresponding to the MW of Ph-6ST in the ten-times concentrated flow-through (figs. 4.10 B and C). Importantly, no Ph-6ST was detected in the elution peak fractions containing rFH. Fractions D6 – E2 (elution volume of 158.5 – 174.5 mL) were deemed to contain the highest concentrations of rFH and so were pooled and dialysed into neutral storage buffer.

Measurement of absorbance readings and assessment of purity of cation-exchange purified sialylated rFH was hampered by the fact that the nucleoside cytidine (present in the substate CMP-sialic acid and the product CMP) absorbs wavelengths of 280 nm and 260 nm. These cytidine derivatives elevate the A_{260}/A_{280} ratio of non-purified sialylated rFH (row two and three of table 4.7) preventing accurate quantification of protein content. But after purification by cation-exchange chromatography, the A_{260}/A_{280} returns to approximately 0.6.

Sample	Absorbance			Concentration		Volume (mL)
	A260	A280	A260/A280	(μM)	(mg/mL)	
DiGal rFH	0.329	0.546	0.603	11.06	1.66	20.8
Nonpurified DiSia rFH	0.935	0.917	1.020	18.58	2.79	75.6
Nonpurified DiSia rFH post dialysis	0.497	0.519	0.958	10.51	1.58	83
Purified DiSia rFH pst dialysis	0.210	0.358	0.587	7.25	1.09	17.5

Table 4.7: Absorbance readings at wavelengths of 280 nm and 260 nm for samples of rFH before sialylation, after sialylation before and after dialysis and after purification of sialylated rFH by cation-exchange chromatography. DiGal rFH is rFH expressed in SuperGal and with yeast glycans removed by Endo H_f. DiSia rFH is the same, but sialylated with Ph-6ST.

An approximate total of 34.5 milligrams rFH was sialylated and loaded onto the cation-exchange chromatography column. After purification and dialysis against PBS, pH 7.4, 19 milligrams of rFH was recovered. Purified Ph-6ST-treated rFH was then assessed for the extent of sialylation by MALDI-TOF mass spectrometry.

4.3.6 Glycan analysis by MALDI-TOF MS of Ph-6ST sialylated rFH

Kang *et al*. (2015) report that sialylation of diantennary glycans, catalysed by Ph-6ST, can yield both monosialylated and disialylated glycans. Binding of ECL lectin to N-acetyllactosamine is blocked upon sialylation. While complete loss of detectable binding by ECL lectin occurred under optimal conditions for Ph-6ST (fig. 4.9 C) it is not known whether both antennae of a diantennary complex glycan need to be sialylated to abolish ECL binding. Put simply, after treatment of rFH with Ph-6ST, are the glycans monosialylated, disialylated or a mixture of the two?

To test this, the glycans were enzymatically liberated from sialylated, purified rFH by incubation with PNGase F. The glycans were then purified and analysed by MALDI-TOF mass spectrometry to assess the extent of sialylation.

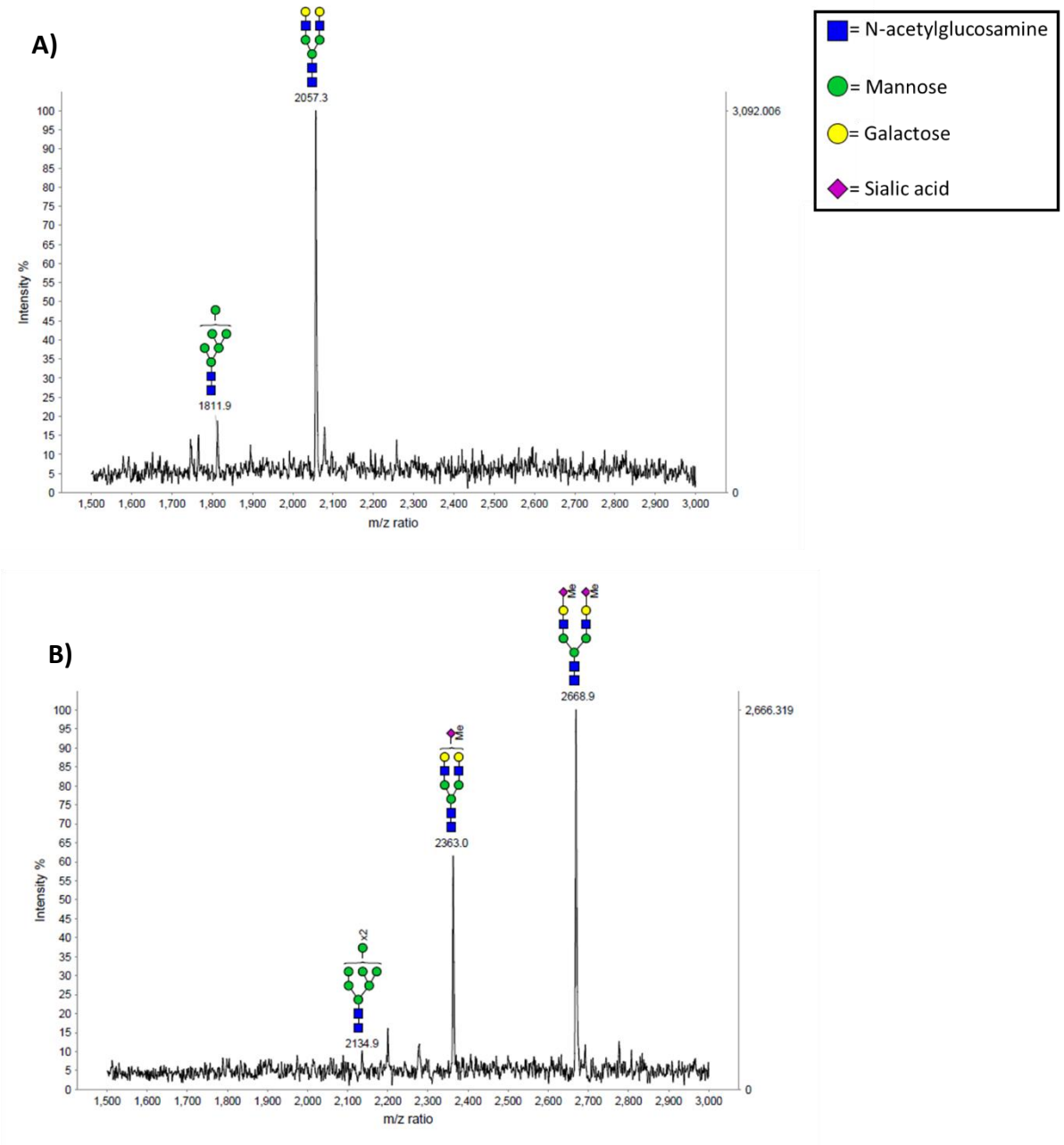


Figure 4.11: MALDI-TOF mass spectra of the glycoprofile of rFH before (A) and after (B) enzymatic sialylation catalysed by Ph-6ST.

The mass spectra show that before sialylation (fig. 4.11 A) the glycans on rFH are almost exclusively digalactosylated diantennary complex-type glycans. After sialylation (fig. 4.11 B) however, the glycoprofile consists of mono- and disialylated diantennary complex-type glycans. The disialylated species has the higher relative abundance. Also, there is no detectable digalactosylated diantennary glycans.

4.4 Discussion

This chapter has demonstrated the successful production of catalytically active, recombinant α(2,6)-sialyltransferase from *Photobacterium sp*. in *E. coli* and, that this sialyltransferase can be used to sialylated rFH expressed in glycoengineered *P. pastoris* strain SuperGal, to yield a protein glycoprofile consisting of mono- and disialylated species. What effect the terminal galactose moieties of monosialylated glycans might have on the half-life of rFH is unknown. But, due to accelerated clearance of non-sialylated glycoproteins from circulation by the asialoglycoprotein receptor (Ashwell and Harford, 1982 and, Schwartz, 1984 and discussed in Chapter 1 section 1.1.3), it is likely that complete sialylation – that is, all glycans in the disialylated form – will be necessary to maximise half-life and thus therapeutic efficacy, so this remains an unmet goal.

Kang *et al*. (2015) first reported sialidase activity of Ph-6ST and, in addition, they compared sialidase activity of Ph-6ST with two other bacterial α(2,6)-sialyltransferases and found that, under the conditions they used, Ph-6ST has the highest sialidase activity.
In light of this, with the view to improving the disialylated glycan content of rFH it will be necessary to test a range of α(2,6)-sialyltransferases under different experimental conditions.

As demonstrated in this chapter, *in vitro* sialylation hinders downstream processing – this would add significant cost to manufacturing processes if rFH expressed in SuperGal were to be developed as a biotherapeutic. With this in mind, a strain of *P. pastoris* glycoengineered to sialylated rFH *in vivo* would be immensely valuable in improving productivity by reducing downstream processing.

Chapter 5

Consolidation of the humanised glycan biosynthetic pathway onto a single plasmid using inABLE multi-part DNA assembly

5.1 Overview

The aim of this work was to engineer *P. pastoris* to biosynthesis a fully humanised terminal sialylated complex-glycan *in vivo*. This was achieved in Chapter 3. However, the GlycoSwitch technology, as well as stopping short of the fully humanised sialylated glycan, had additional draw backs: i) it used all available dominant, antibiotic resistant selectable markers available for use in *P. pastoris*, ii) potential for plasmid integrant instability due to repeated use of the same, limited set of regulatory gene sequences and, iii) potential for bottle necks in the biosynthetic pathway due to this same limited set of regulatory elements. Each draw-back will be discussed below.

i) Dominant, antibiotic resistant selectable markers

Antibiotic selectable markers present several advantages over selection by auxotrophic gene complementation in that they offer straightforward dominant selection without the need for production of auxotrophic strains by gene knockout, this is especially inconveniencing for instances where multiple genes are introduced for biosynthetic pathway manipulation. An additional advantage of selection by antibiotic resistance is that strains that contain multiple copies of a target plasmid can be selected for by increasing concentrations of antibiotic (Sunga *et al*., 2008 and, Gong *et al*., 2013*b*) – a useful tool for increasing protein titres or overcoming biosynthetic pathway bottlenecks by increasing expression of low efficiency enzymes.

However, *P. pastoris* lacks antibiotic resistant selectable markers, just five have been reported. These confer resistance to the antibiotics blasticidin S (Drocourt *et al*, 1990 and, Kimura *et al*, 1994), Zeocin™ (Chen *et al*., 2012), hygromycin (Yang *et al*., 2014), G418 (Scorer *et al*., 1994) and nourseothricin (Chen *et al*., 2012). The glycoengineering technology platform, GlycoSwitch (including the starting glycoengineered strain SuperMan 5), uses all five antibiotic selectable markers available for use in *P. pastoris* (Jacobs *et al*., 2009). To demonstrate the lack of *P. pastoris* antibiotic selectable markers available, in this study the Zeocin™ selectable marker was used twice. This marker was present in both the

pPICZα plasmid containing the *CFH* gene, as well as on the pGlycoSwitch-GnTI plasmid containing the gene *GnTI* encoding GlcNAc transferase I. In this regard, as SuperMan5 *CFH* was already Zeocin resistant, *GnTI* expressing strains could not be selected for on the basis of Zeocin resistance. So, careful selection methods were required, based on identification of the GlcNAc transferase I modification to the glycosylation pathway, to select for strains that expressed both *CFH* and *GnTI*.

One of the principal aims of this study was to attempt the ambitious step of sialylation of rFH *in vivo*. This procedure would involve the introduction into *P. pastoris* of the six genes responsible for biosynthesis of Neu5Ac, charging with cytidine monophosphate (CMP), transport of CMP-Neu5Ac into the golgi and transfer of Neu5Ac onto LacNAc glycan termini. Therefore, making available antibiotic resistant selectable markers would reduce the number of genetic manipulation steps involved for bioengineering in the fully humanised glycan biosynthetic pathway.

ii) Plasmid integrant stability

Demonstration of genetic stability of strains produced by recombinant DNA technology is a key requirement as outlined by the World Health Organisation (WHO, 2013) and is an important factor in ensuring the quality of a recombinant biotherapeutic (Burns, 1993). Lim *et al*. (2000) produced a multi-copy integrant strain of *P. pastoris*, expressing the recombinant protein Gaurmerin, and was able to show that this strain was stable up to 70 generations, exceeding the approximately 20 generations *P. pastoris* goes through in a standard high-cell density, large scale fermentation process. However, Lim *et al*. (2000) showed that the multi-copy integrant strain held just two copies of the expression cassette. In comparison, the SuperGal *CFH* strain produced in this study (Chapter 3) contains 5 copies of the Glyceraldhyde-3-phosphate promoter (GAPp), 4 exogenous and 1 endogenous (the site of integration), 4 exogenous alchohol oxidase I terminator sequences (AOXIt), 3 exogenous translation elongation factor 1-α promoters (TEFp), 3 exogenous terminator sequences from *Ashbya gossypii* and 4 pUC vector plasmid backbone sequences (for plasmid amplification in *E. coli*). Assuming all plasmids integrated in neighbouring genomic GAPp sites, these homologous sequences will be near each other in the genome. Thus, there is a real risk that the SuperGal *CFH* strain could be genetically unstable and that loss of one or more integrated pGlycoSwitch plasmids by homologous recombination could occur during long-term storage and fermentation. Indeed, in confirmation of the significance of ensuring genetic stability, efforts have been made in the literature to

develop novel promoter sequences with the specific intention to reduce the number of repeat sequences in extensively genetically engineered *P. pastoris* (Vogl *et al*., 2015).

iii) Potential for bottlenecks in the engineered glycan biosynthetic pathway

Unlike DNA/RNA or protein synthesis, glycan synthesis is a non-template driven process and the glycan sequences is dependent on the spatial-temporal localisation and relative activities of the glycan biosynthetic enzymes in the secretory pathway. The GlycoSwitch technology incorporates peptide localisation sequences into the exogenous glycan modification enzymes to ensure the spatial-temporal demand for correct glycan sequence is met. For example, the early acting ManI is localised to the ER – golgi by fusion to the HDEL ER retention signal sequence (Callewaert *et al*., 2001). Conversely fusion of GnTI with the transmembrane domain from the *S.cerevisae* α(1,2)-mannose transferase Kre2p localises the enzyme to the cis-medial golgi (Lussier *et* al., 1995 and, Vervecken *et al*., 2004). This is a necessary process to avoid missed steps in the glycan biosynthetic process. However, low enzyme activity in the biosynthetic pathway could also be a cause of missed steps in glycan biosynthesis. For example, a particular enzyme at a specific point in the glycan biosynthetic pathway with low activity may not be able to catalyse the formation of its product as efficiently as an enzyme which acts before it in the pathway. The result will be a build-up of substrate for the low activity enzyme.

Efforts have been made to fully characterise promoter libraries with the aim of fine control of gene expression to optimise protein expression or to manipulate flux of biosynthetic pathways. Most of these libraries have focused on *E. coli* (Alper *et al*., 2005 and, Nevoigt *et al*., 2007), however, a few have sort to develop promoter libraries in *P. pastoris* (Stadlymayr *et al*., 2010; Qin *et al*., 2011; and, Vogl and Glieder, 2013). These libraries have identified promoters of strengths varying from weak to strong, and present a useful tool kit for the intelligent design for the remodelling of the GlycoSwitch glycan biosynthetic pathway. Jacobs *et al*. (2009) noted that the doubling time of the *P. pastoris* strain with humanised glycosylation was double that of the SuperMan 5 strain. It is likely that repeat use of the high expression strength GAPp has burdened the cell with high level gene expression compromising its growth potential. Therefore, consolidation of the GlycoSwitch technology onto a single expression plasmid presented the opportunity to identify genes encoding biosynthetic glycosyltransferases with low activity and to couple these genes to promoters with strong expression profiles, and *vice versa*, with the aim of predicting and

overcoming potential bottlenecks in the glycan biosynthetic pathway to reduce glycan heterogeneity and to ease the protein expression burden of the cells.

5.1.1 *OCH1* encodes α(1,6)-mannosyltransferase

OCH1 encodes α(1,6)-mannosyltransferase in yeast (Nakayama *et al*., 1992). During yeast glycan biosynthesis, covered in Chapter 1 section 1.4.1, the gene product of *OCH1* (OCH1) catalyses the transfer of a single Man onto the 8 Man glycan as it leaves the ER (Nakayama *et al*., 1997). The 9 Man product is the substrate for further mannosyltransferases which lead to the formation of hypermannosylated glycans in yeasts.

Vervecken *et al*. (2004) hypothesised that deletion of the *P. pastoris OCH1* gene would prevent hypermannosylation and that the subsequent 8 Man glycan could serve as a platform for further *in vivo* glycoengineering. Vervecken *et al*. (2004) achieved an *Δoch1* knockout strain by incorporating a truncated version of the *och1* gene into a vector plasmid (fig. 5.1). RE digestion at a site within the truncated *och1* gene and transformation into *P. pastoris* allowed for single site homologous recombination to occur between the endogenous *OCH1* gene and the truncated *och1* sequence. This resulted in a plasmid inserted into the *OCH1* genomic locus and produced two truncated, non-functional *och1* sequences.

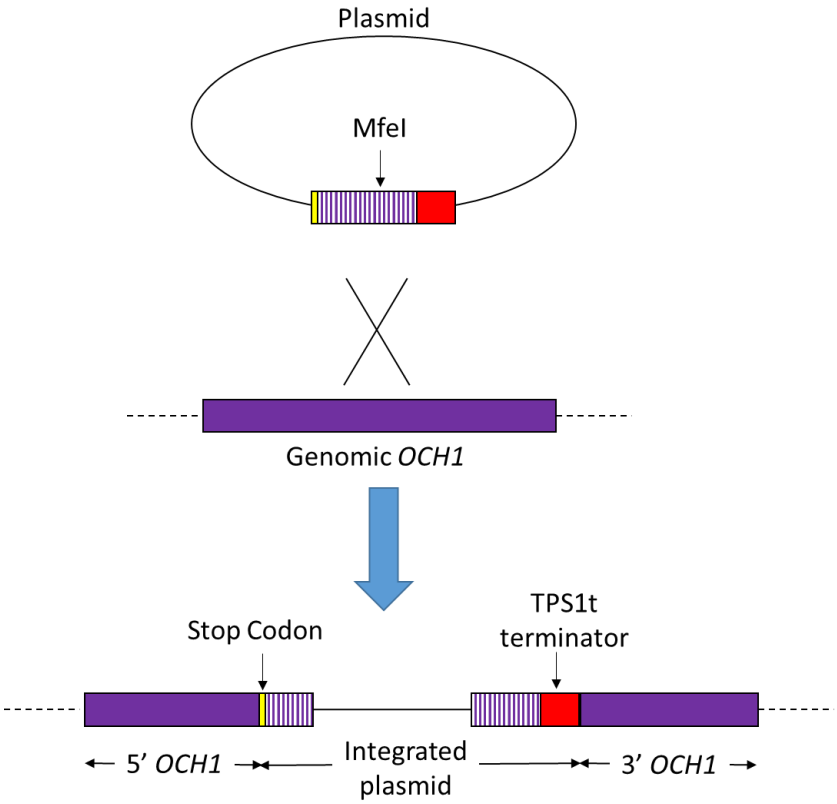


Figure 5.1: Schematic diagram of the simultaneous plasmid integration and *OCH1* gene deletion. The truncated *och1* sequence (purple striped rectangle) has on the 5' end an in frame stop codon (small yellow rectangle) and in frame at the 3' end is the TPS1t terminator sequence (red rectangle). Linearisation in the truncated *och1* by RE digest using the RE MfeI, prior to transformation into *P. pastoris*, facilitates single homologous recombination between the plasmid truncated *och1* sequence and the *P. pastoris* genomic *OCH1* gene. After homologous recombination, the genomic *OCH1* gene is divided into two halves, with the plasmid inserted in the centre. The 5' end of the plasmid places a premature stop codon in frame and downstream of the 5' half of the *OCH1* sequence. The 3' end of the plasmid places the TPS1t terminator sequence upstream and in frame of the 3' end of the *OCH1* gene to prevent transcriptional readthrough from the plasmid to the 3' end of *OCH1*.

The aim of this study was initially to consolidate the first half of the GlycoSwitch technology onto a single plasmid with a view to glycoengineering a *P. pastoris* strain with an unmodified glycosylation pathway, KM71H, to produce a hybrid-type glycan with a single LacNAc antenna, $Gal_1GlcNAc_1Man_5GlcNAC_2$. Included in this plasmid was a truncated *och1* sequence that, when linearised, would direct plasmid integration at the endogenous genomic *OCH1* locus whilst simultaneously deleting the *OCH1* gene. In addition, this plasmid was generated with unique promoter-gene-terminator pairs with a view to predict and overcome bottlenecks in the biosynthetic pathway by coupling promoter strength to enzymatic activity and to reduce potential genetic instability from repeated use of the same promoter and terminator sequences.

5.2 Materials and Methods

5.2.1 InABLE plasmid assembly

5.2.1.1 Truncated part PCR amplification

All primers and oligonucleotide sequences were ordered from Sigma Genosys. *P. pastoris* codon optimised *ManI* and truncated *och1* gene sequences were synthesised by Invitrogen GeneArt gene synthesis service (sequences in the appendix). *ManI* and *och1* were designed as readily useable parts for inABLE assembly and flanked by type II S RE recognition sites for liberation of the part from the vector plasmid backbone. The type II S RE will bind at the recognition site, move along the DNA strand and cut at a defined distance from the recognition site. *In silico* design ensures the type II S RE cuts the TP to generate the 5' and 3' TP overhangs necessary for ligation with PO and LO sequences. For use in inABLE the following DNA sequence had to be PCR amplified using primers designed to incorporate type II S RE sites at either end of the DNA sequence: *GnTI*, *GalT* (including shuttle vector plasmid backbone), AODt and ILV5p. PCR reactions were carried out using KOD Hot Start DNA Polymerase (Merck Millipore) (*och1*, *GnTI*, AODt and ILV5p) or Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) (*GalT*) according to the conditions in table 5.1.

KOD Hot Start DNA polymerase	
Component	Final concentration
10x KOD Hot start Buffer	x1
$MgSO_4$	1.5 mM
dNTPs	200 µM
ddH_2O	Up to 50 µL
Forward primer	0.3 µM
Reverse primer	0.3 µM
Template DNA	10 ng
KOD Hot Start DNA polymerase	0.02 U/µL

Temperature (°C)	Time (s)	Cycles
95	120	
95	20	x30
55	10	
70	25 per kb	
70	300	

Q5 Hot Start HF DNA polymerase	
Component	Final concentration
5x Q5 reaction buffer	1x
ddH_2O	Up to 50 µL
dNTPs	200 µM
Forward primer	0.5 µM
Reverse primer	0.5 µM
Template DNA	1 ng
Q5 Hot Start HF DNA polymerase	0.02 U/µL

Temperature (°C)	Time (s)	Cycles
98	30	
98	10	x30
55	30	
72	30 per kb	
72	300	

Table 5.1: Reaction components and PCR cycle conditions for PCR amplification of inABLE truncated parts using either KOD Hot Start DNA polymerase or Q5 Hot Start High-Fidelity (HF) DNA polymerase.

The following primers were used for PCR amplification, type II S RE recognition sites are capitalised:

5.i) *GnTI*_For: 5' – gagagaagcggccgcGCTCTTCgcaggtgcggttattgttc – 3'
5.ii) *GnTI*_Rev: 5' – ttagatctcgagGCTCTTCgggcctaattccagtaggatcatag – 3'
5.iii) *GalT*_bb_For: 5' – gaggaatctagaGCTCTTCgtagttttgatattgtgcgggc – 3'
5.iv) *GalT*_bb_Rev: 5' – ttagatttaattaaGCTCTTCgggcttgagatccttttttctgcg – 3'
5.v) AODt_For: 5' – gaggaagcggccgcGCTCTTCgaaccaattcaggcaatgtg – 3'
5.vi) AODt_Rev: 5' – ttagatctcgagGCTCTTCgggcatgtatacggatgttttattatctatttatgc – 3'
5.vii) ILV5p_For: 5' – gaggaagcggccgcGCTCTTCgttgacttcgtgaaagtttctttag – 3'
5.viii) ILV5p_Rev: 5' – ttttttcggataatttttttaaagcgttagatctcgagGCTCTTCgggcaaaa – 3'

The PCR products were separated by 1.0% (w/v) agarose gel electrophoresis (100 V, 45 minutes) (fig. 5.1 A) and gel extracted using Zymoclean gel DNA recovery kit (Zymo Research), procedure in Chapter 4 section 4.2.1.1. The remaining parts ADH1p, GAPp, GAL7t, and TPS1t existed as pre-generated parts, flanked by type II S RE recognition sites and cloned into plasmids for selection and amplification in *E. coli*.

5.2.1.2 Linker and Part Oligonucleotide phosphorylation and annealing: first round

In matched pairs, single-stranded oligonucleotides were subjected to a phosphorylation and annealing reaction, procedure in Chapter 4 section 4.2.1.2, to yield LO/PO_X.1, LO/PO_A.2, LO/PO_A.3, LO/PO_X.2, LO/PO_X.3, LO/PO_B.2, LO/PO_B.3, LO/PO_X.4, LO/PO_X.5, LO/PO_C.2, LO/PO_C.3 and LO/PO_X.6 for the assembly reaction (table 5.2).

Oligo length	LO strands	Sequence 5' - 3'	PO strands	Sequence 5' - 3'
Long	A.2	gccatgagatttccttctagttccgtcttggcac	A.2	ctggtccaatcaatccaagtgccaagacggaact
Short	A.2	agaaggaaatctcat	A.2	ttggattgattggac
Long	A.3	gcctgcagctaaggtaatcagatccaagtttccc	A.3	gtttgccccggaagattggggaaacttggatctg
Short	A.3	attaccttagctgca	A.3	caatcttccggggca
Long	B.2	gccatggccctctttctcagtaagagactgttga	B.2	ctgcaatgacggtaaatctcaacagtctcttact
Short	B.2	gagaaagagggccat	B.2	gatttaccgtcattg
Long	B.3	gccaaagaaagtggaatattcattcatatcatat	B.3	aggcagctaatagaaaaaatatgatatgaatgaa
Short	B.3	tattccactttcttt	B.3	tttttctattagctg
Short	C.2	gccatgctgcttaccaaaaggttttcaaagctgt	C.2	ctatgaacgtcagcttgaacagctttgaaaacct
Short	C.2	tttggtaagcagcat	C.2	tcaagctgacgttca
Long	C.3	gcctggtaagctatattcgccgtttctgtcattt	C.3	agggtccgtacaaaacgcaaatgacagaaacggc
Short	C.3	gaatatagcttacca	C.3	gcgttttgtacggac
Long	X.1/3/5	gccatgaaacattcttccgatatctgcattgttg	X.1	gtacgaagagaccagcaccaacaatgcagatatc
Short	X.1/3/5	ggaagaatgtttcat	X.1	gtgctggtctcttcg
Long	X.2/4/6	gcccgaagagacaggtggcacttttcggggaaatgtgcgcggaa	X.3	agacgaagagaccagcaccaacaatgcagatatc
Short	X.2/4/6	ccgaaaagtgccacctgtctcttcg	X.3	gtgctggtctcttcg
Long			X.5	caacgaagagaccagcaccaacaatgcagatatc
Short			X.5	gtgctggtctcttcg
Long			X.2/4/6	agaaaaataaacaaataggggttccgcgcacatttcc
Short			X.2/4/6	cccctatttgtttatttt

Table 5.2: Pairs of oligonucleotide sequences annealed and phosphorylated to produce LO and POs associated with truncated parts A.1-4, B.1-4 and C.1-4 required for the assembly of "nested" parts A, B and C.

5.2.1.3 Part-linker fusion reactions: A1-3, B1-3 and C1-3

Twelve separate part-linker fusion reactions were set up, conditions in Chapter 4 section 4.2.1.3. PO, LO and parts were combined according to table 5.3.

Part-linker reaction	PO	Part	LO	Part-linker concentration (ng/µL)
1	PO_X.1	A.1	LO_A.2	7.60
2	PO_A.2	A.2	LO_A.3	18.50
3	PO_A.3	A.3	LO_X.2	12.20
4	PO_X.2	A.4	LO_X.1	12.20
5	PO_X.3	B.1	LO_B.2	2.40
6	PO_B.2	B.2	LO_B.3	16.10
7	PO_B.3	B.3	LO_X.4	2.30
8	PO_X.4	B.4	LO_X.3	12.20
9	PO_X.5	C.1	LO_C.2	8.50
10	PO_C.2	C.2	LO_C.3	2.24
11	PO_C.3	C.3	LO_X.6	3.20
12	PO_X.6	B.4	LO_X.5	58.60

Table 5.3: Twelve part-linker fusion reactions were set up. Each PO and LO associates with a truncated part via overhanging complementary triplet sequences at the 3' end of PO, 5' end of LO and 5' and 3' ends of the part. Part-linker concentrations are DNA concentration of part-linkers after isolation and gel extraction.

Part-linkers were isolated by 1% (w/v) agarose gel electrophoresis (100 V, 45 minutes) (fig. 6.2), gel extracted and purified using a Zymoclean gel DNA recovery kit (Zymo Research) according to conditions in Chapter 4 section 4.2.1.1. DNA concentration of gel extract part-

linkers was measured by fluorimetry (Invitrogen, Qubit) (fig. 5.3) and DNA molecular weights were calculated using the part-linker fusion DNA sequences and the online "Sequence Manipulation Suite: DNA Molecular Weight" software.

5.2.1.4 inABLE assembly of "nested" parts A, B and C

The inABLE DNA assembly reaction was carried out by combining the part-linker fusion products in the proportions outlined in table 6.4 and incubating at 37 °C for 30 minutes.

Assembly Reaction	Reagants	Volume (µL)	Assembly Reaction	Reagants	Volume (µL)	Assembly Reaction	Reagants	Volume (µL)
A	PO_X.1-A.1-LO_A.2	2.6	B	PO_X.3-B.1-LO_B.2	5.6	C	PO_X.5-C.1-LO_C.2	1.3
	PO_A.2-A.2-LO_A.3	2.6		PO_B.2-B.2-LO_B.3	2.5		PO_C.2-C.2-LO_C.3	2.8
	PO_A.3-A.3-LO_X.2	1.0		PO_B.3-B.3-LO_X.4	6.1		PO_C.3-C.3-LO_X.6	4.9
	PO_X.2-A.4-LO_X.1	4.5		PO_X.4-B.4-LO_X.3	4.5		PO_X.6-C.4-LO_X.5	4.3
	Buffer 2 (10x)	1.5		Buffer 2 (10x)	2.5		Buffer 2 (10x)	1.8
	PEG 8000 (50%[w/v])	1.5		PEG 8000 (50%[w/v])	2.5		PEG 8000 (50%[w/v])	1.8
	SapI	1.0		SapI	1.0		SapI	1.0
	Total Volume	14.6		Total Volume	24.6		Total Volume	17.9

Table 5.4: Three assembly reactions were set up, using the part-linkers generated previously (table 5.3), to yield three "nested" parts A, B and C. Each assembly reaction contained a backbone plasmid part-linker encoding an antibiotic resistance gene and an origin of replication for selection and amplification in *E. coli*.

The assembled plasmids A, B and C were transformed into *E. coli* electrocompetent TOP10 (One Shot TOP10 Electrocomp *E. coli*, ThermoFisher Scientific, C404050) by electroporation (1.7 kV, 25 µF) and transformed cells were spread plated onto LB agar containing 100 µg/mL ampicillin (ThermoFisher Scientific) (A and B) or 50 µg/mL nourseothricin sulphate (Carbosynth) (C) and incubated at 37°C for approximately 16 hours.
Transformation colonies, containing the assembled vector plasmids, were selected by colony PCR followed by DNA agarose electrophoresis (data not shown), procedure in chapter 4 section 4.2.1.4. Primers and extension times for colony PCR reactions are listed in table 5.5.

Plasmid	Colony PCR Primer	Extension Time (s)
A	A_For: gccatgagatttccttctagttccgtcttggcac	33
	A_Rev: gtttgccccggaagattggggaaacttggatctg	
B	B_For: gccatggccctctttctcagtaagagactgttga	33
	B_Rev: aggcagctaatagaaaaaatatgatatgaatgaa	
C	C_For: gccatgctgcttaccaaaaggttttcaaagctgt	133
	C_Rev: agggtccgtacaaaacgcaaatgacagaaacggc	

Table 5.5: PCR primers and extension times of colony PCRs carried out on colonies of *E. coli* TOP10 electocompetant cells transformed with either plasmid A, B or C.

DNA, from two clones each identified by colony PCR to contain the assembled plasmids A, B or C, was prepared using Mini-prep (Qaigen) plasmid preparation kit, procedure in chapter 4 section 4.2.1.4. Correct assembly of plasmids was further confirmed by restriction mapping of purified plasmids A, B and C using the REs AlwnI, BsaB I and Cla I (plasmid A), Xba I and Xcm I (plasmid B) and, Mfe I, Pvu I and Sph I (plasmid C) (fig. 5.3). The RE mapping reaction was as follows, 500 ng plasmid was combined with 0.5 µL of each RE and 2 µL Cutsmart buffer and reactions were made up to 20 µL with ddH_2O. RE and Cutsmart buffer are from New England Biolabs. RE map DNA fragments were isolated and visualised by DNA 1% (w/v) agarose electrophoresis (100 V, 45 minutes).

5.2.1.5 Linker and Part Oligonucleotide phosphorylation and annealing: second round

For the second step in the "nested" assembly, oligonucleotides, in matched pairs, were annealed and phosphorylated, procedure in Chapter 4 section 4.2.1.2, to yield the LO and POs associated with parts A.1, B.1 and C.1 (LO/PO_A.1, LO/PO_B.1, LO/PO_C.1) (table 5.6).

Oligo length	LO strands	Sequence 5' - 3'	PO strands	Sequence 5' - 3'
Long	A.1	gcctagggggtgtacaatatggacttcctctttt	A.1	gtatgggtttggttgccagaaaagaggaagtccat
Short	A.1	attgtacacccccta	A.1	ctggcaaccaaaccca
Long	B.1	gccttttgtagaaatgtcttggtgtcctcgtcc	B.1	agagatggctacctgattggacgaggacaccaag
Short	B.1	acatttctacaaaaa	B.1	aatcaggtagccatc
Long	C.1	gccgatccttcagtaatgtcttgtttctttgtt	C.1	caaaatggctcaccactgcaacaaaagaaacaaga
Short	C.1	cattactgaaggatc	C.1	gcagtggtgagccatt

Table 5.6: Pairs of oligonucleotide sequences annealed and phosphorylated to produce LO and POs associated with truncated parts A.1, B.1 and C.1 required for the assembly of pPpGalHy from "nested" parts A, B and C.

5.2.1.6 Part-linker fusion reactions: A, B and C

Three separate part-linker fusion reactions were set up, according to conditions in Chapter 4 section 4.2.1.3. PO, LO and parts were combined according to table 5.7 to yield “nested” part-linkers A, B and C.

Part-linker reaction	PO	"Nested" Part	LO	Part-linker concentration (ng/μL)
1	PO_A.1	A	LO_B.1	366
2	PO_B.1	B	LO_C.1	924
3	PO_C.1	C	LO_A.1	924

Table 5.7: Three part-linker fusion reactions were set up to generate “nested” part-linkers for the assembly of pPpGalHy. Part-linker concentrations are DNA concentration of part-linkers after isolation and gel extraction.

Part-linkers were isolated by 1% (w/v) agarose gel electrophoresis (fig. 5.5), gel extracted and purified using a Zymoclean gel DNA recovery kit (Zymo Research) according to conditions in Chapter 4 section 4.2.1.1. DNA concentration of gel extract part-linkers was measured by fluorimetry (Invitrogen, Qubit) and DNA molecular weights were calculated using the part-linker fusion DNA sequences and the online “Sequence Manipulation Suite: DNA Molecular Weight” software.

5.2.1.7 inABLE assembly of pPpGalHy

The inABLE DNA assembly reaction was carried out by combining the part-linker fusion products in the proportions outlined in table 5.8 and incubated at 37 °C for 30 minutes.

Assembly Reaction	Reagants	Volume (μL)
pPpGalHy	PO_A.1-A-LO_B.1	1.0
	PO_B.1-B-LO_C.1	5.0
	PO_C.1-C-LO_A.1	5.0
	Buffer 2 (10x)	1.5
	PEG 8000 (50%[w/v])	1.5
	EarI	1.0
	Total Volume	15.0

Table 5.8: An assembly reaction was set up, using the “nested” part-linkers generated previously (table 5.7), to yield the plasmid pPpGalHy.

The assembled plasmid pPpGalHy was transformed into *E. coli* electrocompetent TOP10 (1.7 kV, 25 µF) and transformed cells were spread plated onto LB agar containing 50 µg/mL nourseothricin and incubated at 37°C for approximately 16 hours.
Transformation colonies, containing the assembled vector plasmid, were selected by colony PCR followed by DNA 1% (w/v) agarose electrophoresis (100 V, 45 minutes) (data not shown), colony PCR procedure in chapter 4 section 4.2.1.4. Primers and extension times for colony PCR reactions are listed in table 5.9.

Plasmid	Colony PCR Primer	Extension Time (s)
pPpGalHy	pPpGalHy_For: gcctaggggtgtacaatatggacttcctctttt	330
	pPpGalHy_Rev: caaaatggctcaccactgcaacaaaagaaacaaga	

Table 5.9: PCR primers and extension times for the colony PCR carried out on colonies of *E. coli* TOP10 electocompetant cells transformed with pPpGalHy.

DNA, from two clones identified by colony PCR to contain the assembled plasmid pPpGalHy, was purified using Mini-prep (Qaigen) plasmid preparation kit, procedure in chapter 4 section 4.2.1.4. Correct assembly of the plasmid was further confirmed by restriction mapping using the REs EcoR V and Xba I in Cutsmart buffer (New England Biolabs) (fig 5.6), procedure in section 5.2.1.4. RE map DNA fragments were isolated and visualised by DNA 1% (w/v) agarose electrophoresis (100 V, 45 minutes).

5.2.2 pPpGalHy transformation, selection and shake-flask culture of P. pastoris *KM71H* CFH

5.2.2.1 pPpGalHy transformation of P. pastoris *KM71H* CFH

Purified pPpGalHy was sequenced using Edinburgh genomics gene sequencing service. After analysis of sequencing data and confirmation of correct assembly, 40 µg of purified pPpGalHy (40 ng/µL) was linearised by digestion with the RE MfeI-HF (7.5 µL, 20kU/mL), cutsmart buffer (75 µL) and made up to 1.5 mL with ddH_2O and incubated at 37 °C for approximately 16 hours. Linearisation of pPpGalHy was assessed by separation of an aliquot by 1% (w/v) agarose gel electrophoresis (100 V, 45 minutes). Linear pPpGalHy was concentrated by isopropanol precipitation, procedure in Chapter 3 section 3.2.1.2, and 4 µg, in a 10 µL volume of H_2O, was used to transform electocompetant *P. pastoris* strain KM71H *CFH*, competency procedure in Chapter 3 section 3.2.1.3. Transformed cells were

plated on YPD agar containing nourseothricin (100 µg/mL) and incubated at 30 °C for 6 days.

5.2.2.2 Selection of transformed P. pastoris *by colony PCR*

Approximately 150 colonies grew on the nutrient agar plates. 46 of these were analysed by two different colony PCRs to detect for the present of the pPpGalHy plasmid. Colonies were culture and crude genomic DNA prepared and subjected to PCR using OneTaq DNA polymerase (New England Biolabs) using the primers and extension times in table 5.10 and the procedure in Chapter 3 section 3.2.2.

Plasmid	Colony PCR Primer	Extension Time (s)
pPpGalHy	ADH1p_For: ctggcaaccaaaccca	120
	ADH1p_Rev: ctggtccaatcaatccaagtgccaagacggaact	
	NrsR _For: gaggaagcggccgcgctcttcgttccagattttctcggactc	120
	NrsR _Rev: ttagatggatccgctcttcgggcgacactggatggcgg	

Table 5.10: PCR primers and extension times for the colony PCR carried out on colonies of *P. pastoris* KM71H *CFH* cells transformed with pPpGalHy.

DNA products of PCR reactions were visualised by 1% (w/v) agarose gel electrophoresis (100 V, 45 minutes) (fig. 5.7 and 5.8).

5.2.2.3 Shake-flask culture and expression of FH in P. pastoris *transformed with pPpGalHy*

Nine colonies of KM71H *CFH* transformed with pPpGalHy were randomly selected for shake flask fermentation and expression of FH. The nine colonies, and a tenth taken from a glycerol stock of parental strain KM71H *CFH* (not transformed with pPpGalHy), were used to inoculate 5 mL BMGY media containing 100 µg/mL nourseothricin sulphate in 50 mL conical tubes (Falcon). These starter cultures were left to incubate at 30 °C, 250 rpm shaking for approximately 16 hours. OD_{600} of the starter cultures was monitored and when it reached 0.8 – 1.5 25 mL BMGY media in 100 mL baffled shake flasks was inoculated with 50 µL of the starter cultures and left to incubate at 30 °C, 250 rpm for approximately 60 hours. Cell cultures were transferred to 50 mL conical tubes (falcon) and centrifuged at 4 °C, 1600 rpm for 20 minutes. The supernatant was removed and discarded and the pellet was resuspended in 5 mL induction media (BMMY). Cultures were transferred to fresh 50

mL baffled flask and incubated at 16 °C, 250 rpm for approximately 96 hours. Cultures were fed with 50 µL pure methanol at 24 hours, 48 hours, 72 hours and 80 hours post induction. Cultures were transferred to 50 mL conical tubes and centrifuged at 4 °C, 1600 rpm for 20 minutes. Supernatant was removed and transferred to fresh 50 mL conical tubes on ice and 1% (v/v) EDTA (250 mM stock) and 0.5% (v/v) PMSF (100 mM stock) protease inhibitors were added. 1 mL aliquot of each culture supernatant was removed for SDS-PAGE and western blot analysis, the remainder was stored frozen at -20 °C. 10 µL of each culture supernatant was added to 40 µL SDS loading buffer and 10 µL of this was loaded onto three 4-12% bis-tris polyacrylamide gels (Bolt Plus, 15-well, ThermoFisher Scientific) along with FH protein standards and the gel was run (200 V, 35 minutes). One gel was stained with Coomasie Instant*Blue* stain (Expedeon) the other two gels were subjected to lectin-based western blot procedure in chapter 2 section 2.2.2.1. One western blot was exposed to DAB substrate (metal enhanced, ThermoFisher Scientific) for 1 minute 30 seconds, the other to substrate for 8 minutes.

5.3 Results

5.3.1 Promoter-gene pairing

There are 3 genes involved in the GlycoSwitch hybrid glycan biosynthesis pathway. The specific activities of these enzymes were compared to identify if any have a significantly low specific activity (Table 5.11).

Enzyme	Specific Activity (U/mg)	Molecular Weight (kDa)	Specific Activity per Mole (U/nmol)
Mannosidase I (ManI)	6.70	56.2	0.38
N-acetylglucosamine transferase I (GnTI)	0.48	50.9	0.03
Galactose transferase (GalT)	4.7 - 8.4	44.1	0.21 - 0.37

Table 5.11: Specific activities of enzymes used in humanising the glycosylation biosynthetic pathway in *P. pastoris*. U (µmol/min) is the enzyme activity as measured by the amount of substrate used in the enzymatic reaction per minute.

Using the MW of the enzymes and the specific activities per unit mass enzyme, the molar specific activity of each enzyme was calculated (last row table 5.2). Consideration of differences in MW allowed for a fairer comparison of specific activity. GnTI has the lowest enzymatic activity (Chen *et al*., 2008) whilst ManI (Maras *et al*., 2000) and GalT (Krezdorn *et al*., 1993 and, Malissard *et al*., 1996) have similar activities.

The *P. pastoris* TEFp is a high-level expression, constitutive promoter (Ahn *et al*., 2007) that has routinely been used as the standard for comparison of *P. pastoris* promoter controlled gene expression levels (promoter strength or activity). Literature data for the promoter strength of 3 *P. pastoris* promoters were compared (Table 5.12).

Promoter	Activity relative to TEF1 promoter
GAPp	100%
ADH1p	20 - 27%
ILV5p	15%

Table 5.12: Strength of the influence of gene expression of three *P. pastoris* promoters. Promoter strength (activity) is measured as a percentage relative to the activity of the strong constitutive *P. pastoris* promoter TEFp.

All promoters in table 5.12 are constitutive promoters. The GAPp has the highest activity with an activity equal to the TEFp promoter (Partow *et al*., 2010 and, Stadlmyer *et al*., 2010), the alcohol dehydrogenase 1 promoter (ADH1p) has an activity approximately one quarter of TEFp (Partow *et al*., 2010) whilst the acetohydroxy acid isomeroreductase promoter (ILV5p) had the lowest relative activity (Vogl and Glieder, 2013).

The low activity of GnTI compared to the other enzymes (table 5.11) could cause a bottleneck in the glycan biosynthetic pathway. Therefore, to overcome this potential bottleneck, the GnTI encoding gene was placed under the control of GAPp, the strongest available promoter. The remaining promoters were randomly assigned to the other genes.

5.3.2 "Nested" inABLE assembly: incorporating the hybrid glycan biosynthetic pathway onto a single plasmid

The vector plasmid (pPpGalHy) (fig. 5.2), encoding the hybrid glycan biosynthetic pathway, contains a total of 9 DNA sequences, or parts. However, due to the inefficiency of assembling plasmids containing greater than 6-7 parts, a "nested" approach was employed to construct this plasmid.

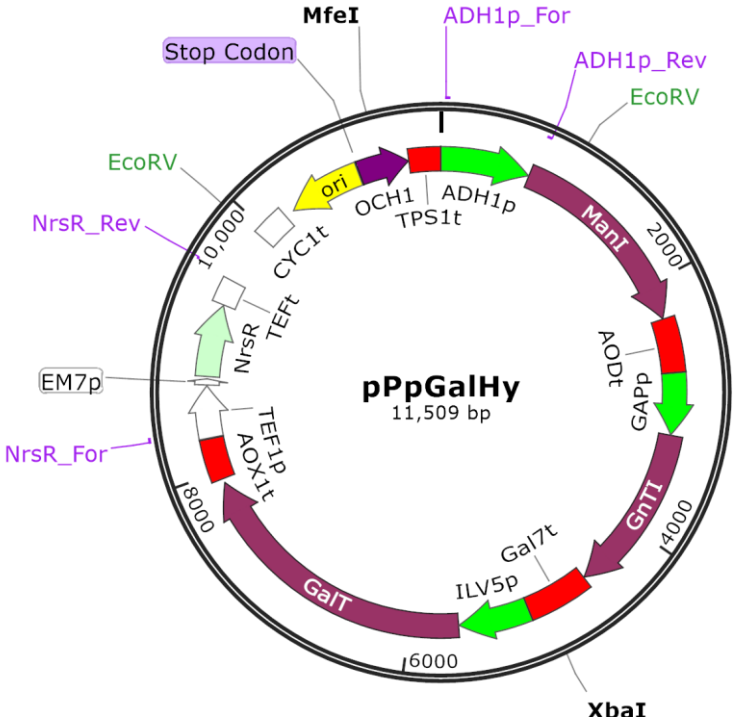


Figure 5.2: The shuttle vector plasmid, pPpGalHy, containing genes for the biosynthesis of a hybrid-type glycan in *P. pastoris* and the truncated *och1* sequence (purple) for simultaneous integration of the plasmid in to the *P. pastoris* genome at the *OCH1* gene locus. Marked on the *och1* sequence is the RE Mfe I recognition site for linearisation of the plasmid and the in frame stop codon is marked at the 5' end of the *och1* sequence. Also encoded in the plasmid is a gene encoding nourseothricin resistance (NrsR, light green arrow) for selection in *E. coli* and *P. pastoris* and, an origin of replication for plasmid amplification in *E. coli* (yellow arrow). Green arrows are promoters, maroon arrows are gene sequences, red rectangles are terminators and clear arrows and rectangles are *E. coli* regulatory elements. PCR primers (NrsR_For and _Rev and, ADH1p_ For and _Rev) are in purple writing and positions of annealing to plasmid are marked by purple lines. Positions of recognition sites are marked for RE EcoR V and Xba I, these RE were used in restriction mapping of pPpGalHy.

In the first round of the "nested" approach the 9 DNA sequences were divided into groups of three (Table 5.13). Each group of three was assembled into a single contiguous part, yielding three "nested" parts.

"Nested" Part	Part	Sequence contained in part
A	A.1	*S.cerevisiae* Alcohol dehydrogenase I promoter (ADH1p)
	A.2	C-terminal HDEL- tagged α(1,2)-mannosidase from *Trichoderma reesei*
	A.3	*S.cerevisiae* Alcohol oxidase terminator (AODt)
	A.4	Backbone vector containing ampicillin resistance selectable marker and origin of replication for selection and amplification in *E.coli*
B	B.1	*P.pastoris* Glyceraldehyde-3-phosphate dehydrogenase promoter (GAPp)
	B.2	N-acetylglucosamine transferase I (GnTI) from *H.sapiens* with N-terminal *S.cerevisiae* Kre2p membrane localisation domain
	B.3	Galactose-1-phosphate uridylyltransferase terminator (GAL7t)
	B.4	Backbone vector containing ampicillin resistance selectable marker and origin of replication for selection and amplification in *E.coli*
C	C.1	Acetohydroxy acid isomeroreductase promoter (ILV5p)
	C.2	DNA sequence containing *S.pombe* Galactose epimerase/*H.sapiens* Galactose transferase fusion construct (GalT) in frame with Alcohol oxidase I terminator (AOXIt) as well as a backbone shuttle vector containing a gene encoding nourseothricin resistance for selection in *P.pastoris* and *E.coli* and, an origin of replication for amplification in *E.coli*
	C.3	Truncated *och1* sequence for targeted intregration into *P.pastoris* genome
	C.4	*S.cerevisiae* Trehalose-6-phosphate synthetase I terminator (TPS1t)

Table 5.13: The plasmid pPpGalHy was made up of three "nested" parts (A, B and C). Each "nested" part was made up of four parts (A.1-4, B.1-4 and C.1-4). In "nested" parts A and B the fourth part (A.4 and B.4) is a vector plasmid backbone for selection and amplification in *E. coli*. "Nested" part C contains part C.2 which encodes a shuttle backbone vector plasmid for selection in *P. pastoris* as well as *E. coli*, and an origin of replication of amplification in *E. coli*.

Each part was either synthesised (*ManI* (A.2) and truncated *och1* (C.3)), PCR amplified (AODt (A.3), GnTI (B.2), ILV5p (C.1) and GalT containing backbone plasmid (C.2)) or, existed as a pre-generated part cloned into a plasmid for selection and amplification in *E. coli*. In all cases the parts were produced such that they were flanked by the same type II S RE recognition site, one site was in the reverse orientation (fig. 1.17 Chapter 1 section 1.6.2). The parts were then digested with the type II S RE which liberated the truncated part sequence. In independent, separate part-linker fusion reactions each truncated part was ligated to the truncated parts associated PO and the LO of the next part in the sequence to generate part-linkers (fig. 5.3). For example, the truncated part A.1 was ligated to PO_A.1 at the 5' end and LO_A.2 at the 3' end; truncated part A.2 was ligated to PO_A.2 at the 5' end and LO_A.3 at the 3' end; etc. Once the truncated part is fused to an LO and a PO it is referred to as a part-linker.

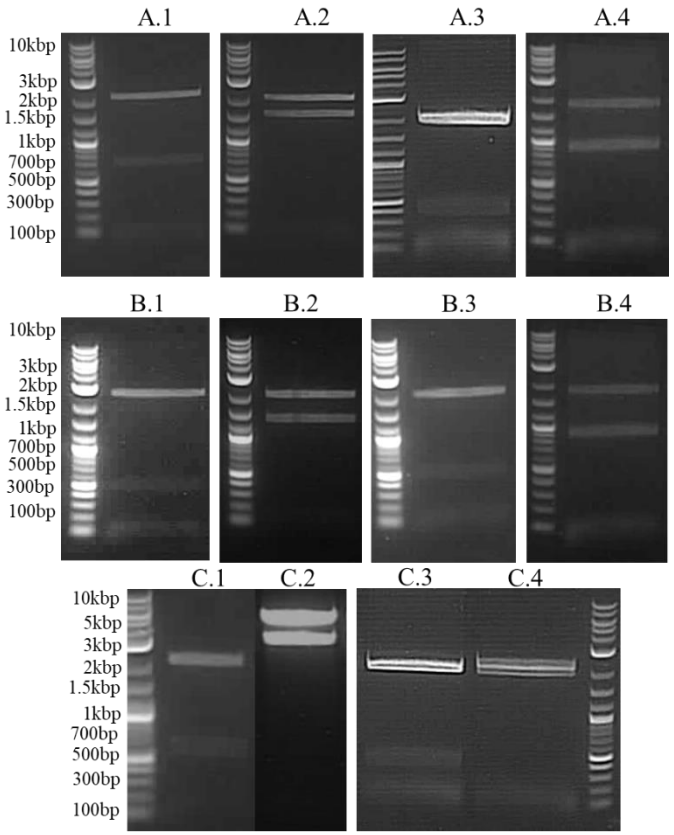


Figure 5.3: Products of part-linker fusion reactions. The expected band positions of part-linker fusion products are as follows: A.1 (ADH1p) ~700 bp, A.2 (*ManI*) ~1.6 kbp, A.3 (AODt) ~400 bp, A.4 (plasmid backbone) ~1.8 kbp, B.1 (GAPp) ~500 bp, B.2 (*GnTI*) ~1.3kbp, B.3 (GAL7t) ~500 bp, B.4 (plasmid backbone) ~1.8 kbp, C.1 (ILV5p) ~550 bp, C.2 (*GalT* and shuttle vector plasmid backbone) ~5.2 kbp, C.3 (*och1*) ~500 bp and C.4 (TPS1t) ~250 bp.

Agarose electrophoresis of part-linkers was carried out to separate part-linkers from other contaminating DNA sequences (fig. 5.3) e.g. backbone plasmids or free LO and PO sequences. Part-linkers were then gel extracted using Zymoclean DNA recovery kit.
All part-linkers are visible as bands at expected base pair positions except for C.4. However, gel extraction of the region between 200 and 400 bp yielded a measurable DNA concentration. It was therefore assumed that this contained part-linker C.4 and was used in the assembly of "nested" part C. DNA sequencing analysis of pPpGalHy (data not shown) later confirmed the presence of C.1 in the plasmid.

Part-linkers A.1-4, B.1-4 and C.1-4 were incubated, separately in a one-pot reaction to produce the plasmids A, B and C (see plasmid maps in figure 5.5). In this procedure, the 16 bp complimentary sequences of associated LO and POs anneal by base pairing. This joined the part-linkers into a single, contiguous sequence. For example, in the part-linker fusion process for A the part-linkers LO_A.4-TP_A.1-PO_A.1 and LO_A.1-TP_A.2-PO_A.2 were

formed. In the assembly process PO_A.1 and LO_A.1 can join to each other by complimentary base pairing between the 16 bp 5' overhand of LO_A.1 and 16 bp 3' overhang on PO_A.1. The result is that ADH1p (A.1) is joined, in frame, with *ManI* (A.2).

Plasmids A, B and C (see plasmid maps in figure 5.5) were used to transform *E. coli* TOP10 electrocompetent cells by electroporation. Transformed *E. coli* was selected for on nutrient agar containing the appropriate selection antibiotic (for plasmid A and B ampicillin and, for plasmid B nourseothricin). Colonies containing plasmid A, B and C were selected for by colony PCR (data not shown) and 2 colonies containing each plasmid were cultured in shake flasks. Plasmids were extracted and purified using a plasmid extraction procedure (Qaigen, plasmid mini-prep kit) and correct assembly of plasmids A, B and C was confirmed by restriction mapping (fig. 5.4).

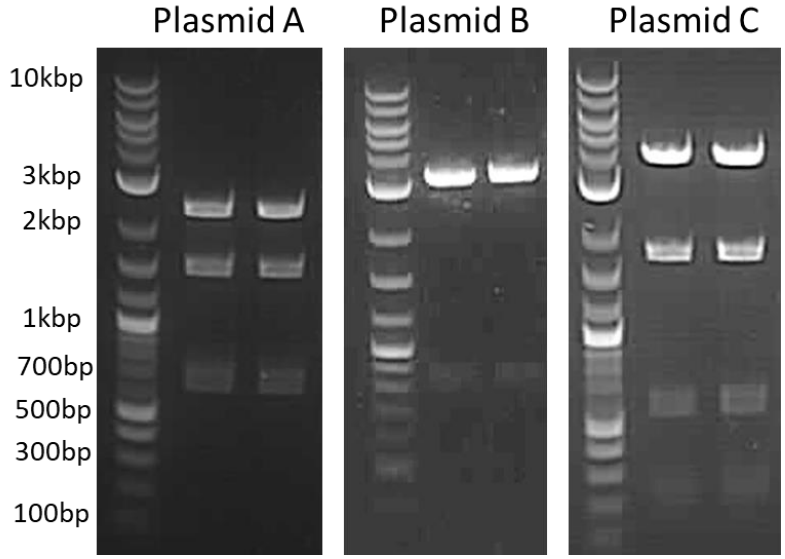


Figure 5.4: Restriction mapping of plasmids isolated from 2 colonies of *E. coli* transformed with either plasmid A, B or C. Plasmid A (4550 bp) was digested with the REs AlwN I, BsaB I and Cla I to yield three DNA fragments of 2379 bp, 1524 bp and 647 bp. Plasmid B (4156 bp) was digested with the REs Xba I and Cla I to yield DNA fragments of 3291 bp and 865 bp. Plasmid C (6467 bp) was digested with the REs Mfe I, Pvu I and Sph I to yield DNA fragments of 3926 bp, 1710 bp, 575 bp and 256 bp.

Purified plasmids A, B and C were subjected to restriction mapping (fig. 5.4) to confirm that all part-linkers were assembled into each plasmid in the correct order and orientation. The REs and associated positions of recognition sites are labelled on plasmids A, B and C (fig. 5.5). The restriction maps show that all plasmids were assembled correctly. For example, digestion of plasmid A with AlwN I, BsaB I and Cla I was expected to yield three DNA fragments 2379 bp, 1524 bp and 647 bp in length. The agarose gel electrophoresis image of the restriction map of plasmid A, from both *E. coli* colonies transformed with this plasmid,

show three bands of lengths approximately 2500 bp, 1500 bp and 650 bp confirming the correct assembly of plasmid A.

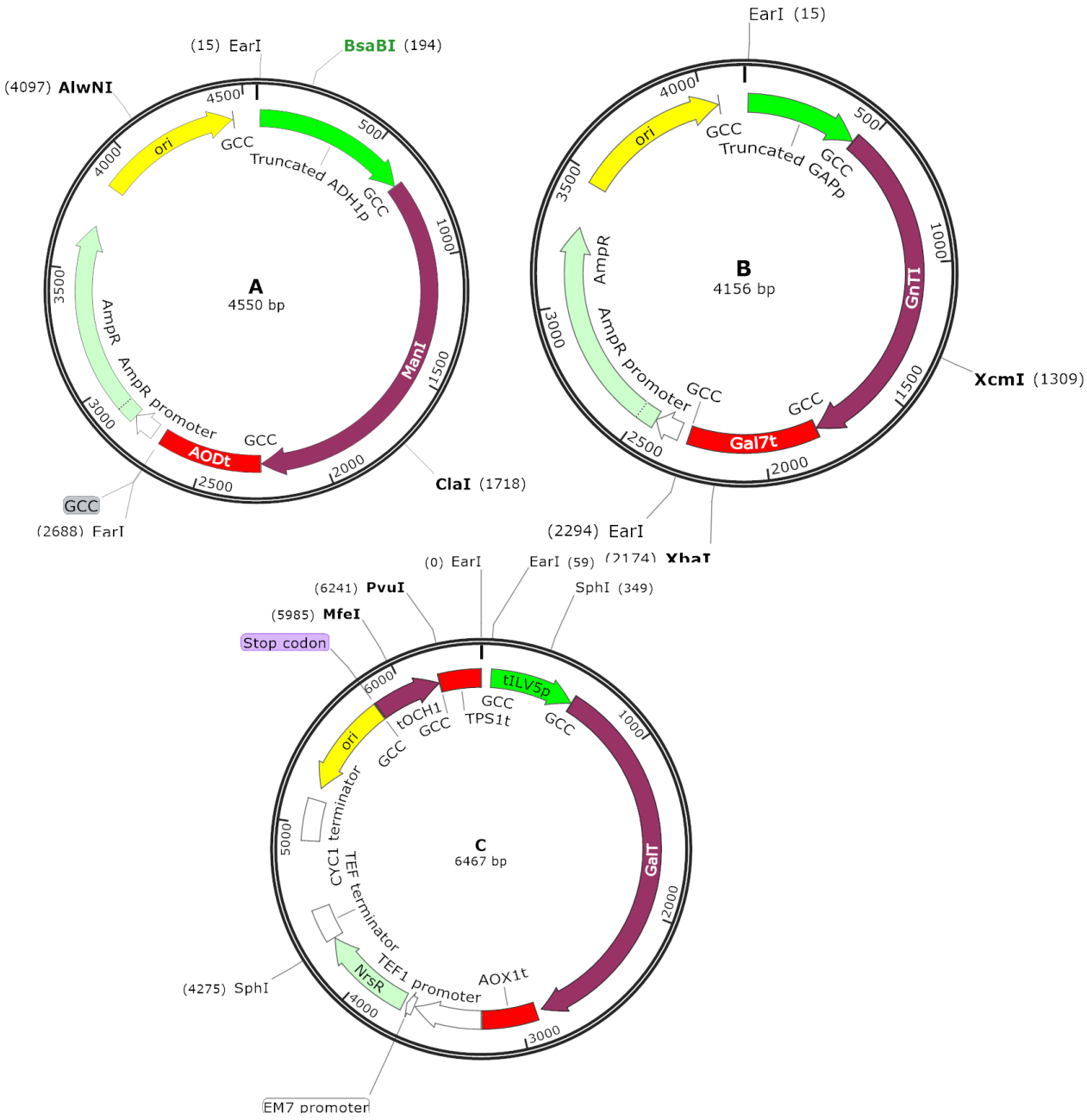


Figure 5.5: Plasmids A, B and C containing "nested" parts A, B and C. The GCC triplet sequence marks the intersection between each part. The recognition site for the type II S RE Ear I is marker on each map. Digestion of each plasmid with Ear I liberates the "nested" parts to allow for production of part-linkers of A, B and C for the second round of inABLE assembly. Also marked on the plasmid maps are the RE sites use for confirmation of correct plasmid assembly by restriction mapping (fig. 5.4). The exact base pair position of the RE cut site is labelled in brackets next to the RE.

The "nested" parts A, B and C are contained within plasmids A, B and C. The "nested" parts were liberated from the plasmids by RE digest with the type II S RE Ear I. The RE sites flank the "nested" part in opposite orientation (or mirror images) and create an overhanging

triplet sequence for the attachment of LOs and POs in part-linker fusion reactions (fig. 1.17 Chapter 1 section 1.6.2). The Ear I digested parts A, B and C were isolated by agarose gel electrophoresis (data not shown) and gel extracted and purified (zymoclean gel extraction kit). Part-linker fusion reactions were carried out on "nested" parts A, B and C (fig. 5.6). In contrast to the part-linker fusion reactions for parts A.1-4, B.1-4 and C.1-4, in which "nested" LOs and POs were used for parts A.1 and A.4, B1. and B.4 and, C.1 and C.4, for this set of part-linker fusion reactions the true LOs and POs for parts A.1, B.1 and C.1 were used for the part-linker fusion reactions of "nested" parts A, B and C.

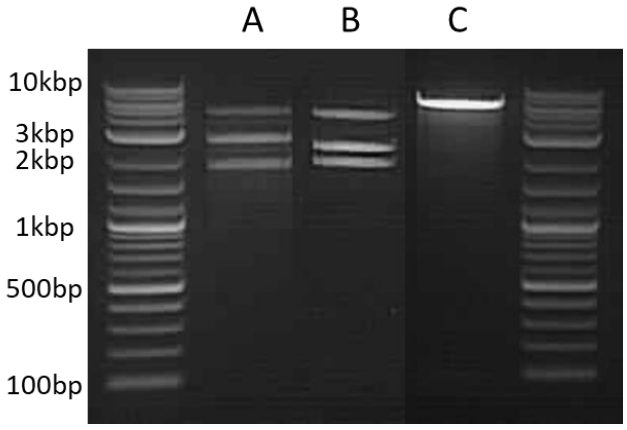


Figure 5.6: Products of part-linker fusion reactions carried out on "nested" parts A, B and C, previously digested with Ear I type II S RE. The expected band for A is ~2700 bp, for B is ~2300 bp and for C is ~6400 bp.

After part-linker fusion reactions were carried out, "nested" part linkers A, B and C were isolated by agarose gel electrophoresis (fig. 5.6), gel extracted and purified (zymoclean gel extraction kit). Figure 5.6 shows that there are bands at the expected sizes for part-linkers A, B and C. The lowest band in lanes A and B is the digest plasmid backbone (band at approximately 2 kbp). Considering the lengths of plasmid A and B are 4550 bp and 4156 bp, respectively (fig. 5.5), then the highest band at approximately 5 kbp is likely to be undigested plasmid A and B. Digestion of plasmid C with Ear I (sites marked on plasmid map in figure 5.5) produces two DNA sequences, a long sequence comprising *och1*, *GalT* as well as a nourseothricin resistance selectable marker gene and origin of replication in *E. coli*, the second sequence is very short (the gap between the two Ear I sites on the plasmid map) and does not show on the agarose gel images (fig. 5.6 lane C).

Purified "nested" part-linkers were place together in a one-pot assembly reaction. Associated LOs and POs joined neighbouring parts together by base pair annealing between the 16 bp complimentary overhand sequences. The assembled plasmid, pPpGalHy, was

transformed into TOP10 electrocompetent *E. coli* by electroporation. Transformed colonies were selected for on agar containing the antibiotic nourseothricin and colony PCR was carried out to identify colonies that contained pPpGalHy (data not shown). 2 colonies, identified to contain pPpGalHy, were cultured in shake flasks and the pPpGalHy plasmid was isolated and purified using a plasmid extraction procedure (Qaigen, plasmid mini-prep kit). Correct assembly of purified plasmids was confirmed by restriction mapping (fig. 5.7).

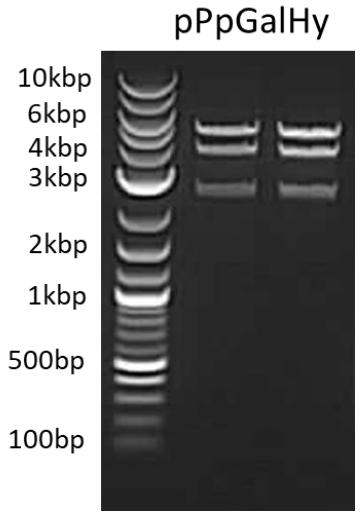


Figure 5.7: Restriction mapping of plasmids isolated from 2 colonies of *E. coli* transformed with pPpGalHy. Purified plasmid pPpGalHy (11509 bp) was digested with the REs EcoR V and Xba I. There were two RE sites for EcoR V and one for XbaI in pPpGalHy. Digestion with these RE yielded three DNA fragments at positions approximately 3000 bp, 4000 bp and 5000 bp.

Restriction mapping, of the two purified plasmids from two colonies transformed with pPpGalHy, was expected to yield three DNA fragments of lengths 5009 bp, 3948 bp and 2552 bp. Digestion of pPpGalHy with the REs EcoR V and Xba I (RE recognition sites are marked the plasmid map in figure 5.2) did yield three DNA fragments of approximate lengths 5000 bp, 4000 bp and 3000 bp which fits with the expected DNA fragment lengths. Restriction mapping confirmed the correct assembly of pPpGalHy, this was further confirmed by DNA sequencing analysis (data not shown).

5.3.3 Transformation and expression of pPpGalHy in *P. pastoris*

The purified pPpGalHy plasmid was linearised by digestion with the RE Mfe I. Approximately 10 µg linear pPpGalHy was used to transform *P. pastoris* strain KM71H *CFH*. This strain has not been glycoengineered, and therefore has a wild-type glycosylation profile and an intact, full length *OCH1* gene in the genome. Additionally, KM71H *CFH* had previously been transformed with a plasmid encoding human FH (*CFH*).
After transformation clones containing pPpGalHy were selected for on yeast nutrient agar containing the antibiotic nourseothricin. Colony PCR was carried out using PCR primers that bind either side of the ADH1p and amplify across this promoter (ADH1p_For and ADH1p_Rev, in figure 6.2). PCR was expected to yield a PCR product 722 bp in length (fig. 5.8).

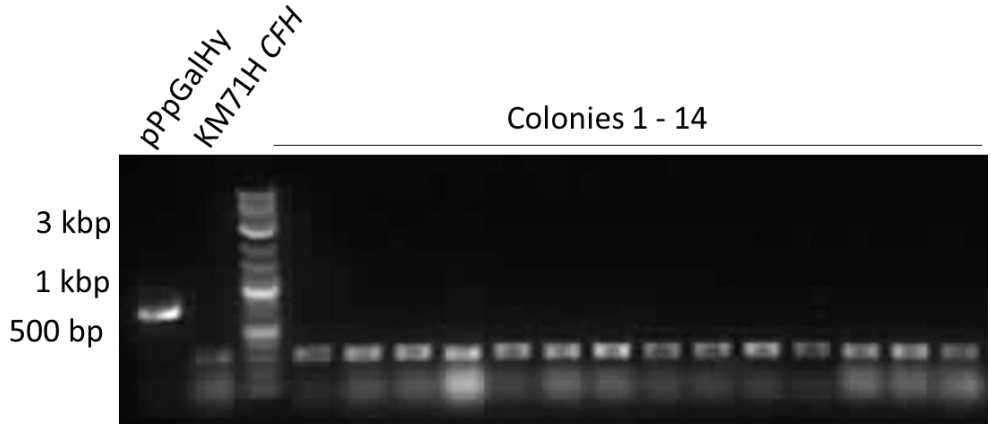


Figure 5.8: Colony PCR of *P. pastoris* KM71H *CFH* colonies transformed with pPpGalHy and selected for on yeast nutrient agar containing the antibiotic nourseothricin. PCR primers ADH1p_For and _Rev were designed to amplify across the ADH1p in pPpGalHy to yield a 722 bp product. The lane labelled pPpGalHy is the positive control and contains the product from PCR amplification of the purified plasmid pPpGalHy, KM71H *CFH* is the negative control as it was not transformed with pPpGalHy and does not contain the *S. cerevisiae* ADH1 promoter. To the right of the DNA ladder are colonies 1 – 14 of KM71H *CFH* transformed with pPpGalHy.

None of the 14 colonies analysed by colony PCR gave a PCR product of the expected length (fig. 5.8). The lack of PCR product cannot be explained by a failure in PCR as the positive control, purified pPpGalHy (first lane in figure 6.7), yielded a product of the expected length. There is a band in all 14 colonies at approximately 300bp and at 100 bp. The 100 bp band is likely due to dimerisation of PCR primers (this has been observed in other, unrelated, PCR reactions). The 300 bp product is also present in the negative control, KM71H *CFH* without pPpGalHy. Therefore, this band is due to off target binding of the PCR primers to a site within the genome of *P. pastoris* strain KM71H.

However, the transformed cells were plated onto nutrient agar containing the antibiotic nourseothricin. If colonies were to grow on this media they must contain a gene conferring resistance to nourseothricin. Therefore, a second round of colony PCR (fig. 5.9) was performed on the 14 selected colonies. In this PCR, primers were designed (NrsR_For and _Rev) to bind the pPpGalHy plasmid at sites that flanked, and would amplify across, the gene conferring nourseothricin resistance (*NrsR*) (see figure 5.2 for sites of primer binding) to yield a 1249 bp product. The aim here was to determine if *P. pastoris* transformed with pPpGalHy contains the *NrsR* gene.

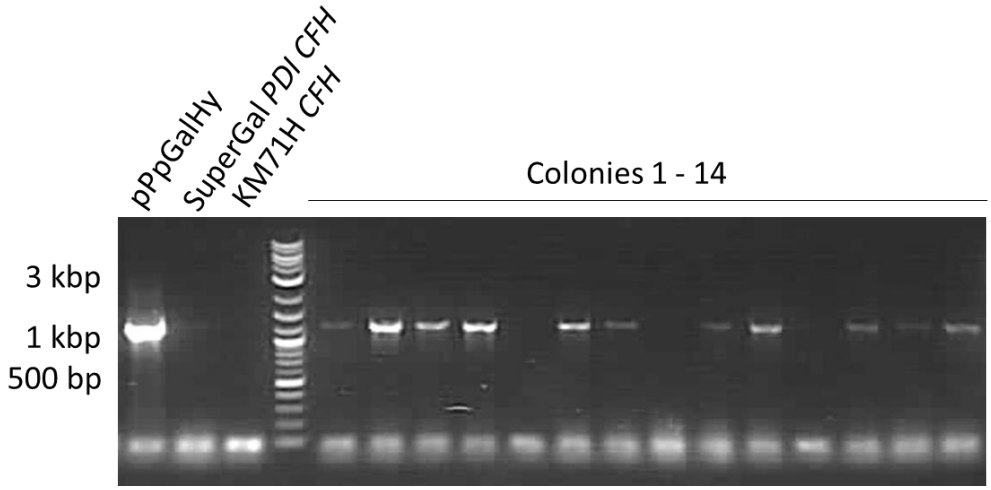


Figure 5.9: Colony PCR of *P. pastoris* KM71H *CFH* colonies transformed with pPpGalHy and selected for on yeast nutrient agar containing the antibiotic nourseothricin. PCR primers NrsR_For and _Rev were designed to amplify across the *NrsR* gene in the pPpGalHy plasmid to yield a 1249 bp product. The lane labelled pPpGalHy is a positive control and contains the product from PCR amplification of the purified plasmid pPpGalHy, SuperGal *PDI CFH* is also a positive control as it had been transformed with the plasmid pGlycoSwitch-GalT, a plasmid that contains the *NrsR* gene. KM71H *CFH* is the negative control as it was not transformed with pPpGalHy and does not contain the *NrsR* gene. To the right of the DNA ladder are colonies 1 – 14 of KM71H *CFH* transformed with pPpGalHy.

A band is present at ~1200 bp in the positive control lanes pPpGalHy and, although faint, SuperGal *CFH* (which was transformed with the plasmid pGlycoSwitch-Galt and contains the *NrsR* gene) and there is no band at ~1200 bp in the negative control lane KM71H *CFH*. To the right of the DNA ladder, colony PCR using NrsR_For and _Rev yields a band at ~1200 bp for 11 out of 14 of the colonies analysed. This shows that these colonies contain the *NrsR* gene. The colony PCRs shown in figure 5.8 and 5.9 are samples of a much larger colony screening procedure. Approximately 150 colonies were present on nutrient agar plates after transformation with pPpGalHy, 46 of these were screened by colony PCR. The data in figures 5.8 and 5.9 reflects a trend that was seen across all 46 colonies analysed.

The results in figure 5.8 and 5.9 from both colony PCRs show that colonies of *P. pastoris* KM71H transformed with pPpGalHy do not contain the ADH1 promoter but 11 of these colonies do contain the *NrsR* gene. Taken together, this would suggest that the cells were successfully transformed with pPpGalHy but, however, part of the pPpGalHy plasmid containing the ADH1p promoter was lost at some point between transformation and selection for PCR analysis. As selection occurred in the presence of nourseothricin, complete loss of pPpGalHy did not occur as it was advantageous to retain part of the plasmid that conferred resistance to the antibiotic.

However, there was still the possibility that other parts of the pPpGalHy plasmid were present in the transformed colonies. Therefore, to test to see if the glycan biosynthesis genes on pPpGalHy were present and active in *P. pastoris* KM71H *CFH* transformed with pPpGalHy, nine colonies were cultured in shake flasks, expression of FH was induced and the supernatants of these cultures was analysed for the presence of FH by SDS-PAGE and the presence of LacNAc terminal hybrid glycans on FH by ECL western blot (fig. 5.10).

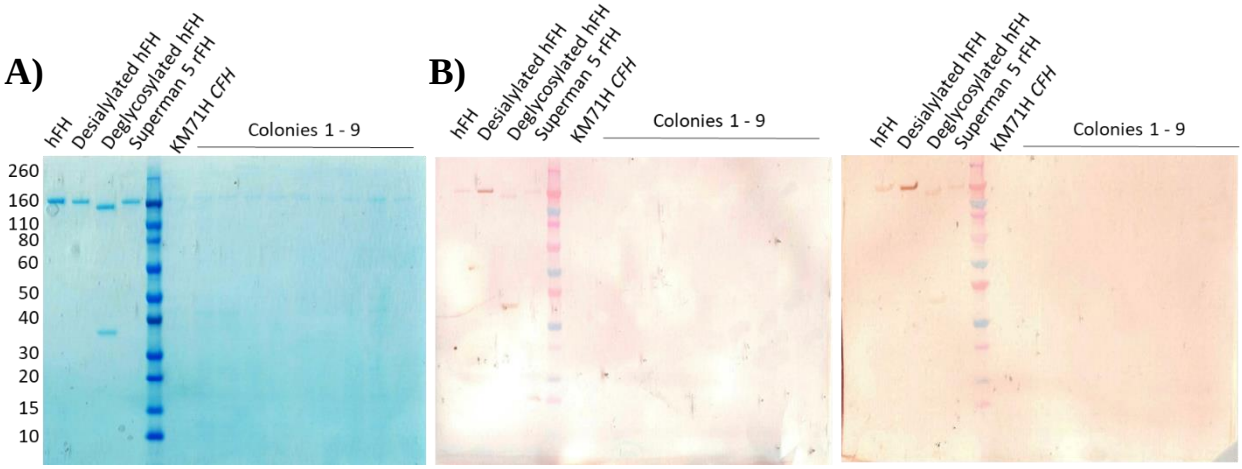


Figure 5.10: SDS-PAGE (A) and ECL lectin western blots (B) of FH expression in nine colonies of *P. pastoris* strain KM71H *CFH* transformed with pPpGalHy. The ECL lectin western blots are duplicates, the western blot on the right was incubated in metal enhanced DAB substrate for an extended time period to enhance low level signal – i.e. to detect any low abundance galactose in the rFH samples. To the left of the molecular weight ladder are control samples. hFH is FH purified from human blood plasma, desialylated hFH is hFH desialylated with α(2-3,6,8)-neuraminidase (desialylated hFH is the positive control for the ECL western blots), deglycosylated hFH is hFH deglycosylated with PNGase F. SuperMan 5 rFH is purified recombinant FH expressed in *P. pastoris* strain SuperMan 5 *CFH*. KM71H *CFH* is the parental strain and had not been transformed with pPpGalHy. This strain was cultured in the same way as colonies 1-9.

SDS-PAGE analysis (fig. 5.10 A) shows that all 10 colonies expressed rFH. However, there are no bands indicating the presence of galactose according to the ECL lectin western blot images (fig. 5.10 B). The faint bands on SDS-PAGE indicates a low concentration of FH in the samples loaded onto the gels, therefore a second, duplicate ECL western blot was carried

out with an extended incubation time with the detection substrate (1 min 30s incubation time for left image in fig. 6.9 B compared to 8 min incubation time for the right image). This was done to try and enhance any low-level signal from any galactose that may be present on rFH. However, galactose was still not detectable on rFH expressed in any of the 10 colonies analysed.

5.4 Discussion

The aim of this study was to consolidate the hybrid glycan biosynthetic pathway onto a single plasmid in order to make available antibiotic selectable markers for *in vivo* glycan sialylation. Additional aims were to reduce the potential for genetic instability by replacing the repeatedly used regulatory elements for unique sequences and, in doing so, to couple promoter activity to the specific activity of the enzyme gene product. However, although the genes could be consolidated onto a single plasmid, pPpGalHy (fig. 5.2), this plasmid was not retained within *P. pastoris*. The detection of the *NrsR* gene in *P. pastoris* KM71H *CFH* transformed with pPpGalHy, but absence of this gene in KM71H *CFH* not transformed with the plasmid, suggests that (at least part of) the plasmid had integrated into the genome of this strain. However, ADH1p was not detectable by colony PCR in this strain. This suggests that at some point before or after genome integration recombination occurred resulting in the loss of part of the plasmid that contained ADH1p but, retention in the genome of part of the plasmid bearing the *NrsR* gene.

Efforts were made in this study to replace the regulatory elements used in the GlycoSwitch technology, GAP promoter and AOX1 terminator, with alternative regulatory elements in an attempt to reduce the number of homologous sequences in close proximity – potentially reducing genetic instability caused by homologous recombination. Homologous recombination is a process that has been exploited in *P. pastoris* to generate strains with plasmids integrated into the genome (Cregg *et al*., 1985). However, non-homologous end joining (NHEJ) is also known to occur in *P. pastoris* and deletion of the *P. pastoris KU70* homologue, the protein product of which facilitates NHEJ, was shown to improve plasmid-genome targeting efficiency in *P. pastoris* (Näätsaari *et al*., 2012). There is the possibility that loss of the pPpGalHy plasmid, but retention of the *NrsR* gene, could have occurred by homologous recombination between two homologous sites within the plasmid. However, as care was taken to ensure no large homologous sites were present in pPpGalHy it

therefore seems more likely that plasmid loss was due to NHEJ. DNA sequencing of the site of insertion of pPpGalHy into the genome would allow for identification of the DNA sequence at the 5' end and 3' end of the fragment of pPpGalHy that was retained in the genome. This would allow for identification of any homologous sequence present in the plasmid which could then be removed to improve stability. Additionally, DNA sequencing may identify that plasmid loss occurred by NHEJ in which case retention of genomically integrated pPpGalHy may be achieved by using a *P. pastoris Δku70* knockout strain with reduced NHEJ activity (Näätsaari *et al*., 2012).

Since the development of the GlycoSwitch technology efforts have been made to expand the molecular toolbox of *P. pastoris*. As well as the development of novel promoters (Vogl *et al*., 2015) other efforts to improve the molecular toolbox include: investigation into promoter strengths to produce natural promoter libraries (Partow *et al*., 2010, Stadlmyer *et al*., 2010, Vogl and Glieder, 2013 and Love *et al*., 2016), as well as short synthetic promoter libraries for fine-tuned gene expression (Hartner *et al*., 2008 and, Vogl *et al*., 2014), the importance and effect of terminators in gene expression has been demonstrated in yeast (Curran *et al*., 2013 and, Vogl *et al*., 2016), the development of recyclable selectable markers (Nett and Gerngross, 2003), genome editing techniques such as the Cre-loxP system (Marx *et al.*, 2008) and CRISPR/Cas9 (Weninger *et al*., 2016) and, finally, the full genome sequence of the background host *P.pastoris* strain GS115 has been reported (De Schutter *et al*., 2009). This study was a small exploration of the potential for using more modern techniques to improve upon the existing GlycoSwitch *P. pastoris* glycoengineering technology. However, utilisation of some of the recent advances in the molecular toolbox listed above it may now be possible to generate a stable glycoengineered strain of *P. pastoris* with precisely controlled, site specific plasmid-genome integration, a normal doubling time and devoid of selectable markers.

Chapter 6

Investigating the function of sialylated and non-sialylated recombinant factor H *in vitro* and *in vivo*

6.1 Overview

The overarching aim of this study was to engineer *P. pastoris* to produce recombinant FH with humanised glycosylation that would be suitable for therapeutic trials. It was important to characterise the resulting material, and in so doing compare the activity of sialylated (Sia-) and non-sialylated, terminal galactosylated (Gal-) IYD variant rFH, with human plasma-derived FH (hFH) and deglycosylated (DeGly-) VYE variant rFH

In previous chapters of this study different glycoforms of FH had been produced. Firstly, human plasma derived FH (hFH) carrying, predominantly disialylated diantennary complex-type glycans, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$. Secondly, recombinant FH expressed in *P. pastoris* strain SuperMan 5 and subsequently deglycosylated (DeGly-rFH). Thirdly, recombinant FH expressed in the glycoengineering strain SuperGal and carrying diantennary digalactosylated complex-type glycans, $Gal_2GlcNAc_2Man_3GlcNAc_2$ (Gal-rFH). Fourthly, Gal-rFH sialylated with the α(2,6)-sialyltransferase from the bacterium *Photobacterium sp*. to yield rFH carrying diantennary disialylated, $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$, as well as monosialylated, $Sia_1Gal_2GlcNAc_2Man_3GlcNAc_2$, complex-type glycans (Sia-rFH).

As discussed in Chapter 1 section 1.2.2, the role of FH is to regulate complement activation in fluid-phase and on self-surface by binding to and suppressing C3b amplification. FH destabilises the C3bBb complex; this is its DAA. FH also has CA for cleavage of C3b by the serine protease FI. Cleavage of C3b by FI sets off a chain of proteolytic events converting C3b to iC3b then to C3dg and, eventually, to C3d.

In vitro function of different glycoforms of FH produced either recombinantly (DeGly-, Gal- and Sia-rFH) or purified from human blood plasma (hFH) were compared. SPR Biacore based assays were used to compare the ability of these different glycoforms to bind to C3b and C3d and to compare the DAA activity. In solution assays were used to compare the CA of these different glycoforms.

Complement is an ancient, evolutionarily conserved, crucial part of the immune system. As such, the general features of complement are conserved in mice and humans and, like in humans, mice rely upon regulators of complement activation to protect host cells from complement. FH is structurally and functionally conserved in mice and humans (Pouw *et al*., 2015) as the key regulator of the alternative pathway of complement.
A FH knockout (*cfh*$^{-/-}$) mouse model mimics the disease phenotype observed in factor H deficient humans and pigs (Høgåsen *et al*., 1995, Rougier *et al*., 1998 and, Pickering *et al*., 2002). Lack of FH allows unhindered activation of fluid-phase C3 causing depletion of plasma C3, as it is converted to C3b and deposits on surfaces, and leads to spontaneous membranoproliferative glomerulonephritis in both pigs, humans and, crucially, *cfh*$^{-/-}$ mice. Supplementation of *cfh*$^{-/-}$ mice with either human (Fakhouri *et al*., 2010, Nichols *et al*., 2015 and, Michelfelder *et al., 2017*) or mouse FH (Paixão-Cavalcante *et al*., 2009) successfully restores complement regulation which can be observed by either an increase in plasma C3 levels and / or reduction in C3b staining of kidney sections.
In this study, comparison of *in vivo* function of rFH with humanised glycosylation was carried out in mice. The focus of these assays was two-fold. Firstly, the four glycoforms (hFH, DeGly-rFH, Gal-rFH and Sia-rFH) were compared for the ability to recover plasma C3 levels and, thus, regulate complement *in vivo*. Secondly, due to the known effects of differential glycosylation on circulatory half-life of glycoproteins (Morell *et al*., 1971 and, Fukuda *et al*., 1989), the clearance rate of the four glycoforms of FH were compared in mice.

In the first of the two *in vivo* mice assays, the mice used were wild-type, i.e. with no genetic modifications. Wild-type mice were used as the assay was designed to analyse exclusively the plasma half-life of FH independently of the role of FH as a complement regulator.
In the second mouse assay the aim was to evaluate the ability of the four glycoforms to regulate complement. Therefore, for this assay, *cfh*$^{-/-}$ mice were used.

6.2 Materials and Methods

6.2.1 Production of deglycosylated recombinant factor H

6.2.1.1 Deglycosylation of partially purified recombinant factor H

A sample of 54.6 mg of rFH expressed in *P. pastoris* strain SuperMan 5 and partially purified by cation exchange chromatography before dialysis against PBS, pH 7.4 (see methods in Chapter 3 sections 3.2.1.1 and 3.2.1.2), was deglycosylated by incubating with 1 kU Endo H_f per milligram of rFH at 37 °C for 45 minutes. Deglycosylation was assessed using samples of approximately 2.5 µg protein by SDS-PAGE analysis on 4-12% bis-tris polyacrylamide gels (NuPAGE 4-12% bis-tris, ThermoFisher Scientific, NP03222) and run (200 V, 1 hour) before staining with Coomasie Brilliant Blue G250.

6.2.1.2 Concanavalin A resin chromatography

A 5-mL concanavalin A (ConA) sepharose (Sigma-aldrich, C9017) resin gravity-flow column was washed with 5 CV of a solution containing 1 M NaCL, 5 mM $MgCl_2$ and 5 mM $CaCl_2$. The resin was equilibrated with 5 CV of PBS, pH 7.4, before loading the deglycosylated rFH sample (DeGly-rFH). Flow-through fractions were collected, and then the column was washed with 2 CV of PBS and wash fractions were collected. The protein concentration of flow-through and wash fractions was estimated from absorbance at 280 nm. SDS-PAGE analysis of approximately 1 µg protein samples from flow-through and wash fractions was carried out on 4-12% bis-tris polyacrylamide gels (200 V, 1 hour).

6.2.1.3 Anion-exchange chromatography of deglycosylated recombinant factor H

Flow-through and wash fractions were pooled, placed in dialysis tubing (10 kDa molecular weight cut off, 35 mm dry internal diameter, Snakeskin, ThermoFisher Scientific, 88245) and dialysed against anion-exchange chromatography buffer A (20 mM glycine, 80 mM NaCl, pH 9.5) for 2 hours at 4 °C followed by transfer into fresh buffer and dialysis for 16 hours at 4 °C.

Anion-exchange chromatography was performed using a 6 mL Resource Q column (GE Healthcare Life Sciences, 17117901) according to the conditions in Chapter 3 section 3.2.1.3. Briefly, at a flow-rate of 6 mL/min, the column was equilibrated with 6 CV of buffer A, protein sample was loaded onto the column and washed with 5 CV buffer A. Protein was eluted in a linear gradient from 0-50 % buffer B (20 mM glycine, 1 M NaCl, pH 9.5). Fractions of 2 mL were collected and 2 µL samples from various fraction were analysed by

SDS-PAGE. Peak rFH-containing fractions were pooled, placed in dialysis tubing (10 kDa molecular weight cut off, 35 mm dry internal diameter, Snakeskin, ThermoFisher Scientific, 88245) and dialysed against PBS for 2 hours at 4 °C followed by transfer into fresh buffer and dialysis for 16 hours at 4 °C. Protein concentration and purity was assessed by absorbance at 280 and 260 nm. The purified sample of ~14 mg DeGly-rFH was aliquoted and stored frozen at -80 °C.

6.2.2 In vitro functional characterisation of four glycoforms of factor H

6.2.2.1 SPR-based C3b and C3d-binding assays

Protein samples were assessed for purity by SDS-PAGE prior to SPR-based binding assays. Samples of 3 µg of each glycoform of FH, C3b (CompTech, A114) and C3d (CompTech, A117) were prepared under reducing and non-reducing conditions, applied to 4-12% bis-tris polyacrylamide gels and polyacrylamide gel electrophoresis was carried out (200 V, 1 hour) and stained with Coomasie Brialliant blue G250.

C3b- and C3d-binding assays were performed using a Biacore T200 instrument at 25 °C. For these studies, 150 RUs of C3b or 50 RUs of C3d were attached by standard amine coupling to the flow cells of a C1 sensor-chip (GE Healthcare Life Sciences), the other two available flow cells were left blank for background subtraction. A two-fold dilution series of each glycoform of FH was performed in 10 mM HEPES, 150 mM NaCl, 0.05% (v/v) surfactant P20 (HBS-P^+ buffer, GE Healthcare Life Sciences). Glycoform dilutions were injected for 150 s at a flow rate of 50 µL/min followed by a dissociation time of 600 s. For each dilution series, a negative control sample containing PBS diluted in HBS-P^+ buffer was also injected. The sensor-chip was regenerated by three injections of 1 M NaCl for 30 s at a flow-rate of 50 µL/min. Data were analysed using the Biacore Evaluation software and a 1:1 steady state binding model.

6.2.2.2 SPR-based decay acceleration assay

Decay-accelerating activity was measured in an SPR-based assay performed using a Biacore T200 instrument at 25 °C. 1000 RUs of C3b was attached to a single flow-cell of a Biacore CM5-sensor chip (GE Healthcare Life Sciences) using standard amine coupling. A second

flow-cell was left blank for background subtraction. C3bBb was assembled on the chip by a 180 s 10 µL/min injection of a solution containing 500 nM Factor B (CompTech, A135) and 50 nM Factor D (CompTech, A136). The formation of the C3 convertase complex C3bBb was monitored (~150 RUs), and then allowed to decay naturally, during a dissociation phase lasting 240 s, to observe intrinsic convertase decay. Glycoforms of FH (20 nM), or PBS, were injected for 180 s at a flow-rate of 10 µL/min and a further dissociation phase was monitored. The sensor-chip was regenerated between cycles by a single 30 s injection of 0.1 M hFH followed by three 30 s injections of 1 M NaCl at a flow rate of 10 µL/min.

6.2.2.3 Cofactor assay

The activity of FH as a cofactor for C3b cleavage, catalysed by FI, was measured by detection of C3b 63 and 39 kDa α'-chain proteolytic products following SDS-PAGE. The non-cleaved β-chain served as an internal control. The 16 µl reaction mixture contained 1.7 µM C3b, 14 nM FI (CompTech, A138), and a range of concentrations (80, 40, 20, 10 nM) of each glycoform of FH. The reaction mixture was incubated for 30 minutes at 37°C, then stopped by addition of SDS-PAGE loading buffer. The mixture was boiled for 5 minutes and subjected to polyacrylamide gel electrophoresis on 4-12% bis-tris polyacrylamide gels (200 V, 1 hour) and stained with Coomasie Brilliant Blue G250. Densitometry values were obtained using ImageJ software.

6.2.3 In vivo characterisation of four glycoforms of factor H in mice

6.2.3.1 Endotoxin removal

Endotoxin was removed from 10 mg samples of each glycoform of FH using Pierce High-Capacity Endotoxin Removal 0.5 mL Spin Columns (ThermoFisher Scientific, 88274). Briefly, columns were regenerated overnight in 0.2 M NaOH followed by washing once each with 2 M NaCl (Sigma-aldrich, 71386-1L), endotoxin-free ultra-purified H_2O (Merck, TMS-011-A) and endo-toxin free PBS (Merck, TMS-012-A) before adding sample and end-over-end mixing at 4 °C for 1 hour. The columns were centrifuged at 500 g for 1 minute after each wash step and for sample collection the columns were centrifuged at 500 g for 1 minute in a Sorvall Lynx 4000 centrifuge using an F14-14X50cy rotor (ThermoFisher Scientific).

6.2.3.2 Factor H half-life assay

Wild-type and *cfh*$^{-/-}$ mice were injected intraperitoneally with 1 mg of each glycoform of FH in identical volumes of PBS, or PBS alone. Three mice per glycoform of FH were used, and two mice for administration of PBS. Blood was collected into EDTA, *via* tail venesection before injection and at 2, 24, 48, 72 and 96 hours post-injection. Plasma was separated by centrifugation before storage of plasma samples at −80 °C.

Sandwich enzyme-linked immunosorbent assays (ELISA) were carried out on plasma samples from mice injected with glycoforms of FH. Each well of a 96-well plate was incubated in 100 µL 2 µg/mL Ox24 monoclonal anti-FH antibody in carbonate buffer. Wells were washed thrice with wash buffer (PBS containing 0.1% (v/v) tween-20) and then blocked with 200 µL blocking buffer (1% (v/v) bovine serum albumin (BSA) in wash buffer) for 1 hour at room temperature. Plates were washed with wash buffer and then plasma samples from mice were diluted 1:1000 and 1:5000 with blocking buffer and 100 µL added to wells in the 96-well plate. For calibration, a 1:2 dilution series for each sample was set-up with a starting concentration of 100 ng/mL and containing a total of seven dilutions. Then 100 µL of each dilution (plus a zero FH sample) were added to the 96-well plate before incubation for 1 hour at room temperature. The plates were washed thrice with wash buffer. Then 100 µL Sheep anti-FH antibody diluted 1:20,000 in blocking buffer was added to each well in the 96 well-plate and left to incubate for 1 hour at room temperature. The plate was washed thrice with wash buffer. 100 µL Donkey anti-sheep horseradish-peroxidase conjugate antibody diluted 1:20,000 in blocking buffer was added to each well in the 96-well plate and left to incubate for 1 hour at room temperature. The plate was developed by the addition of 100 µL of TMB HRP Microwell Substrate (Leinco Technologies, T118). The reaction was stopped after 5 minutes by the addition of 100 µL 10% (v/v) H_2SO_4. Spectrophotometry was performed at a wavelength of 450 nm.

6.2.3.3 Plasma C3 concentration in cfh$^{-/-}$ *mice*

Three mice per glycoform of FH were administered with 1 mg intraperitoneal injections of FH and plasma harvested in a process identical to that detailed for FH half-life analysis in wild-type mice, except two additional injections of 1 mg of each glycoform were

performed at 48 hours and 96 hours and mice were sacrificed at 144 hours after the first injection.

The plasma concentration of C3 was measured by sandwich ELISA in a procedure similar to the sandwich ELISA for measurement of FH concentration detailed in section 6.2.3.2., except, plates were pre-coated with 1:250 dilution anti-C3 11H9 antibody, the blocking buffer used was 2% (w/v) BSA plus 2 mM EDTA in wash buffer, the 1:2 serial dilution used C3 (CompTech), plasma was diluted 1:1,000 and 1:2,000, samples, dilution series, control, antibody and TMB volumes used were all 50 µL, and detection was with 1:25,000 dilution of HRP-conjugated anti-C3 antibody.

6.3 Results

6.3.1 Deglycosylation and purification of partially purified recombinant FH

To prepare DeGly-rFH, a stock of rFH, previously produced in the *P. pastoris* strain SuperMan5 *CFH* and partly purified from crude supernatant by cation-exchange chromatography, was used as starting material. This was deglycosylated with Endo H_f and then subjected to a two-step purification procedure.

Partially purified rFH (54.6 mg) was enzymatically deglycosylated by incubating with endo H_f (1 kU per mg FH). Deglycosylation was assessed by MW shift on SDS-PAGE (fig. 6.1).

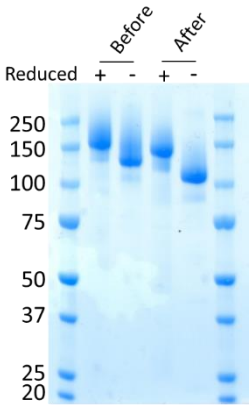


Figure 6.1: SDS-PAGE of rFH that had been produced in *P. pastoris* strain SuperMan5 *CFH* before and after deglycosylation with endo H_f. Approximately 2.5 µg of protein from each sample of rFH was loaded onto the polyacrylamide gel. +/- signifies whether or not samples were treated with reducing agent. MW markers are measured in kDa.

SDS-PAGE analysis of rFH after incubation with Endo H_f shows a significant loss in MW. For example, prior to the addition of Endo H_f reduced rFH migrates to a position just above the 150 kDa MW marker and non-reduced rFH migrates to a position equivalent to approximately 125 kDa. After the addition of Endo H_f, reduced rFH migrates to a position equivalent to approximately 140 kDa and non-reduced rFH migrates to a position equivalent to approximately 110 kDa. The ExPASy predicted MW of the amino acid sequence of human FH is approximately 137 kDa. Deglycosylation with Endo H_f leaves a GlcNAc (MW= 203 Da) at each of eight N-linked glycosylation sites. Therefore, the predicted MW of Endo H_f-deglycosylated rFH is 138.6 kDa. Thus, under reduced conditions, the Endo H_f-treated rFH is migrating at the expected MW. In addition, the band for Endo H_f-treated rFH appears less smeared. Smearing of the untreated sample is likely the result of variation in MW caused by heterogeneity in the *P. pastoris* glycans, as shown by MALDI-TOF mass spectrometry (Chapter 3 section 3.3.3 figure 3.5). Thus, the shift in MW and reduction in band smearing suggest that rFH has been successfully deglycosylated by Endo H_f. In support of this, the MW loss of approximately 10 kDa after treatment with Endo H_f is consistent with the loss of 8 five mannose glycans released from rFH by Endo H_f cleavage between the two GlcNAc residues of the N-linked glycan chitobiose core. The released glycans, $Man_5GlcNAc_1$, would have a molecular weight of approximately 1 kDa each.

Following deglycosylation, DeGly-rFH was applied to a gravity-flow chromatography column containing concanavalin A (ConA) sepharose resin. As discussed in Chapter 1 section 1.5.2, concanavalin A is a lectin which binds to the mannose residues in the trimannosyl core of N-linked glycans. Application of DeGly-rFH to this resin resulted in the removal of free glycans and any rFH that had not been deglycosylated. After application of Endo H_f treated rFH to the ConA column, flow-through and wash fractions were collected and SDS-PAGE analysis was carried out to assess for the presence of rFH (fig. 6.2).

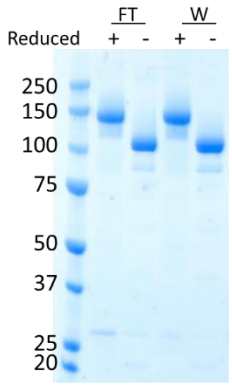


Figure 6.2: SDS-PAGE of flow-through (FT) and wash (W) fractions of Endo H_f treated rFH after application to a ConA chromatography column. Approximately 1 µg of protein of each sample was loaded. +/– signifies whether or not samples were reduced. MW markers are measured in kDa.

As expected, rFH was eluting both in the flow-through and the wash fractions (fig. 6.2), suggesting that rFH was either not interacting, or interacting weakly, with the Con A resin. Flow-through and wash fractions were pooled together, dialysed against anion-exchange buffer A (20 mM glycine, 80 mM NaCl, pH 9.5) and the concentration was measured by UV-Vis spectroscopy absorbance at 280 nm wavelength. Approximately 55 mg of rFH was applied to the column but only 28.7 mg (table 6.1) of protein was collected in the pooled and dialysed flow-through and wash fractions, yielding a 48% loss of protein.
To maximise enzymatic deglycosylation of rFH an extended incubation time with Endo H_f was needed (> 1 hour to overnight to ensure complete deglycosylation).
However, *P. pastoris* is known to secrete proteases (Sinha *et al*., 2005 and, Salamin *et al*., 2010) and, as rFH used in the enzymatic deglycosylation reaction is only partially purified, there was a risk of contamination from *P. pastoris* secreted proteases. So, the incubation time used for deglycosylation of rFH (45 – 60 minutes) was a trade-off between achieving deglycosylation whilst reducing proteolysis. Therefore, the loss of protein could be attributed to incomplete deglycosylation and retention of glycosylated rFH on the ConA resin.

DeGly-rFH was then purified by anion-exchange chromatography (6 mL resource Q column) and peak fractions were analysed by SDS-PAGE (fig. 6.3).

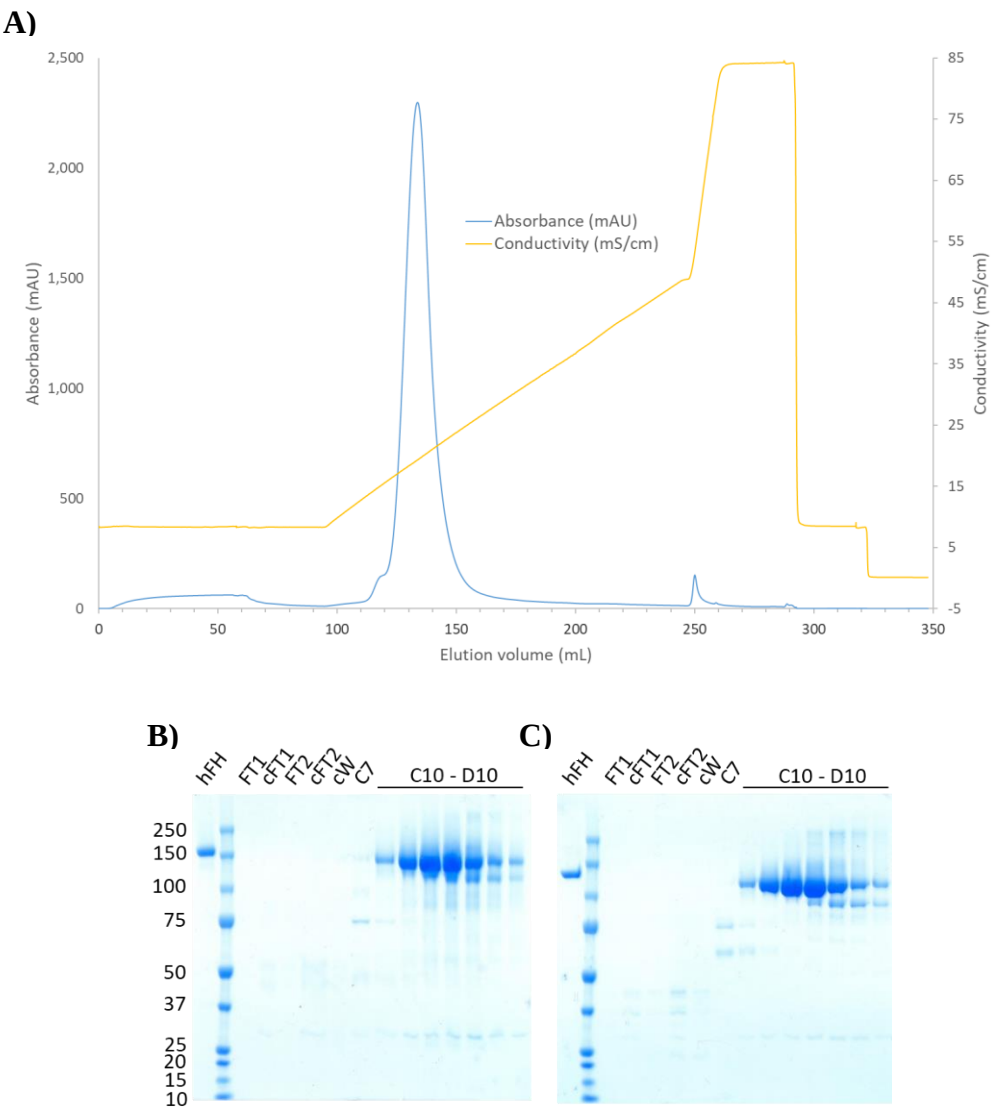


Figure 6.3: Chromatogram (A) and SDS-PAGE analysis of reduced (B) and non-reduced (C) samples of flow-through fractions 1 (FT1) and 2 (FT), ten-times concentrated flow-through fractions (cFT1 and cFT2), ten-times concentrated wash fraction (cW) and peak fractions C7 (elution volume 116 mL) and C10 – D10 (elution volume 122 mL – 146 mL). Fractions were collected in 2 mL volumes. Every second fraction is loaded onto the gel. 2 µL of each fraction was diluted in 14 µL SDS loading buffer, and all 16 µL of each sample was loaded onto SDS-PAGE gel. MW markers are measured in kDa.

A total of 28.5 mg of deglycosylated rFH was loaded onto the anion-exchange column and fractions were collected in 2 mL volumes. The chromatogram shows a large A_{280} peak eluting at 120 – 150 mL with a shoulder at ~110 – 120 mL (fig. 6.3 A). Fractions were analysed by SDS-PAGE (fig. 6.3 B and C). This showed that there is undetectable rFH in the flow-through or wash fractions, even when concentrated ten times (cFT1, cFT2 and cW in figure 6.3 B and C) and the peak shoulder also appear devoid of rFH (C7). Instead, rFH eluted within the main peak (fraction C10 – D10, 122 mL – 146 mL elution volume) evidenced by intense bands at ~140 kDa for reduced samples and 120 kDa for non-reduced

samples. However, rFH appears to be co-eluting with a protein with a MW ~30 kDa. There is also a band just below the main rFH band, ~120 kDa in the reduced samples and approximately 90 kDa in the non-reduced samples. These are likely to be proteolytically "clipped" rFH. Fractions C12 – D6 (elution volume 126 – 138 mL) had the highest concentration of purified rFH, assessed by SDS-PAGE (fig. 6.3 B and C), and were pooled and dialysed against PBS, pH 7.4.

Purification Step	Extinction coefficient FH ($M^{-1}cm^{-1}$)	Molecular weight FH (kDa)	Absorbance			Concentration		Volume (mL)	Yeild (mg)	Purity (A260/A280)
			280	260	320	(μM)	(mg/mL)			
Before deglycosylation	246,800	139					1.3	42	54.6	
After deglycosylation, ConA chromatography and dialysis against anion exchange buffer A			0.170	0.136	0.080	3.44	0.479	60	28.7	0.8
Post anion exchange chromatography pooled peak fractions dialysed against phosphate buffered saline			0.352	0.219	0.029	7.13	0.991	14	13.9	0.6

Table 6.1: Table containing protein absorbance readings, concentrations, volumes, protein yields and purity at each stage in the deglycosylation and purification of rFH.

After dialysis, the nucleic acid contamination and concentration of protein was measured (table 6.1), by absorbance at 260 nm and 280 nm, to yield 13.9 mg protein of low nucleic acid contamination (A260/A280 of 0.6).

6.3.2 SDS-PAGE analysis of factor H glycoforms and C3b and C3d

Prior to functional characterisation, the four samples of FH glycoforms prepared in this study, together with the commercially produced samples of C3b and C3d, were evaluated for purity by SDS-PAGE (fig. 6.4).

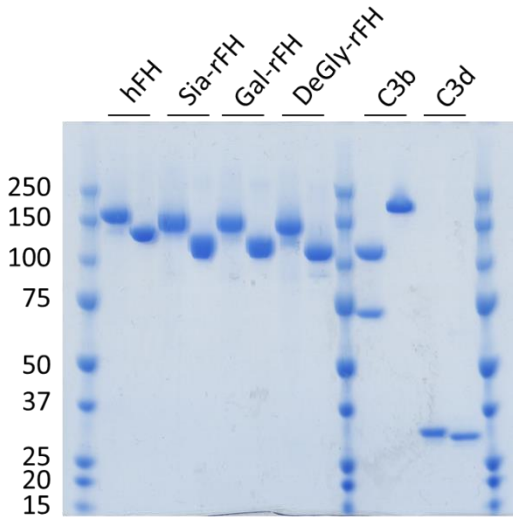


Figure 6.4: SDS-PAGE analysis of five glycoforms of FH, C3b and C3d to evaluated purity. Replicate samples were prepared. The first replicate was prepared under reducing conditions (left) and the other replicate prepared under non-reducing conditions (right). 3 µg of protein was applied to a 4-12% bis-tris polyacrylamide gel. MW markers are measured in kDa. The expected MW of hFH is 155 kDa, C3b is 177 kDa and of C3d is 35 kDa).

Samples were prepared under reducing (left lane) and non-reducing (right lane) conditions to assess for partial protein degradation by protease "clipping" (fig. 6.4). Under reducing conditions all samples of FH display evidence of a small amount of "clipping" as there is a faint band just below the main FH band at ~150 kDa. This band is not present in the non-reduced samples hFH, Sia-rFH and Gal-rFH but is present in the non-reduced samples of DeGly-rFH. This may be the result of DeGly-rFH having undergone deglycosylation in only a partially purified state and, therefore, was exposed to *P. pastoris* secreted proteases. Additionally, DeGly-rFH had been stored frozen after the initial cation-exchange purification from crude *P. pastoris* culture media, and it underwent a freeze-thaw cycle which may have contributed to increased degradation. All three groups of protein display a mobility shift on SDS-PAGE (fig. 6.4) when reduced consistent with the presence of the numerous (40) disulfides expected to form in FH.

C3b (177 kDa) migrates as a single band when not reduced (fig 6.4), but as two bands when reduced. This is expected, as C3b is composed of two subunits, an α'-chain and a β-chain, which are covalently linked via a disulphide bond (Janssen *et al*., 2006). Reducing agent results in the appearance of two bands corresponding to these two chains. The C3 fragment C3d (35 kDa) also has a disulphide bond (Nagar *et al*., 1998) that could explain the shift in mobility on SDS-PAGE following reduction (fig. 6.4).

Overall, it appears from SDS-PAGE that all of the samples are of acceptable purity. Approximately 3 μg of each protein was loaded and the detection limit of the Coomassie G250 stain used is in the region of 30 ng (although this is protein sequence dependent) therefore any protein contaminants are likely to be less than 1% of the total protein content.

6.3.3 *in vitro* functional characterisation of sialylated and non-sialylated recombinant factor H: Binding to C3b and C3d

The interactions of the four different glycoforms of FH with C3b and with C3d, a breakdown product of C3b that remains surface bound after degradation of C3b, were assessed using SPR-based binding assays. Duplicate concentrations of each FH glycoform were flowed over either C3b (150 RUs) or C3d (50 RUs), that had been fixed to a C1 sensor-chip by standard amine coupling, to obtain steady-state sensorgrams (figs. 6.5 and 6.6).

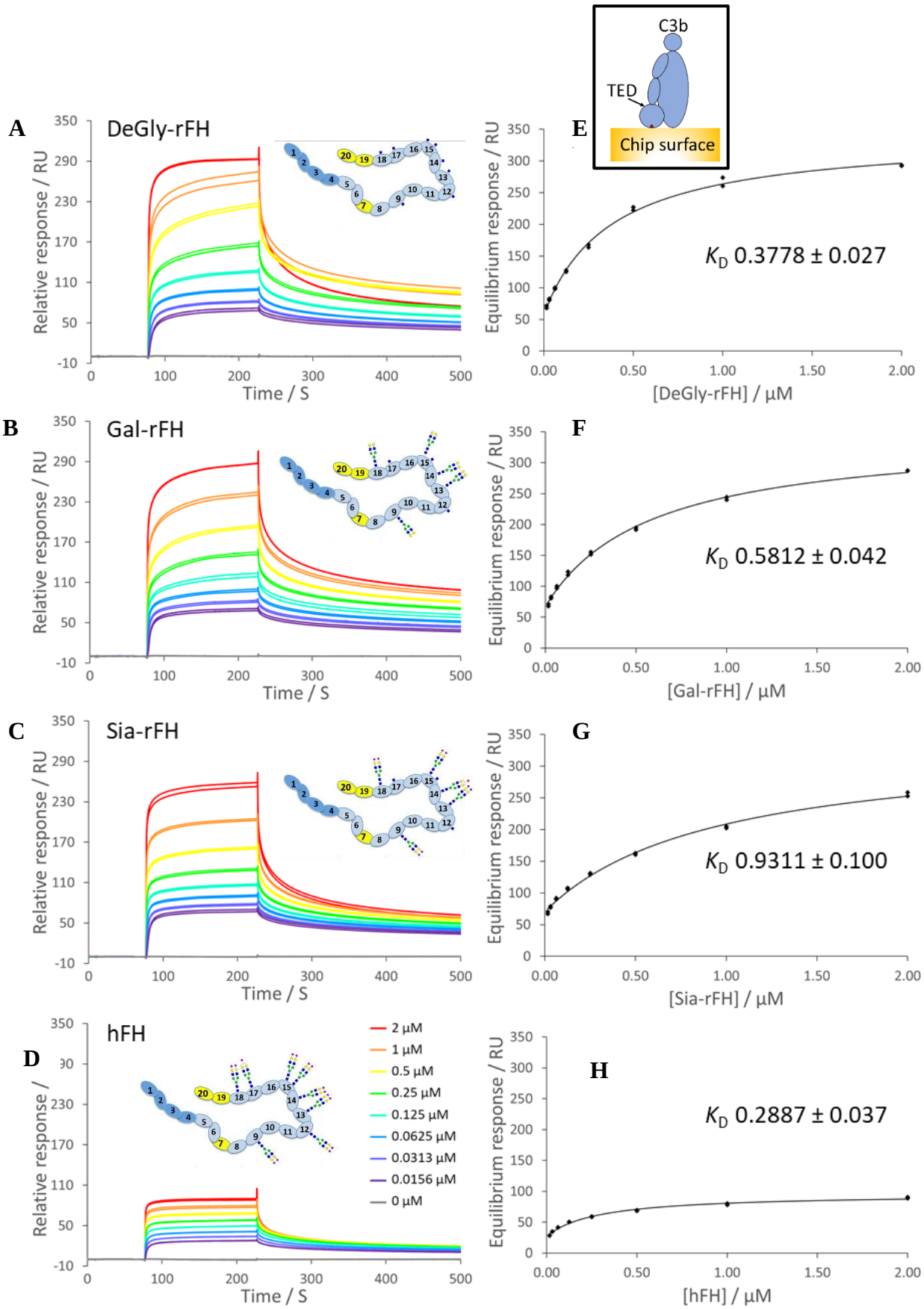


Figure 6.5: C3b binding sensorgrams (A), B), C) and D)) and affinity curves (E), F) G) and H)) of recombinant deglycosylated FH (hFH) (A) and E)), humanised terminally galactosylated rFH (Gal-rFH) (B) and F)), humanised terminally sialylated rFH (Sia-rFH) (C) and G)) and, human plasma derived FH (hFH) (D) and H)) binding to 150 RUs of C3b. Sensograms were generated from duplicate injections of eight concentrations of each glycoform of FH and a blank for 150 s at 50 µL/min followed by dissociation for 600 s. Sensorgrams were referenced and blank subtracted, and 1:1 steady state kinetic model was fitted to sensorgrams to generate the affinity curves. R_{max} and χ^2 values in table 6.2.

The dissociation phase of the sensorgrams for all glycoforms of FH binding to C3b, including FH purified from plasma, does not return to base line. This has been observed previously by colleagues in the Barlow group and is thought to be due to “stickiness” of the sensor-chip surface *i.e.* FH is pulled down to the sensor-chip surface by binding to C3b, but when FH dissociates from C3b it interacts with and sticks to the sensor-chip surface. The sensorgrams, however, show that steady-state has been reached for all glycoforms of FH-C3b interactions. This allowed for the production of affinity curves and calculation of K_D values (Table 6.2).

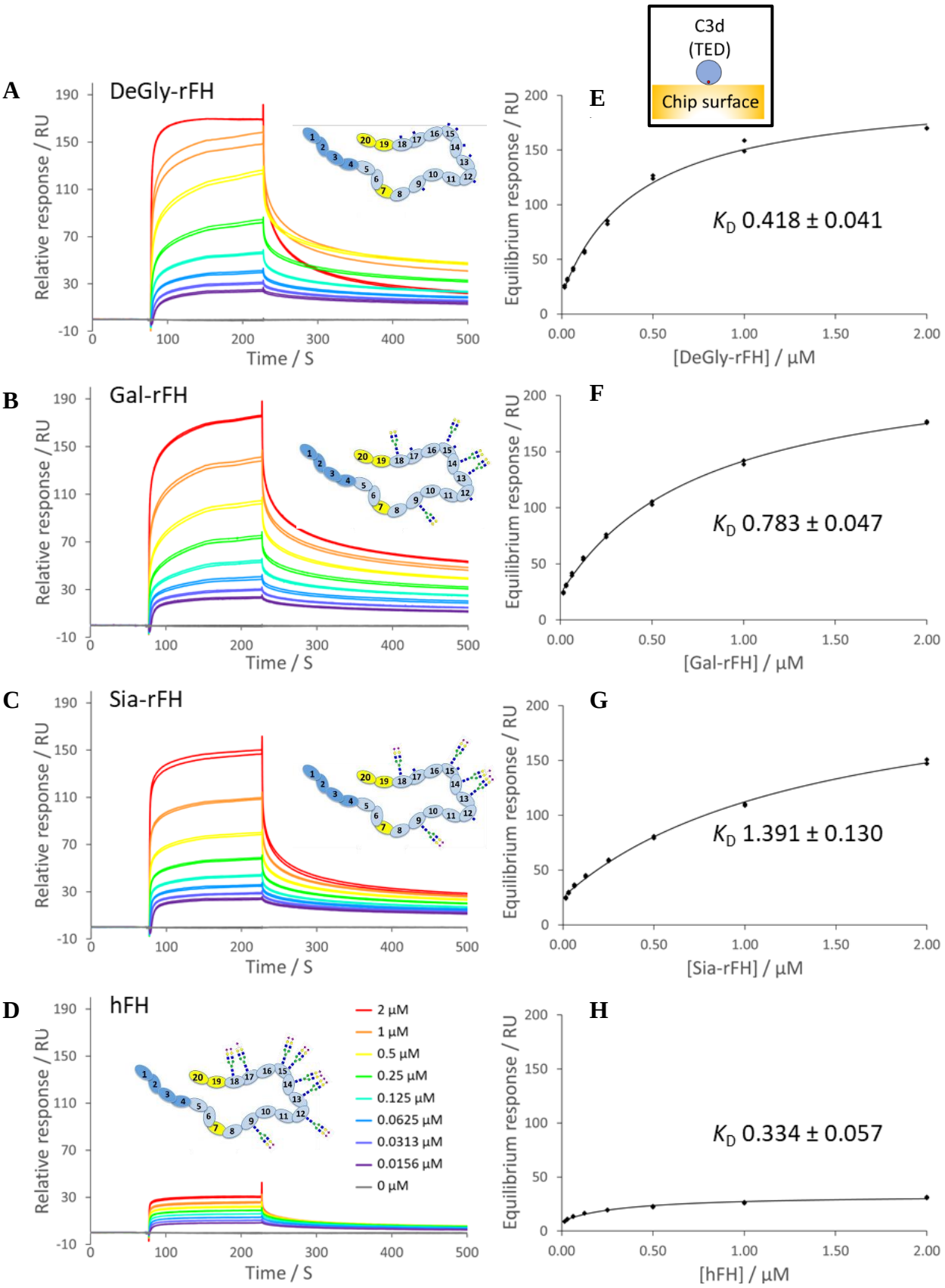


Figure 6.6: C3d binding sensorgrams (A), B), C) and D)) and affinity curves (E), F) G) and H)) of recombinant deglycosylated FH (hFH) (A) and E)), humanised terminally galactosylated rFH (Gal-rFH) (B) and F)), humanised terminally sialylated rFH (Sia-rFH) (C) and G)) and, human plasma derived FH (hFH) (D) and H)) binding to 50 RUs of C3d. Sensograms were generated from duplicate injections of eight concentrations of each glycoform of FH and a blank for 150 s at 50 µL/min followed by dissociation for 600 s. Sensograms were referenced and blank subtracted, and 1:1 steady state kinetic model was fitted to sensograms to generate the affinity curves. R_{max} and χ^2 values in table 6.2.

The sensograms for C3d binding (fig. 6.6) also show that the trace is not returning to baseline during FH dissociation. This is not surprising because C3d and C3b binding assays were conducted in different flow-cells of the same sensor-chip.

The affinities displayed in Table 6.2 were calculated by fitting the steady-state sensograms to a simple 1:1 interaction model. As well as the calculated affinities (dissociation constant, K_D), theoretical maximum response (R_{max}) and chi-squared (χ^2) values are also displayed in table 6.2.

An immediately striking feature of the data is the variation in the values of R_{max} for both C3b and C3d. The R_{max} vales are in line with expectations for hFH, but they are higher than expected for the other versions. C3b and FH have similar molecular weights so it would be expected that the R_{max} for FH binding to a chip with 150 RUs of C3b would not exceed 150 RUs. A value of 68.5 for hFH seems quite low, for example in Chapter 2 section 2.3.5 table 2.2 the R_{max} for hFH binding 150 RUs C3b was 76.5, but could be explained if some C3b molecules are immobilised in a way that renders the FH-binding site(s) inaccessible. The other glycoforms clearly behave differently. One explanation, consistent with the non-return of baseline to zero, is that FH molecules first bind C3b then migrate on to the surface from where they are hard to dislodge. Thus, more than one FH molecule could be present per C3b, e.g. one bound to the C3b and a second bound to the surface nearby producing an R_{max} that exceeds expectations. Another possibility – more consistent with the acceptable fit obtained when assuming a 1:1 interaction is that these non-native FH molecules may self-associate (Nan *et al.*, 2008) and bind to C3b and C3d as dimers.

	150RUs C3b					50RUs C3d				
Sample	K_D (µm)	SE K_D	R_{max} (RU)	χ^2 (RU^2)	Offset (RU)	K_D (µm)	SE K_D	R_{max} (RU)	χ^2 (RU^2)	Offset (RU)
DeGly-rFH	0.3778	0.027	283.10	22.4	58.22	0.418	0.041	189.4	17.50	16.95
Gal-rFH	0.5812	0.042	279.90	17.4	67.58	0.783	0.047	211.6	5.06	23.15
Sia-rFH	0.9311	0.100	267.10	22.1	70.46	1.391	0.130	207.5	5.21	25.46
hFH	0.2887	0.037	68.53	4.73	27.54	0.334	0.057	24.52	1.02	8.891

Table 6.2: Summary of the parameters fitted for the SPR-based binding studies of four different glycoforms of FH to the ligands C3b and C3d. Binding to 150 RUs of C3b and, 50 RUs of C3d was measured. SE = standard error.

The data in table 6.2 suggest that all four glycoforms bind to surface-bound C3b with similar affinities, in the approximate range of 0.3-0.9 µM. hFH has a slightly higher affinity

for C3b (K_D of ~0.29 μM) than the others (the mean K_D for all four is 0.5 μM) but this interaction has a markedly lower value of R_{max} (69 RUs compared to a mean of 225 RUs). A similar picture emerges for binding to C3d except that the R_{max} is even lower, relative to the others, (for hFH 25 RUs compared to an average for all four of 160 RUs). Of the four, Sia-rFH has the weakest binding for both C3b and C3d. Indeed, the ratio of C3b-binding affinity to C3d-binding affinity is remarkably consistent for all four FH glycoforms. In all cases the χ^2 values are less than 10% of the R_{max} values indicating an acceptable fit of the data to 1:1 steady state kinetics.

Taken together, hFH is behaving as expected but with a slightly lower response and K_D for C3b seen previously in this study and by others (Herbert *et al*., 2015 and, Kerr *et al*., 2017) and a higher than expected response for C3d has yielded a measurable K_D for C3d. The recombinant glycoforms all behave similarly, with similar responses and K_Ds for C3b and C3d binding. However, the assay needs repeating to determine if any individual differences between the recombinant glycoforms is statistically significant. The key finding is that humanisation of the glycosylation of recombinant FH does not appear to make rFH behave more like hFH.

6.3.4 *in vitro* functional characterisation of sialylated and non-sialylated recombinant factor H: decay acceleration activity

The ability of hFH and the three different glycoforms of recombinant FH to accelerate the decay of the bimolecular C3-convertase complex, C3bBb, was investigated using an SPR-based decay-acceleration activity (DAA) assay. In these experiments, 1000 RUs of C3b was immobilised on the surface of a CM5 sensor-chip by standard amine coupling. Then the C3bBb complex, C3-convertase, was generated by the addition of complement factors B and D. Formation of the C3-convertase complex was monitored (~ 150 RUs) and then allowed to decay naturally (in the absence of FB and FD) for a short time to observe intrinsic convertase decay, prior to addition, in duplicate, of the different glycoforms of FH to be tested. Plotting several such experiments on the same graph allowed a direct comparison of the dissociation rate of the C3-convertase complex on its own and in the presence of the different glycoforms of FH (fig. 6.7). Each glycoform DAA assay was carried out in duplicate. However, duplicate sensorgrams traces were omitted for clarity.

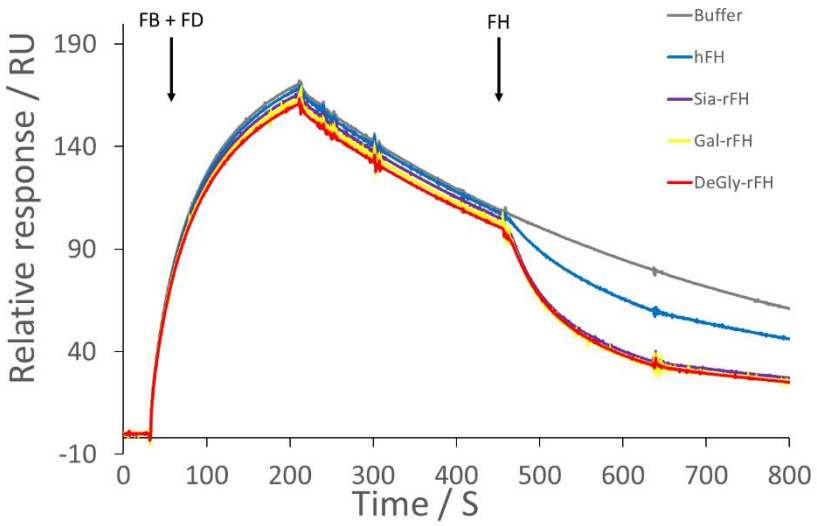


Figure 6.7: Sensorgram showing the results of a decay acceleration activity (DAA) assay of various glycoforms of FH on the C3-convertase complex, C3bBb. Approximately 1000 RUs C3b was amine coupled to the surface of a CM5 sensor-chip, C3bBb was assembled by an injection of 500 nM and FB 50 nM FD lasting 180 s. After formation of C3bBb, the complex was allowed to naturally decay for 240 s followed by the addition of FH glycoforms, in duplicate, at 10 µL/min for 180 s. Duplicate traces have been omitted for clarity. hFH is human plasma derived FH, Sia-rFH is terminal sialylated recombinant FH, Gal-rFH is terminal galactosylated recombinant FH and DeGly-rFH is deglycosylated recombinant FH variant VYE.

A rising response occurs over the range of ~20 s – 200 s (fig. 6.7), which is consistent with the binding of Factor B to surface immobilised C3b, and the subsequent cleavage of Factor B by Factor D to form the C3-convertase complex, C3bBb. An exponential decline in response units between ~200 s – 450 s represents the natural dissociation of the C3bBb complex. At ~460 s, the FH samples (20 nM) were injected, resulting in an accelerated decline in RU. This corresponds to a more rapid decay of the complex catalysed by the presence of FH – its DAA.

	Sample	Half-life (s)	Mean half-life (s)	SD	Change in half-life (approx.)
Pre-FH	DeGly-rFH 1	372.136	370.1	2.9	
	DeGly-rFH 2	367.966			
	Sia-rFH 1	385.475	381.6	5.4	
	Sia-rFH 2	377.823			
	Gal-rFH 1	372.181	371.1	1.5	
	Gal-rFH 2	370.008			
	hFH 1	385.126	385.0	0.1	
	hFH 2	384.916			
	Buffer 1	386.176	396.2	14.2	
	Buffer 2	406.229			
Post-FH	DeGly-rFH 1	179.61	180.0	0.5	x2
	DeGly-rFH 2	180.348			
	Sia-rFH 1	187.473	188.7	1.8	x2
	Sia-rFH 2	190.021			
	Gal-rFH 1	178.424	180.1	2.3	x2
	Gal-rFH 2	181.694			
	hFH 1	292.891	292.4	0.8	x1.3
	hFH 2	291.815			
	Buffer 1	415.086	435.3	28.6	x0.9
	Buffer 2	455.576			

Table 6.3: Calculated half-life (s) for decay of the C3bBb complex before and after the addition of FH. The change in half-life was calculated by dividing the mean half-life after the addition of FH by the mean half-life before the addition of FH.

Duplicates of each glycoform were assessed for DAA of C3bBb and decay half-life of the C3bBb complex was calculated before and after the addition of each glycoform. Averages of each duplicate were taken to find the mean half-life for each glycoform (Table 6.3). Changes in half-life were calculated for each sample relative to the initial natural decay of the C3bBb complex. All glycoforms of FH accelerate the decay of the C3bBb complex (fig. 6.7). However, hFH induces the smallest reduction in half-life (table 6.3) whilst the recombinant glycoforms all induce similar, but larger (then hFH) reductions in half-life (1.3 and 2-fold less than that of the C3bBb complex, respectively). This study previously found (see Chapter 2 section 2.3.5 table 2.3) that hFH had a 1.5-fold, DeGly-rFH a 2.5-fold and desialylated hFH a 2-fold decrease in half-life compared to natural C3bBb complex decay and this indicated that deglycosylation significantly improves DAA and that desialylation of hFH slightly improves DAA of FH. In comparison, this study (table 2.3) suggests that there is no difference between the DAA of the recombinant glycoforms of FH, whether sialylated or not, and that all three recombinant glycoforms have greater DAA compared to hFH.

6.3.5 *in vitro* functional characterisation of sialylated and non-sialylated recombinant factor H: cofactor activity

The ability for the different glycoforms of FH to act as a cofactor for FI-mediated proteolysis of the α'chain of C3b was investigated. Four concentrations (80, 40, 20 and 10 nM) of each glycoform of FH were incubated with C3b (1.9 μM) and FI (1.6 nM), and the extent of cleavage of the α'-chain of C3b was assessed by SDS-PAGE (fig. 6.8 A). Each reaction was carried out as independent duplicates. ImageJ software densitometry function was used to measure density of each band. The β-chain was used as an internal standard against which band intensities for the α'-chain-derived 63-kDa and 39-kDa fragments were normalised. The mean of normalised duplicate band intensities were plotted against concentration of FH (fig. 6.7 B, C and D).

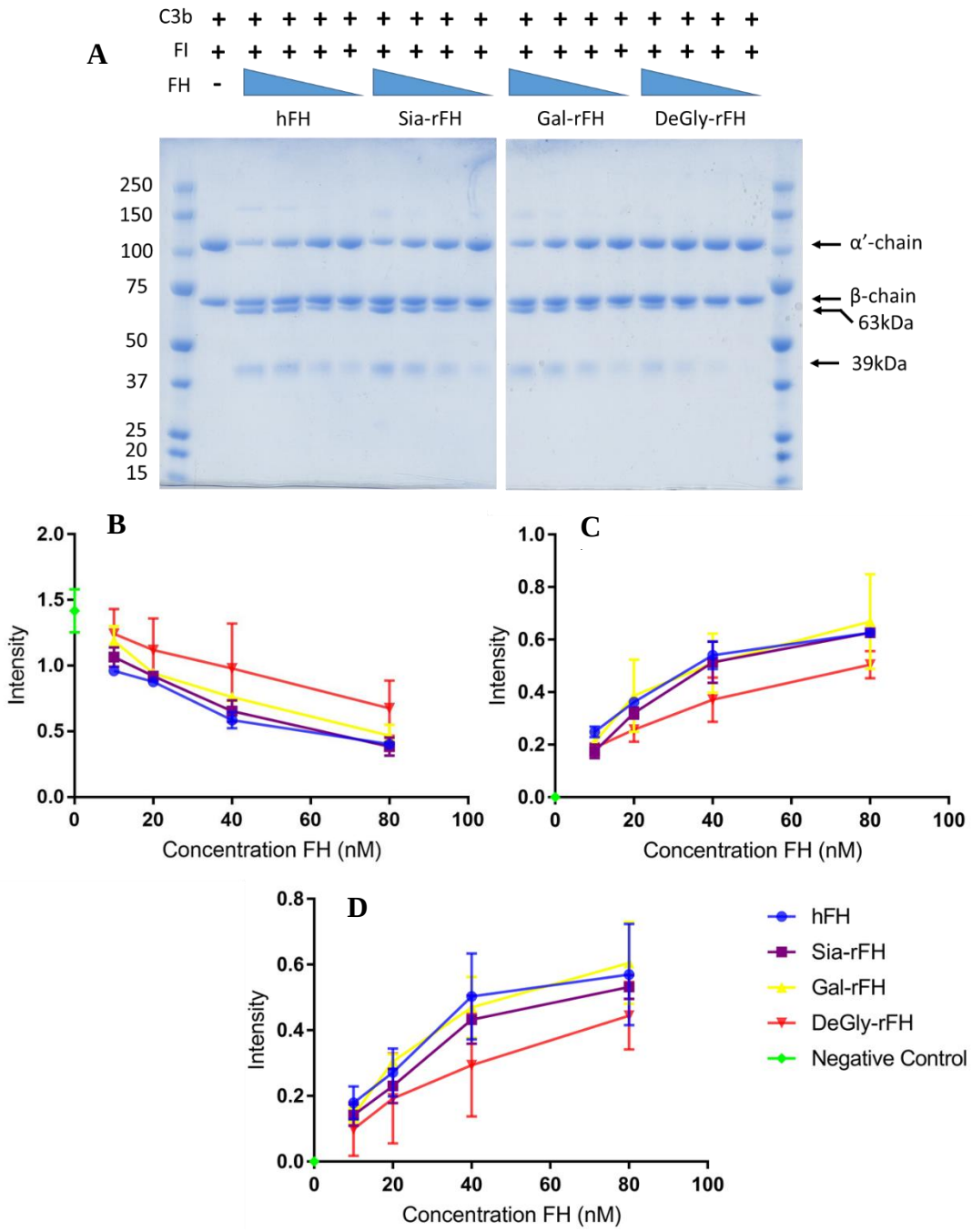


Figure 6.8: A) SDS-PAGE of the cofactor assay of four glycoforms of FH incubated with C3b and Factor I. B) Band intensity of the α'-chain, C) band intensity of the 63k-Da chain, D) band intensity of the 39-kDa chain. Band intensities were measured using ImageJ software and normalised against the β-chain in the same lane as the measured band. Each sample was run in duplicate (duplicates not shown in A)). Error bars are standard deviation. hFH is human plasma derived FH, Sia-rFH is terminally sialylated recombinant (r)FH with complex-type glycosylation, Gal-rFH is terminally galactosylated rFH with complex-type glycosylation, DeGly-rFH is deglycosylated rFH and the negative control is C3b and FI incubated in the absence of FH. MW markers measured in kDa.

Inspection of the SDS-PAGE gels (fig 6.12 A) shows that, as expected, in the presence of FH, FI cleaves the α'-chain of C3b to form two products of molecular weight 63 kDa and 39 kDa whilst leaving the β-chain of C3b intact. In the absence of FH the α'-chain is not cleaved by

FI. Densitometry performed on the gels (fig 6.8) suggests that Sia-rFH, Gal-rFH and hFH facilitate cleavage of the α'-chain to similar extents and, in comparison, DeGly-rFH has a slightly reduced cofactor activity. However, overlap of error bars, e.g. between hFH, Sia-rFH and DeGly-rFH in figure 6.8 D, indicate that any difference observed may not be significant.

6.3.6 *in vivo* functional characterisation of sialylated and non-sialylated recombinant factor H: circulatory half-life in wild-type mice

The four glycoforms of FH were assessed for retention in blood circulation of wild-type mice, capable of producing endogenous mouse FH (fig. 6.9). Blood circulation retention times of endotoxin-free glycoforms were assessed in triplicate, with three mice per glycoform of FH. A sample of 1 milligram of each glycoform was injected into wild-type mice and blood samples were taken pre-injection (zero-time point), and at 2, 24, 48, 72 and 96 hours post-injection. The concentration of human FH was measured in blood plasma samples by sandwich enzyme-linked immunosorbent assay (ELISA) in which FH was captured on the surface of a 96-well plate using the human FH-specific OX-24 monoclonal antibody. This allowed for the selective quantification of human, as opposed to mouse, FH.

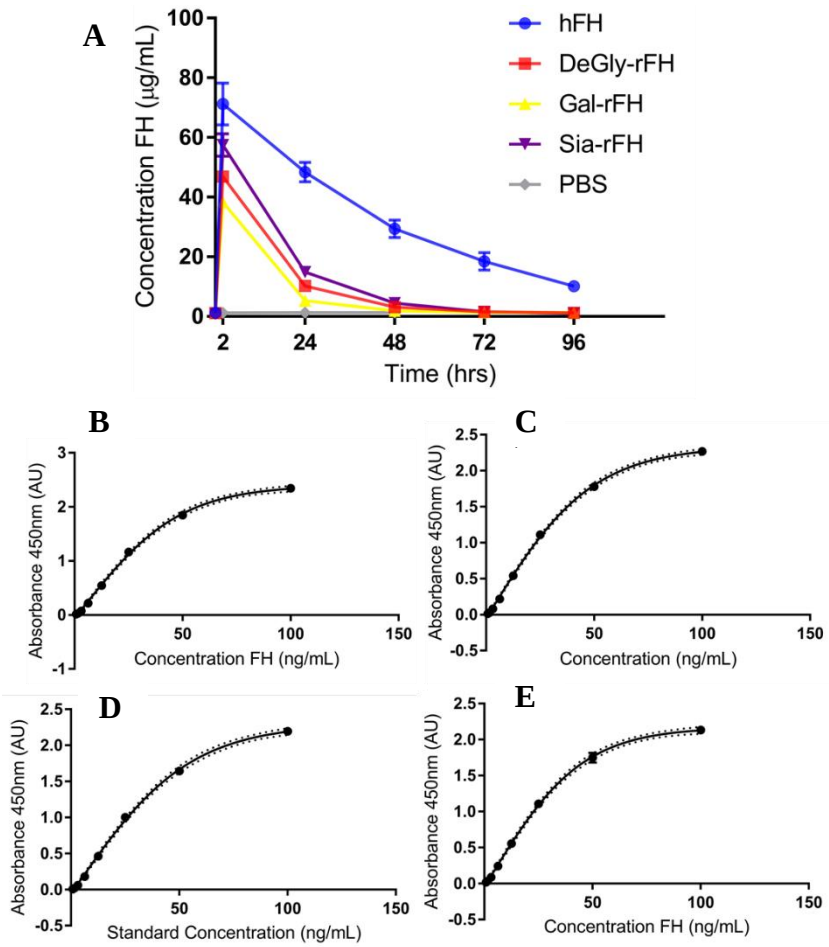


Figure 6.9: A) Concentration of four glycoforms of FH in samples of mouse plasma before, at two hours, and every 24 hours after, a single intraperitoneal injection of 1 mg of FH. Standard curves generated by plotting absorbance at wavelength of 450 nm against concentration of FH, using purified plasma derived human FH (B), DeGly-rFH (C), Gal-rFH (D) and Sia-rFH (E). PBS is phosphate buffered saline. Error bars in A) are standard error and the dotted lines in B), C), D) and E) trace the standard deviation.

The plots of concentration of FH as a function of time (fig 6.9 A) show a steep increase two hours post-injection; the hFH injections yielded the highest plasma concentration of FH (71.2 ± 7.0 µg/mL), Sia-rFH the next highest (57.5 ± 3.7 µg/mL), then DeGly-rFH (46.8 ± 0.5 µg/mL) and Gal-rFH has the lowest (38.3 ± 0.4 µg/mL). Between two hours and 24 hours, there was a steep decline in the concentration of FH. This continued and by 96 hours the concentration of the recombinant glycoforms of FH was close to or below the detection limit. Conversely, there is still a measurable concentration of hFH at 96 hours post-injection (10.1 ± 1.5 µg/mL).

Sample	Mean Concentation at 24 hours (µg/mL)	SEM	Mean Concentation at 96 hours (µg/mL)	SEM	Time (hrs)	Mean Half-life (hrs)	SEM
hFH	71.227800	7.03295	10.119330	1.529989	94	34.096	3.947
DeGly-rFH	46.849170	0.542914	1.230804	0.018902		17.903	0.031
Gal-rFH	38.306270	0.366129	1.269851	0.007739		19.126	0.041
Sia-rFH	57.463990	3.719782	1.060030	0.02574		16.334	0.365

Table 6.4: Calculated half-life (hours) for decrease of concentration of four glycoforms of FH in mouse blood plasma between 24 and 96 hours (hours) post-injection.

Measurements of half-life in blood plasma (table 6.4) allow for a comparison of the rate of clearance of each glycoform from mouse blood circulation and are independent of any potential differences in concentration of injected FH. hFH has the longest half-life in mouse blood plasma. In comparison, the half-life of the recombinant glycoforms of FH are similar and approximately half that of hFH. This suggests that the recombinant glycoforms are cleared more rapidly than hFH from mouse blood circulation.

Enzymatic deglycosylation with Endo H_f leaves behind a GlcNAc at each N-linked glycosylation site. Therefore, there are potentially three ways in which the clearance of DeGly-rFH, might be accelerated compared to hFH in mice. First, the remaining GlcNAc residue is a ligand of the mannose-binding receptor (Taylor and Drickamer, 1992). Second, the glycans mask putative protease recognition sites (Russell *et al*., 2009) therefore removal of the glycan exposes these sites for proteolysis. Third, removal of the glycans could expose novel antigenic sites for binding by antibodies and T-cell receptors (Szabó *et al*., 2009). Gal-rFH and Sia-rFH show a similar half-life to DeGly-rFH. Gal-rFH and Sia-rFH had been enzymatically deglycosylated with Endo H_f to remove yeast and hybrid-type glycans. Therefore, these glycoforms will have some GlcNAc residues and be at risk of accelerated blood plasma clearance by the same three mechanisms that applied to clearance of DeGly-rFH. In addition, terminally exposed galactose on Gal-rFH and on the monosialylated complex-type glycans on Sia-rFH present ligands for the asialoglycoprotein receptor (Ashwell and Harford, 1982 and, Schwartz, 1984).

6.3.7 *in vivo* functional characterisation of sialylated and non-sialylated recombinant factor H: Plasma C3 levels in *cfh*$^{-/-}$ mice

Next, the ability of the four glycoforms of FH to restore plasma C3 levels was investigated in *cfh-/-* deficient mice. A dose of 1 milligram of each glycoform of FH was administered to mice every 48 hours. A total of three injections were administered in a 144 hour period and

blood plasma samples were taken 0, 48, 96 and 144 hrs pre-injection. Each glycoform was administered to three mice and triplicate blood plasma samples were analysed at each time point for C3 concentration by sandwich ELISA. A standard curve of C3 absorbance at 450 nm wavelength against concentration of purified C3 was generated (fig. 6.10 inset) and used to estimate the concentration of C3 in blood plasma (fig. 6.10).

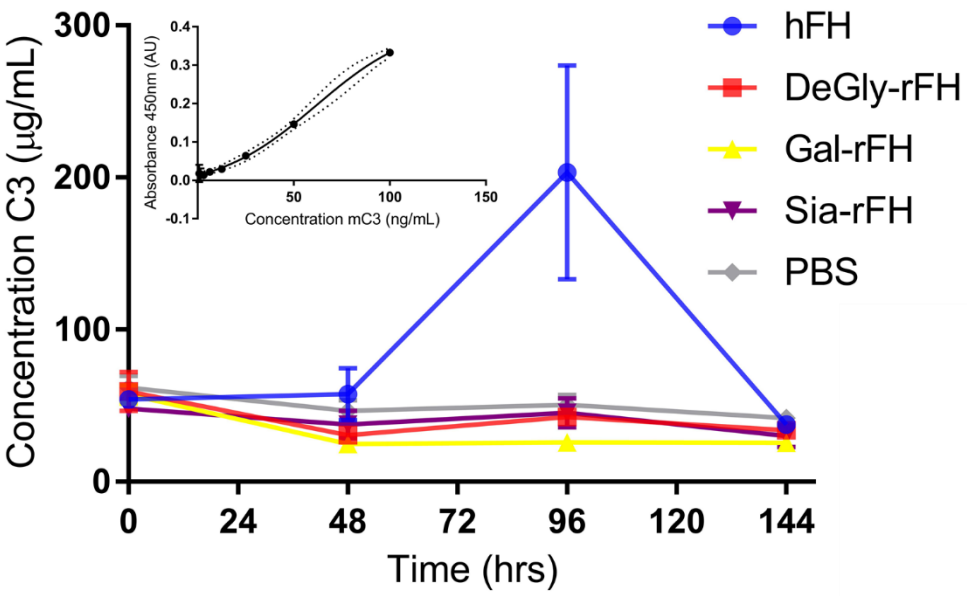


Figure 6.10: Concentration of C3 in mouse blood circulation after repeated intraperitoneal injections of 1 milligram of four different glycoforms of FH. Blood samples were taken before each FH injection at 0, 48, 96 and 144 hours. A standard curve, inset, of absorbance at 450 nm wavelength against concentration of purified C3 was used to estimate blood plasma C3 concentration. Error bars are standard error of the mean.

Figure 6.10 shows that plasma C3 levels remain low for the first 48 hours after the first injection of FH for all glycoforms of FH. After 48 hours, however, there is an increase in plasma C3 for hFH administered mice, but no change in C3 levels for the rFH administered mice. Pickering *et al*. (2002) observed that the normal C3 plasma concentration of wild-type mice is approximately 450 – 500 µg/mL and Paixão-Cavalcante *et al*. (2009) measured normal C3 plasma levels to be 387 µg/mL. Additionally, the plasma C3 levels in *cfh*$^{-/-}$ were measured to be 30 µg/mL. The peak C3 concentration observed, in this study, at 96 hours for hFH, 203.5 ± 70.5 µg/mL, is lower than previously observed (Pickering *et al*., 2002 and, Paixão-Cavalcante *et al*., 2009) but well above the low level of C3, 25 – 60 µg/mL, observed for all recombinant glycoforms of FH, PBS and hFH between 0 – 24 hours and at 144 hours.

The C3 plasma concentrations (fig. 6.10) were statistically evaluated using two-way analysis of variance (ANOVA) with Bonferroni multiple comparison test (table 6.5).

Sample	Time (hrs)	Mean Difference	95% CI of difference	Adjusted P-value	Significant difference
hFH	0	7.56	-34.97 to 50.09	>0.9999	No
DeGly-rFH		2.278	-40.25 to 44.81	>0.9999	No
Gal-rFH		4.489	-38.04 to 47.02	>0.9999	No
Sia-rFH		13.54	-28.99 to 56.07	>0.9999	No
hFH	48	-10.99	-58.53 to 36.56	>0.9999	No
DeGly-rFH		15.78	-26.75 to 58.31	>0.9999	No
Gal-rFH		21.64	-20.89 to 65.17	0.7600	No
Sia-rFH		8.874	-33.66 to 51.4	>0.9999	No
hFH	96	-152.6	-200.2 to -105.1	<0.0001	Yes
DeGly-rFH		8.161	-34.37 to 50.69	>0.9999	No
Gal-rFH		24.76	-17.77 to 67.28	0.5408	No
Sia-rFH		5.175	-37.35 to 47.7	>0.9999	No
hFH	144	4.26	-38.27 to 46.79	>0.9999	No
DeGly-rFH		7.983	-34.55 to 50.51	>0.9999	No
Gal-rFH		16.28	-26.25 to 58.8	>0.9999	No
Sia-rFH		11.65	-30.88 to 54.18	>0.9999	No

Table 6.5: Two-way analysis of variance (ANOVA) with Bonferroni multiple comparison test. The C3 concentration for each glycoform at each time point was analysed relative to the C3 concentration for PBS injected mice at each time point. P values were taken to be <0.05 and <0.01 vs. PBS. CI is confidence interval.

Although the plasma C3 levels at 96 hours for hFH are lower than observed in the literature, statistical analysis (table 7.4) of C3 concentration variance shows that the elevated C3 concentration at 96 hours is statistically significant. Statistical analysis also shows that there is no significant difference in plasma C3 concentration for any glycoform of rFH at any time point, relative to PBS administered mice. These data suggest that only hFH is elevating C3 plasma levels and that none of the rFH glycoforms have a measurable effect.

6.3.8 *in vivo* functional characterisation of sialylated and non-sialylated recombinant factor H: Plasma factor H levels in *cfh*$^{-/-}$ mice

The lack of response observed for plasma C3 levels in mice administered with the recombinant glycoforms of FH could be due to their low half-life resulting in accelerated clearance. In addition, Fakhouri *et al*., (2010) have observed the spontaneous development of anti-FH antibodies in mice administered with FH, therefore the lack of response seen for the recombinant glycoforms may be due also in part to the development of an anti-FH antibody. Thus, the FH concentration in blood plasma samples taken from *cfh*$^{-/-}$ mice was measured (fig 6.11) using the same sandwich ELISA assay employed previously for quantification of FH plasma concentration in wild-type mice.

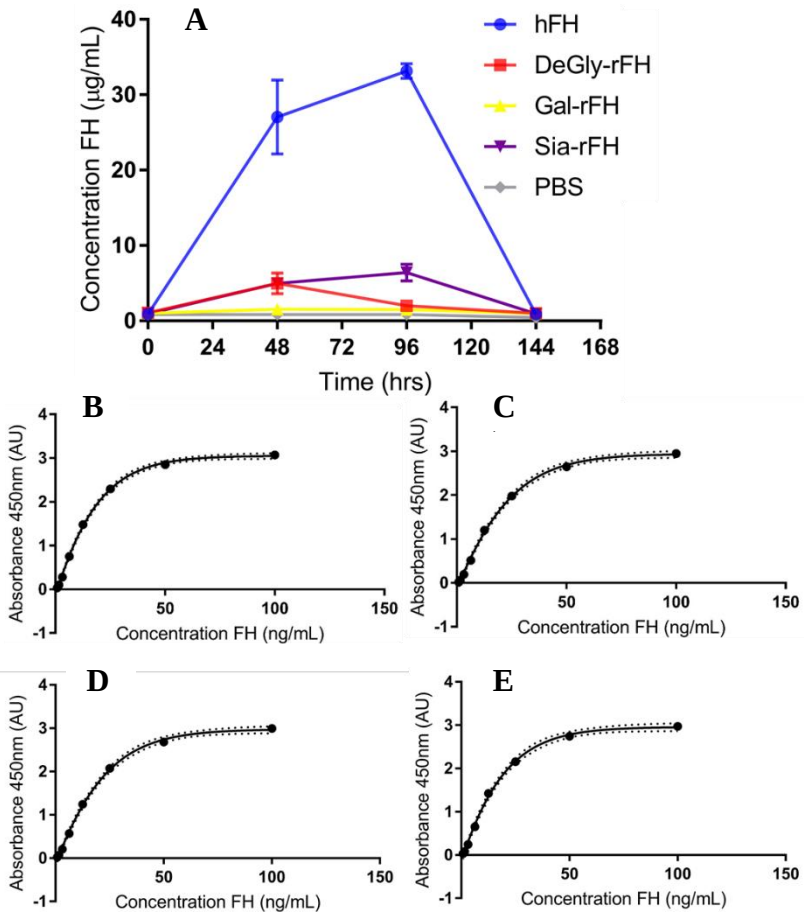


Figure 6.11: A) Concentration of FH in mouse blood circulation after repeated injection of 1 milligram of four different glycoforms of FH. Blood samples were taken before each FH injection at 0, 48, 96 and 144 hours. Standard curves of absorbance at 450 nm wavelength against concentration of purified hFH B), DeGly-rFH C), Gal-rFH D) and Sia-rFH E) were used to estimate blood plasma concentration of each of the four glycoforms of FH. Error bars are standard error of the mean.

The blood plasma concentration of hFH increases steadily to a maximum of 33.16 ± 0.99 µg/mL at 96 hours. The concentration of hFH decreases rapidly after 96 hours to 0.88 ± 0.01 µg/mL at 144 hours. The observed steady increase in concentration of hFH for 96 hours followed by a rapid decline between 96 to 144 hours is consistent with the findings of Fakhouri *et al*. (2010). Therefore, the observed rapid decline in plasma hFH concentration may indeed be due to the production in mice of an anti-hFH antibody.
In comparison, mice administered with Gal-rFH show no detectable Gal-rFH in blood plasma and mice administered with Sia-rFH and DeGly-rFH show a slight increase in plasma concentration at 48 hours, which is maintained at 96 hours in Sia-rFH administered mice. Statistical analysis by two-way ANOVA with Bonferroni multiple comparison test (table 6.6) was carried out to assess whether the differences seen in plasma concentrations of FH glycoforms compared to PBS administered mice (background) are significant.

Sample	Time (hrs)	Mean Difference	95% CI of difference	Adjusted P-value	Significant difference
hFH	0	-0.06999	-3.676 to 3.536	No	>0.9999
DeGly-rFH		-0.2519	-3.858 to 3.354	No	>0.9999
Gal-rFH		-0.1883	-3.794 to 3.417	No	>0.9999
Sia-rFH		-0.1149	-3.721 to 3.491	No	>0.9999
hFH	48	-26.2	-30.14 to -22.25	<0.0001	Yes
DeGly-rFH		-4.136	-7.742 to -0.5302	0.019	Yes
Gal-rFH		-0.7103	-4.316 to 2.895	>0.9999	No
Sia-rFH		-4.142	-8.092 to -0.192	0.0366	Yes
hFH	96	-32.32	-36.27 to -28.37	<0.0001	Yes
DeGly-rFH		-1.18	-4.786 to 2.426	>0.9999	No
Gal-rFH		-0.6885	-4.294 to 2.917	>0.9999	No
Sia-rFH		-5.569	-9.519 to -1.619	0.003	Yes
hFH	144	0.003153	-4.558 to 4.564	>0.9999	No
DeGly-rFH		-0.1228	-4.684 to 4.438	>0.9999	No
Gal-rFH		-0.09057	-4.652 to 4.47	>0.9999	No
Sia-rFH		-0.08116	-4.642 to 4.48	>0.9999	No

Table 6.6: Two-way analysis of variance (ANOVA) with Bonferroni multiple comparison test. The FH concentration for each glycoform at each time point was analysed relative to the FH concentration for PBS injected mice at each time point. P values were taken to be <0.05 and <0.01 vs. PBS. CI is confidence interval.

Statistical analysis (table 6.6) shows that the observed differences in plasma concentrations of DeGly-rFH (48 hours) and Sia-rFH (48 and 96 hours), relative to background, are significant. The analysis also shows that there is no detectable protein observed in plasma of all mice by the 144 hour time point. This would suggest that all glycoforms are essentially cleared by 144 hours.

The fact that DeGly- and Sia-rFH plasma were detectable in the plasma of FH knockout mice at significant concentrations for at least 48 hours is consistent with the half-life data for FH obtained in wild-type mice (fig. 6.9). The concentration of DeGly-rFH was measured to be 3.15 ± 0.29 µg/mL in wild-type mice and 4.98 ± 1.37 µg/mL in *cfh*$^{-/-}$ mice. Likewise, in wild-type mice the concentration of Sia-rFH was measured to be 4.49 ± 0.44 µg/mL and in *cfh*$^{-/-}$ mice 4.99 ± 0.76 µg/mL. Importantly, a measurable increase in C3 plasma concentration was observed in the case of hFH administration only after 96 hours when, compared to DeGly-rFH and Sia-rFH, the concentration of hFH in blood plasma was much higher. Therefore, although DeGly-rFH and Sia-rFH were present in blood plasma for up to 48 hours (or longer in the case of Sia-rFH), these glycoforms may not persist long enough or at high enough a concentration to elicit a recovery of normal C3 plasma concentration.

6.4 Discussion

The binding affinities of hFH and DeGly-rFH have been assessed previously in this study (Chapter 2 section 2.3.5 table 2.2) and in the literature (Schmidt *et al*., 2008, Schmidt *et al*., 2011, Herbert *et al*., 2015 and, Kerr *et al*., 2017), where it was consistently shown that DeGly-rFH binds to C3b with greater affinity compared to hFH, and that DeGly-rFH binds to C3d with measurable affinity whereas hFH does not. But, previous studies (Kerr *et al*., 2017) also noted lower responses for hFH compared to DeGly-rFH. In particular, the low responses for C3d binding by hFH probably explain why this binding was, in previous work, considered too weak to measure K_D.

In addition, in earlier findings (Chapter 2 section 2.3.5 table 2.2) and those of Kerr *et al*. (2017) the R_{max} for DeGly-rFH binding to 150 RUs C3b was approximately 50% higher than for hFH whereas, in this study the R_{max} for DeGly-rFH is approximately four-times higher than hFH. This could be due to subtle batch-to-batch variations in the Biacore chip surface and reflect a general "stickiness" of the proteins that is very sensitive to such variations.

The different recombinant glycoforms of FH bound C3b and C3d with similar K_Ds and R_{max} values. Nonetheless, the data appears to show that there may be a relationship between the size of the glycans carried by rFH and binding affinity and estimated R_{max} for C3b and C3d. For example, Sia-rFH has a lower affinity and estimated R_{max} for binding C3b and C3d compared to Gal-rFH and DeGly-rFH, and Gal-rFH is intermediate between the other two. But, the experiment must be repeated several times to clarify if the slight observed differences are statistically significant.

However, this trend, that glycan size inversely correlates with FH affinity for C3b and C3d, was observed previously in Chapter 2 (section 2.3.5 table 2.2) for different glycoforms of hFH. In addition, Michelfelder *et al*. (2017) expressed and purified rFH with diantennary di-N-acetylglucosamine terminal complex-type glycans, $GlcNAc_2Man_3GlcNAc_2$, (GlcNAc-rFH) (as well as terminal mannose glycans) in glycoengineered moss. Michelfelder *et al*. (2017) showed that more GlcNAc-rFH bound immobilised C3b than hFH. Although a direct comparison cannot be drawn between the Biacore based C3b binding assay, reported here (table 6.2), and the ELISA conducted by Michelfelder *et al*. (2017), it is interesting to note that an independent study has demonstrated a link between reduced glycan length and increased C3b binding.

DAA assays comparing the four glycoforms of FH show that all recombinant glycoforms outperform hFH in DAA. However, the in-solution CA assays, conducted in this study, show that there is no difference in CA between Gal-, Sia-rFH and hFH and, possibly, DeGly-rFH. In regard to hFH and DeGly-rFH, the findings are consistent with the literature in which Schmidt *et al*. (2011) and Kerr *et al*. (2017) report that DeGly-rFH has enhanced DAA compared to hFH. In addition, Kerr *et al*. (2017) also show that there is no detectable difference in CA between DeGly-rFH and hFH.

Given that the DAA assay is carried out on a sensor-chip surface and C3b is randomly orientated on the chip surface, molecules of FH that can bind to both FH binding sites on C3b would be able to engage a larger proportion of C3bBb, resulting in a greater rate of turnover of the C3-convertase complex.

In comparison, CA is carried out in solution and therefore there is not the same ligand orientation constraint. Thus, in the DAA assay the ability of FH to bind C3bBb may be a limiting factor but, this may not be the case in the CA.

The glycoprofile of glycans released from Gal-rFH expressed in SuperGal *CFH* before and after enzymatic removal of yeast and hybrid-type glycosylation show that there is a significant contribution from these glycans to the overall glycoprofile of rFH expressed in SuperGal *CFH* and, that removal of these glycan species will leave some N-linked glycosylation sites unoccupied. Therefore, the main difference between hFH and Sia-rFH, which both have glycoprofiles composed of predominantly mono- and disialylated complex-type glycans, $Sia_1Gal_2GlcNAc_2Man_3GlcNAc_2$ and $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$ respectively, is that some of the N-linked glycosylation sites on Sia-rFH are unoccupied, whilst hFH potentially has all N-linked glycosylation sites occupied. In comparison to hFH, Sia- and Gal-rFH therefore have less glycans and may have higher conformational freedom allowing for a greater proportion of molecules to adopt the active, open conformation in the absence of self-surface markers. This would account for the higher C3b, C3d affinities and DAA of Sia- and Gal-rFH compared to hFH. If conformational freedom is the main factor influencing the difference in C3b, C3d binding affinities and DAA of the rFH glycoforms and hFH then it would be interesting to see how non-deglycosylated rFH functions in a comparative study. Non-deglycosylated rFH would have all glycosylation sites occupied, albeit with five mannose glycans, and, therefore, would presumably be under a similar conformational constraint as hFH. As both non-deglycosylated rFH and hFH would both

have fully occupied N-linked glycosylation sites, any differences observed between these two glycoforms would be due to the specific glycans the two proteins carry.

This study (fig. 6.9) showed that hFH has a longer circulatory half-life in plasma of wild-type mice compared to the recombinant glycoforms of FH. Additionally, there does not appear to be a difference in plasma concentrations of hFH, Sia-rFH and DeGly-rFH between wild-type and *cfh*$^{-/-}$ mice, as FH concentrations at 48 hours post-injection are consistent in the two strains of mice (compare figs. 6.9 and 6.11).

Potentially therapeutic FH or FH-like molecules have been developed and analysed *in vivo* in the literature (Nichols *et al*., 2015 and, Michelfelder *et al*., 2017). Nichols *et al*. (2015) developed a "mini-FH" which consisted of the complement regulatory domains CCPs 1-5 joined to the self-surface recognition domains CCPs 18-20. Although displaying CA and DAA comparable to hFH *in vitro*, *in vivo* this "mini-FH" was cleared rapidly from plasma of *cfh*$^{-/-}$ mice. Animals that were given a 0.5 milligram intraperitoneal injection of this "mini-FH" had a plasma concentration of 200 μg/mL at a time point two hours post-injection which then fell dramatically to become almost undetectable six hours post-injection. Intraperitoneal injections of 0.5 milligram glycoengineered moss-produced GlcNAc-rFH into *cfh*$^{-/-}$ mice (Michelfelder *et al*., 2017) had an equally rapid clearance rate – ~420 μg/mL GlcNAc-rFH was measured in plasma at ten minutes post-injection but, was not detectable 6 hours post-injection. Therefore, although the recombinant glycoforms studied here (fig 6.9 and 6.11) appear to be cleared more rapidly than hFH, they do appear to persist longer in circulation than comparative amounts of potentially therapeutic FH and FH-like molecules developed by others.

As discussed in Chapter 1 section 1.1.3, glycoproteins carrying sialylated glycans have prolonged circulatory half-life (Morell *et al*., 1971) because sialic acids mask glycoproteins preventing clearance by the asialoglycoprotein receptor (Ashwell and Harford, 1982, Schwartz, 1984 and, Fukuda *et al*., 1989). It is intriguing, therefore, that Sia-rFH does not have an extended half-life compared to Gal- and DeGly-rFH (table 6.4). As previously discussed, the unoccupied glycosylation sites present on Sia-rFH, because of hybrid- and yeast-type glycan enzymatic removal by Endo H_f, may present sites for binding by antibodies and clearance receptors.

Normal plasma C3 levels in *CFH*$^{+/+}$ mice are ~450 – 500 µg/mL and *cfh* $^{-/-}$ mice have plasma C3 levels ≤ 30 µg/mL (Pickering *et al*., 2002). Intraperitoneal injections of hFH to *cfh*$^{-/-}$ mice (fig. 6.10) result in a rise in plasma C3 levels, reaching a high of ~200 µg/mL at 96 hours after a second intraperitoneal injection of hFH at 48 hours. Although this is not a full restoration of wild-type plasma C3 levels, the rise in C3 levels in hFH injected *cfh*$^{-/-}$ mice is consistent with findings in the literature for mice administered with murine FH (Paixão-Cavalcante *et al*., 2009) and hFH (Fakhouri *et al*., 2010, Nichols *et al*., 2015 and, Michelfelder *et al*., 2017). It is significantly different from the responses of the three recombinant glycoforms of FH, which yielded no change in measurable plasma C3 levels in *cfh*$^{-/-}$ mice.

In contrast to the possible inability of the rFH glycoforms to increase plasma C3 levels observed here (fig. 6.10), the potentially therapeutic GlcNAc-rFH expressed in glycoengineered moss showed an increase in plasma C3 levels, in mice administered with a single 0.5 milligram intraperitoneal injection of GlcNAc-rFH. This peaked six hours post-injection (~200 µg/mL GlcNAc-rFH compared to ~450 µg/mL hFH) and fell slightly, but remained higher than PBS-treated mice, by 24 hours post-injection (Michelfelder *et al*., 2017). Additionally, potentially therapeutic "mini-FH", composed of CCPs 1-5 linked to 18-20, also showed an increase in plasma C3 levels six hours after a single injection of 0.5 milligram of "mini-FH" (~75 µg/mL compared to ~100 µg/mL hFH) however, the assay was not continued past six hours (Nichols *et al*., 2015). Both studies (Nichols *et al*., 2015 and, Michelfelder *et al*., 2017) assessed plasma C3 levels within a shorter time frame than studied here (fig. 6.10) (6-24 hours post-injection compared to 48 hours post-injection). Considering this, and knowing that the rFH glycoforms have a reduced half-life compared to hFH, it cannot be excluded that the rFH glycoforms may influence plasma C3 levels within the 48 hours post-injection.

All literature studies referenced here (Paixão-Cavalcante *et al*., 2009, Fakhouri *et al*., 2010, Nichols *et al*., 2015 and, Michelfelder *et al*., 2017), in addition to measuring the change in plasma C3 levels, also evaluated kidney sections for C3 staining of the glomerular basement membrane (GBM). However, GBM C3 staining was not measured in this study.

The opsonin C3b binds to surfaces via an exposed thioester in the TED domain. FI mediated cleavage of C3b converts surface-associated C3b to iC3b, further cleavage events produce C3dg and eventually C3d which remains tethered to the surface to facilitate an adaptive immune response.

It was observed here, and elsewhere (Herbert *et al*., 2015 and, Kerr *et al*., 2017), that the recombinant glycoforms of FH have a measurable affinity for C3d (table 6.2) but hFH binds C3d poorly. Considering this, and the fact that C3d persists on surfaces, a possible reason for the reduced plasma FH concentrations of rFH glycoforms observed here (figs. 6.9 and 6.11), compared to hFH, is that the rFH glycoforms are pulled out of solution by binding to surface associated C3d. Analysis of FH staining of kidney sections, coupled with C3 and C3d staining, would clarify this hypothesis.

Ultimately, the data presented here shows that glycoengineered Sia-rFH has an enhanced ability to bind C3b, a measurable affinity for C3d and enhanced DAA compared to hFH. However, glycoengineering and sialylating the glycans carried by rFH does not appear to improve the *in vivo* half-life or function compared to DeGly-rFH and this may be due to unoccupied glycosylation sites on Gal-rFH and Sia-rFH. Future work would seek to improve the glycan site occupancy on rFH by complex-type glycans.

Chapter 7

Future perspectives

7.1 Glycosidic linkage characterisation

Although the characterisation of the glycosylation of hFH has been reported in several publications (Ritchie *et al*., 2002, Fenaille *et al*., 2007*a*, Benicky *et al*., 2014, and, Pompach *et al*., 2014), the work of Fenaille *et al*. (2007*a*) remains the most detailed assessment to date. Fenaille *et al*. (2007*a*) use a multiplex approach to quantify the glycoprofile of hFH and characterise the glycans at each N-linked glycosylation site. However, the types of glycosidic bonds that link the glycan monosaccharide constituents are not currently known.

The glycosidic bond that links two monosaccharides is known to influence conformational entropy (Striegel and Boone, 2011), viscosity (Kath *et al*., 1999) and solubility (Best *et al*. 2001). The different biophysical properties of glycosidic linkages affect glycoconjugate function. For example, differential infectivity and species discrimination of human and animal influenza viruses is mediated by the type of sialic acid glycosidic linkage present on glycoconjugates in the epithelia of the upper respiratory tract (Rogers and Paulson, 1983 and, Shinya *et al*., 2006). Another example is that α(2,6)-linked sialic acid, but not α(2,3)-linked sialic acid, protects T-cells from galectin-1 induced cell death (Toscano *et al*., 2007).

Different tissues are known to carry different proportions of α(2,3)- and α(2,6)-linked sialic acid. For example, Baum and Paulson (1990) were able to show that human ciliated tracheal epithelial cells carry α(2,6)-sialic acid whereas goblet cell secreted mucus carries exclusively α(2,3)-sialic acid.

Differential sialic acid glycosidic linkages have even been observed on the level of glycoproteins. By exploiting the glycosidic linkage specificity of sialic acid binding lectins SNA – which has specificity for α(2,6)-sialic acid – and *Maakia amurensis* lectin – which has specificity for α(2,3)sialic acid – Heerze and Armstrong (1990) have shown that human fibrinogen, haptoglobin and transferrin is exclusively sialylated with α(2,6)-sialic acid whilst fetuin carries sialic acid with both α(2,6)- and α(2,3)-glycosidic linkages.

Glycosidic linkage analysis can be achieved by gas chromatography-mass spectrometry (Lv, *et al*., 2015) as well as by exploiting the glycosidic-linkage specificities of different lectins. Applying these methodologies to hFH could elucidate the terminal sialic acid glycosidic linkage. Given that the sialic acid linkages on hFH are not known, this raises some interesting questions: are the sialic acid linkages all one type, e.g. α(2,6) or α(2,3), and if so, why? Or are they a mixture of both? Answers to these questions have practical relevance for the use of recombinant hFH as a potential molecular replacement therapeutic and would better inform the choice of sialyltransferase used to cap rFH.

7.2 Quantification and site-specific glycan functional characterisation

Absent in this study was glycan quantification and site-specific glycan characterisation. Glycan quantification can be achieved by high-performance capillary electrophoresis with laser-induced fluorescence detection, as used by Fenaille *et al*. (2007*a*) to quantify the glycosylation of hFH, or by high-performance liquid chromatography (HPLC) (Roth *et al*., 2012). In Chapter 3 of this study, the glycoprofile of rFH expressed in glycoengineered strain SuperGal was shown to be heterogenous and included hybrid- and yeast-type glycans as well as the target humanised complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$. Anticipating that the yeast- and hybrid-glycans would affect *in vivo* half-life, these glycan-types were removed enzymatically with the yeast- and hybrid-glycan specific endoglycosidase Endo H_f. The removal of yeast- and hybrid-type glycans was qualified by MALDI-TOF mass spectrometry but not quantified. Therefore, it is not known exactly what proportion of the glycan content of FH is missing and how many glycosylation sites are unoccupied.

Site-specific glycan characterisation has been carried out in the literature by tryptic digest coupled with liquid-chromatography electrospray ionisation tandem mass spectrometry (MS/MS) and MALDI-TOF MS/MS (Fenaille, *et al*., 2007*a* and Pompach *et al*., 2014). Fenaille *et al*. (2007*a*) found that the complexity of the glycans on hFH decreases toward the C-terminus. The glycosylation site closest to the N-terminus, located on CCP9, had 40% triantennary trisialylated species occupancy and no monosialylated diantennary species. In comparison, the glycosylation site closest to the C-terminus had just 3% occupancy by triantennary trisialylated species whilst the diantennary monosialylated species accounted for 20% of the glycan occupancy at this site. In addition, the glycosylation site on CCP13

had the highest site occupancy (50%) of the diantennary disialylated species. Fenaille *et al*. (2007*a*) hypothesised that the folded back, compact structure predicted for FH (Aslam and Perkins, 2001 and, Oppermann *et al*., 2006) reduces accessibility of glycosyltransferases to the central, compact region of FH. Thus, there is clear site-specific glycan variance on hFH that is potentially caused by differential glycosyltransferase accessibility. If this is the case with hFH, is it also the case with rFH expressed in glycoengineered strain SuperGal? Do certain N-linked glycosylation sites on rFH expressed in SuperGal have a higher site occupancy by the target complex-type glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$, and do others have a higher occupancy by yeast-type and hybrid-type glycans? Site-specific glycan characterisation of rFH expressed in SuperGal would help answer these questions.

7.3 Exposed N-linked glycosylation sites

We observed an accelerated clearance of Sia- and Gal-rFH compared to hFH in mice. This likely contributed to the lack of recovery of plasma C3 levels. However, what was not tested *in vivo* was SuperGal-expressed rFH that had not been treated with glycosidase to selectively remove "unwanted" hybrid- and yeast-type glycans. As discussed in chapter 6, the accelerated clearance of Sia-rFH and, in part, Gal-rFH is likely due to the single GlcNAc left on each glycosylation site, due to Endo H_f cleavage of hybrid- and yeast-type glycans, and exposure of potentially immunogenic sites that would otherwise be masked by the presence of a glycan at that site (Skehel *et* al., 1984 and, Sun *et al*., 2012). Therefore, there are two parallel clearance pathways to consider: on the one hand the mannose-binding lectin clearance mechanism will be at work and, on the other hand, an antibody based clearance mechanism is likely to operate. Both of these mechanisms are likely to be responsible for the accelerated clearance observed for Sia-rFH and Gal-rFH. However, in theory, SuperGal-expressed and sialylated rFH not deglycosylated with Endo H_f should have all N-linked glycosylation sites occupied and, therefore, will not be cleared by the second antibody-based clearance mechanism. It would be interesting to test what effect this has on clearance rate?

7.4 Optimisation of glycoengineered *P. pastoris* strain SuperGal

As discussed, the yeast- and hybrid-type glycans thought to be present on rFH expressed in SuperGal would accelerate clearance and reduce *in vivo* half-life due to interaction with the

asialoglycoprotein receptor and mannose-binding lectins. It was shown in chapter 6 of this study that enzymatic removal of yeast- and hybrid-type glycans does not significantly improve *in vivo* half-life and, as speculated in the previous section, might even reduce it. Therefore, an alternative strategy to improve *in vivo* half-life is to increase complex-type N-linked glycan glycosylation site occupancy.

Different clones of *P. pastoris* can have markedly different heterologous protein expression yields (Hu *et al*., 2011) and therefore it is necessary to characterise the expression of a large number of clones to identify those with highest expression. This variation is due in part to the ability of *P. pastoris* to undergo random gene duplication, creating multiple copies of the heterologous gene (Higgins and Cregg, 1998). Variation could be assessed by taking advantage of advances in flow cytometry, in combination with glycan motif lectin recognition, which has allowed for cell surface (Batisse *et al*., 2004) and site-specific cell surface glycoprotein (Jayakumar *et al*., 2009) glycoprofiles to be monitored. Mannoproteins are heavily glycosylated components of the outer cell wall of yeasts with the glycan component extending out into the culture media (Klis *et al*., 2006). Cell wall mannoproteins of *P. pastoris* glycoengineered with the GlycoSwitch technology were used to characterise clones that had undergone a modification to the glycan biosynthetic pathway (Jacobs *et al*., 2009).

Considering that cell wall mannoproteins can be used as a marker for successfully glycoengineered clones of SuperGal, it may be possible, using cell sorting flow cytometry with the galactose binding lectin ECL, to exploit the natural clonal variation in expression levels of heterologous proteins to select for clones of SuperGal that are more efficient at complex-type diantennary digalactosylated glycan, $Gal_2GlcNAc_2Man_3GlcNAc_2$, biosynthesis. In this way, it may be possible to express rFH in SuperGal that has a much higher complex-type N-linked glycan site occupancy and would potentially abrogate the need to remove non-complex-type glycan.

Bibliography

Ahn, J., Hong, J., Lee, H., Park, M., Lee, E., Kim, C., Choi, E., Jung, J., & Lee, H. (2007). Translation elongation factor 1-α gene from Pichia pastoris: Molecular cloning, sequence, and use of its promoter. *Applied Microbiology and Biotechnology*, *74*(3), 601–608.

Alper, H., Fischer, C., Nevoigt, E., & Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences*, *102*(36), 12678–12683.

Appel, G. B., Cook, H. T., Hageman, G., Jennette, J. C., Kashgarian, M., Kirschfink, M., Lambris, J. D., Lanning, L., Lutz, H. U., Meri, S., Rose, N. R., Salant, D. J., Sathi, S., Smith, R. J. H., Smoyer, W., Tully, H. F., Tully, S. P., Walker, P., Welsh, M., Würzner, R., & Zipfel, P. F. (2005). Membranoproliferative glomerulonephritis type II (dense deposit disease): an update. *Journal of the American Society of Nephrology: JASN*, *16*(5), 1392–403.

Ashwell, G., & Harford, J. (1982). Carbohydrate-specific receptors of the liver. *Annu. Rev. Biochem. 51*, 531–554.

Aslam, M., & Perkins, S. J. (2001). Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. *Journal of Molecular Biology*, *309*(5), 1117–1138.

Ault, B. H., Schmidt, B. Z., Fowler, N. L., Kashtan, C. E., Ahmed, A. E., Vogt, B. A., & Colten, H. R. (1997). Human factor H deficiency: Mutations in framework cysteine residues and block in H protein secretion and intracellular catabolism. *Journal of Biological Chemistry*, *272*(40), 25168–25175.

Baenziger, J. U., & Fiete, D. (1979). Structure of the complex oligosaccharides of fetuin. *Journal of Biological Chemistry*, *254*(3), 789–795.

Bailey, L. A., Hatton, D., Field, R., & Dickson, A. J. (2012). Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnology and Bioengineering*, *109*(8), 2093–2103.

Bardor, M., Nguyen, D. H., Diaz, S., & Varki, A. (2005). Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. *Journal of Biological Chemistry*, *280*(6), 4228–4237.

Barlow, P. N., Baron, M., Norman, D. G., Day, A. J., Willis, A. C., Sim, R. B., & Campbell, I. D. (1991). Secondary structure of a complement control protein module by two-dimensional 1H NMR. *Biochemistry*, *30*(4), 997-1004.

Barlow, P. N., Norman, D. G., Steinkasserer, A., Horne, T. J., Pearce, J., Driscoll, P. C., Sim, R. B., & Campbell, I. D. (1992). Solution structure of the fifth repeat of factor H: a second example of the complement control protein module. *Biochemistry*, *31*(14), 3626-3634.

Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim R. B., & Campbell, I. D. (1993). Soltuion structure of a pair of complement modules by nuclear magnetic resonance. *J. Mol. Biol.*, *232*(1), 268-284.

Batisse, C., Marquet, J., Greffard, A., Fleury-Feith, J., Jaurand, M. C., & Pilatte, Y. (2004). Lectin-based three-color flow cytometric approach for studying cell surface glycosylation changes that occur during apoptosis. *Cytometry Part A*, *62*(2), 81–88.

Baum, L. G., & Paulson, J. C. (1990). Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity. *Acta Histochem Suppl*, *40*, 35–38.

Benicky, J., Sanda, M., Pompach, P., Wu, J., & Goldman, R. (2014). Quantification of fucosylated hemopexin and complement factor H in plasma of patients with liver disease. *Analytical Chemistry*, *86*(21), 10716–10723.

Best, R. B., Jackson, G. E., & Naidoo, K. J. (2001). Molecular dynamics and NMR study of the α(1→4) and α(1→6) glycosidic linkages: maltose and isomaltose. *Journal of Physical Chemistry B*, *105*(20), 4742–4751.

Blackmore, T. K., Sadlon, T. a, Ward, H. M., Lublin, D. M., & Gordon, D. L. (1996). Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H. *Journal of Immunology*, *157*(12), 5422–7.

Blackmore, T. K., Hellwage, J., Sadlon, T. a, Higgs, N., Zipfel, P. F., Ward, H. M., & Gordon, D. L. (1998). Identification of the second heparin-binding domain in human complement factor H. *Journal of Immunology*, *160*(7), 3342–3348.

Blaum, B. S., Hannan, J. P., Herbert, A. P., Kavanagh, D., Uhrín, D., & Stehle, T. (2015). Structural basis for sialic acid-mediated self-recognition by complement factor H. *Nature Chemical Biology*, *11*(1), 77–82.

Bobrowicz, P., Davidson, R. C., Li, H., Potgieter, T. I., Nett, J. H., Hamilton, S. R., Stadheim, S. R., Miele, R. G., Bobrowicz, P., Mitchell, T., Rausch, S., Renfer, E., & Wildt, S. (2004). Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast *Pichia pastoris*: Production of complex humanized glycoproteins with terminal galactose. *Glycobiology*, *14*(9), 757–766.

Bosques, C. J., Collins, B. E., Meador, J. W., Sarvaiya, H., Murphy, J. L., Dellorusso, G., Bulik, D. A., Hsu, I. H., Washburn, N., Sipsey, S. F., Meyette, J. R., Raman, R., Shriver, Z., Sasisekharan, R., & Venkataraman, G. (2010). Chinese hamster ovary cells can produce galactose-α-1,3-galactose antigens on proteins. *Nature Biotechnology*.

Brooimans, R. A., Ark, A. A. van der, Buurman, W. A., Es, L. A. van, & Daha, M. R. (1990). Differential regulation of complement factor H and C3 production in human umbilical vein endothelial cells by IFN-gamma and IL-1. *The Journal of Immunology*, *144*(10), 3835–3840.

Büll, C., Heise, T., Adema, G. J., & Boltje, T. J. (2016). Sialic Acid Mimetics to Target the Sialic Acid-Siglec Axis. *Trends in Biochemical Sciences*.

Burns, N. (1993). Genetic Stability: An issue of product quality. *Biologicals*, *21*, 145–146.

Büttner-Mainik, A., Parsons, J., Jérôme, H., Hartmann, A., Lamer, S., Schaaf, A., Schlosser, A., Zipfel, P. F., Reski, R., & Decker, E. L. (2011). Production of biologically active recombinant human factor H in Physcomitrella. *Plant Biotechnology Journal*, *9*(3), 373–383.

Caffaro, C. E., & Hirschberg, C. B. (2006). Nucleotide sugar transporters of the Golgi apparatus: From basic science to diseases. *Accounts of Chemical Research*, *39*(11), 805–812.

Callewaert, N., Laroy, W., Cadirgi, H., Geysens, S., Saelens, X., Min Jou, W., & Contreras, R. (2001). Use of HDEL-tagged *Trichoderma reesei* mannosyl oligosaccharide 1,2-α-D-mannosidase for N-glycan engineering in *Pichia pastoris*. *FEBS Letters*, *503*(2–3), 173–178.

Ceciliani, F., Pocacqua, V., Lecchi, C., Fortin, R., Rebucci, R., Avallone, G., Bronzo, F., Cheli, F., & Sartorelli, P. (2007). Differential expression and secretion of alpha1-acid glycoprotein in bovine milk. *J.Dairy Res.*, *74*, 374–380.

Cereghino, J. L., & Cregg, J. M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiology Reviews*, *24*(1), 45–66.

Chapman, A., Trowbridge, I. S., Hymant, R., & Kornfeld, S. (1979). Structure of the lipid-linked oligosaccharides that accumulate in class E thy-1-negative mutant lymphomas. *Cell*, *17*, 509–515.

Che, A., Knight, T., Canton, B., Kelly, J., & Shetty, R. (2010). Method for Assembly of Polynucleic Acid Sequences. *World Intellectual Property Organization*.

Chen, M., Forrester, J. V., & Xu, H. (2007). Synthesis of complement factor H by retinal pigment epithelial cells is down-regulated by oxidized photoreceptor outer segments. *Experimental Eye Research*, *84*(4), 635–645.

Chen, M.-T., Lin, S., Shandil, I., Andrews, D., Stadheim, T. A., & Choi, B.-K. (2012). Generation of diploid *Pichia pastoris* strains by mating and their application for recombinant protein production. *Microbial Cell Factories*, *11*(1), 91.

Chen, R., Pawlicki, M. A., Hamilton, B. S., & Tolbert, T. J. (2008). Enzyme-catalyzed synthesis of a hybrid N-linked oligosaccharide using N-acetylglucosaminyltransferase I. *Advanced Synthesis and Catalysis*, *350*(11–12), 1689–1695.

Chen, X., Liu, Z., Wang, J., Fang, J., Fan, H., & Wang, P. G. (2000). Changing the donor cofactor of bovine α1,3-galactosyltransferase by fusion with UDP-galactose 4-epimerase. More efficient biocatalysis for synthesis of alpha-Gal epitopes. *Journal of Biological Chemistry*, *275*(41), 31594–31600.

Chen, Z. A., Pellarin, R., Fischer, L., Sali, A., Nilges, M., Barlow, P. N., & Rappsilber, J. Structure of complement $C3_{(H2O)}$ revealed by quantitative cross-linking/mass spectrometry and modelling. *Molecular* & Cellular Proteomics, *15*(8).

Cheng, J., Huang, S., Yu, H., Li, Y., Lau, K., & Chen, X. (2009). Trans-sialidase activity of *Photobacterium damsela* α2,6-sialyltransferase and its application in the synthesis of sialosides. *Glycobiology*, *20*(2), 260–268.

Chiba, Y., Suzuki, M., Yoshida, S., Yoshida, A., Ikenaga, H., Takeuchi, M., Jigami, Y., & Ichishima, E. (1998). Production of human compatible high mannose-type (Man5GlcNAc2) sugar chains in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *273*(41), 26298–26304.

Choi, B.-K., Bobrowicz, P., Davidson, R. C., Hamilton, S. R., Kung, D. H., Li, H., Miele, R. G., Nett, J. H., Wildt, S., & Gerngross, T. U. (2003). Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast *Pichia pastoris*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5022–7.

Clark, S. J., Higman, V. A., Mulloy, B., Perkins, S. J., Lea, S. M., Sim, R. B., & Day, A. J. (2006). His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form. *Journal of Biological Chemistry*, *281*(34), 24713–24720.

Clark, S. J., Perveen, R., Hakobyan, S., Morgan, B. P., Sim, R. B., Bishop, P. N., & Day, A. J. (2010). Impaired binding of the age-related macular degeneration-associated complement factor H 402H allotype to Bruch's membrane in human retina. *Journal of Biological Chemistry*, *285*(39), 30192–30202.

Clark, S. J., Ridge, L. A., Herbert, A. P., Hakobyan, S., Mulloy, B., Lennon, R., Wurzner, R., Morgan, B. P., Uhrin, D., P. N. Bishop, & Day, A. J. (2013). Tissue-Specific Host Recognition by Complement Factor H Is Mediated by Differential Activities of Its Glycosaminoglycan-Binding Regions. *The Journal of Immunology*, *190*(5), 2049–2057.

Collins, B. E., Blixt, O., DeSieno, A. R., Bovin, N., Marth, J. D., & Paulson, J. C. (2004). Masking of CD22 by cis ligands does not prevent redistribution of CD22 to sites of cell contact. *Proceedings of the National Academy of Sciences*, *101*(16), 6104–6109.

Couderc, R., & Baratti, J. (1980). Oxidation of Methanol by the Yeast, *Pichia pastoris*. Purification and Properties of the Alcohol Oxidase. *Agric. Biol. Chem.*, *44*(10), 2279–2289.

Cregg, J. M., Barringer, K. J., Hessler, A. Y., & Madden, K. R. (1985). *Pichia pastoris* as a host system for transformations. *Molecular and Cellular Biology*, *5*(12), 3376–3385.

Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P., & Stillman, C. A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cellular Biology*, *9*(3), 1316–23.

Crocker, P. R. (2002). Siglecs: Sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Current Opinion in Structural Biology, 12*(5), 609-615.

Cukan, M. C., Hopkins, D., Burnina, I., Button, M., Giaccone, E., Houston-Cummings, N. R., Jiang, Y., Li, F., Mallem, M., Mitchell, T., Moore, R., Nylen, A., Prinz, B., Rios, S., Sharkey, N., Zha, D., Hamilton, S., Li, H., & Stadheim, T. A. (2012). Binding of DC-SIGN

to glycoproteins expressed in glycoengineered *Pichia pastoris*. *Journal of Immunological Methods*, *386*(1–2), 34–42.

Cunningham, L. W., Nuenke, J. & Strayhorn, W. D. (1957). Sulfhydryl content and tryptic susceptibility of thermally denatured ovalbumin. *The Journal of Biological Chemistry*, *228*, 835–846.

Curran, K. A., Karim, A. S., Gupta, A., & Alper, H. S. (2013). Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering*, *19*, 88–97.

Decker, E. L., Parsons, J., & Reski, R. (2014). Glyco-engineering for biopharmaceutical production in moss bioreactors. *Front. Plant Sci.*, *5*, 346.

De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., van de Peer, Y., & Callewaert, N. (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology*, *27*(6), 561–566.

Debray, H., Montreuil, J., Lis, H., & Sharon, N. (1986). Affinity of four immobilized *Erythrina* lectins toward various n-linked glycopeptides and related oligosaccharides. *Carbohydrate Research*, *151*, 359–370.

Delorme, E., Lorenzini, T., Giffin, J., Martin, F., Jacobsen, F., Boone, T., & Elliott, S. (1992). Role of Glycosylation on the Secretion and Biological Activity of Erythropoietin. *Biochemistry*, *31*(41), 9871–9876.

Devine, D. V, & Rosse, W. F. (1987). Regulation of the activity of platelet-bound C3 convertase of the alternative pathway of complement by platelet factor H. *Proceedings of the National Academy of Sciences of the United States of America*, *84*(16), 5873–7.

Ding, L., Yu, H., Lau, K., Li, Y., Muthana, S., Wang, J., & Chen, X. (2011). Efficient chemoenzymatic synthesis of sialyl Tn-antigens and derivatives. *Chemical Communications*, *47*(30), 8691–3.

Ding, N., Yang, C., Sun, S., Han, L., Ruan, Y., Guo, L., Hu, X., & Zhang, J. (2017). Increased glycosylation efficiency of recombinant proteins in *Escherichia coli* by auto-induction. *Biochemical and Biophysical Research Communications*, *485*(1), 138–143.

Dissing-Olesen, L., Thaysen-Andersen, M., Meldgaard, M., Hojrup, P., & Finsen, B. (2008). The function of the human interferon-beta 1a glycan determined in vivo. *J Pharmacol Exp Ther*, *326*(1), 338–347.

Drocourt, D., Calmels, T., Reynes, J. P., Baron, M., & Tiraby, G. (1990). Cassettes of the *Streptoalloteichus hindustanus* ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. *Nucleic Acids Research*, *18*(13), 4009.

Duarte-Vázquez, M. A., García-Almendárez, B. E., Rojo-Domínguez, A., Whitaker, J. R., Arroyave-Hernández, C., & Regalado, C. (2003). Monosaccharide composition and

properties of a deglycosylated turnip peroxidase isozyme. *Phytochemistry*, *62*(1), 5–11.

Dunkelberger, J. R., & Song, W-C. (2010). Complement and its role in innate and adaptive immune responses. *Nature Cell Research*, *20*, 34-50.

Edwards, A. O., Ritter, R., Abel, K. J., Manning, A., Panhuysen, C., & Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science*, *308*(5720), 421–4.

Ezekowitz, R. A., Sastry, K., Bailly, P., & Warner, A. (1990). Molecular characterization of the human macrophage mannose receptor: demonstration of multiple carbohydrate recognition-like domains and phagocytosis of yeasts in Cos-1 cells. *The Journal of Experimental Medicine*, *172*(6), 1785–94.

Fakhouri, F., De Jorge, E. G., Brune, F., Azam, P., Cook, H. T., & Pickering, M. C. (2010). Treatment with human complement factor H rapidly reverses renal complement deposition in factor H-deficient mice. *Kidney International*, *78*(3), 279–286.

Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., & Aebi, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *102*(8), 3016–3021.

Fenaille, F., Le Mignon, M., Groseil, C., Ramon, C., Riandé, S., Siret, L., & Bihoreau, N. (2007*a*). Site-specific N-glycan characterization of human complement factor H. *Glycobiology*, *17*(9), 932–944.

Fenaille, F., le Mignon, M., Groseil, C., Siret, L., Bihoreau, N. (2007*b*). Combined use of 2,4,6-trihydroxyacetophenone as matrix and enzymatic deglycosylation in organic-aqueous solvent systems for the simultaneous characterization of complex glycoproteins and N-glycans by matric-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *21*, 812–816.

Fenouillet, E., Fayet, G., Hovsepian, S., Bahraoui, E. M., & Ronin, C. (1986). Immunochemical evidence for a role of complex carbohydrate chains in thyroglobulin antigenicity. *Journal of Biological Chemistry*, *261*(32), 15153–15158.

Ferreira, V. P., Herbert, A. P., Hocking, H. G., Barlow, P. N., & Pangburn, M. K. (2006). Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. *Journal of Immunology*, *177*(9), 6308–6316.

Ferreira, V. P., Herbert, A. P., Cortés, C., McKee, K. A., Blaum, B. S., Esswein, S. T., Uhrín, D., Barlow, P. N., Pangburn, M. K., & Kavanagh, D. (2009). The binding of factor H to a complex of physiological polyanions and C3b on cells is impaired in atypical hemolytic uremic syndrome. *Journal of Immunology*, *182*(11), 7009–18.

Fukuda, M. N., Sasaki, H., Lopez, L., & Fukuda, M. (1989). Survival of recombinant erythropoietin in the circulation: the role of carbohydrates. *Blood*, *73*(1), 84–9.

Furukawa, J. I., Shinohara, Y., Kuramoto, H., Miura, Y., Shimaoka, H., Kurogochi, M., Nakano, M., & Nishimura, S. I. (2008). Comprehensive approach to structural and functional glycomics based on chemoselective glycoblotting and sequential tag conversion. *Analytical Chemistry*, *80*(4), 1094–1101.

Ghaderi, D., Taylor, R. E., Padler-Karavani, V., Diaz, S., & Varki, A. (2010). Implications of the presence of N-glycolylneuraminic acid in recombinant therapeutic glycoproteins. *Nature Biotechnology*, *28*(8), 863–867.

Giannakis, E., Jokiranta, T. S., Male, D. A., Ranganathan, S., Ormsby, R. J., Fischetti, V. A., Mold, C., & Gordon, D. L. (2003). A common site within factor H SCR 7 responsible for binding heparin, C-reactive protein and streptococcal M protein. *European Journal of Immunology*. *33*(4), 962-969.

Goldstein, I. J., Reichert, C. M., Misaki, A., & Gorin, P. A. J. (1973). An "extension" of the carbohydrate binding specificity of concanavalin A. *BBA - Protein Structure*, *317*(2), 500–504.

Gomathinayagam, S., Hoyt, E., Thompson, A. M., Brown, E., Karaveg, K., Hamilton, S. R., & Li, H. (2012). High-throughput multimodal strong anion exchange purification and N-glycan characterization of endogenous glycoprotein expressed in glycoengineered *Pichia pastoris*. *Methods in Molecular Biology*, *899*, 315–323.

Gomathinayagam, S., Laface, D., Houston-Cummings, N. R., Mangadu, R., Moore, R., Shandil, I., Sharkey, N., Li, H., Stadheim, T. A., & Zha, D. (2015). In vivo anti-tumor efficacy of afucosylated anti-CS1 monoclonal antibody produced in glycoengineered *Pichia pastoris*. *Journal of Biotechnology*, *208*, 13–21.

Gong, B., Burnina, I., Stadheim, T. A., & Li, H. (2013). Glycosylation characterization of recombinant human erythropoietin produced in glycoengineered *Pichia pastoris* by mass spectrometry. *Journal of Mass Spectrometry*, *48*(12), 1308–1317.

Gong, X., Ding, C., Liu, L., & Wu, J. (2013*b*). Enhancement of human insulin precursor production by increasing the copy number in *Pichia pastoris*. *Wei Sheng Wu Xue Bao*, *53*(6), 545–552.

Gordon, D. L., Kaufman, R. M., Blackmore, T. K., Kwong, J., & Lublin, D. M. (1995). Identification of complement regulatory domains in human factor H. *Journal of Immunology*, *155*(1), 348 – 356.

Green, E. D., Adelt, G., Baenziger, J. U., Wilson, S., & Van Halbeek, H. (1988). The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy. *Journal of Biological Chemistry*, *263*(34), 18253–18268.

Ha, S., Wang, Y., & Rustandi, R. R. (2011). Biochemical and biophysical characterization of humanized IgG1 produced in *Pichia pastoris*. *mAbs*, *3*(5), 453-460.

Hageman, G. S., Anderson, D. H., Johnson, L. V, Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Brochardt, J. D., Gehrs, K. M., Smith, R. J. H., Silvestri, J., Russell, S. R., Klaver, C. C. W., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R.,

Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, J., Zernant, J., Merriam, J. E., Gold, B., Dean, M., & Allikmets, R. (2005). A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(20), 7227–32.

Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schnetz-Boutaud, N., Agarwal, A., Postel, E. A., & Pericak-Vance, M. A. (2005). Complement factor H variant increases the risk of age-related macular degeneration. *Science*, *308*(5720), 419–21.

Hakobyan, S., Harris, C. L., Tortajada, A., Dejorge, E. G., Garcia-Layana, A., Fernandez-Robredo, P., de Cordoba, S. R., & Paul Morgan, B. (2008). Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: Application to assessing risk of age-related macular degeneration. *Investigative Ophthalmology and Visual Science*, *49*(5), 1983–1990.

Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., Nett, J. H., Rausch, S., Stadheim, T. A., Wischnewski, H., Wildt, S., & Gerngross, T. U. (2003). Production of complex human glycoproteins in yeast. *Science*, *301*(5637), 1244–6.

Hamilton, S. R., Davidson, R. C., Sethuraman, N., Nett, J. H. Jiang, Y., Rios, S., Bobrowicz, P., Stadheim, T. A., Li, H., Choi, B. K., Hopkins, D., Wischnewski, H., Roser, J., Mitechell, T., Strawbridge, R. R., Hoopes, J., Wildt, S., & Gerngross, T. U. (2006). Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science*, *313*(5792), 1441-1443.

Hamilton, S. R., Cook, W. J., Gomathinayagam, S., Burnina, I., Bukowski, J., Hopkins, D., Schwartz, S., Du, M., Sharkey, N. J., Bobrowicz, P., Wildt, S., Li, H., Stadheim, T. A., & Nett, J. H. (2013). Production of sialylated O-linked glycans in *Pichia pastoris*. *Glycobiology*, *23*(10), 1192–1203.

Hartnell, A., Steel, J., Turley, H., Jones, M., Jackson, D. G., & Crocker, P. R. (2001). Characterization of human sialoadhesin, a sialic acid binding receptor expressed by resident and inflammatory macrophage populations. *Blood*, *97*(1), 288–296.

Hartner, F. S., Ruth, C., Langenegger, D., Johnson, S. N., Hyka, P., Lin-Cereghino, G. P., Lin-Cereghino, J., Kovar, K., Cregg, J. M., & Glieder, A. (2008). Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Research*, *36*(12), e76.

Hebecker, M., Alba-Dominguez, M., Roumenina, L. T., Reuter, S., Hyvarinen, S., Dragon-Durey, M.-A., Jokiranta, T. A., Sanchez-Corral, P., & Jozsi, M. (2013). An Engineered Construct Combining Complement Regulatory and Surface-Recognition Domains Represents a Minimal-Size Functional Factor H. *The Journal of Immunology*, *191*(2), 912–921.

Hebert, D. N., Foellmer, B., & Helenius, a. (1996). Calnexin and calreticulin promote folding, delay oligomerization and suppress degradation of influenza hemagglutinin in microsomes. *The EMBO Journal*, *15*(12), 2961–2968.

Heerze, L. D., & Armstrong, G. D. (1990). Comparison of the lectin-like activity of pertussis toxin with two plant lectins that have differential specificities for α(2-6) and α(2-3)-linked sialic acid. *Biochemical and Biophysical Research Communications*, *172*(3), 1224–1229.

Hegasy, G. A., Willhoeft, U., Majno, S. A., Seeberger, H., Zipfel, P. F., & Hellwage, J. (2003). Pig complement regulator factor H: Molecular cloning and functional characterization. *Immunogenetics*, *55*(7), 462–471.

Herbert, A. P., Uhrín, D., Lyon, M., Pangburn, M. K., & Barlow, P. N. (2006). Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. *Journal of Biological Chemistry*, *281*(24), 16512–16520.

Herbert, A. P., Makou, E., Chen, Z. A., Kerr, H., Richards, A., Rappsilber, J., & Barlow, P. N. (2015). Complement Evasion Mediated by Enhancement of Captured Factor H: Implications for Protection of Self-Surfaces from Complement. *The Journal of Immunology*, *195*(10), 4986–4998.

Higgins, D. R., & Cregg, J. M. (1998). *Pichia* protocols. *Methods in Molecular Biology*, *103*, 1-272.

Hocking, H. G., Herbert, A. P., Kavanagh, D., Soares, D. C., Ferreira, V. P., Pangburn, M. K., Uhrín, D., & Barlow, P. N. (2008). Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. *Journal of Biological Chemistry*, *283*(14), 9475–9487.

Høgasen, K., Jansen, J. H., Mollnes, T. E., Hovdenes, J., & Harboe, M. (1995). Hereditary porcine membranoproliferative glomerulonephritis type II is caused by factor H deficiency. *Journal of Clinical Investigation*, *95*(3), 1054–1061.

Hollister, J. R., Shaper, J. H., & Jarvis, D. L. (1998). Stable expression of mammalian β1,4-galactosyltransferase extends the N-glycosylation pathway in insect cells. *Glycobiology*, *8*(5), 473–480.

Hopkins, D., Gomathinayagam, S., Rittenhour, A. M., Du, M., Hoyt, E., Karaveg, K., Mitchell, T., Nett, J. H., Sharkey, N. J., Stadheim, T. A., Li, H., & Hamilton, S. R. (2011). Elimination of β-mannose glycan structures in *Pichia pastoris. Glycobiology*, *21*(12), 1616–1626.

Hu, F., Li, X., Lü, J., Mao, P., Jin, X., Rao, B., Zheng, P., Zhou, Y., Liu, S., Ke, T., Ma, X., & Ma, L. (2011). A visual method for direct selection of high-producing *Pichia pastoris* clones. *BMC Biotechnology*, *11*(1), 23.

Huffaker, T. C., & Robbins, P. W. (1982). Temperature-sensitive yeast mutants deficient in asparagine-linked glycosylation. *Journal of Biological Chemistry*, *257*(6), 3203–3210.

Iglesias, J. L., Lis, H., & Sharon, N. (1982). Purification and Properties of a d-Galactose/N-Acetyl-d- galactosamine-Specific Lectin from *Erythrina cristagalli*. *European Journal of Biochemistry*, *123*(2), 247–252.

Ikeda, K., Sannoh, T., Kawasaki, N., Kawasaki, T., & Yamashina, I. (1987). Serum lectin with known structure activates complement through the classical pathway. *J. Biol. Chem.*, *262*(16), 7451-7454.

Imai, Y., Singer, M. S., Fennie, C., Lasky, L. A., & Rosen, S. D. (1991). Identification of a carbohydrate-based endothelial ligand for a lymphocyte homing receptor. *Journal of Cell Biology*, *113*(5), 1213–1221.

Inan, M., & Meagher, M. M. (2001). Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of *Pichia pastoris*. *Journal of Bioscience and Bioengineering*, *92*(6), 585–589.

Isenman, D. E., Kells, D. I., Cooper, N. R., Müller-Eberhard, H. J., & Pangburn, M. K. (1981). Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry*, *20*(15), 4458-4467.

Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R., Callewaert, N. (2009). Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. *Nature Protocols*, *4*, 58-70.

Jaeken, J. (2011). Congenital disorders of glycosylation (CDG): It's (nearly) all in it! *Journal of Inherited Metabolic Disease*, *34*(4), 853-858.

Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D., & Gros, P. (2006). Structure of C3b reveals conformational changes that underlie complement activity. *Nature*, *444*(7116), 213–216.

Jarvis, D. L., & Finn, E. E. (1996). Modifying the Insect Cell N-Glycosylation Pathway with Immediate Early Baculovirus Expression Vectors. *Nature Biotechnology*, *14*(10), 1288–1292.

Jarvis, D. L., Howe, D., & Aumiller, J. J. (2001). Novel baculovirus expression vectors that provide sialylation of recombinant glycoproteins in lepidopteran insect cells. *J Virol*, *75*(13), 6223–6227.

Jayakumar, D., Marathe, D. D., & Neelamegham, S. (2009). Detection of site-specific glycosylation in proteins using flow cytometry. *Cytometry Part A*, *75*(10), 866–873.

Jiang, Y., Li, F., Zha, D., Potgieter, T. I., Mitchell, T., Moore, R., Cukan, M., Houston-Cummings, N. R., Nylen, A., Drummond, J. E., McKelvey, T. W., D'Anjou, M., Stadheim, T. A., Sathuraman, N., & Li, H. (2011). Purification process development of a recombinant monoclonal antibody expressed in glycoengineered *Pichia pastoris*. *Protein Expression and Purification*, *76*(1), 7–14.

Johansen, P., Marshall, R. D. and Neuberger, A. (1958). Carbohydrate peptide complex from egg albumin. *Nature*, *181*, 1345–1346.

Johansen, P. G., Marshall, R. D., & Neuberger, A. (1961). Carbohydrates in protein. 3. The preparation and some of the properties of a glycopeptide from hen's-egg albumin. *The Biochemical Journal*, *78*, 518–27.

Kajander, T., Lehtinen, M. J., Hyvärinen, S., Bhattacharjee, A., Leung, E., Isenman, D. E., Meri, S., Goldman, A., & Jokiranta, T. S. Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. *Proc. Natl. Acad. Sci. U.S.A*, *108*(7), 2897-2902.

Kang, J.-Y., Lim, S.-J., Kwon, O., Lee, S.-G., Kim, H. H., & Oh, D.-B. (2015). Enhanced Bacterial α(2,6)-Sialyltransferase Reaction through an Inhibition of Its Inherent Sialidase Activity by Dephosphorylation of Cytidine-5'-Monophosphate. *PLOS ONE*, *10*(7), 1–14.

Kath, F. A., Lange, S., & Kulicke, W. M. (1999). Influence of the glycosidic linkage on the solution conformation of glucans. *Angewandte Makromolekulare Chemie*, *271*, 28–36.

Katz, Y. and Strunk, R. C. (1988). Synthesis and regulation of complement protein factor H in human skin fibroblasts. *The Journal of Immunology*, *141*, 559–563.

Kelly, U., Yu, L., Kumar, P., Ding, J.-D., Jiang, H., Hageman, G. S., Arshavsky, V. Y., Frank, M. M., Hauser, M. A., & Rickman, C. B. (2010). Heparan sulfate, including that in Bruch's membrane, inhibits the complement alternative pathway: implications for age-related macular degeneration. *Journal of Immunology*, *185*(9), 5486–94.

Kerr, H., Wong, E., Makou, E., Yang, Y., Marchbank, K., Kavanagh, D., Richards, A., Herbert, A. P., & Barlow, P. N. (2017). Disease-linked mutations in factor H reveal pivotal role of cofactor activity in Self-surface–selective regulation of complement activation. *Journal of Biological Chemistry*, *292*(32), 13345–13360.

Kimura, M., Kamakura, T., Zhou Tao, Q., Kaneko, I., & Yamaguchi, I. (1994). Cloning of the blasticidin S deaminase gene (BSD) from *Aspergillus terreus* and its use as a selectable marker for *Schizosaccharomyces pombe* and *Pyricularia oryzae*. *MGG Molecular & General Genetics*, *242*(2), 121–129.

Kitagawa, H., & Paulson, J. C. (1994). Differential expression of five sialyltransferase genes in human tissues. *J. Biol. Chem*., *269*(27), 17872-17878.

Klein, B. E. K., Klein, R., & Lee, K. E. (2002). Incidence of age-related cataract over a 10-year interval: The Beaver Dam Eye Study. *Ophthalmology*, *109*(11), 2052–2057.

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Henning, A. K., san Giovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barstable C., & Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, *308*(5720), 385–9.

Klis, F. M., Boorsma, A., & De Groot, P. W. J. (2006). Cell wall construction in Saccharomyces cerevisiae. *Yeast*, *23*(3), 185-202.

Koprivova, A., Stemmer, C., Altmann, F., Hoffmann, A., Kopriva, S., Gorr, G., Reski, R., & Decker, E. L. (2004). Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic N-glycans. *Plant Biotechnology Journal*, *2*(6), 517–523.

Krezdorn, C. H., Watzele, G., Kleene, R. B., Ivanov, S. X., & Berger, E. G. (1993). Purification and characterization of recombinant human beta 1-4 galactosyltransferase expressed in *Saccharomyces cerevisiae*. *Eur J Biochem*, *212*(1), 113–120.

Kühn, S., & Zipfel, P. F. (1996). Mapping of the domains required for decay acceleration activity of the human factor H-like protein 1 and factor H. *European Journal of Immunology*, *26*(10), 2383–2387.

Lalonde, M. E., & Durocher, Y. (2017). Therapeutic glycoprotein production in mammalian cells. *Journal of Biotechnology*, *251*, 128-140.

Lee, C. L., Pang, P. C., Yeung, W. S. B., Tissot, B., Panico, M., Lao, T. T. H., Chu, I. K., Lee, K. F., Chung, M. K., Lam, K. K. W., Koistinen, R., Koistinen, H., Seppälä, M., Morris, H. R., Dell, A., & Chiu, P. C. N. (2009). Effects of differential glycosylation of glycodelins on lymphocyte survival. *Journal of Biological Chemistry*, *284*(22), 15084–15096.

Lehle, L. (1980). Biosynthesis of the Core Region of Yeast Mannoproteins. *European Journal of Biochemistry*, *109*(2), 589–601.

Lehtinen, M. J., Rops, A. L., Isenman, D. E., van der Vlag, J., & Jokiranta, T. S. (2009). Mutations of factor H impair regulation of surface-bound C3b three mechanisms in atypical hemolytic uremic syndrome. *Journal of Biological Chemistry*, *284*(23), 15650–15658.

Li, H., Sethuraman, N., Stadheim, T. A., Zha, D., Prinz, B., Ballew, N., Bobrowicz, P., Choi, B. K., Cook, W. J., Cukan, M., Houston-Cummings, N. R., Davidson, R., Gong, B., Hamilton, S. R., Hoopes, J. P., Jiang, Y., Kim, N., Mansfield, R., Nett, J. H., Rios, S., Strawbridge, R., Wildt, S., & Gerngross, T. U. (2006). Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nature Biotechnology*, *24*(2), 210–215.

Liang, C. J., Yamashita, K., & Kobata, a. (1980). Structural study of the carbohydrate moiety of bovine pancreatic ribonuclease B. *Journal of Biochemistry*, *88*(1), 51–8.

Licht, C., Weyersberg, A., Heinen, S., Stapenhorst, L., Devenge, J., Beck, B., Waldherr, R., Kirschfink, M., Zipfel, P. M., & Hoppe, B. (2005). Successful plasma therapy for atypical hemolytic uremic syndrome caused by factor H deficiency owing to a novel mutation in the complement cofactor protein domain 15. *American Journal of Kidney Diseases*, *45*(2), 415–421.

Licht, C., Heinen, S., Józsi, M., Löschmann, I., Saunders, R. E., Perkins, S. J., Waldherr, R., Skerka, C., Kirschfink, M., Hoppe, B., & Zipfel, P. F. (2006). Deletion of Lys224 in regulatory domain 4 of Factor H reveals a novel pathomechanism for dense deposit disease (MPGN II). *Kidney International*, *70*(1), 42–50.

Licht, C., Schlötzer-Schrehardt, U., Kirschfink, M., Zipfel, P. F., & Hoppe, B. (2007). MPGN II - Genetically determined by defective complement regulation? *Pediatric Nephrology*, *22*(1), 2–9.

Lim, H. K., Kim, K. Y., Lee, K. J., Park, D. H., Chung, S. I., & Jung, K. H. (2000). Genetic stability of the integrated structural gene of guamerin in recombinant *Pichia pastoris*. *Journal of Microbiology and Biotechnology*, *10*(4), 470–475.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2012). SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, *11*(56), 1–16.

Love, K. R., Shah, K. A., Whittaker, C. A., Wu, J., Bartlett, M. C., Ma, D., Leeson, R. L., Priest, M., Borowsky, J., Young, S. K., & Love, J. C. (2016). Comparative genomics and transcriptomics of *Pichia pastoris*. *BMC Genomics*, *17*(1), 550.

Lussier, M., Sdicu, A. M., Ketela, T., & Bussey, H. (1995). Localization and targeting of the *Saccharomyces cerevisiae* Kre2p/Mnt1p α1,2-mannosyltransferase to a medial-Golgi compartment. *Journal of Cell Biology*, *131*(4), 913–927.

Lv, G., Hu, D., Zhao, J., & Li, S. (2015). Quality control of sweet medicines based on gas chromatography-mass spectrometry. *Drug Discoveries & Therapeutics*, *9*(2), 94–106.

Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H., & Jarvis, D. L. (2014). A novel baculovirus vector for the production of nonfucosylated recombinant glycoproteins in insect cells. *Glycobiology*, *24*(3), 325–340.

Makou, E., Mertens, H. D. T., Maciejewski, M., Soares, D. C., Matis, I., Schmidt, C. Q., Herbert, A. P., Svergun, P. I., & Barlow, P. N. (2012). Solution structure of CCP modules 10-12 illuminates functional architecture of the complement regulator, factor H. *Journal of Molecular Biology*, *424*(5), 295–312.

Maley, F., Trimble, R. B., Tarentino, A. L., & Plummer, T. H. (1989). Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Analytical Biochemistry*.

Malhotra, R., Ward, M., Sim, R. B., & Bird, M. I. (1999). Identification of human complement Factor H as a ligand for L-selectin. *Biochem J*, *341*, 61–69.

Malissard, M., Borsig, L., Di Marco, S., Grutter, M. G., Kragl, U., Wandrey, C., & Berger, E. G. (1996). Recombinant soluble beta-1,4-galactosyltransferases expressed in *Saccharomyces cerevisiae*. Purification, characterization and comparison with human enzyme. *European Journal of Biochemistry*, *239*(2), 340–348.

Mallem, M., Warburton, S., Li, F., Shandil, I., Nylen, A., Kim, S., Jiang, Y., Meehl, M., D'Anjou, M., Stadheim, T. A., & Choi, B. K. (2014). Maximizing recombinant human serum albumin production in a MutS *Pichia pastoris* strain. *Biotechnology Progress*, *30*(6), 1488–1496.

Maras, M., Callewaert, N., Piens, K., Claeyssens, M., Martinet, W., Dewaele, S., Contreras, H., Dewerte, I., Pentillä, M., & Contreras, R. (2000). Molecular cloning and enzymatic characterization of a *Trichoderma reesei* 1,2-α-D-mannosidase. *Journal of Biotechnology*, *77*(2–3), 255–263.

Marx, H., Mattanovich, D., & Sauer, M. (2008). Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. *Microbial Cell Factories*, *7*(1), 23.

Matthews, C. B., Wright, C., Kuo, A., Colant, N., Westoby, M., & Love, J. C. (2017). Reexamining opportunities for therapeutic protein production in eukaryotic microorganisms. *Biotechnology and Bioengineering*, *114*(11), 2432-2444.

Meri, S., & Pangburn, M. K. (1990). Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding

site on factor H. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(10), 3982–3986.

Michelfelder, S., Parsons, J., Bohlender, L. L., Hoernstein, S. N. W., Niederkru ger, H., Busch, A., Krieghoff, N., Koch, J., Fode, B., Schaaf, A., Frischmuff, T., Pohl, M., Zipfel, P. F., Reski, R., Decker, E. L., & Häffner, K. (2017). Moss-Produced, Glycosylation-Optimized Human Factor H for Therapeutic Application in Complement Disorders. *Journal of the American Society of Nephrology*, 1462–1474.

Milder, F. J., Gomes, L., Schouten, A., Janssen, B. J., Huizinga, E. G., Romijn, R. A., Hemrika, W., Roos, A., Daha, M. R., & Gros, P. (2007). Factor B provides insights into activation of the central protease of the complement system. *Nat. Struct. Mol. Biol.*, *14*(3), 224-228.

Mine, T., Katayama, S., Kajiwara, H., Tsunashima, M., Tsukamoto, H., Takakura, Y., & Yamamoto, T. (2009). An α2,6-sialyltransferase cloned from *Photobacterium leiognathi* strain JT-SHIZ-119 shows both sialyltransferase and neuraminidase activity. *Glycobiology*, *20*(2), 158–165.

Miura, M., Hirose, M., Miwa, T., Kuwae, S., & Ohi, H. (2004). Cloning and characterization in *Pichia pastoris* of PNO1 gene required for phosphomannosylation of N-linked oligosaccharides. *Gene*, *324*(1–2), 129–137.

Miura, Y., Shinohara, Y., Furukawa, J. I., Nagahori, N., & Nishimura, S. I. (2007). Rapid and simple solid-phase esterification of sialic acid residues for quantitative glycomics by mass spectrometry. *Chemistry - A European Journal*, *13*(17), 4797–4804.

Morell, A. G., Gregoriadis, G., Scheinberg, H., Hickman, J., & Ashwell, G. (1971). The role of sialic acid in determining the survival of glycoproteins in the circulation. *The Journal of Biological Chemistry*, *246*(5), 1461–1467.

Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D., Svergun, D. I., Johansson, C. M., Uhrín, D., Barlow, P. N., & Hannan, J. P. (2011). Structural basis for engagement by complement factor H of C3b on a self surface. *Nat. Struct. Mol. Biol*., *18*(4), 463-470.

Morgan, H. P., Mertens, H. D., Guariento, M., Schmidt, C. Q., Soares, D. C., Svergun, D. I., Herbert, A. P., Barlow, P. N., & Hannan, J. P. (2012). *PLoS One*, *7*(2).

Morris, K. M., Aden, D. P., Knowles, B. B., & Colten, H. R. (1982). Complement biosynthesis by the human hepatoma-derived cell line HepG2. *Journal of Clinical Investigation*, *70*(4), 906–913.

Morris, H. R., Dell, A., Easton, R. L., Panico, M., Koistinen, H., Koistinen, R., Oehninger, S., Patanker, M. S., Seppälä, M., & Clark, G. F. (1996). Gender-specific glycosylation of human glycodelin affects its contraceptive activity. *Journal of Biological Chemistry*, *271*(50), 32159–32167.

MyAssays: Analysis Software Solutions. (2017). [ONLINE] Available at: https://www.myassays.com/home.aspx

Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F. S., & Glieder, A. (2012). Deletion of the *Pichia pastoris* ku70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS ONE*, *7*(6).

Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E., & Rini, J. M. (1998). X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science*, *280*(5367), 1277–1281.

Nakanishi-Shindo, Y., Nakayama, K. I., Tanaka, A., Toda, Y., & Jigami, Y. (1993). Structure of the N-linked oligosaccharides that show the complete loss of α-1,6-polymannose outer chain from och1, och1 mnn1, and och1 mnn1 alg3 mutants of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *268*(35), 26338–26345.

Nakayama, K., Nagasu, T., Shimma, Y., Kuromitsu, J., & Jigami, Y. (1992). OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. *The EMBO Journal*, *11*(7), 2511–2519.

Nakayama, K. I., Nakanishi-Shindo, Y., Tanaka, A., Haga-Toda, Y., & Jigami, Y. (1997). Substrate specificity of α-1,6-mannosyltransferase that initiates N-linked mannose outer chain elongation in *Saccharomyces cerevisiae*. *FEBS Letters*, *412*(3), 547–550.

Nan, R., Gor, J., & Perkins, S. J. (2008). Implications of the Progressive Self-association of Wild-type Human Factor H for Complement Regulation and Disease. *Journal of Molecular Biology*, *375*(4), 891–900.

Nett, J. H., & Gerngross, T. U. (2003). Cloning and disruption of the PpURA5 gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. *Yeast*, *20*(15), 1279–1290.

Nevoigt, E., Fischer, C., Mucha, O., Matthäus, F., Stahl, U., & Stephanopoulos, G. (2007). Engineering promoter regulation. *Biotechnology and Bioengineering*, *96*(3), 550–558.

Nichols, E. M., Barbour, T. D., Pappworth, I. Y., Wong, E. K. S., Palmer, J. M., Sheerin, N. S., Pickering, M. C., & Marchbank, K. J. (2015). An extended mini-complement factor H molecule ameliorates experimental C3 glomerulopathy. *Kidney International*, *88*(6), 1314–1322.

Nita-Lazar, M., Wacker, M., Schegg, B., Amber, S., & Aebi, M. (2005). The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology*, *15*(4), 361–367.

Novagen. (2011). pET System Manual 11 th Edition, 2–63.

Nuenke, R. H. and Cunningham, L. W. (1961). Preparation and structural studies of ovalbumin glycopeptides. *The Journal of Biological Chemistry*, *236*(9), 2452–2460.

Nwosu, C. C., Seipert, R. R., Strum, J. S., Hua, S. S., An, H. J., Zivkovic, A. M., German, B. J., & Lebrilla, C. B. (2011). Simultaneous and extensive site-specific N- and O-glycosylation analysis in protein mixtures. *Journal of Proteome Research*, *10*(5), 2612–2624.

Oh-Eda, M., Hasegawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., Yamazaki, T., & Ochi, N. (1990). O-linked sugar chain of human granulocyte colony-stimulating factor protects it against polymerization and denaturation allowing it to retain its biological activity. *Journal of Biological Chemistry*, *265*(20), 11432–11435.

Oppermann, M., Manuelian, T., Józsi, M., Brandt, E., Jokiranta, T. S., Heinen, S., Meri, S., Skerka, C., Götze, O., & Zipfel, P. F. (2006). The C-terminus of complement regulator Factor H mediates target recognition: Evidence for a compact conformation of the native protein. *Clinical and Experimental Immunology*, *144*(2), 342–352.

Ormsby, R. J., Jokiranta, T. S., Duthy, T. G., Griggs, K. M., Sadlon, T. A., Giannakis, E., & Gordon, D. L. (2006). Localization of the third heparin-binding site in the human complement regulator factor H. *Molecular Immunology*, *43*(10), 1624–1632.

Ormsby, R. J., Ranganathan, S., Tong, J. C., Griggs, K. M., Dimasi, D. P., Hewitt, A. W., Burdon, K. P., Craig, J. E., Hoh, J., & Gordon, D. L. (2008). Functional and structural implications of the complement factor H Y402H polymorphism associated with age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *49*(5), 1763–1770.

Ortiz-Soto, M. E., & Seibel, J. (2016). Expression of functional human sialyltransferases ST3Gal1 and ST6Gal1 in *Escherichia coli*. *PLoS ONE*, *11*(5).

Owsley, C., Huisingh, C., Clark, M. E., Jackson, G. R., & McGwin, G. (2015). Comparison of Visual Function in Older Eyes in the Earliest Stages of Age-related Macular Degeneration to Those in Normal Macular Health. *Current Eye Research*, *3683*, 1–7.

Padler-Karavani, V., & Varki, A. (2011). Potential impact of the non-human sialic acid N-glycolylneuraminic acid on transplant rejection risk. *Xenotransplantation*, *18*(1), 1-5.

Paixão-Cavalcante, D., Hanson, S., Botto, M., Cook, H. T., & Pickering, M. C. (2009). Factor H facilitates the clearance of GBM bound iC3b by controlling C3 activation in fluid phase. *Molecular Immunology*, *46*(10), 1942–1950.

Pangburn, M. K., Atkinson, M. A. L., & Meri, S. (1991). Localization of the heparin-binding site on complement factor H. *Journal of Biological Chemistry*, *266*(25), 16847–16853.

Pangburn, M. K. (2002). Cutting edge: localization of the host recognition functions of complement factor H at the carboxyl-terminal: implications for hemolytic uremic syndrome. *J Immunol*, *169*(9), 4702–4706.

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., & Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast*, *27*(11), 955–964.

Patnaik, S. K., & Stanley, P. (2006). Lectin-Resistant CHO Glycosylation Mutants. *Methods in Enzymology*, *416*, 159-182.

Paulson, J. C., Weinstein, J., & Schauer, A. (1989). Tissue-specific expression of sialyltransferase. *J Biol. Chem.*, *264*(19), 10931-10934.

Pechtl, I. C., Kavanagh, D., Mcintosh, N., Harris, C. L., & Barlow, P. N. (2011). Disease-associated N-terminal complement factor H mutations perturb cofactor and decay-accelerating activities. *Journal of Biological Chemistry*, *286*(13), 11082–11090.

Pelham, H. R. (1988). Evidence that luminal ER proteins are sorted from secreted proteins in a post-ER compartment. *The EMBO Journal*, *7*(4), 913–8.

Pickering, M. C., Cook, H. T., Warren, J., Bygrave, A. E., Moss, J., Walport, M. J., & Botto, M. (2002). Uncontrolled C3 activation causes membranoproliferative glomerulonephritis in mice deficient in complement factor h. *Nature Genetics*, *31*(4), 424–428.

Pickering, M. C., Goicoechea de Jorge, E., Martinez-Barricarte, R., Recalde, S., Garcia-Layana, A., Rose, K. L., Moss, J., Walport, M. J., Cook, H. T., Rodriguez de Córdoba, S., & Botto, M. (2007). Spontaneous hemolytic uremic syndrome triggered by complement factor H lacking surface recognition domains. *The Journal of Experimental Medicine*, *204*(6), 1249–1256.

Perkins, S. J., Haris, P. I., Sim, R. B., & Chapman, D. (1988). A study of the structure of human complement component factor H by Fourier transform infrared spectroscopy and secondary structure averaging methods. *Biochemistry*, *27*(11), 4004-4012.

Pilobello, K. T., & Mahal, L. K. (2007). Deciphering the glycocode: the complexity and analytical challenge of glycomics. *Current Opinion in Chemical Biology*, *11*(3), 300-305.

Plummer, T. H., & Hirs, C. H. W. (1964). On the Structure of Bovine Pancreatic Ribonuclease B. *Journal of Biological Chemistry*, *239*(8), 2530–2538.

Plummer, T. H., Tarentino, A., & Maley, F. (1968). The glycopeptide linkage of ribonuclease B. *Journal of Biological Chemistry*, *243*(19), 5158–5164.

Pompach, P., Ashline, D. J., Brnakova, Z., Benicky, J., Sanda, M., & Goldman, R. (2014). Protein and site specificity of fucosylation in liver-secreted glycoproteins. *Journal of Proteome Research*, *13*(12), 5561–5569.

Pouw, R. B., Vredevoogd, D. W., Kuijpers, T. W., & Wouters, D. (2015). Of mice and men: The factor H protein family and complement regulation. *Molecular Immunology*, *67*(1), 12-20.

Powell, A. K., & Harvey, D. J. (1996). Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry*, *10*(9), 1027–32.

Prien, J. M., Ashline, D. J., Lapadula, A. J., Zhang, H., & Reinhold, V. N. (2009). The High Mannose Glycans from Bovine Ribonuclease B Isomer Characterization by Ion Trap MS. *Journal of the American Society for Mass Spectrometry*, *20*(4), 539–556.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., & Chu, J. (2011). GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. *Applied and Environmental Microbiology*, *77*(11), 3600–3608.

Raju, T. S., & Davidson, E. A. (1994). Role of sialic acid on the viscosity of canine tracheal mucin glycoprotein. *Biochem Biophys Res Commun*, *205*(1), 402–409.

Ripoche, J., Day, A. J., Harris, T. J., & Sim, R. B. (1988). The complete amino acid sequence of human complement factor H. *Biochem. J.*, *249*(2), 593-602.

Ritchie, G. E., Moffatt, B. E., Sim, R. B., Morgan, B. P., Dwek, R. A., & Rudd, P. M. (2002). Glycosylation and the complement system. *Chemical Reviews*, *102*(2), 305–319.

Rodríguez De Córdoba, S., Díaz-Guillén, M. A., & Heine-Suñer, D. (1999). An integrated map of the human regulator of complement activation (RCA) gene cluster on 1q32. In *Molecular Immunology*, *36*, 803–808.

Rodríguez De Córdoba, S., Esparza-Gordillo, J., Goicoechea De Jorge, E., Lopez-Trascasa, M., & Sánchez-Corral, P. (2004). The human complement factor H: Functional roles, genetic variations and disease associations. *Molecular Immunology*, *41*(4), 355-367.

Rogers, G. N., & Paulson, J. C. (1983). Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology*, *127*(2), 361–373.

Roth, Z., Yehezkel, G., & Khalaila, I. (2012). Identification and Quantification of Protein Glycosylation. *International Journal of Carbohydrate Chemistry*, *2012*, 1–10.

Rougier, N., Kazatchkine, M. D., Rougier, J. P., Fremeaux-Bacchi, V., Blouin, J., Deschenes, G., Soto, B., Baudouin, V., Pautard, B., Proesmans, W., Weiss, E., & Weiss, L. (1998). Human complement factor H deficiency associated with hemolytic uremic syndrome. *J Am Soc Nephrol*, *9*(12), 2318–2326.

Roumeliotis, G. (2006). GlycoFi proves too sweet for Merck to resist. [ONLINE] *in-Pharma Technologists.com*. Available at: https://www.in-pharmatechnologist.com/Article/2006/05/16/Glyco-Fi-proves-too-sweet-Merck-to-resist.

Runge, K. W., Huffaker, T. C., & Robbins, P. W. (1984). Two yeast mutations in glucosylation steps of the asparagine glycosylation pathway. *Journal of Biological Chemistry*, *259*(1), 412–417.

Runge, K. W., & Robbins, P. W. (1986). A new yeast mutation in the glucosylation steps of the asparagine-linked glycosylation pathway. Formation of a novel asparagine-linked oligosaccharide containing two glucose residues. *Journal of Biological Chemistry*, *261*(33), 15582–15590.

Russell, D., Oldham, N. J., & Davis, B. G. (2009). Site-selective chemical protein glycosylation protects from autolysis and proteolytic degradation. *Carbohydrate Research*, *344*(12), 1508–1514.

Sahu, A., Kozel, T. R., & Pangburn, M. K. (1994). Specificity of the thioester-containing reactive site of human C3 and its significance to complement activation. *The Biochemical Journal*, (Pt 2), 429–36.

Salamin, K., Sriranganadane, D., Léchenne, B., Jousson, O., & Monod, M. (2010). Secretion of an endogenous subtilisin by *Pichia pastoris* strains GS115 and KM71. *Applied and Environmental Microbiology*, *76*(13), 4269–4276.

Sánchez-Corral, P., Bellavia, D., Amico, L., Brai, M., & Rodríguez de Córdoba, S. (2000). Molecular basis for factor H and FHL-1 deficiency in an Italian family. *Immunogenetics*, *51*(4–5), 366–369.

Sánchez-Corral, P., Pérez-Caballero, D., & Huarte, O. (2002). Structural and functional characterization of factor H mutations associated with atypical hemolytic uremic syndrome. *The American Journal of Genetics*, *71*(6), 1285–1295.

Schmidt, C. Q., Herbert, A. P., Kavanagh, D., Gandy, C., Fenton, C. J., Blaum, B. S., Lyon, M., Uhrín, D., & Barlow, P. N. (2008). A New Map of Glycosaminoglycan and C3b Binding Sites on Factor H. *The Journal of Immunology*, *181*(4), 2610–2619.

Schmidt, C. Q., Herbert, A. P., Mertens, H. D. T., Guariento, M., Soares, D. C., Uhrin, D., Rowe, A. J., Svergun, D. I., & Barlow, P. N. (2010). The Central Portion of Factor H (Modules 10-15) Is Compact and Contains a Structurally Deviant CCP Module. *Journal of Molecular Biology*, *395*(1), 105–122.

Schmidt, C. Q., Slingsby, F. C., Richards, A., & Barlow, P. N. (2011). Production of biologically active complement factor H in therapeutically useful quantities. *Protein Expression and Purification*, *76*(2), 254–263.

Schmidt, C. Q., Bai, H., Lin, Z., Risitano, A. M., Barlow, P. N., Ricklin, D., & Lambris, J. D. (2013). Rational Engineering of a Minimized Immune Inhibitor with Unique Triple-Targeting Properties. *The Journal of Immunology*, *190*(11), 5712–5721.

Scholl, H. P. N., Issa, P. C., Walier, M., Janzer, S., Pollok-Kopp, B., Börncke, F., Fritsche, L. G., Chong, N. V., Fimmers, R., Wienker, T., Holz, F. G., Weber, B. H. F., & Oppermann, M. (2008). Systemic complement activation in age-related macular degeneration. *PLoS ONE*, *3*(7).

Schwaeble, W., Zwirner, J., Schulz, T. F., Linke, R. P., Dierich, M. P., & Weiss, E. H. (1987). Human complement factor H: expression of an additional truncated gene product of 43 kDa in human liver. *European Journal of Immunology*, *17*(10), 1485–1489.

Schwartz, A. L. (1984). The hepatic asialoglycoprotein receptor. *CRC Critical Reviews in Biochemistry*, *16*(3), 207–233.

Schwertz, R., de Jong, R., Gretz, N., Kirschfink, M., Anders, D., & Scharer, K. (1996). Outcome of idiopathic membranoproliferative glomerulonephritis in children. Arbeitsgemeinschaft Padiatrische Nephrologie. *Acta Paediatr*, *85*(3), 308–312.

Scorer, C. A., Clare, J. J., McCombie, W. R., Romanos, M. A., & Sreekrishna, K. (1994). Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. *Biotechnology*, *12*(2), 181-184.

Sequence Manipulation Suite: DNA Molecular Weight. (2017). [ONLINE] Available at: https://www.bioinformatics.org/sms2/dna_mw.html

Sharma, A. K., & Pangburn, M. K. (1994). Biologically active recombinant human complement factor H: synthesis and secretion by the baculovirus system. *Gene*, *143*, 301–302.

Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B., & Peumans, W. J. (1987). The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *Journal of Biological Chemistry*, *262*(4), 1596–1601.

Shields, R. L., Lai, J., Keck, R., O'Connell, L. Y., Hong, K., Gloria Meng, Y., Weikert, S. H. A., & Presta, L. G. (2002). Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *Journal of Biological Chemistry*, *277*(30), 26733–26740.

Shinkawa, T., Nakamura, K., Yamane, N., Shoji-Hosaka, E., Kanda, Y., Sakurada, M., Uchida, K., Anazawa, H., Satoh, M., Yamasaki, M., Hanai, S., & Shitara, K. (2003). The absence of fucose but not the presence of galactose or bisecting N-acetylglucosamine of human IgG1 complex-type oligosaccharides shows the critical role of enhancing antibody-dependent cellular cytotoxicity. *Journal of Biological Chemistry*, *278*(5), 3466–3473.

Shinohara, Y., Furukawa, J. I., Niikura, K., Miura, N., & Nishimura, S. I. (2005). Direct N-glycan profiling in the presence of tryptic peptides on MALDI-TOF by controlled ion enhancement and suppression upon glycan-selective derivatization. *Analytical Chemistry*, *76*(23), 6989–6997.

Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., & Kawaoka, Y. (2006). Avian flu: Influenza virus receptors in the human airway. *Nature*, *440*(7083), 435–436.

Sinha, J., Plantz, B. A., Inan, M., & Meagher, M. M. (2005). Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-τ. *Biotechnology and Bioengineering*, *89*(1), 102–112.

Skehel, J. J., Stevens, D. J., Daniels, R. S., Douglas, A. R., Knossow, M., Wilson, I. A., & Wiley, D. C. (1984). A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Cell*, *81*, 1779–1783.

Skerka, C., Lauer, N., Weinberger, A. A. W. A., Keilhauer, C. N., Sühnel, J., Smith, R., Schlötzer-Schrehardt, U., Fritsche, L., Heinen, S., Hartmann, A., Weber, B. H. F., & Zipfel, P. F. (2007). Defective complement control of Factor H (Y402H) and FHL-1 in age-related macular degeneration. *Molecular Immunology*, *44*(13), 3398–3406.

Sleat, D. E., Zheng, H., Qian, M., & Lobel, P. (2006). Identification of Sites of Mannose 6-Phosphorylation on Lysosomal Proteins. *Molecular & Cellular Proteomics*, *5*(4), 686–701.

Snider, M. D., Huffaker, J. R., Couto, J. R., Robins, P. W., and Feizi, T. (1982). Genetic and biochemical studies of asparagine-linked oligosaccharide assembly. *Phil. Trans. R. Soc. Lond. B*, *300*, 207–223.

Sofat, R., Mangione, P. P., Gallimore, J. R., Hakobyan, S., Hughes, T. R., Shah, T., Goodship, T., D'Aiuto, F., Langenberg, C., Wareham, N., Morgan, B. P., Pepys, M. B., & Hingorani, A. D. (2013). Distribution and determinants of circulating complement factor H concentration determined by a high-throughput immunonephelometric assay. *Journal of Immunological Methods*, *390*(1–2), 63–73.

Solá, R. J., & Griebenow, K. (2010). Glycosylation of Therapeutic Proteins. *BioDrugs*, *24*(1), 9–21.

Somers, W., Tang, J., Shaw, G., & Camphausen, R. (2000). Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P-and E-selectin bound to SLe X and PSGL-1. *Cell*, *103*, 467–479.

Spiro, R. G. (1960). Studies on fetuin, a glycoprotein of fetal serum. I. Isolation, chemical composition, and physiochemical properties. *The Journal of Biological Chemistry*, *235*(10), 2860–2869.

Spiro, R. G., & Bhoyroo, V. D. (1974). Structure of the O glycosidically linked carbohydrate units of fetuin. *Journal of Biological Chemistry*, *249*(18), 5704–5717.

Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., & Gasser, B. (2010). Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. *Journal of Biotechnology*, *150*(4), 519–529.

Stenberg, E., Persson, B., Roos, H., & Urbaniczky, C. (1991). Quantitative determination of surface concentration of protein with surface plasmon resonance using radiolabeled proteins. *Journal of Colloid And Interface Science*, *143*(2), 513–526.

Strasser, R., Altmann, F., Mach, L., Glössl, J., & Steinkellner, H. (2004). Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking β1,2-linked xylose and core α1,3-linked fucose. *FEBS Letters*, *561*(1–3), 132–136.

Striegel, A. M., & Boone, M. A. (2011). Influence of glycosidic linkage on solution conformational entropy of oligosaccharides: Malto- vs. isomalto- and cello- vs. laminarioligosaccharides. *Biopolymers*, *95*(4), 228–233.

Sumer-Bayraktar, Z., Kolarich, D., Campbell, M. P., Ali, S., Packer, N. H., & Thaysen-Andersen, M. (2011). N-glycans modulate the function of human corticosteroid-binding globulin. *Molecular & Cellular Proteomics*, *10*(8).

Sun, S., Wang, Q., Zhao, F., Chen, W., & Li, Z. (2012). Prediction of biological functions on glycosylation site migrations in human influenza H1N1 viruses. *PLoS ONE*, *7*(2).

Sunga, A. J., Tolstorukov, I., & Cregg, J. M. (2008). Posttransformational vector amplification in the yeast *Pichia pastoris*. *FEMS Yeast Research*, *8*(6), 870–876.

Sunstrom, N. A., Hunt, S., Bailey, C., Baig, M., & Sleigh, M. (2000). Regulated autocrine growth of CHO cells. *Cytotechnology*, *34*(1-2), 39-46.

Szabó, T. G., Palotai, R., Antal, P., Tokatly, I., Tóthfalusi, L., Lund, O., Nagy, G., Falus, A., & Buzás, E. I. (2009). Critical role of glycosylation in determining the length and structure of T cell epitopes. *Immunome Research*, *5*(1).

Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., & Guerry, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology*, *32*(5), 1022–1030.

Szymanski, C. M., Burr, D. H., & Guerry, P. (2002). *Campylobacter* protein glycosylation affects host cell interactions. *Infection and Immunity*, *70*(4), 2242–2244.

Taylor, M. E., & Drickamer, K. (1993). Structural requirements for high affinity binding of complex ligands by the macrophage mannose receptor. *Journal of Biological Chemistry*, *268*(1), 399–404.

Terashima, M., Amano, M., Onodera, T., Nishimura, S.-I., & Iwasaki, N. (2014). Quantitative glycomics monitoring of induced pluripotent- and embryonic stem cells during neuronal differentiation. *Stem Cell Research*, *13*(3), 454–464.

Thompson, R. A., & Winterborn, M. H. (1981). Hypocomplementaemia due to a genetic deficiency of beta 1H globulin. *Clinical and Experimental Immunology*, *46*(1), 110–9.

Toomey, C. B., Johnson, L. V., & Bowes Rickman, C. (2017). Complement factor H in AMD: Bridging genetic associations and pathobiology. *Progress in Retinal and Eye Research*.

Torreira, E., Tortajada, A., Montes, T., Rodríguez de Córdoba, S., & Llorca, O. (2009). Coexistence of closed and open conformations of complement factor B in the alternative pathway C3bB(Mg^{2+}) proconvertase. *J. Immunol.*, *183*(11), 7347-7351.

Toscano, M. A., Bianco, G. A., Ilarregui, J. M., Croci, D. O., Correale, J., Hernandez, J. D., Zwirner, N. W., Poirier, F., Riley, E. M., Baum, L. G., & Rabinovich, G. A. (2007). Differential glycosylation of TH1, TH2 and TH-17 effector cells selectively regulates susceptibility to cell death. *Nature Immunology*, *8*(8), 825–834.

Tretter, V., Altmann, F., & März, L. (1991). Peptide-N4-(N-acetyl-β-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached α1 → 3 to the asparagine-linked N-acetylglucosamine residue. *European Journal of Biochemistry*, *199*(3), 647–652.

Tschopp, J. F., Brust, P. F., Cregg, J. M., Stillman, C. A., & Gingeras, T. R. (1987). Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. *Nucleic Acids Research*, *15*(9), 3859–3876.

Tsukamoto, H., Takakura, Y., Mine, T., & Yamamoto, T. (2008). *Photobacterium sp*. JT-ISH-224 Produces Two Sialyltransferases, α-/β-Galactoside α2,3-Sialyltransferase and β-Galactoside α2,6-Sialyltransferase. *The Journal of Biochemistry*, *143*(2), 187–197.

Uematsu, R., Furukawa, J., Nakagawa, H., Shinohara, Y., Deguchi, K., Monde, K., & Nishimura, S. (2005). High throughput quantitative glycomics and glycoform-focused proteomics of murine dermis and epidermis. *Mol Cell Proteomics*, *4*(12), 1977–1989.

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y.-Y., Reading, C. A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., & DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nature Chemical Biology*, *8*(5), 434–436.

Vervecken, W., Kaigorodov, V., Callewaert, N., Geysens, S., De Vusser, K., & Contreras, R. (2004). In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. *Applied and Environmental Microbiology*, *70*(5), 2639–2646.

Vik, D. P., Keeney, J. B., Muñoz-Cánoves, P., Chaplin, D. D., & Tack, B. F. (1988). Structure of murine complement factor H gene. *J. Biol. Chem.*, *263*(32), 16720-1674.

Vogl, T., & Glieder, A. (2013). Regulation of *Pichia pastoris* promoters and its consequences for protein production. *New Biotechnology*, *30*(4), 385-404.

Vogl, T., Ruth, C., Pitzer, J., Kickenweiz, T., & Glieder, A. (2014). Synthetic core promoters for *Pichia pastoris*. *ACS Synthetic Biology*, *3*(3), 188–191.

Vogl, T., Kickenweiz, T., Sturmberger, L., Glieder, A. (2015). Bidirectional promoter. *United States Patent Application Publication*.

Vogl, T., Sturmberger, L., Kickenweiz, T., Wasmayer, R., Schmid, C., Hatzl, A. M., Gertsmann, M. A., Pitzer, J., Wagner, M., Thallinger, G. G., Geier, M., & Glieder, A. (2016). A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in *Pichia pastoris*. *ACS Synthetic Biology*, *5*(2), 172–186.

Wallis, R., Dodds, A. W., Mitchell, D. A., Sim, R. B., Reid, K. B., Schwaeble, W. J. (2007). Molecular interactions between MASP-2, C4, and C2 and their activation fragments leading to complement activation via the lectin pathway. *J. Biol. Chem.*, *282*(11), 7844-7851.

Wang, C., Eufemi, M., Turano, C., & Giartosio, A. (1996). Influence of the carbohydrate moiety on the stability of glycoproteins. *Biochemistry*, *35*(23), 7299–7307.

Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V., & Cregg, J. M. (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene*, *186*(1), 37–44.

Weis, J. H., Morton, C. C., Bruns, G. a, Weis, J. J., Klickstein, L. B., Wong, W. W., & Fearon, D. T. (1987). A complement receptor locus: genes encoding C3b/C4b receptor and C3d/Epstein-Barr virus receptor map to 1q32. *Journal of Immunology*, *138*(1), 312–5.

Weiler, J. M., Daha, M. R., Austen, K. F., & Fearon, D. T. (1976). Control of the amplification convertase of complement by the plasma protein beta1H. *Proc. Natl. Acad. Sci. U.S.A.*, *73*(9), 3268-3272.

Weninger, A., Hatzl, A. M., Schmid, C., Vogl, T., & Glieder, A. (2016). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *Journal of Biotechnology*, *235*, 139-149.

Werten, M. W. T., Van Den Bosch, T. J., Wind, R. D., Mooibroek, H., & De Wolf, F. A. (1999). High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast*, *15*(11), 1087–1096.

Whaley, K. (1980). Biosynthesis of the complement components and the regulatory proteins of the alternative complement pathway by human peripheral blood monocytes. *The Journal of Experimental Medicine*, *151*(3), 501–516.

WHO. (2013). Guidelines on the quality, safety and efficacy of biotherapeutic protein products prepared by recombinant DNA technology, (814), 1–91.

Windwarder, M., & Altmann, F. (2014). Site-specific analysis of the O-glycosylation of bovine fetuin by electron-transfer dissociation mass spectrometry. *Journal of Proteomics*, *108*, 258–268.

Wu, J., Wu, Y. Q., Ricklin, D., Janssen, B. J. C., Lambris, J. D., & Gros, P. (2009). Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. *Nature Immunology*, *10*(7), 728–733.

Xue, X., Wu, J., Ricklin, D., Forneris, F., Di Crescenzio, P., Schmidt, C. Q., Granneman, J., Sharp, T. H., Lambris, J. D., & Gros, P. (2017). Regulator-dependent mechanisms of C3b processing by factor I allow differentiation of immune responses. *Nature Structural and Molecular Biology*, *24*(8), 643–651.

Yamamoto, T., Nakashizuka, M., & Terada, I. (1998). Cloning and Expression of a Marine Bacterial β-Galactoside α2,6-Sialyltransferase Gene from *Photobacterium damsela* JT0160. *Journal of Biochemistry*, *123*(1), 94–100.

Yamamoto, T., Hamada, Y., Ichikawa, M., Kajiwara, H., Mine, T., Tsukamoto, H., & Takakura, Y. (2007). A β-galactoside α2,6-sialyltransferase produced by a marine bacterium, *Photobacterium leiognathi* JT-SHIZ-145, is active at pH 8. *Glycobiology*, *17*(11), 1167–74.

Yang, C., Mookerjea, S., & Nagpurkar, A. (1992). Clearance of rat C-reactive protein in vivo and by perfused liver. *Glycobiology*, *2*(1), 41–48.

Yang, J., Nie, L., Chen, B., Liu, Y., Kong, Y., Wang, H., & Diao, L. (2014). Hygromycin-resistance vectors for gene expression in *Pichia pastoris*. *Yeast*, *31*(4), 115–125.

Zipfel, P. F., Skerka, C., Hellwage, J., Jokiranta, S. T., Meri, S., Brade, V., Kraiczy, P., Noris, M., & Remuzzi, G. (2002). Factor H family proteins: on complement, microbes and human diseases. *Biochemical Society Transactions*, *30*(6), 971–978.

Appendix

Och1

Gctatattcgccgtttctgtcatttgcgttttgtacggaccctcacaacaattatcatctccaaaaatagactatgatccattgacgct
ccgatcacttgatttgaagactttggaagctccttcacagttgagtccaggcaccgtagaagataatcttcgaagacaattggagtt
tcattttccttaccgcagttacgaaccttttccccaacatatttggcaaacgtggaaagtttctccctctgatagttcctttccgaaaa
acttcaaagacttaggtgaaagttggctgcaaaggtccccaaattatgatcattttgtgatacccgatgatgcagcatgggaactta
ttcaccatgaatacgaacgtgtaccagaagtcttggaagcttt

ManI

atgagatttccttctagttccgtcttggcacttggattgattggaccagccttggcataccctaaaccaggagcaacaaagagaggt
tctccaaaccctactagagctgctgctgttaaggctgcttttcaaacttcttggaatgcttaccatcactttgctttccctcatgatgat
ttgcacccagtttctaactctttcgatgatgaaagaaatggttggggttcttctgctattgatggtttggatactgctattttgatggga
gatgctgatatcgttaacactatcttgcaatacgttcctcaaattaacttcactactactgctgttgctaatcaaggttcttctgttttcg
agactaacatcagatacttgggtggtttgttgtctgcttatgatttgttgagaggtccattttcttctttggctactaaccaaactttggt
taactctttgttgagacaagctcaaactttggctaacggtttgaaagttgctttcactactccttctggtgttccagatcctactgttttc
tttaatccaactgttagaagatctggtgcttcttctaacaatgttgctgaaattggttctttggttttggagtggactagattgtctgat
ttgactggtaaccctcaatacgctcagttggctcaaaagggtgaatcttatttgttgaatccaaaaggttctcctgaggcttggccag
gtttgatcggtactttcgtttctacttctaacggtactttccaagattcttctggttcttggtctggtttgatggattctttctacgaatatt
tgattaagatgtacttgtatgatccagttgctttcgctcattacaaagatagatgggttttgggtgctgattctactattggtcatttgg
gttctcacccttctactagaaaggatttgactttcttgtcttcttacaacggtcaatctacttctccaaattctggtcacttggcttctttt
ggtggtggtaacttcattttgggtggtatcttgttgaacgaacaaaagtacatcgatttcggtattaaattggcttcttcttacttcggt
acttatactcaaactgcttctggtattggtcctgagggtttcgcttgggttgattctgttactggtgctggtggttctccaccttcttctc
aatctggtttttattcttctgctggtttctgggttactgctccttactatattttgagaccagaaactttggagtctttgtactatgcttat
agagttactggagattctaaatggcaagatttggcttgggaagctttgtctgctattgaggatgcttgtagagctggttctgcttattc
ttctattaacgatgttactcaagctaacggtggtggtgcttctgatgatatggaatcttttggttcgctgaggctttgaagtacgctta
tttgatcttcgctgaagaatctgatgttcaagttcaagctactggtggtaacaaatttgttttcaatactgaagctcatccattttcaat
cagaagtagtagtagaaggggaggacatttggctcacgacgagttgtaa

Ph-6ST

atgaaaaacttttattattaactttaatattacttactgcttgtaataattcagaagaaaatacacaatctattattaaaaatgatat
taataaaactattattgatgaggagtatgttaatttagagccaattaatcaatcaaacatctcttttacaaaacactcttgggtacaa
acttgtggtacgcaacaactattaacagaacaaaataaagagtcaatatcattatctgtagtggcgccacgattagatgacgatga
aaagtactgctttgattttaatggtgttagtaataaaggtgaaaaatatataacaaaagtaacattaaacgtagtggctccatctttt
agaggtttatgttgatcatgcatctcttccaactcttcagcagctaatggatattattaaatcggaagaagaaaatcctacagcaca
aagatatatagcttgggggagaatagttccgactgatgagcaaatgaaagagttaaatattacatcgtttgcattgataaataacc
atacaccagctgacttagtacaagaaattgttaagcaagcacaaacaaagcatagattgaatgttaaacttagctctaacactgct
cattcatttgataatttagtgccaatactaaaagaattaaattcgtttaataacgttacggtaacaaatatagatttatatgatgatg
gttcagcagaatatgtaaatttatataactggagagatacattaaataaaacagataatttaaaaattggtaaagattatcttgag
gatgtcattaatggtatcaatgaagacacttcaaatacaggaacatcatctgtttataactggcaaaaactatatccagctaactac
cattttttaagaaaagattatttaactttagaaccatcattacatgagttacgagactatattggtgatagtttaaagcaaatgcaat
gggatggtttcaaaaaattcaatagcaaacaacaagaattattcttatcgattgttaattttgacaaacaaaaattacaaaatgaa
tataattcatctaatttaccaaactttgtgtttacaggtacgactgtatgggctggtaaccatgaaagagagtattatgcgaaacaa

caaattaatgtcattaataatgcaattaatgaatcgagcccacattatttaggcaatagttatgatttgttcttcaaaggtcaccctg
gtggcggtatcattaatacattaataatgcaaaactatccttcaatggttgatattccatcaaaaatatcatttgaagttttgatgatg
acagatatgcttcctgatgcagttgctggtatagcgagctctttatatttcacgataccagctgaaaaaattaaatttatagtttttac
atcgacagaaactataactgatcgtgaaactgctttgagaagtcctttagttcaagtaatgataaaactaggtattgtaaaagaag
agaatgtacttttttgggctgatctgccaaattgtgaaacaggtgtttgtattgcagtctag