An *in silico* approach to predict alterations in the Golgi *N*-glycosylation machinery

Peter Fisher (MChem Hons)

PhD

University of York

Biology

September 2018

Abstract

Protein glycosylation is responsible for modulating the numerous properties of secreted and membrane proteins. The glycan structures that are produced by the cell in the Golgi apparatus are only partially a result of the transcriptome of the enzymes responsible. As the activity of glycosylation enzymes is specific and sequential, a complementary factor in regulating glycan structure is the organisation of these enzymes within the Golgi apparatus. However, the organisation of glycosylation enzymes, including levels and localizations within the Golgi apparatus has only been accessible with advanced microscopy techniques and is limited to a small number of enzymes. In this work a computational model capable of predicting changes in Golgi enzyme localizations and activities informed by glycan profiles has been developed. This model is used to predict the localization of *N*-linked glycan-modifying enzymes in three human cell lines and then to predict changes in enzyme homeostasis in Cog4-deficient cell lines and upon osteogenic differentiation. The results demonstrate the ability to infer Golgi enzyme organisation from glycan profiles and how this can be used to predict alterations to enzyme levels and localizations as a result of perturbation to the Golgi apparatus trafficking machinery or differentiation. This methodology could be used for the generation of tailor-made treatments for sufferers of congenital disorders of glycosylation, for example by suggesting appropriate monosaccharide supplements. The model also has the potential to link glycan function and Golgi-enzyme organisation to differentiation, allowing for targeted glyco-engineered bone regeneration therapies.

Table of contents

List of figures

List of tables

Acknowledgements

I owe a great debt of thanks to my three supervisors Dani, Jane and Jamie who have complemented each other in their supervision, given their support and shared their extensive knowledge with me throughout my time in York. I have always found a willingness to offer help and doors always open. I would also like to thank Nia and Julie for their input and analytical minds that made my thesis advisory panels so helpful. Many thanks also to Adam Dowle, Graeme Park and Ed Bergström for their help with instrument training and with technical issues. Thanks, is also due to the BBSRC for funding this work.

Huge thanks must also go to Jeni who has had to listen to my problems, triumphs and mistakes for the last four years. Undoubtedly this work would not have been completed without her. She has been supportive throughout the whole process and in her company my stress levels would decrease ten-fold. I would also like to thank my family, who have supported and encouraged me through every stage of education. Many thanks must also go to all members of the Cell Biology Laboratories at the University of York for their help, patience and advice. In particular I would like to thank Matthew Walker for always helping me out in the laboratory when required. Thanks also to Dimi for his excellent conversations which kept me sane throughout my PhD. I would also like to thank David and Paul for their friendships and rounds of golf.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. The work presented in this thesis was performed by the author between October 2014 and August 2018 in the Department of Biology, University of York, in the laboratory of Dr Dani Ungar or in the Department of Chemistry, University of York, in the laboratory of Professor Jane Thomas-Oates. All experiments were performed by the author, with the exception of a few glycan profiling experiments, performed by other members of the academic group. Figures that contain data from their work are clearly marked. All other sources are acknowledged as references. Neither this thesis nor any part of it has previously been submitted for acceptance of a higher degree.

Chapter 1: Introduction

1.1 Membrane trafficking

Mammalian cells are compartmentalised into membrane-bound organelles. Proteins and other biological material are often required to move between organelles and across lipid membranes. One manner in which macromolecules can be transported within the cell is through membrane trafficking. During membrane trafficking, molecules are encased in a lipid-based membrane such as a vesicle which is then transported to its destination. Vesicles contain specific cargo that does not mix with cytosolic proteins, and have designated sites of delivery. The secretory pathway makes use of membrane trafficking to transport *de novo* synthesised proteins from the ER to the cell exterior (figure 1.1) in order to modify the cell's external environment.
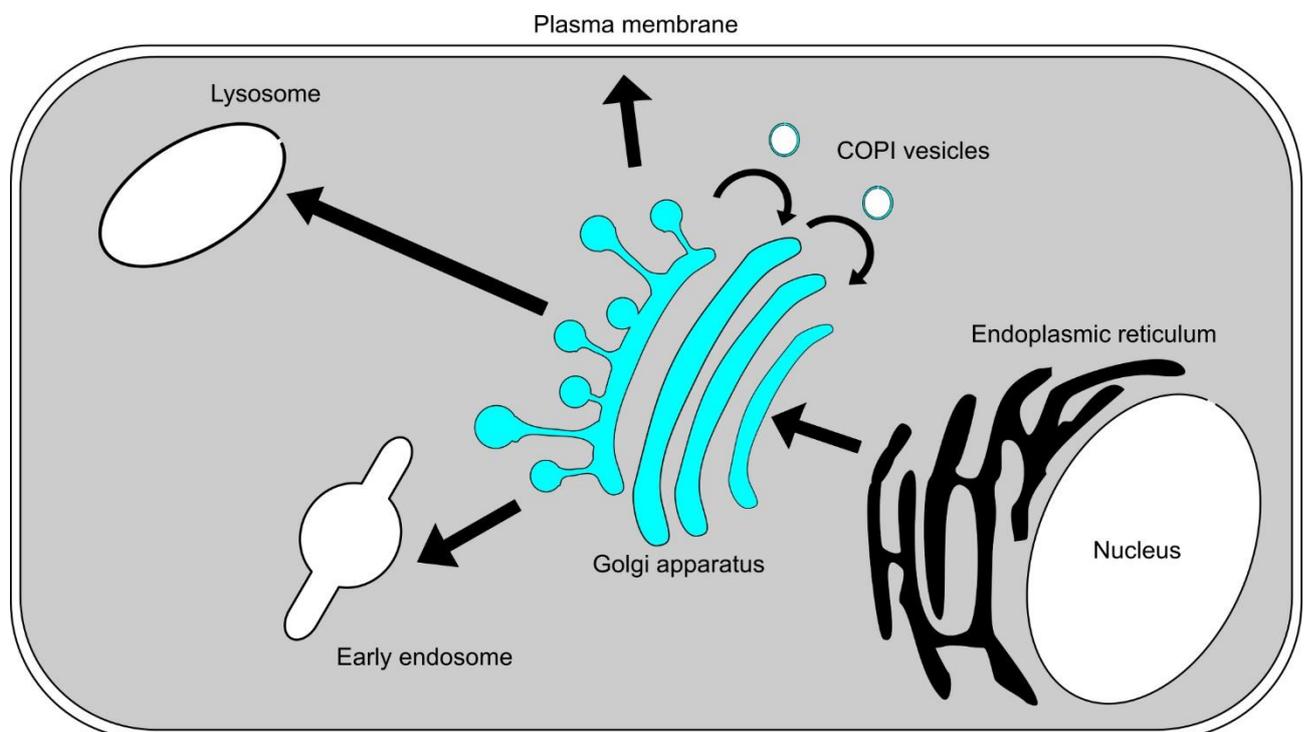
Figure 1.1 | Mammalian secretory pathway

Showing the route of secretory proteins from the ER to the Golgi apparatus which acts as the protein-sorting hub of the cells. After traversing the Golgi apparatus proteins are sorted and sent to other organelles where the proteins perform their function. Golgi-resident proteins are maintained in the Golgi by the action of coatomer protein complex-I (COPI) vesicles.

1.1.1 Golgi apparatus

The Golgi apparatus is a dynamic cellular organelle that was discovered by and named after the Italian biologist Camillo Golgi in the 1890s. The Golgi apparatus is made from flattened disc-shaped cisternae (1) that form the Golgi stack. Mammalian Golgi stack formation involves the complementary action of GRASP55 and GRASP65 (Golgi reassembly stacking protein of 55 kDa/65 kDa) (2). The Golgi stack may be further compartmentalised into the *cis* cisterna, *medial* cisterna, *trans* cisterna and the *trans* Golgi network (TGN). The *cis* cisterna of the Golgi stack faces the ER and receives newly synthesised proteins, while the *trans* cisterna and *trans* Golgi network are the exit sites from the Golgi apparatus. Typically for a mammalian Golgi stack, each cisterna is approximately 1 µm across (3) and the stack consists of between 3 and 8 cisternae (4).

In mammalian cells, the Golgi apparatus is peri-nuclear, is located near the centrosome and forms an interconnected twisted ribbon-like network, termed the Golgi ribbon (5). Vesicle traffic from the ER to the Golgi apparatus is dependent on the microtubule network. During mitosis the microtubule network becomes depolymerised, resulting in dispersion of the Golgi ribbon into mini-stacks (6). Golgi ribbon fragmentation can also be achieved with the microtubule depolymerising drug, nocodazole (7). The structure of the Golgi can vary between organisms. For example, in mammalian cells the Golgi stacks fuse with adjacent stacks to form the Golgi ribbon (8, 9). Ribbon formation has been demonstrated to be driven through GRASP65 interacting with the mammalian protein Enabled, to promote actin polymerization and facilitate GRASP65 oligomerization between adjacent Golgi stacks (10). The Golgi ribbon in mammalian cells differs from the Golgi of *Saccharomyces cerevisiae*, which appears as single dispersed cisternae (11), and also from the Golgi of *Drosophila melanogaster* and plants, which exist as individual Golgi mini-stacks (12, 13). Golgi morphology can also vary across cell types and pathologies (14, 15).

1.1.2 Intra-Golgi trafficking

The Golgi apparatus is the trafficking and sorting hub within the cell, that must be able to move secretory and cell surface proteins concurrently in the *cis* to *trans* direction (anterograde), while retaining and sorting resident Golgi proteins into the correct cisternae. A number of Golgi trafficking models have been suggested based on varying

amounts of evidence to describe the nature of intra-Golgi trafficking. Two of the most well-known and well-studied models are the anterograde vesicular model (16, 17) and the cisternal maturation model (8, 18). Other models of Golgi trafficking have been advocated (19, 20), but lack the volume of evidence compared with the vesicular and cisternal maturation models. That said, segregation of cargo prior to arrival at the TGN seems feasible (21) which could support a rapid partitioning-based model. The lack of consensus in the field on a Golgi trafficking mechanism is in part down to the inability of each model to explain all the experimental observations.

The anterograde vesicular model of Golgi trafficking is perhaps the most intuitive Golgi trafficking model. This model envisages the Golgi cisternae as static vessels and that each cisterna contains a unique set of Golgi-resident proteins that include glycosidases and glycosyltransferases (22), thus maintaining the polarity of the organelle. Proteins traversing the secretory pathway to the TGN are transported in COPI-coated vesicles that bud off from cisternae and move in an anterograde direction to adjacent cisternae, with which they fuse. Proteins that are to be retained in the Golgi cisternae are excluded from the COPI vesicles, thus preventing their exit from the Golgi apparatus. This model cannot explain analysis of COPI vesicles, which have been found to contain significant amounts of retrograde cargo including MAN2 (23). The vesicular model also cannot explain the secretion of larger cargoes, such as procollagen, which do not fit into COPI vesicles (24).

The cisternal maturation model addresses some of the weaknesses of the vesicular transport model. The cisternal maturation model views the Golgi cisternae as the transport carriers of proteins in the anterograde direction. The COPI-coated vesicles containing Golgi-resident proteins move in a retrograde direction in the face of the anterograde travelling cisterna, thus retaining these proteins in the Golgi apparatus (figure 1.2). A new *cis* cisterna is formed through the fusion of coatomer protein complex-II (COPII) vesicles leaving the ER. The newly formed *cis* cisterna matures into *medial*, then *trans*, then into the TGN, thus carrying proteins forward in the secretory pathway. The non-linear, unique distributions of Golgi-resident proteins can be explained either by differential recycling efficiencies (25) or by the targeting of specific cargo-containing vesicles to specific cisternae (26). From a modelling perspective the distinction is irrelevant as long as traffic functions.
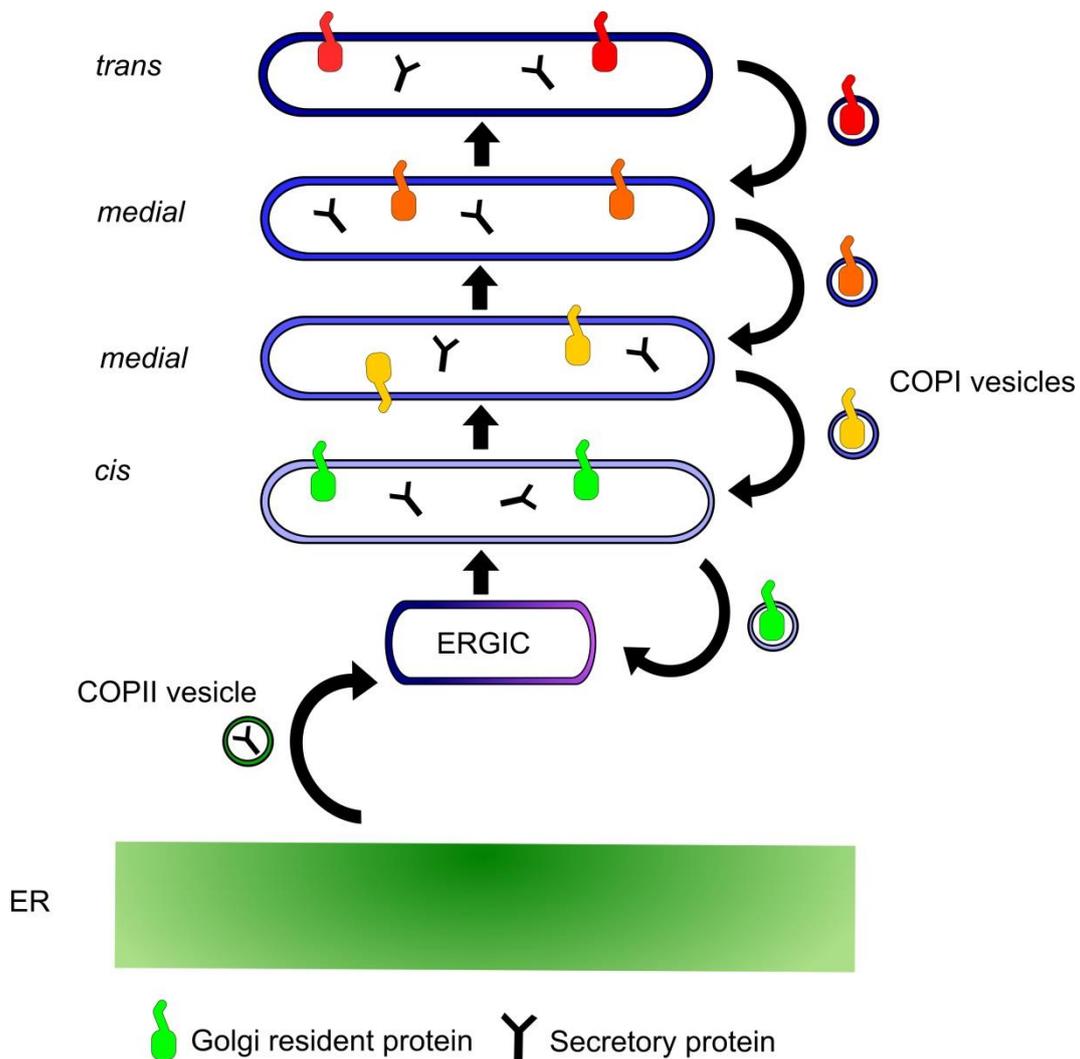
Figure 1.2 | Cisternal maturation model of Golgi trafficking

Secretory proteins are packaged into COPII vesicles and fuse with COPI vesicles to form the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). The ERGIC facilitates trafficking to the *cis* cisterna of the Golgi apparatus. The secretory proteins then remain within the Golgi cisternae which mature through the stack thus moving the secretory proteins to the *medial* and then *trans* face of the Golgi apparatus. In order to maintain and recycle Golgi-resident proteins such as glycosyltransferases, COPI vesicles transport these proteins in a retrograde fashion.

1.1.3 Conserved oligomeric Golgi complex

The conserved oligomeric Golgi (COG) complex is a hetero-octameric protein complex, consisting of subunits termed Cog1-8, that has been implicated in retrograde intra-Golgi vesicle trafficking. The COG complex is a member of the complexes associated with tethering containing helical rods (CATCHR) family of proteins (27, 28).

Early images of the COG complex revealed a structure with two similarly sized lobes (29) composed of two heterotrimeric assemblies, Cog2/3/4 and Cog 5/6/7 bridged by a Cog1/8 dimer (30). Following the elucidation of this bi-lobed structure of the COG complex, Cog1-4 subunits are referred to as lobe A and Cog5-8 as lobe B (figure 1.3A). More recent electron microscopy images suggest that a bi-lobed architecture of the COG complex may be simplistic, as the yeast lobe A resembles a "y" with flexible legs (31). The structure of the complete yeast COG complex displays multiple legs with curvature well suited for vesicle capture (figure 1.3B) (32). Crystal structures of the whole COG complex do not exist at the time of writing, however, the crystal structure of the Cog5-Cog7 sub-complex has been solved and is similar in structure to other multi-subunit tethering complexes (33).

A.

Lobe A **Lobe B**

B.

Lobe A **Lobe B**
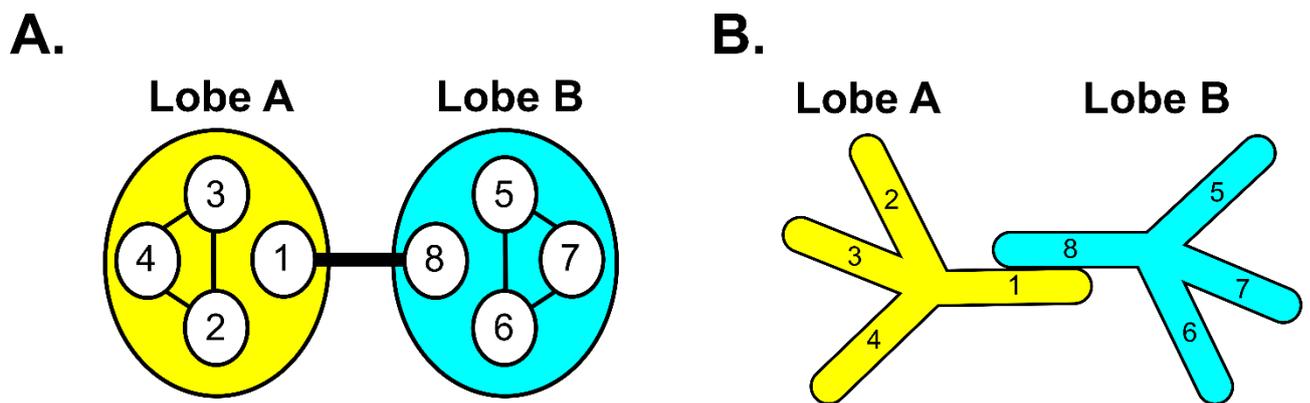
Figure 1.3 | Subunit organisation of the COG complex
Showing the original bi-loped (A) and the splayed "y" shaped (B) structures of the COG complex.

In the context of the cisternal maturation model of Golgi traffic, the COG complex is responsible for tethering COPI-coated vesicles to Golgi cisternae in order to facilitate the delivery and therefore retention of Golgi-resident proteins, including glycosylation enzymes. Interestingly, and in support of the COG complexes' role in retrograde vesicle trafficking, the vesicular stomatitis virus G protein (VSV-G), an anterograde cargo, was not altered as a result of Cog3 depletion, although the trafficking of a retrograde cargo was inhibited (34). It has been hypothesised that the COG complex plays a sorting role in the trafficking process, by ensuring that the glycosylation

machinery is recycled to the correct cisterna, thus maintaining the fidelity of glycosylation (26). Mitochondrial re-localization assays have been used to demonstrate the ability of different COG subunits to recruit different intra-Golgi vesicles. For example, Cog4 or Cog8 recruited vesicles containing different cargo to the mitochondria (35), supporting a role for the COG complex in sorting.

The COG complex has been shown to interact with a plethora of other proteins that are involved in Golgi trafficking (table 1). One family of proteins involved in Golgi vesicle trafficking are the golgins. Golgins are long coiled-coil proteins that extend out from the Golgi into the cytoplasm. The ability of golgins to extend into the cytoplasm makes them ideal for the initial tethering of COPI-coated vesicles, a function that has been demonstrated *in vitro* (36). Furthermore, different golgins have also been shown to target different subsets of vesicles (37) complementing the hypothesized role of the COG complex in vesicle sorting. The Rab family of proteins are small GTPases and can be found distributed throughout the cell, however particular Rab GTPases are found on specific parts of the Golgi apparatus (38). The role of Rab GTPases in membrane trafficking is considered to be in defining the discrete steps of the trafficking pathway by cycling between a membrane associated GTP-bound state and a GDP-bound off state (39, 40). Soluble *N*-ethylmaleimide attachment protein receptors (SNAREs) play a vital role in the final stage of vesicle trafficking, vesicle fusion. SNARE proteins are active on vesicles (v-SNARE) or target membranes (t-SNARE) (41). The coming together of cognate pairs of SNARE proteins and the resulting assembly of a SNARE complex drives the fusion of the two opposing membranes (42, 43). The interactions between the COG complex and different subsets of golgins, Rab GTPases and SNAREs result in the delivery of vesicles with specific cargo to target cisternae.

The mechanism of COG-dependent tethering or the existence of multiple mechanisms has not yet been proven. However, there is evidence that lobe A and B are present on cisternae or vesicular membranes respectively as independent sub-complexes (44). The assembly of the full COG complex drives vesicle tethering (44). The COG complex has also been shown to interact with the distal end of the long coiled-coil golgin Tata Modulatory Factor (TMF). This interaction may suggest a mechanism in which the initial tethering event is facilitated by TMF, but further docking of the vesicle with the

Golgi membrane is achieved by the COG complex (45), potentially through the coming together of lobes A and B.

Cog subunit	Interacting partner	Reference
Cog2	p115 CASP GM130 TMF Giantin Golgin84	(45–47)
Cog3	p115 GM130 Giantin Syntaxin-5 GS28	(47, 48)
Cog4	Syntaxin-5 Syntaxin-16 GS28 Vtila Vps45 Sly1 Rab1 Rab4 Rab30	(35, 49–51)
Cog5	Rab1 Rab2 Rab39	(45)
Cog6	TATA element modulatory factor (TMF) Syntaxin-5 Syntaxin-6 GS27 SNAP29 Rab1 Rab2	(45, 48, 51, 52)

	Rab4 Rab6 Rab10 Rab14 Rab36 Rab41	
Cog7	Golgin84 Syntaxin-6 Syntaxin-16 GS15 GS28 Vps45	(47, 49)
Cog8	CASP Syntaxin-5 Syntaxin-16 GS27	(35, 45, 49)

Table 1.1 | Interactions of the mammalian COG complex
Cog1 is known only to interact with Cog8.

The depletion or deletion of Cog subunits can have drastic effects in organisms. For example, in *D. melanogaster* the absence of *four way stop* (Cog5 homologue) results in male sterility (53), whereas Cog3 knockdown is embryonic lethal (54). Cog3 or Cog8 depletion in *A. thaliana* results in pollen tube growth defects (55) and in humans, mutations to the COG complex result in congenital disorders of glycosylation (discussed below). However, in *S. cerevisae* lobe B is expendable, with lobe A providing essential function (31), which suggests distinct functions for each lobe.

1.1.4 The COG complex and glycosylation

The COG complex has no known enzymatic activity, yet disruptions to the COG complex have been shown to distress the glycosylation machinery (56, 57) and result in alterations to glycosylation (58–61). Depletion of Cog subunits has been shown to result in the degradation of Golgi-resident proteins, including glycosylation enzymes

(62). Cog3 knockdown also resulted in the accumulation of untethered vesicular structures, which contained Golgi proteins involved in trafficking (34). Furthermore, in Cog4 KD HeLa cells myc-tagged MGAT1 was sequestered within the untethered vesicles (57).

There is much evidence suggesting that disruption of the COG complex leads to the mislocalization of Golgi-resident proteins away from the Golgi. Indeed, mislocalization of Golgi proteins through genetic disruption of the COG complex results in altered glycosylation. For example, CHO cells lacking the Cog1 or 2 subunit are unable to efficiently process glycans, resulting in significant alterations in the glycans produced (60, 63). Disruption of the binding interface between Cog5 and Cog7 has also resulted in alterations to glycosylation (33). Decreases in the sialylation of *N*-glycans were found to be the result of knocking down Cog3 or Cog4, which is in contrast to Cog6- or Cog8- knockdown, which led to slight increases in sialylation (57). This suggests different roles for each lobe and/or Cog subunit. This appears to be true in HeLa cells, where it has been demonstrated that lobe A contributes to the maintenance of Golgi morphology, whereas lobe B regulates *trans* localised Golgi-resident proteins such as β4GalT and ST6Gal1 (64). Other phenotypes of Cog-subunit deficiencies, such as abnormal Golgi morphology and defective sorting, cannot be explained by glycosylation defects (65).

1.1.5 Congenital disorders of glycosylation

Defects in glycosylation can lead to a rare set of genetic disorders known as congenital disorders of glycosylation (CDG). Due to the highly heterogenic nature of glycan biosynthesis, symptoms of CDGs are highly pleiotropic. The glycosylation machinery is encoded by 2% of the human genome (66) and, as such, several examples of mutations in the monosaccharide-nucleotide transporters and glycosyltransferases are known to result in CDGs (67–69).

The more common CDGs are those that come from defects in proteins that are directly involved in glycosylation. Due to improvements in diagnostics and whole exome sequencing, a subset of CDGs has been discovered that all arise as a result of mutations in the COG complex (table 2). It is worth noting that many of the COG mutations that will be discussed are found in single patients or very small numbers of

patients, meaning that phenotypes cannot be taken as a universal feature of the mutation. CDG-causing mutations have been identified in seven of the eight COG subunits, causing symptoms of variable severity, the Cog3 subunit being the exception. When comparing COG-CDGs with CDGs caused by defects in proteins directly involved in glycosylation such as MGAT2-CDG, an important consideration is that COG-CDGs will affect all forms of glycosylation (*N*- linked, O-linked and glycolipids). As previously discussed, the COG complex is involved in tethering and therefore sorting of glycosylation enzymes, thus COG-CDGs highlight the importance of Golgi trafficking in determining the cellular glycan profile.

CDG	Symptoms include:	Refs
CDG-COG1	Hypotonia, growth retardation, progressive microcephaly, hepatosplenomegaly	(70)
CDG-COG2	Microcephaly, seizures, liver dysfunction, psychomotor retardation, hypocupremia, and hypoceruloplasminemia	(71)
CDG-COG4	Mild mental retardation, mild dysmorphia, epilepsy, recurrent respiratory infections, mild ataxia	(59, 72, 73)
CDG-COG5	Moderate mental retardation with cerebellar atrophy, hypotonia	(74)
CDG-COG6	Early fatality, severe neurological impairment, seizures,	(75–77)

	vomiting, intracranial bleeding,	
CDG-COG7	Early death, dysmorphism, hypotonia, seizures, hepatomegaly, recurrent infections, cardiac failure, excessive skin	(78–80)
CDG-COG8	Mitral regurgitation, hypotonia, encephalopathy	(81, 82)

Table 1.2 | Congenital disorders of glycosylation and the COG complex

The severity of the consequences of COG mutations is highly varied, although hypomorphic mutations are more commonly found in lobe A subunits than in lobe B subunits, presumably as loss of function mutations in Cog1-4 subunits are embryonic lethal (31, 54). One example exists of a mutation with a severe outcome that truncates a lobe A subunit (Cog1), which results in reduced expression of Golgi α-mannosidase II and β-1,4 galactosyltransferase (70). Such a severe mutation in Cog1, that is not embryonic lethal may be explained by studies of a CHO cell line that are Cog1 null, in which normal levels of the remaining lobe A subunits are retained (83). In contrast, CDG-causing missense mutations in Cog2 and Cog4 reduce the levels of the other lobe A subunits (71, 72).

Unlike the hypomorphic mutations found in lobe A subunits, mutations that result in a larger loss of Cog subunit function are more commonly found in the lobe B subunits, although missense mutations are also found. A number of different mutations have been discovered in Cog5-CDG patients that result in different clinical phenotypes (74, 84). Disruption of the interaction between Cog5 and Cog7 results in defective

glycosylation in HEK293 cells (33). A weakened interaction between Cog5 and Cog7 may thus play a role in Cog5-CDGs. Mutations in the Cog6 subunit have been identified in a number of patients, each mutation associated with a different clinical outcome (77). For example, a fatal missense mutation (G549V) in Cog6 results in lobe B subunit instability and aberrant glycosylation (75). The same missense mutation is also found to not be fatal in a different patient, despite hyposialyltion of serum transferrin (76).

A complete loss of Cog7 resulted in death within 3 months for a CDG patient (78). The recycling of Golgi-resident proteins is perturbed in Cog7 deficient cells; for example, the mis-localization of giantin and GS15 (85) demonstrates the role of the COG complex in coordinating the molecular players involved in tethering and fusion not just of enzymes. Two truncations of the Cog8 subunit of a CDG patient were found to lead to a reduction in other lobe B subunits and Cog1 (81). The Cog8 subunit has been suggested to be a major SNARE hub. In support of this, cells derived from the Cog8-CDG patient and Cog8 knockdown HeLa cells both have reduced levels of the Golgi-associated SNAREs GS28 and GS15 (86).

This section aimed to give the reader background knowledge into Golgi trafficking and to detail the role of Golgi trafficking in determining the glycans that are produced by a cell. This is far from a complete picture of glycan function, which is beyond the scope of this thesis. In order to investigate how Golgi trafficking can influence glycosylation, the glycan profile of the cell needs to be observed. Mass spectrometry (MS) is a major technique commonly used for the analysis of glycosylation. This work makes extensive use of glycan profiles obtained from MS which will be discussed in the next section.

1.2 Glycosylation

Glycosylation is the process by which oligosaccharides are attached to proteins or lipids. Glycosylation is an immensely complex post-translational modification that is conserved in eukaryotes; complex because the manner of attachment and the structure of the glycan can vary enormously. To simplify the characterisation of protein glycosylation in mammals, the glycans are often classified as *N*-glycans, O-glycans or glycosaminoglycans (GAGs). This classification relies on the type of chemical bond

formed between the oligosaccharide chain and the protein, or in the case of GAGs, the glycan structure. In the case on *N*-linked glycosylation, an amide bond is formed between *N*-acetylglucosamine and an asparagine residue of the protein (figure 1.4A); for O-linked glycans and proteoglycans glycosidic bond is formed between a monosaccharide and a serine or threonine residue (figure 1.4B). Within this simple classification of protein glycosylation, there is a wide range of structural variety. Furthermore, a glycoprotein can have more than one glycosylation site and more than one individual glycan structure glycan attached to it.

A.   **B.**
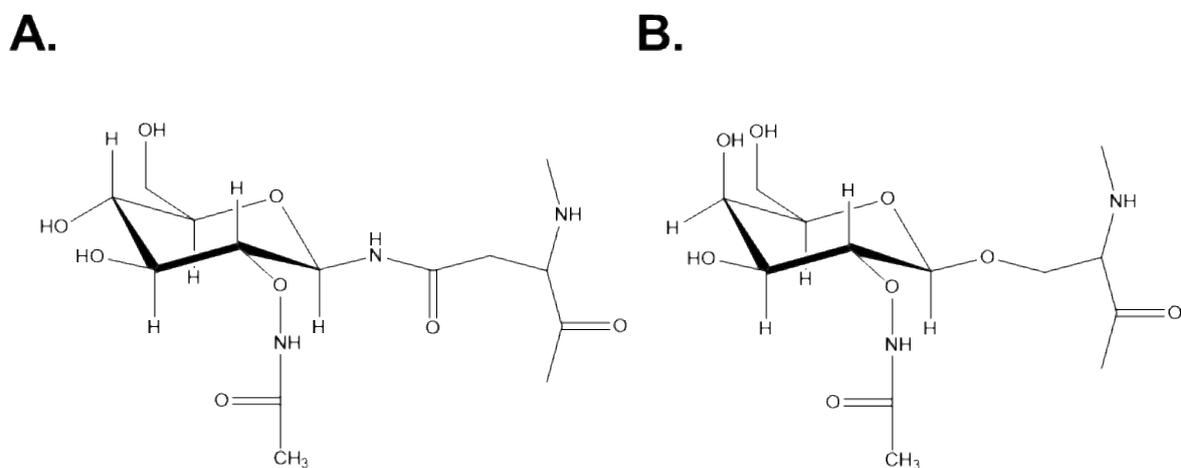
Figure 1.4 | *N*-linked and O-linked glycosylation

Showing the amide bond formed between *N*-acetylglucosamine and asparagine required to generate *N*-glycans (A); and the bond formed between a serine residue and *N* acetylgalactosamine in an example of O-glycosylation (B). Note O-glycans may also be attached to threonine.

The majority of secreted and membrane proteins are glycosylated (87). Given that the presence of glycans can attenuate the stability, alter the higher order structure and change the function of the protein they are attached to (88, 89), protein glycosylation challenges one of the central dogmas of biology: protein function is a result of protein structure which in turn is defined by the coding gene. At least in the case of glycoproteins, protein function and structure are not solely determined by their amino acid sequence. In this sense, protein glycosylation may be considered a post-translational modification that can result in multiple functional outcomes.

1.2.1 Structure and chemical properties of *N*-linked glycans

N-linked glycans found in mammalian cells typically consist of six different monosaccharides, all of which are hexoses and are shown in figure 1.5 in their pyranose form. In solution the majority of monosaccharides exist in the cyclic form but are in equilibrium with their acyclic form. The hexoses that make up *N*-glycans can be related to each other by epimerization or substitution of a hydroxyl group. For example, the addition of an acetamido group to position 2 of glucose (Glc) results in *N*-acetylglucosamine (GlcNAc). Similarly, epimerization (conversion of stereoisomers) of glucose at position 4 gives galactose.
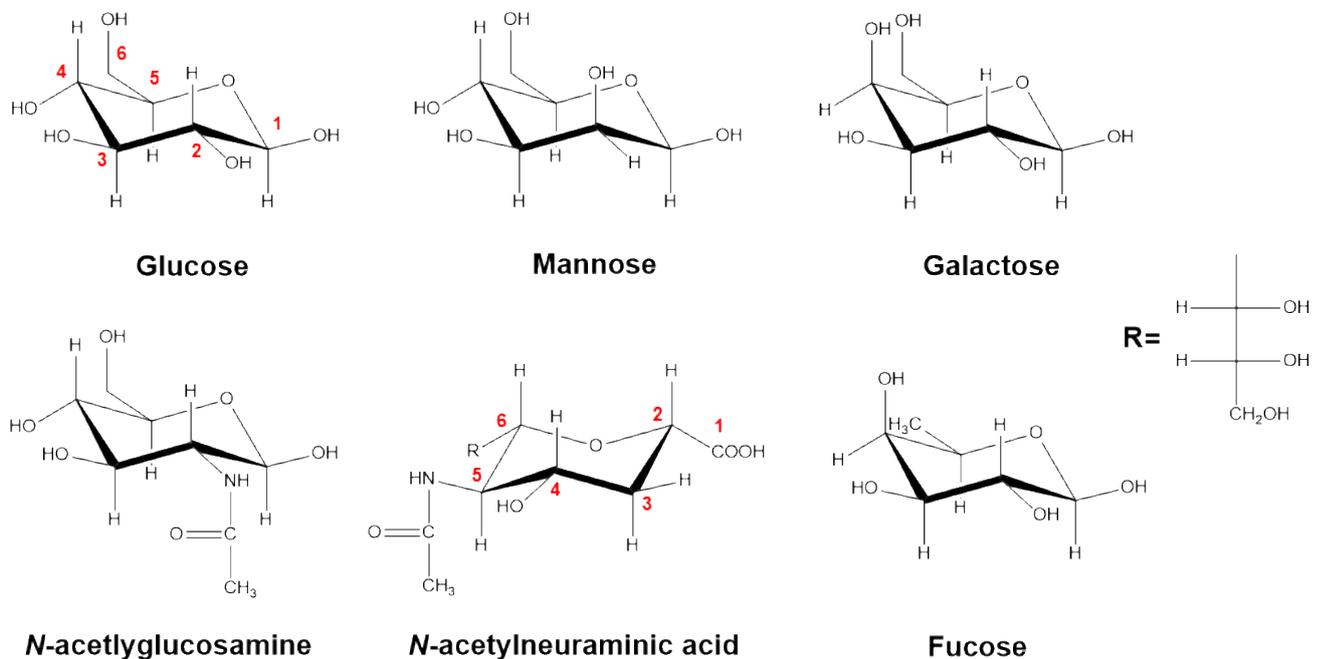
Figure 1.5 | Major constituents of mammalian *N*-glycans

The *N*-glycans of mammalian cells are made predominantly from the six monosaccharides shown. Each monosaccharide may be derived from glycose through epimerization or replacement of hydroxyl groups. The monosaccharides are shown in their beta configuration. The numbering of the carbon atoms in the monosaccharides is also shown.

In order to generate oligosaccharides, and therefore *N*-glycans, glycosidic bonds must be formed between monosaccharides, or between a monosaccharide and oligosaccharide. A glycosidic bond is formed through the reaction of a hemiacetal group and a hydroxyl group. The hemiacetal group at the anomeric carbon may form

a glycosidic bond with any of the hydroxyl groups present on the reacting monosaccharide, resulting in a range of possible regioisomers (figure 1.6A, B). The possibility of regioisomers vastly expands the diversity of oligosaccharides far beyond that of polypeptides. Furthermore, the availability of different hydroxyl groups allows branching to occur from a single residue (figure 1.6C) further increasing the structural diversity of glycans.
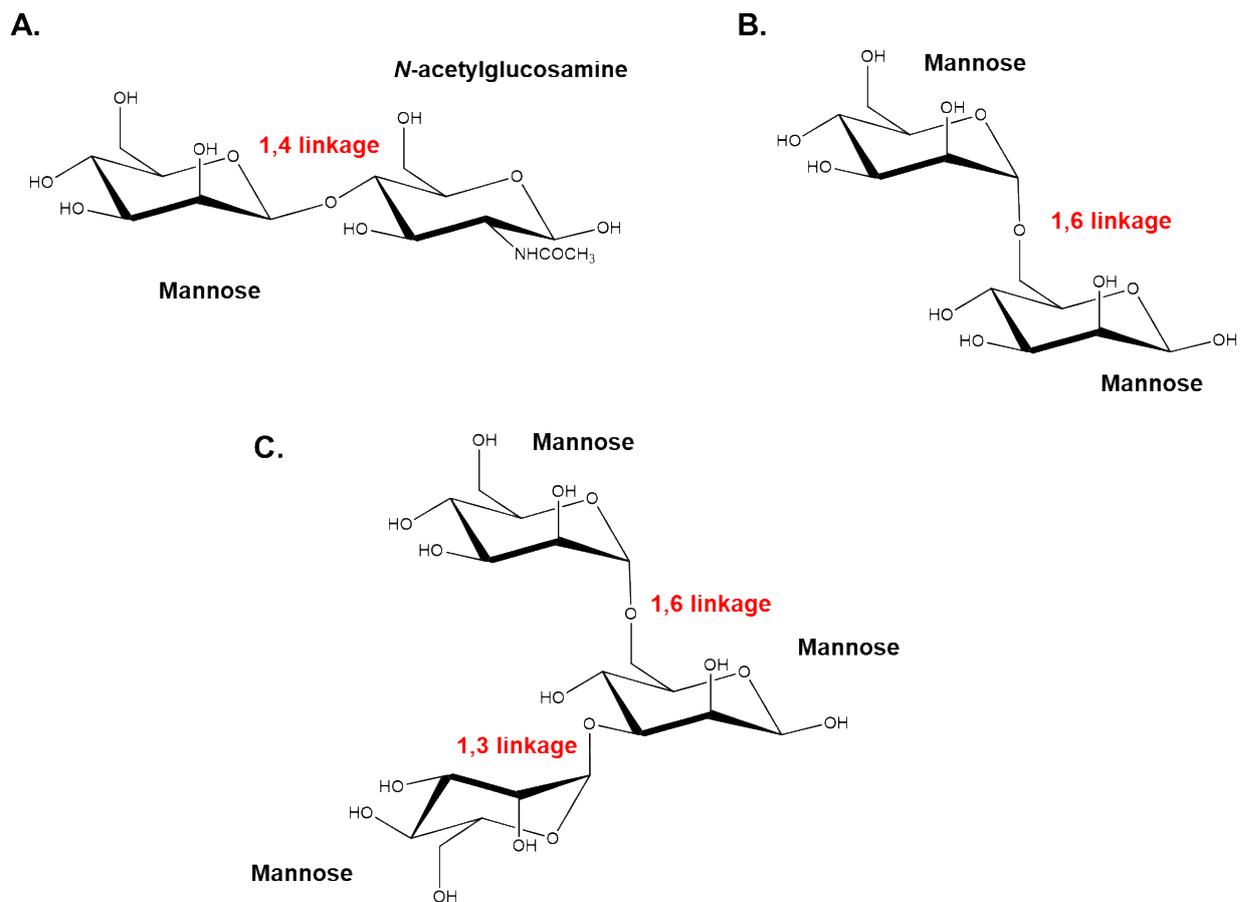
Figure 1.6 | Linkages and branching in oligosaccharides

Monosaccharides can form bonds between any of their hydroxyl groups. For example, a 1,4 linkage between *N*-acetylglucosamine and mannose (A); or a 1,6 linkage between two mannose residues (B). Oligosaccharides are not necessarily linear and may also form branches with two monosaccharides bonded to a common third monosaccharide (C).

In eukaryotic cells, *N*-linked glycans are attached to the side chain of an asparagine residue in an Asn-X-Ser/Thr sequon, where "X" cannot be proline. All *N*-glycans have a common tri-mannosyl chitobiose core structure consisting of two GlcNAc residues and three mannose (Man) residues. This core is derived by the action of glycosidases,

which trim residues from a precursor oligosaccharide, and glycosyltransferases, which add residues. The action of these enzyme families generates three classes of *N*-glycan: oligomannose glycans (figure 1.7A), which contain between two and six mannose residues in addition to the tri-mannosyl chitobiose; hybrid glycans (figure 1.7B), where a GlcNAc antenna is found alongside five mannose residues; and complex glycans (figure 1.7C, D, E), which do not contain any mannose residues beyond those found in the common core but do contain between two and five GlcNAc-initiated antennae. Antennae initiated by GlcNAc residues in either hybrid or complex type glycans may be further modified through the addition of galactose and sialic acid residues.
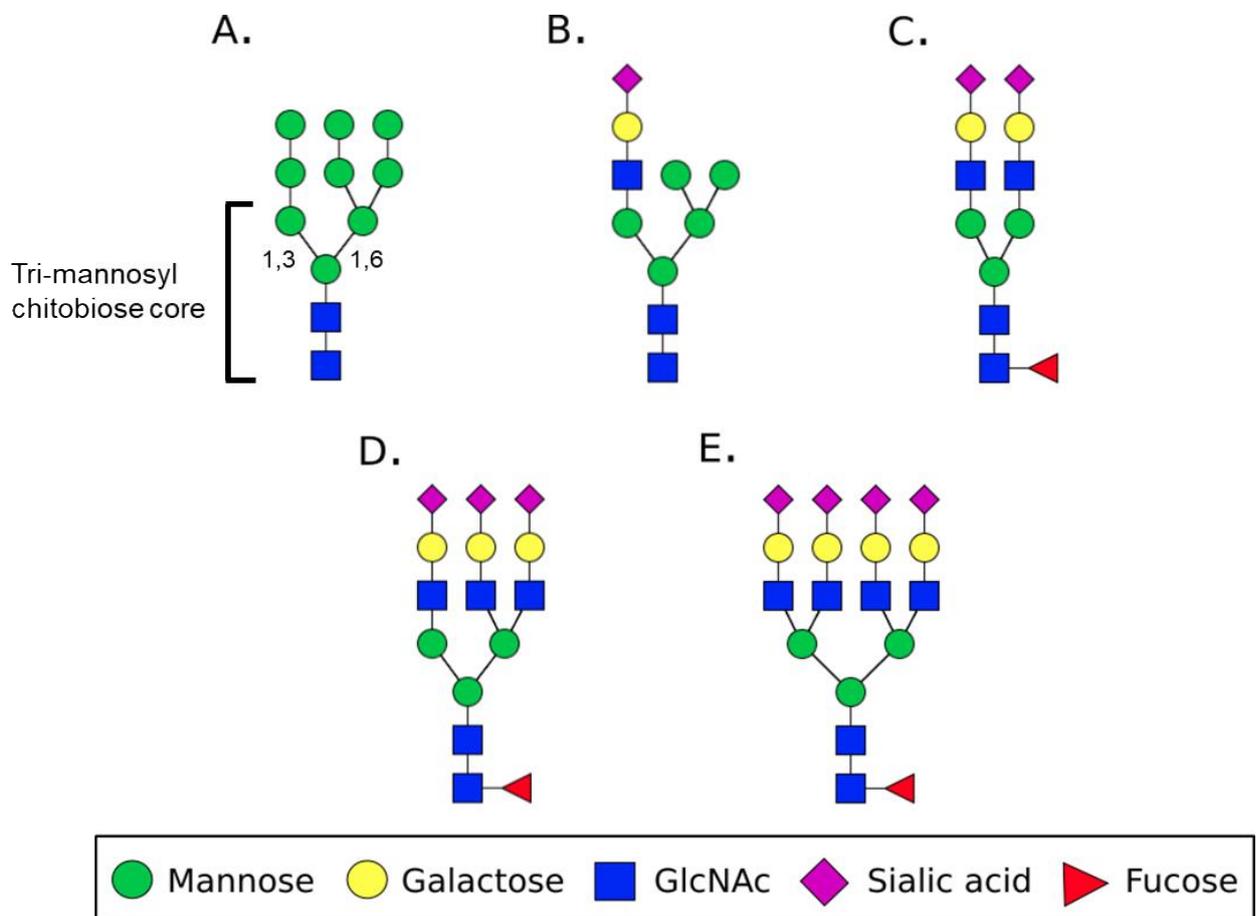
Figure 1.7 | Classes of *N*-linked glycans found in mammalian cells

The three main classes of *N*-linked glycans: oligomannose (A), hybrid (B) and complex (C, D, E). Complex glycans may have multiple GlcNAc seeded branches giving rise to bi-antennary (C), tri-antennary (D) and tetra-antennary (E) glycans. Symbol nomenclature is taken from the Consortium for functional glycomics which is from a version proposed by Kornfeld (90).

1.2.2 Biosynthesis of *N*-glycans

N-glycan biosynthesis is initiated at the endoplasmic reticulum (ER) where the precursor glycan, $GlcNAc_2Man_9Glc_3$ is built up through the addition of individual monosaccharides onto the polyisoprenol lipid dolichol, via a diphosphate bridge. The construction of the precursor glycan begins on the cytoplasmic side of the ER and involves the asparagine-linked glycosylation (ALG) family of enzymes. Following the synthesis of $GlcNAc_2Man_5$-P-P-dolichol on the cytoplasmic face of the ER, the glycan is transferred into the ER lumen by a flippase. The remaining monosaccharide residues are added to the developing precursor glycan in the ER lumen, before the multi-subunit oligosaccharyltransferase complex transfers the complete $Glc_3Man_9GlcNAc_2$ glycan *en bloc* to an asparagine residue of a newly translated protein.

Prior to the trafficking of the nascent glycoprotein to the Golgi apparatus, the terminal Glcs of the precursor glycan are sequentially removed through the action of ER glucosidases. The removal of these three Glc residues may be the first functional role of *N*-glycans in their lifetime: that of glycan-mediated protein quality control. Unfolded glycoproteins that contain more than one terminal glucose are retained in the ER through binding to calnexin and/or calreticulin before being marked for ER-associated degradation, therefore limiting the prevalence of misfolded and potentially aberrant proteins.

Following the trimming of the Glc residues and correct protein folding, the resulting protein carries $Man_9GlcNAc_2$. The $Man_9GlcNAc_2$ glycan may now be trafficked to the Golgi apparatus, or may be trimmed further to $Man_8GlcNAc_2$ through the action of an ER mannosidase I enzyme. Upon entry to the Golgi apparatus, the glycoprotein is exposed to a variety of glycosidases and glycosyltransferases, which modify the *N*-glycan. Early processing of the *N*-glycan involves the sequential removal of mannose residues from the $Man_9GlcNAc_2$/$Man_8GlcNAc_2$ structure to yield $Man_5GlcNAc_2$ (figure 1.8). Oligomannose glycans can also be a target for GlcNAc-phosphotransferase, the action of which, following that of an uncovering enzyme, results in a phosphorylated glycan. Phosphorylation of oligomannose glycans halts further glycan processing and targets the modified glycoprotein to the lysosome rather than the plasma membrane.

Hybrid type glycans are generated through the action of alpha-1, 3-mannosyl-glycoprotein beta-1, 2-*N*-acetylglucosaminyltransferase (MGAT1) on $Man_5GlcNAc_2$ (figure 1.8). MGAT1 adds a GlcNAc residue to the 1, 3 Man arm of the $Man_5GlcNAc_2$ glycan. This GlcNAc residue can be added to further, usually with the addition of galactose (Gal) and sialic acid (NeuAc). Alternatively, the remaining mannose residues, which do not make up the tri-mannosyl chitobiose core, can be cleaved by the enzyme mannosidase II (MAN2), followed by the addition of a second GlcNAc residue by alpha-1, 6-mannosyl-glycoprotein beta-1, 2-*N*-acetylglucosaminyltransferase (MGAT2), to generate a complex type bi-antennary glycan. Fucose (Fuc) may also be added, most usually to the innermost GlcNAc residue of *N*-glycans by fucosyltransferase 8 (FUT8). A complex bi-antennary glycan can be further modified with additional antennae to yield tri- and tetra-antennary glycans through the action of alpha-1, 3-mannosyl-glycoprotein beta-1, 4-*N*-acetylglucosaminyltransferase (MGAT4) and alpha-1, 6-mannosyl-glycoprotein beta-1, 4-*N*-acetylglucosaminyltransferase (MGAT5). Each terminal GlcNAc residue may be galactosylated through the action of a galactosyltransferase (GalT) which, in turn, can be capped with a sialic acid. The process of modifying the precursor glycans that exit the ER produces glycan structural heterogeneity. It is important to note that the promiscuity of the glycosylation enzymes gives rise to a divergent reaction network that is not shown in figure 1.8. For example, following the first few trimming reactions, the number of possible glycan structures increases exponentially, until levelling off upon the addition of terminal sialic acids, usually around the 16th reaction (91). The result of this reaction network is numerous different reaction pathways that end up generating the same glycan, hence only a relatively low number of glycan species are observed when compared with the number of immature intermediate structures that can exist.
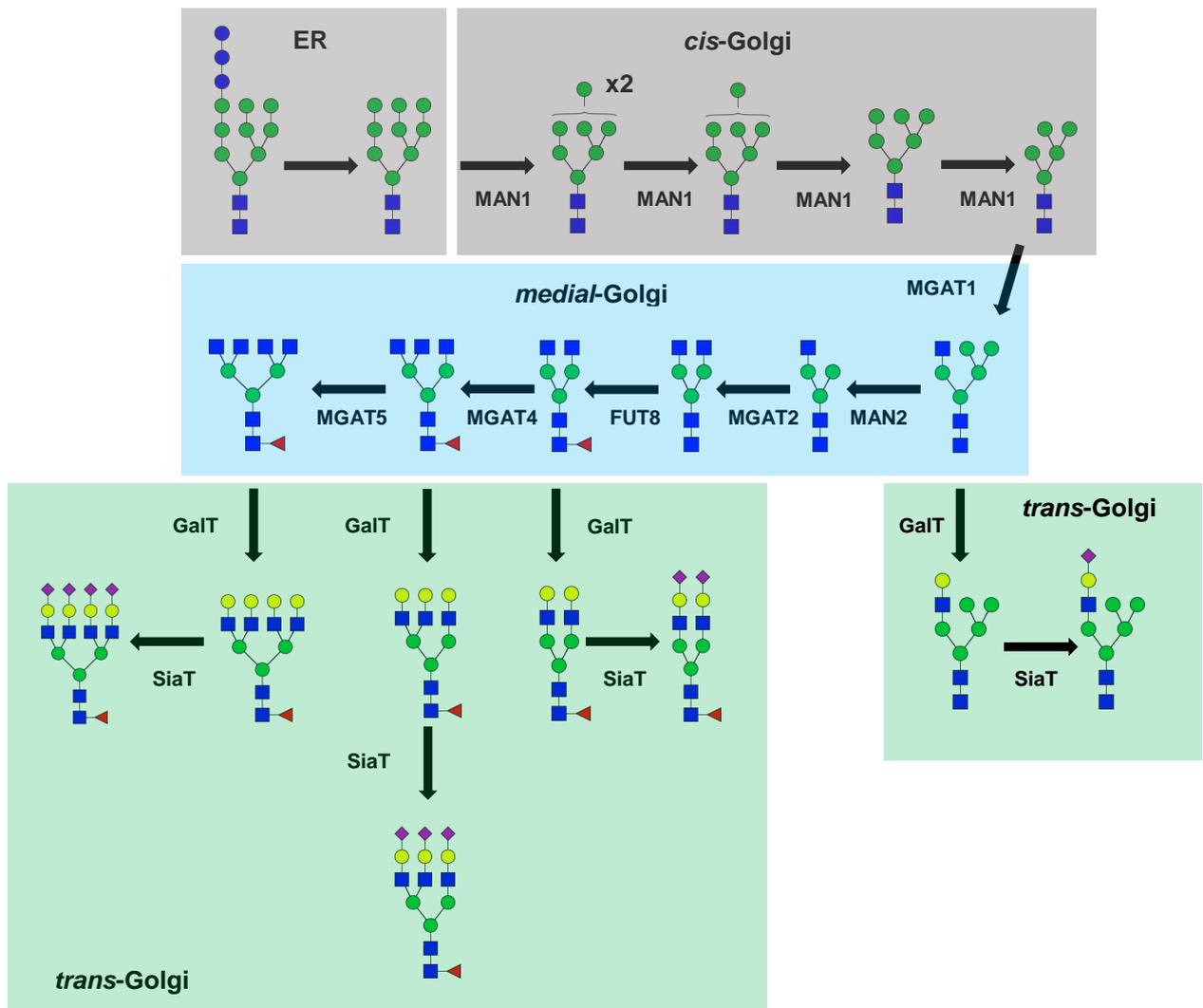
Figure 1.8 | *N*-glycan biosynthesis in the mammalian Golgi apparatus

A 14-monosaccharide precursor is built in the ER and transferred *en bloc* to glycoproteins before their transport to the Golgi. Mannosidase I is active in the *cis* cisternae resulting in mannose trimming. Reactions in the *medial* cisternae include further mannose trimming and GlcNAc addition. Galactose and sialic acid residues are added in the *trans* cisternae.

1.2.3 Glycosidases and glycosyltransferases

The Golgi apparatus is the organelle residence of a variety of enzymes tasked with generating the mature glycan structures required by the cell to function. These enzymes act sequentially on the precursor *N*-glycan that exited the ER and can be classed as either glycosidases or glycosyltransferases. The formation of a new glycosidic bond is catalysed by the action of glycosyltransferases, where each glycosyltransferase is specific for a monosaccharide nucleotide donor and an acceptor glycan. In many cases it is not just the acceptor residue the glycosyltransferase is

specific for, but also the extended *N*-glycan structure (92, 93). Despite the structural substrate specificity of the glycosylation enzymes, the enzymes remain somewhat promiscuous, otherwise there would be limited glycan structural heterogeneity. Glycosidases are responsible for breaking glycosidic bonds and do not require a donor substrate. However, similar to glycosyltransferases, glycosidases are specific in hydrolysing bonds between particular residues. The following section highlights the properties and functions of the major *N*-glycan-modifying enzymes that are relevant to the work described in this thesis.

1.2.3.1 Mannosidase I

The Golgi mannosidase I (MAN1) enzyme acts early in the maturation pathway of *N*-linked glycans. MAN1 cleaves terminal α-1,2-linked mannose residues, which is required for the generation of hybrid and complex type *N*-glycans. An ER member of the same enzyme family is capable of trimming $Man_9GlcNAc_2$ to $Man_8GlcNAc_2$ (94); however, the Golgi MAN1 glycosidases are required for further processing of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$ *in vivo*. There are three isoforms of the Golgi MAN1 enzyme (95–97), each capable of cleaving α-1,2-linked mannose residues, but with different efficiencies for each terminal mannose residue in $Man_8GlcNAc_2$. For the ER MAN1 and Golgi MAN1, the different preferences for each terminal mannose can be explained by distinct modes of branch insertion into the enzyme active site (98).

Interestingly, the oligomannose glycan itself also has an effect on the efficiency of α-1,2 mannose cleavage by the mannosidase enzyme. Particularly relevant to the work presented in this thesis, is that trimming of $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$ occurs at a much slower rate than the preceding cleavage reactions (99). This property of MAN1 is potentially responsible for the "doubly peaked" distribution of the oligomannose glycans observed in glycan profiling experiments (figure 1.9) and is particularly pertinent when trying to model glycan profiles as discussed later (section 3.4).
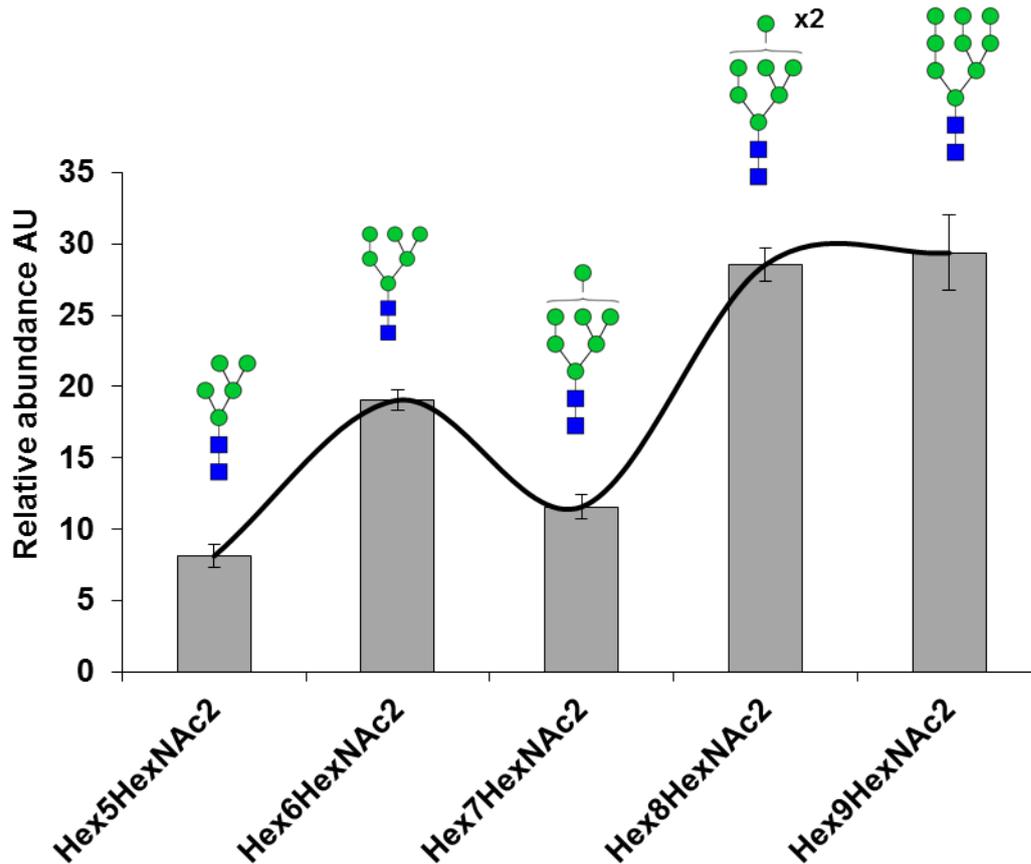
Figure 1.9 | Non-Poisson distribution of oligomannose glycans from HEK293T cells

Relative glycan abundance of the oligomannose glycans from HEK293T cells observed by mass spectrometry. The distribution shows a distinctive two peaked shape centred around Hex8HexNAc2 and Hex6HexNAc2.

1.2.3.2 MGATs

MGAT1, 2, 4 and 5 are responsible for initiating GlcNAc antennae in hybrid and complex-type glycans. MGAT1 is the earliest acting member of the family and is responsible for the transfer of GlcNAc from UDP-GlcNAc to $Man_5GlcNAc_2$ (figure 1.5), generating a hybrid-type glycan (100). Without the presence of MGAT1, neither hybrid nor complex type *N*-glycans can be synthesised in the Golgi apparatus. Unsurprisingly then, the absence of MGAT1 is embryonic lethal to mice (101).

MGAT2 acts after MAN2 to generate bi-antennary complex glycans (102), which may then be followed by MGAT4 and MGAT5 to create tri- and tetra-antennary glycans (figure 1.5). In the case of MGAT4 and MGAT5, the ability of each enzyme to initiate further branching of complex glycans was reduced if a galactose residue was already

present (103, 104). Branching multiplicity, which is dictated by these enzymes, has been shown to be important for a number of functions including galectin binding (105).

1.2.3.3 Mannosidase II

The conversion of hybrid-type glycans to complex glycans is achieved through the removal of two mannose residues from the α-1,6 arm of GlcNAcMan$_5$GlcNAc$_2$ (figure 1.5) by the glycosidase mannosidase II (MAN2). The action of MAN2 allows the addition of a second GlcNAc residue to form a complex bi-antennary glycan. As MAN2 is required for the biosynthesis of complex glycans, MAN2 inhibition using the indolizidine alkaloid, swainsonine (106), can allow researchers to study the role of complex *N*-glycan structures in a number of systems. For example, swainsonine treatment has been used to study the role of *N*-glycans in differentiation and proved ineffective in impacting the ability of mesenchymal stromal cells to differentiate into osteoblasts (107).

1.2.3.4 Galactosyltransferases

The addition of galactose to glycans is performed by galactosyltransferases, which encompasses a number of enzymes that are capable of making different linkages between galactose and other residues. One member of the galactosyltransferase subfamily is β1,4-galactosyltransferase I, which transfers galactose to terminal GlcNAc residues on hybrid and complex *N*-glycans, thus generating the lactosamine (Gal-GlcNAc) signal important for galectin recognition (108). One of the properties of β1,4-galactosyltransferase I, that takes on significance when simulating glycosylation reactions, is it's preferences for the different branches of bi- and tri-antennary complex glycans (92).

1.2.3.5 Sialyltransferases

The human genome encodes 20 sialyltransferases (109) that transfer sialic acid from the donor molecule CMP-sialic acid onto glycoproteins and glycolipids. Each member of the sialyltransferase family has different substrate specificities and/or is capable of catalysing the formation of different linkages between the oligosaccharide and sialic acid (109). In the context of *N*-glycosylation, the addition of sialic acid is often the final capping modification that occurs on a galactose residue.

The terminal position of sialic acids in *N*-glycans often confers functionality. For example, the presence of an α2,6 linked sialic acid in the *N*-glycans of immunoglobulin G (IgG) proteins has been linked to anti-inflammatory properties (110). Interestingly, the enzymatic action of the ST6Gal1 enzyme shows a preference for the galactose that resides on the 1,3Man branch (111). Another interesting property of sialyltransferases is that they are themselves glycoproteins and their activity is influenced by their glycans (112). The inability of a cell line to produce galactosylated glycans resulted in a 60% decrease in the catalytic activity of sialyltransferases (112). Other examples of glycan structure regulating the activity of enzymes have been reported and will be discussed later.

1.2.3.6 Fucosyltransferase 8

FUT8 is responsible for the addition of fucose from the GDP-fucose donor to the tri-mannosyl chitobiose core (113). A comprehensive evaluation of FUT8 specificity *in vitro* was able to demonstrate a greater catalytic efficiency for bi-antennary glycans and a general promiscuity towards modifications such as galactosylation, to the 1,6Man branch (93). Evidence that FUT8 can act to fucosylate oligomannose-type glycans, in particular $Man_5GlcNAc_2$, has been contentious. Several studies have demonstrated that FUT8 cannot fucosylate $Man_5GlcNAc_2$ *in vitro* (93), although other studies have shown that, given the correct circumstances, $Fuc_1Man_5GlcNAc_2$ can be produced *in vivo* (114, 115).

1.2.3.7 Antenna fucosylation

In addition to fucosylation of the tri-mannosyl chitobiose core, the antennae of complex type glycans can also be fucosylated (figure 1.10). For example, fucosyltransferases 3 and 4 are capable of transferring a fucose residue to GlcNAc residues in the antennae or GlcNAc residues within a poly-lactosamine disaccharide repeat. An important function of antenna fucosylation is to create the Lewis X and sialyl-Lewis X epitopes in *N*-glycans (figure 1.10), although the epitope can also be found in O-linked glycans and glycosphingolipids. Sialyl-Lewis X is a ligand for E-selectin, an important player in inflammation (116).
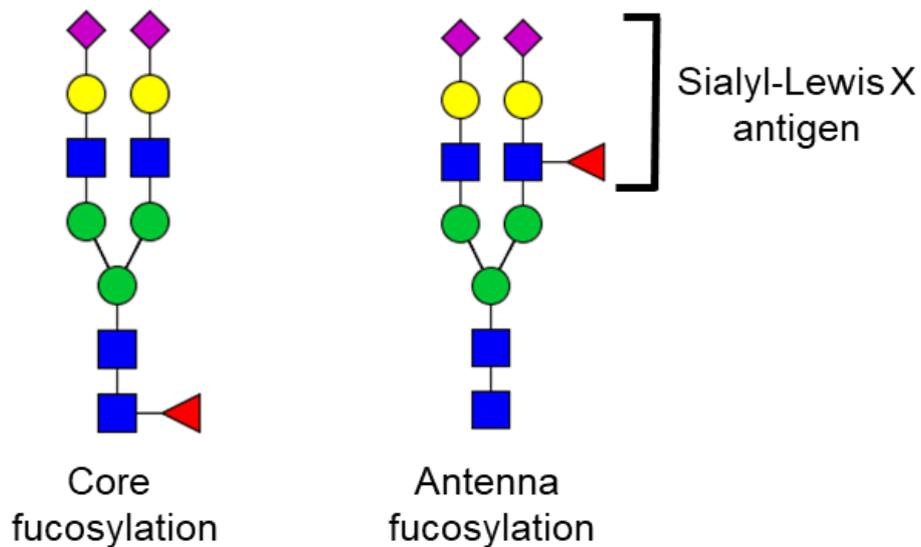
Figure 1.10 | Different types of fucosylated *N*-glycans

Mammalian fucosylation may occur on the reducing GlcNAc of the tri-mannosyl chitobiose core (core fucosylation) through a 1,6 linkage or on the GlcNAc residue of branched complex glycans (antenna fucosylation) which can give rise to the sialyl-Lewis X epitope.

1.2.4 *N*-glycan microheterogeneity

Glycans have a large degree of microheterogeneity; in other words, different glycans may be found at the same glycosite on two different molecules of the same protein (117). Unlike the synthesis of proteins, the biosynthesis of *N*-glycans is not template driven. Yet this heterogeneity is not random, as different cell types show distinct and reproducible glycan profiles. Rather than being determined by a template, the *N*-glycans produced by the cell are influenced by a number of factors, including: enzyme levels, enzyme localization, monosaccharide-nucleotide availability and features of the glycoprotein. The fact that there is only a loose correlation between transcript levels of glycosylation enzymes and the glycan structures produced by the cell, is evidence that glycan structure is not solely determined by enzyme levels (118, 119).

A major source of microheterogeneity is enzyme competition within the Golgi apparatus. In the absence of a glycan template and in order to control heterogeneity, the process of glycosylation is dictated by the unique morphology and sub-compartmentalisation of glycosylation enzymes within the Golgi apparatus (22, 26). It is therefore of great importance to be able to infer the levels of glycosylation enzymes, as well as how they are compartmentalised, to understand the glycosylation process.

Despite glycan microheterogeneity, glycans play important roles in protein functions, discussed in the following section.

1.2.5 Biological roles of glycans

Approximately 1-2% of the encoding human genome produces proteins involved in glycosylation (120). With this significant proportion of the genome dedicated to glycosylation, it follows that glycans must play an important and essential function in biology. Despite the obvious abundance and importance of glycans, very little was known regarding their specific individual biological functions until the late 1980s (121). The full spectrum of glycan functions is slowly being revealed and, with such structural diversity, glycans have been found to play roles in everything from protein folding to cell adhesion and cell signalling. While a complete review of the functional roles of glycans is beyond the scope of this thesis, the following section highlights a handful of processes in which glycans can play a key part.

1.2.5.1 Glycoprotein steady-state levels

Glycans are naturally hydrophilic and can affect the physical properties of the protein they are attached to. Protein instability as a result of improper folding has been demonstrated in cells treated with the *N*-glycosylation inhibitor, tunicamycin (122). Furthermore, deletion of glycosylation sites in the interleukin-6 signal transducer, gpp130, resulted in a more unstable protein without affecting its ability to signal (88). Altered glycosylation can also reduce the steady state levels of cell surface proteins without affecting protein stability. For example, the steady state level of the low density lipoprotein receptor (LDLR) is reduced in a cell line with disrupted glycosylation (63), likely due to increased endocytosis and subsequent degradation.

Glycans can also act in a signalling role for improperly folded proteins, as part of ER quality control. In the ER, $Glc_3Man_9GlcNAc_2$ is trimmed to $Man_9GlcNAc_2$ before re-glucosylation to the monoglucosylated form by an ER glucosyltransferase, that acts as a sensor of misfolded proteins. This glucosyltransferase has been shown to recognise core GlcNAc and hydrophobic protein domains, each potential indicators of an unfolded protein (123). The terminal glucose of the *N*-glycan is recognised by calnexin and calreticulin (124), thus trapping misfolded proteins in the ER (125) prior to ER-associated degradation.

1.2.5.2 Signalling and protein function

In addition to an intra-cellular signalling role in protein quality control, *N*-glycans have also been shown to impact signalling on the cell surface. Glycans can act on signalling pathways in different ways, either by modulating the ability of the glycoprotein receptor to signal, or by the glycan acting as a recognition signal itself. A number of studies have highlighted the role of fucosylation and sialylation in cell signalling. For example, the TGF-beta1 receptor from FUT8-null mice showed reduced binding and signalling as assayed through Smad2 phosphorylation (126). Furthermore, sialylation and antenna fucosylation of the EGF receptor inhibits dimerization and downstream phosphorylation (127). The role of glycan structures in modulating the signalling activity of these receptors is of great relevance in the cancer field, as increased sialylation and fucosylation levels have been implicated in tumour growth and metastasis (128–130).

Glycans can be recognised by lectins, which are capable of binding specific aspects of glycan structures. In mammalian cells the most ubiquitous lectin family is the galectins (131), which recognise the Galβ1,4GlcNAc moiety of glycans. Galectin binding to the glycan can form a galectin lattice on the cell surface which: prevents endocytosis, increases the residence time of the receptor on the cell surface, and alters the degree of signalling. The galectin binding strength to the Galβ1,4GlcNAc moiety is weak, but by increasing the instance of the galectin ligand, either with Galβ1,4GlcNAc repeats or *N*-glycan antenna number, the cell has the ability to alter signalling through galectins (132). Furthermore, the balance between the *N*-glycan antenna number on arrest-promoting proteins such as TGFβR and growth-promoting proteins such as EGFR, has been shown to influence the fate of the cell (133) by regulating galectin binding and signal response.

Enzymatic properties such as activity, binding affinity and specificity may all be influenced by *N*-glycosylation (134). For example, glycan-deficient mutants of the lysosomal hydrolase β-glucosidase, a defect in which can lead to Gaucher disease, showed reduced enzymatic activity (135). Interestingly, the glycosylation of several glycosyltransferases has been found to impact the activity of the enzymes (112, 136, 137). The transferase activity of two members of the sialyltransferase family has been found to be sensitive to *N*-glycosylation; the presence of galactose, for instance, was

found to increase the activity of ST6Gal1 (112). In general, studies into the effects of glycosylation on enzymatic activity compare the effects of the presence or absence of glycans, an unlikely situation in biology, rather than comparing the effects of different glycan structures. More detailed studies altering glycan structure on such enzymes will be more revealing and useful in disease scenarios.

Cell adhesion is an important property of cells and is also of interest to cancer biologists, as it has been linked to the migratory behaviour of cells, an activity increased in cancer metastasis. Epithelial cell-cell adhesion can be manipulated through alteration of the glycosylation of E-cadherin (138). Overexpression of MGAT5 in MKN45 cells resulted in an impairment of E-cadherin-dependent cell-cell adhesion (139) and reduced levels of MGAT5 suppresses tumour metastasis in mice (140). Finally, a comparison of CHO cells predominately synthesising either complex, hybrid or oligomannose glycans, demonstrated that both the strongest cell-cell adhesion and fastest migratory rates were for oligomannose-producing cells (141). This result suggests that *N*-glycan structure plays an important role in determining cell-cell adhesion and migration.

1.2.5.3 Glycans and cell differentiation

The role of glycans in mammalian differentiation processes is not fully understood, although there are large differences between the glycan profiles of undifferentiated and differentiated cells (107, 142). Several alterations to *N*-glycan structures have been observed in induced pluripotent stem cells, including: an increase in oligomannose glycans and a shift from α2,3 to α2,6-linked sialic acid-terminated glycans (143). Likewise, an increase in fucosylated bi-antennary glycans has been shown to accompany adipogenesis (142). Interestingly, the presence of α2,6-linked sialic acid has been identified as a potential marker of the osteogenic and adipogenic differentiation potential of mesenchymal stromal cells (MSC) (144).

Key receptors and ligands that are involved in osteogenesis, such as Wnt3a and BMP are glycoproteins (145, 146). Recently, the mechanistic role of glycosylation in osteogenesis has begun to be uncovered (147). Treatment with a MAN1 inhibitor (kifunensine) of an immortalised MSC line capable of differentiating down a number of lineages (148), promoted mineralisation in osteoblasts (147). Increased mineralisation

in osteoblasts, as a result of early inhibition of *N*-glycan processing, was due to altered PI3K-mediated signalling (147). The authors suggest that glycosylation can act to fine tune PI3K signalling influencing osteogenic potential of MSC.

1.2.5.4 Glycosylation and biologics

Over 60% of biopharmaceuticals are glycosylated in some form (149). Glycosylation can alter the physical and pharmaceutical properties of biologics, which can often vary from batch to batch. Therefore, a key goal in the growing field of biotherapeutics is to be able to glyco-engineer homogeneous glycoforms of proteins with optimal pharmaceutical properties, such as half-life and activity. An early example of this was the creation of a FUT8-knockout CHO cell line, which generated defucosylated antibodies with higher antibody-dependent cellular cytotoxicity (ADCC) than fucosylated glycoforms (150, 151). The half-life of monoclonal antibodies (mAbs) has also been found to be a function of glycan structure. MAbs containing complex glycans have a longer half-life *in vivo* when compared with those exhibiting oligomannose-type glycans (152)

The goal of this section has been to highlight the diverse and varied roles *N*-glycans can play in biology. This is far from a complete picture of glycan function, which is beyond the scope of this thesis. Many observations have been made regarding alterations in glycan structure upon some biological process such as differentiation or disease development. Less clear is how such glycan alterations enable the biological process to occur or whether changes in glycosylation are causal or symptomatic. Fortunately, many groups are pursuing these questions.

1.3 Mass spectrometry

MS is a technique that allows researchers to measure the mass to charge ratio of ions (m/z). In a biological context, the ability to measure the (m/z) of analyte ions has a wide range of applications, from protein characterisation to glycan analysis. The basic components of a modern mass spectrometer are an ionisation source, mass analyser(s) and detector(s). Matrix-assisted laser desorption/ionisation (MALDI)-MS has been used extensively in this body of work to obtain cellular glycan profiles. Other prominent methods of glycan structure determination are liquid chromatography

(LC)-MS and capillary electrophoresis. The aim of the following section is thus to describe the fundamentals of the primary MS technique that was used in this study, and also to demonstrate how the technique of MALDI-MS has advanced the field of glycomics.

1.3.1 Matrix-assisted laser desorption/ionisation (MALDI)

In order to measure the (m/z) of a molecule, the molecule must first be ionised. For biological molecules a soft ionisation method is required as many biomolecules are labile when too much energy is introduced during the ionisation process. As currently practiced, and as demonstrated by Karas and Hillenkamp (153), MALDI is a soft ionisation technique that is commonly used in the study of large non-volatile biological substances such as proteins, peptides and oligosaccharides (154–156). Typically, an analyte is mixed with a large excess of an organic compound, termed the matrix, and allowed to co-crystallise. The most common matrices used are acidic, which promotes protonation of the analyte and can also strongly absorb UV or IR wavelengths (157), which minimises the destruction of the analyte molecules. The properties of the matrix can aid ionisation and desorption of the analyte, although matrix selection remains difficult in the absence of a comprehensive model of the MALDI mechanism (158).The co-crystallised MALDI spot is then irradiated with a laser resulting in an explosive phase transition and charge transfer that produces charged gas-phase ions (figure 1.11).
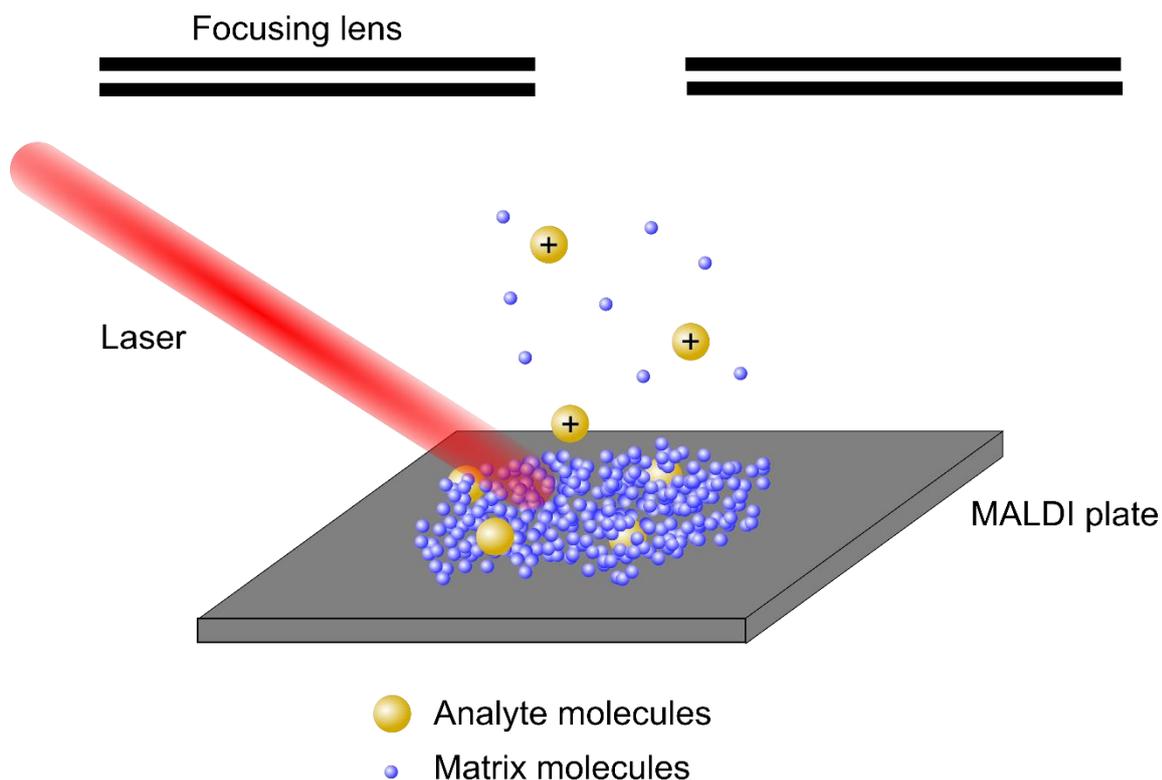
Figure 1.11 | Schematic of MALDI

A laser is fired at the co-crystal of matrix and the sample. Energy from the laser is absorbed by the matrix molecules resulting in ionisation and desorption of both matrix and sample ions which are focussed into the mass spectrometer analyser.

The ionization process that occurs in MALDI mass spectrometry is poorly understood, not least as the analyte is in minority when compared with the excess of matrix molecules (158). The variability in experimental parameters is also a major factor in the ionization process. Any model of ionization must explain the significant features of MALDI-MS including the detection of only singly charged ions (159). Two models of ionization have been proposed, the lucky survivor model (159) and the gas phase protonation model (160).

Both models require the presence of protonated matrix ions although, in the gas phase protonation model, analytes are uncharged and collide with protonated matrix molecules resulting in proton transfer and thus a detectable molecular ionic species. This is in contrast to the lucky survivor model, where any analyte charge states from solution are retained in the analyte/matrix co-crystal. Laser ablation of the co-crystal leads either to the neutralization of charged clusters with counterions, or a lack of

counterions that results in charged molecular species. Experimental evidence for both models has been provided and a unifying model of ionization proposed in which both the gas phase protonation and lucky survivor mechanism occur, the proportion of each depending on several parameters, such as the matrix and laser fluence (161).

Although lagging behind the other -omics, glycomics is a rapidly growing field, in which the development of MALDI-MS has been vital. Indeed, MALDI-MS has been extensively used for the study of glycans since 1991 (162). The application of MS to the study of carbohydrates and glycans is discussed in greater detail below.

1.3.2 Fourier-transform ion cyclotron resonance MS

After ionization in the ion source of the mass spectrometer, a series of ion guides transfers ions into the mass analyser. There are numerous types of mass analysers that are used in mass spectrometers. One such analyser makes use of the cyclotron motion of ions and is called a Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. An FT-ICR mass spectrometer consists of an ICR cell within a superconducting magnet under an ultra-high vacuum (10^{-10} torr). FT-ICR MS was used exclusively in this thesis for *N*-glycan analysis.
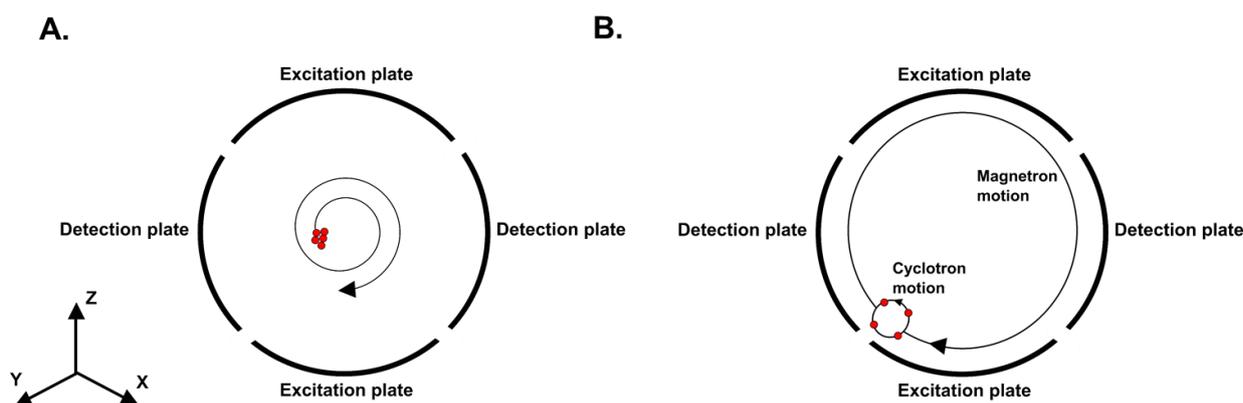
Figure 1.12 | Schematic of cyclotron motion within the ICR cell of an FT-ICR mass spectrometer

A packet of ions is excited by radio frequency pulses from the excitation plates (A). The cyclotron frequency of each ion packet is detected in the detection phase of FT-ICR MS (B).

FT-ICR MS is capable of high mass accuracy detection of ions as it is the cyclotron frequency of ions that is measured within the ICR cell (also known as a Penning trap). Cyclotron motion occurs when an ion orbits in a static magnetic field due to the Lorentz force. A schematic of the ICR cell of an FT-ICR mass spectrometer is shown in figure 1.12. The ICR cell consists of a pair of trapping plates, a pair of excitation plates, and a pair of detector plates, with the three pairs orthogonally arranged with respect to each other. The magnetic field is projected along the Z-axis. The ions are trapped within the ICR cell by applying DC potentials to the pair of trapping plates. The magnetic field and DC potential of the trapping plates cause the ions to orbit within the ICR cell in the XY plane (figure 1.12A) The excitation plates emit a range of frequencies to excite all ions into ion packets, as each ion with unique (m/z) absorbs energy at a different RF. The RF is halted when ions reach their maximum radii. The cyclotron motion of the ion packet can be detected by the pair of detection plates to generate an image current. The cyclotron frequency is dependent on the (m/z) of the ions (equation 1). Thus, by measuring the cyclotron frequency of ions the (m/z) can be derived.

$$\omega = \frac{zB}{m} \qquad (1)$$

Where, ω is the cyclotron frequency, B is the magnetic field, z is the ion charge and m is the mass of the ion.

After the ion packets reach their maximum radii, the excitation RF pulses are stopped and the signal gradually decays with time. This is known as the free induction decay (FID), which plots signal intensity in the time domain. The FID can be converted into the frequency domain using Fourier transformation, which can then be used to plot the mass spectrum ((m/z) versus intensity). An important aspect of FT-ICR detection is that the FID signal is obtained in a non-destructive manner meaning many measurements can be made, increasing mass accuracy and molecule identification confidence.

1.3.3 Mass spectrometry and *N*-glycans

Over the last 30 years, mass spectrometry has played an increasingly vital role in the structural elucidation of glycans. Mass spectrometry-based methods have advantages over other techniques used to study glycans, such as lectin-based assays. These advantages include detailed structural information and quantitation. 2, 5 dihydroxybenzoic acid (DHB) is typically the matrix used for glycan MALDI-MS (58, 60, 61, 107). The conversion of the glycan hydroxyl groups to methyl ethers, in a process known as permethylation, is often undertaken prior to co-crystallisation with DHB. Importantly, the use of permethylation to derivatise glycans has been shown to result in accurate and consistent quantification (163). In addition to the quantification potential of analysing permethylated glycans with MALDI, permethylation also equalises the chemical properties of glycans (164). Glycans are also more likely to be detected as sodiated molecules as opposed to their protonated forms. The sodiated form further stabilises the glycan through interacting with hydroxyl groups and the oxygen atom in the glycosidic bond (165).

Although glycopeptides are amenable to mass spectrometry, *N*-glycans are commonly released enzymatically, which provides the simplest and most direct way of investigating *N*-glycans. The enzymatic release of *N*-glycans may be achieved with a variety of enzymes, the most common being peptide-*N*-glycosidase F (PNGase F), which cleaves the C-N bond between the asparagine residue and the innermost core GlcNAc residue (figure 1.14). Once the glycans are released, mass spectrometry can be used to obtain the mass of the free glycan structures. The determination of the glycan mass with high accuracy that is achievable with FTICR-MS, along with the sequential nature of *N*-glycan biosynthesis, allows for the confident proposal of appropriate glycan structures. Structural ambiguity may be addressed typically using product ion (tandem) MS approaches to generate structurally-informative fragment ions.
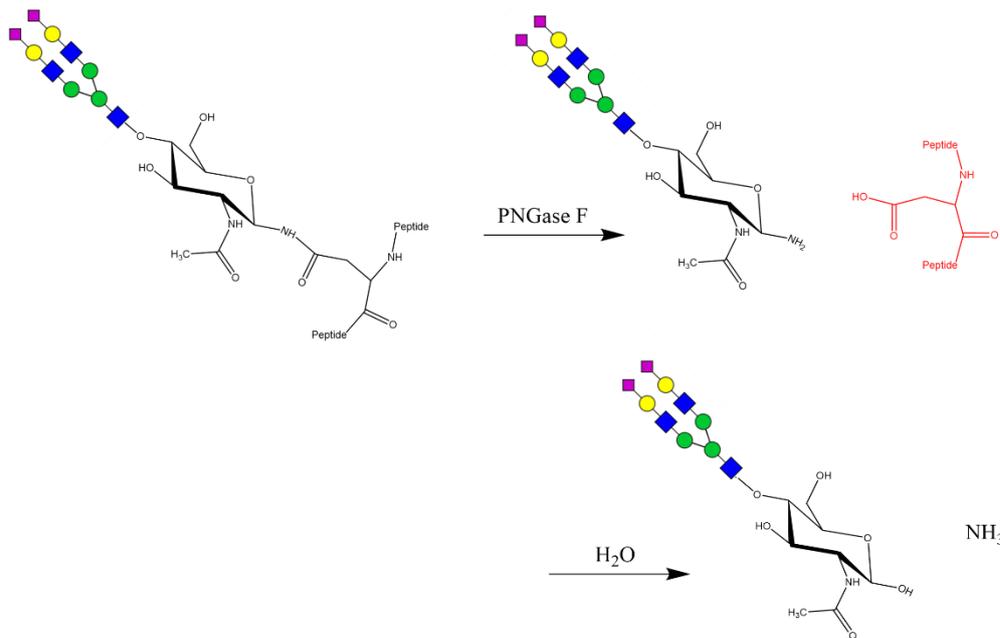
Figure 1.13 | Enzymatic release of *N*-glycan with PNGase F

PNGase F cleaves the bond linking *N*-glycans to the innermost GlcNAc residue, leaving a free *N*-glycan that may be separated from the protein using biochemical techniques.

Although mass spectrometry has been used to elucidate the structures of glycans from numerous biological sources for many decades (166, 167), mass spectrometric glycan profiling of cells grown in culture has proved most useful to cell biologists. Low-throughput methods to obtain the glycan profiles of whole cells have been used to study the Lec family of CHO cells (168). The Lec cells are mutants which have been isolated based on their resistance to lectin binding (169). The glycan profiles of these cell lines were found to be consistent with known genetic data (168). Low-throughput glycan profiling has also been used to study the role of GRASP55/65 in glycosylation (170). Depletion of both GRASP55 and 65 together, showed decreased levels of *N*-glycan complexity, despite the expression level and localization of glycosylation enzymes remaining unaltered (170).

Medium-throughput methods to obtain the glycan profiles using as few as 500,000 cells do not result in the detection of low abundance glycans, but can be readily multiplexed to enable analysis of replicate samples. A comparison of WT HeLa and Cog4KD HeLa cell *N*-glycan profiles, using a medium-throughput method with

replicates, revealed an increase in hybrid glycans but a decrease in the sialylated glycans detected, as a result of Cog4KD (60). The same methodology has been used to assess the effect of a range of COG subunit knockout cell lines, in which alterations to the oligomannose distribution and decreases in total sialylation and fucosylation were observed in Cog2-, Cog4- and Cog7KO HEK293T cells, when compared with WT (58). Mass spectrometric glycan profiling has also shown the significant changes in whole cell glycosylation after MSCs undergo osteogenesis (107). The expression level of $NeuAc_1Fuc_1Hex_5HeNAc_4$ was significantly increased in the osteoblasts, when compared with MSCs. Furthermore, there was a general increase in the proportion of complex glycans as a result of osteogenesis (107). *N-* and O-linked glycans can also be studied in parallel using medium-throughput methods (61). These studies demonstrate the application of mass spectrometry in elucidating changes in the whole cell glycome as a result of mutations, drug treatments and other cellular events, such as differentiation.

Mass spectrometry has been extensively used in this work to obtain whole cell glycan profiles. The goal of this section has been to familiarise the reader with the techniques used to obtain these glycan profiles, namely MALDI and FT-ICR and to highlight how mass spectrometry has been used to investigate biological processes by assessing glycosylation. Determining glycan profile is relatively straightforward but the problem of ascertaining the organization of the Golgi glycosylation machinery that produced the glycan profile is difficult. The following section aims to describe the types of algorithms used to answer this problem as well as offering a review of previous computational models of glycosylation.

1.4 Computational glycobiology

Comprehending complex cellular biological processes has often been helped by the use of mathematical modelling including, cancer invasion (reviewed in 164), signal transduction (reviewed in 165), gene expression (173) and glycosylation (133, 174). Computational modelling of biological systems is recognized as being able to make important contributions to experimental biology. Models may be used to test and disprove hypothesises, suggest additional experiments, make predictions and can lead to greater biological insight. The work described in this thesis utilises several

computational algorithms to achieve its ultimate goal of simulating glycan profiles and predicting alterations in the glycosylation machinery. This section illustrates how modelling can widen our understanding of biology and describes the algorithms that have been used in this study, as well as providing a rationale for their use. The development and use of mathematical models in the context of glycosylation is then outlined. A more detailed description of the algorithms used in this study is provided in the method section.

1.4.1 Simulating biological noise

Many biochemical reactions that occur within a cell involve a small number of molecules; such systems are subject to biological noise and it is important to capture this noise for biological understanding of the system (175–177). Perhaps the first example of the importance of stochasticity was demonstrated by Spudich and Koshland, who showed that bacterial cells, with identical chromosomes and grown in homogenous conditions, displayed individual behavioural characteristics. The authors proposed that certain molecules in low amounts would be subject to probabilistic fluctuations. Noise has also been a significant enough factor as to influence evolution. For example, yeast genes that are particularly sensitive to changes in expression and reduce fitness are associated with low noise (178). A comparison between a deterministic and a stochastic-based model of the mammalian circadian clock demonstrated qualitatively different behaviours between the two models, with the stochastic model closer to reality (179). To quote Elder and Elowitz, " Noise is not merely a quirk of biological systems, but a core part of how they function and evolve" (175).

The glycosylation reactions that are the subject of this work are confined to the Golgi apparatus where the number of enzymes is not large enough to approximate them as concentrations, thus making a stochastic approach more appropriate. In order to simulate a process that can vary randomly with certain probabilities, a stochastic simulation algorithm (SSA) is required. One such algorithm that accounts for noise and is capable of simulating a chemical system as it evolves in time is called the Gillespie algorithm (180). The Gillespie algorithm has been used in this work to simulate the reactions that occur in the Golgi apparatus and is described in generic terms in the next section.

1.4.1.1 Gillespie algorithm

The Gillespie algorithm allows a stochastic simulation of reactions and is particularly suited to biological systems where the number of molecules is not sufficiently large so that concentrations can be used. Rather than considering reaction rates in the traditional sense, the rate constant of a particular reaction is taken to represent the reaction probability per unit time (180). This allows a degree of background noise to arise from the simulation which is often necessary in biological systems. The probabilistic nature of the stochastic simulation of reactions also means that additional factors, such as the availability of monosaccharide-nucleotides, are subsumed within the probabilities, thus eliminating the need for estimating other kinetic parameters (181). In the case of simulating glycan processing, the result of the Gillespie algorithm may be the number of different glycan species produced by the system after a given amount of time. The sum of many of these trajectories is akin to the simulation of a glycan profile.

The Gillespie stochastic simulation algorithm simulates the time evolution of a chemical system by simulating each reaction. This is achieved through a two-step process that is iterated until the allotted time has expired or reactions can no longer continue. The first step is to select which reaction is to occur next. This is done by representing the propensities of two competing reactions as probabilities (equation 2). For two reactions:

$$C + E_A \xrightarrow{k_A} X + E_A \qquad \text{(A)}$$

$$D + E_B \xrightarrow{k_B} Y + E_B$$
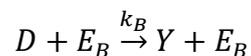
We can calculate the probability of each reaction occurring,

$$P(A|\tau) = \text{rate}_A / R_T = n_{E_A} k_A / R_T \qquad \text{(2)}$$

$$P(B|\tau) = rate_B / R_T = n_{E_B} k_B / R_T$$

Where R_T is the sum of all competing reaction propensities,

$$R_T = n_{E_A}k_A + n_{E_A}k_B = \sum_i rate_i \qquad (3)$$

Following the reaction selection based on the calculated probabilities, the number of molecules in the system is updated and the two stages are repeated. The second step is to simulate the time interval τ, in which the next reaction would occur by randomly drawing from an exponential distribution with probability density function defined by equation 4. The mean of this exponential distribution is $1/R_T$ where R_T is the total reaction rate.

$$P(\tau) = R_T e^{(-R_T\tau)} \qquad (4)$$

1.4.2 Fitting methodologies and the Approximate Bayesian Computation algorithm

The ease at which data can be generated and collected is increasing, leading to more complex models. Standard statistical analysis is often not compatible to these complex, high dimensional data sets. Other statistical methods that estimate parameter values, such as maximum likelihood estimation (MLE) or the rejection method are not suitable to complex data sets because the likelihood function is intractable. The probability of observing the data (D) given the parameter values (θ) is known as the likelihood function, $p(D|\theta)$. In the accept-reject methodology, parameter values are generated from a proposed distribution and are accepted with a probability based on the likelihood function.

Simulated annealing is another method used to estimate parameter values. Simulated annealing attempts to iteratively find the global minimum of a system by sampling

parameter values. The method of simulated annealing does not account for any prior knowledge of the system which the model is trying to replicate. For example, many of the glycosylation enzymes are well characterized, this information is discarded in a simulated annealing approach. One approach that does not require direct calculation of the likelihood function and incorporates prior knowledge of the biological system is based on Bayesian statistics.

The use of Bayesian statistics is well established in many fields, including genetics, and its potential is now being realised in computational systems biology (182). Bayesian methodology is now being applied to many areas of study, such as; drug design (183), models of human evolution (184), double-strand DNA break repairs (185) and tuberculosis transmission rates (186). The goal of Bayesian methods is to compute the posterior probability $p(\theta|D)$, for a set of uncertain parameters θ (enzyme activities and localizations in this case), given some experimentally observed data D. The classic incarnation of this formula is given in equation 5, where $p(\theta)$ is the prior probability of our beliefs about the system before D is observed. As the parameters are considered random variables, they may be treated as having a prior probability distribution $p(\theta)$. Bayes' theorem (equation 5) affords the computation of the posterior probability distribution $p(\theta|\pi)$. Once calculated, the posterior probability distribution yields values for the previously uncertain parameters. These values can be expressed in terms of the posterior expectation. In the case of glycan processing in the Golgi apparatus, the uncertain parameters may be the distribution and/or activity of resident Golgi enzymes. The glycan profiles of cell lines are the observed data.

$$P(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{P(D)} \tag{5}$$

In many cases the likelihood function may not be known or cannot easily be computed leading to the development of approximate Bayesian computation (ABC) methods. ABC-based algorithms all approximate the likelihood function by simulating outcomes from the prior distribution. In the ABC rejection sampler, parameter values are sampled from the prior distribution and used to simulate data. The simulated data is reduced to a summary statistic(s) which is then compared with the equivalent experimental

summary statistic(s). The sampled parameter values are either accepted or rejected based on the similarity between the simulated and experimental summary statistic(s), which must reach a certain user-defined threshold (figure 1.15). This process can be repeated for *n* accepted parameter values. Accepted parameter values can then be plotted to generate the posterior probability distribution. The ABC rejection sampler is capable of inferring posterior distributions when the likelihood function is intractable. However, the acceptance rate can be low if the prior and posterior distributions are dissimilar (187).
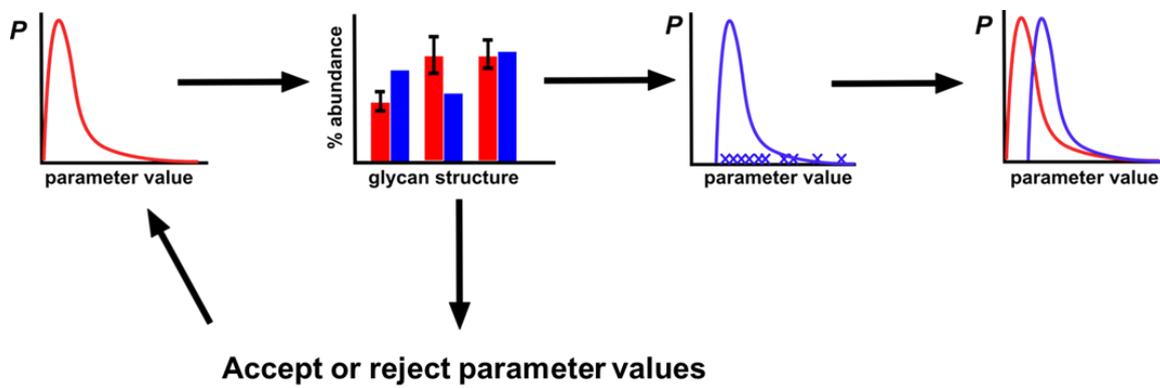
Figure 1.14 | Pictorial representation of an ABC rejection sampler

Sampled values from the prior distribution for a parameter (red) are used to generate a simulated glycan profile which is compared with the observed glycan profile. The value from the prior distribution is either accepted or rejected based on the simulated glycan profile's similarity to the observed profile. The accepted parameter values are then used to construct the posterior distribution (blue).

Integrations over uncertain parameters are computationally demanding and, in many cases, intractable. However, they do lend themselves to Markov Chain Monte Carlo (MCMC)-based methods (188), for example the ABC-MCMC method proposed by Marjoram et al. A Markov chain is a sequence of random variables, $\{X_0, X_1, X_2, \ldots\}$, in which the next state depends, not on the history of the chain, but only on its current state. The ABC-MCMC algorithm differs from the ABC rejection sampler because it proposes new parameter values based on a transition kernel and accepting parameter values based on the prior probability. The similarity between the observed and simulated data is also considered. Parameter values that result in simulated data more closely resembling the observed data are visited more often in ABC-MCMC methods.

For this reason, the ABC-MCMC method avoids the disadvantage of the low acceptance rates, but may get stuck in regions of low probabilities (187).

1.4.3 Modelling glycosylation

The development of systems biology-based methods in the study of glycosylation and the glycome in general has lagged behind that in the fields of genomics and proteomics. This may be a direct consequence of the structural complexity and microheterogeneity of carbohydrates compared with other biological macromolecules. Microheterogenity is the result of incompleteness in the biosynthetic reactions. This can lead to a heterogeneous product with the same site on two different molecules of the same protein having different glycans attached. Attempts to model glycosylation have been driven by the desire to identify critical reactions that dictate the subsequent reaction pathways and thus the final glycan product. This information has particular relevance in the biopharmaceutical industry (91). In addition to the structural complexity of glycosylation, the number of enzymes involved, the enzyme distributions in the Golgi and the degree of uncertainty in these parameters add another layer of complexity.

One of the first examples of a mathematical model describing *N*-linked glycosylation was presented by Umaña & Bailey 1997. This model consisted of mass balance equations for 33 *N*-glycan structures and could simulate the glycoform distributions of some proteins produced by CHO cells (174). The Umaña & Bailey model was elaborated upon to include additional galactosylation, fucosylation and sialylation reactions, generating a model that encompasses 22,871 reactions (190) (KB2005). The modelling of glycosylation has also been used to investigate the anterograde vesicle transport and cisternal maturation models of transport through the Golgi apparatus. The modelling of the two transport mechanisms was achieved from an engineering perspective, with vesicle transport being represented as continuously mixing tanks and cisternal maturation as plug flow reactors in series (191). The cisternal maturation model was found to be the more likely mechanism of transport through the Golgi when simulated glycan profiles were compared with experimental glycan profiles.

Kinetic models of glycosylation have also been used to predict enzyme activity levels based on MALDI-MS data (192) (KB2009). This model, which built upon the previous Krambeck & Betenbaugh model of 2005, predicted alterations to glycosylation enzymes based on MS data from human monocytes and monocytic leukemia cells (192). Gene expression data have been integrated into the KB2009 model to predict enzyme levels in a prostate cancer progression model (193).

A study of the control of CHO cell line branching using the KB2005 model has also been conducted (194). In addition to the branching enzymes MGAT4/5, this study found that by decreasing galactosylation, the number of antennae on *N*-glycans could be increased (194). A more recent study modelling glycan profiles of the Lec CHO cell line family included a method in which the enzyme distribution in each Golgi cisterna could be adjusted (195). Predicted changes to enzyme activities were consistent with the phenotype of each cell line. However, unexpected shifts were also predicted (195). The inclusion of cellular metabolism, such as the availability of monosaccharide-nucleotides, has been shown to improve models of monoclonal antibody glycosylation (196). This approach allows the effects of metabolic factors on glycosylation to be modelled, and has the potential to increase control over the production of glycoforms. Indeed, this methodology has been used to identify GalT as a key factor in determining antibody glycosylation variability under conditions of mild hypothermia (197).

Finally, due to the difficulty in parametrizing the KB2005 model, a stochastic model that describes glycosylation as a Markov chain has recently been developed (91). Furthermore, deterministic approaches to modelling glycosylation, such as the KB2005 model do not account for the biological noise that arises from the inherently heterogenous process of glycosylation. Spahn *et al*. highlight the importance of enzyme compartmentalisation within the Golgi apparatus. By modelling the Golgi apparatus as a non-compartmentalised organelle, it shows that computational simulations no longer matched experimental glycan profiles of monoclonal antibodies (91). Simulating the glycan profiles of a single protein means predictions regarding the whole Golgi glycan processing pathway cannot be made. For example, the processing of oligomannose glycans cannot be easily inferred from the glycan profile of an antibody as the presence of oligomannose glycans is minimal. Furthermore, the simulation of a single protein's glycan profile is likely to be dominated by the intrinsic

properties, such as tertiary structure, of the glycoprotein (198). Such dominance of these properties may mask alterations in the organization of the glycosylation machinery.

Chapter 2: Materials and methods

2.1 Chemicals

Chemicals were obtained from Sigma Aldrich unless stated otherwise.

2.2 Cell culture

Components used for culturing cell lines are all from Invitrogen unless otherwise stated. All cell lines were cultured in Dulbecco's Modified Eagle Media supplemented with fetal bovine serum (10%), penicillin/streptomycin (1%) and concentrated glutamine (1%). All cell lines were grown at 37℃ with 5% CO_2 in a humidified incubator. The method of cell harvesting varied depending on the downstream requirements, therefore the cell harvesting process is described in more detail in the relevant sections.

2.2.1 HEK293T cell culture

Wild type (WT) HEK293T cells were a kind gift from Vladimir Lupashin (University of Arkansas for Medical Sciences). Cog4 knockout (KO) HEK293T cells were generated using CRISPR-Cas9 technologies by Jessica Blackburn-Bailey (University of Arkansas for Medical Sciences) (58). Both WT and Cog4 KO HEK293T cells were cultured in tissue culture flasks (Sarstedt) and passaged 1/10 or as required with trypsin/EDTA when cells reached ~90% confluency.

2.2.2 MSC cell culture

Immortalised mesenchymal stromal cells (MSCs) (148) were provided by the Genever group (University of York). Immortalisation was achieved through lentivirus transduction of human telomerase reverse transcriptase (hTERT) into primary mesenchymal stromal cells before single cell colony isolation and expansion. Cells were cultured in tissue culture flasks (Sarstedt) and passaged 1/6 or as required with trypsin/EDTA when cells reached ~70% confluency.

2.2.3 Transfection

HEK293T and derived cell lines were transfected using Xfect (Clontech). Cells were seeded in a 24-well tissue culture plate at 50% confluency and allowed to adhere. For each 24-well to be transfected, the plasmid (0.75 µg) was diluted in the Xfect reaction buffer to a final volume of 25 µL and mixed by vortexing. The Xfect polymer (0.3 µL per 1 µg of plasmid) was next added to the mixture, then vortexed and the mixture incubated at room temperature for ten minutes. The entire plasmid/Xfect solution was then added dropwise to the cells and left at 37 ℃ overnight. Cells were then passaged onto poly-L-lysine coated glass coverslips.

2.2.4 Drug treatments

2.2.4.1 Swainsonine treatment

Cells were incubated for 48 hours in cell culture media containing swainsonine (10 µg/mL) (Cayman Chemicals) prior to harvesting and downstream experiments.

2.2.4.2 Nocodazole treatment

Cells were incubated with cell culture media containing nocodazole (5 µM) for 3 hours. This treatment was based on the work by Minin (7) and the duration optimised by observation of dispersed Golgi stacks in WT HEK293T cells (figure 2.1).
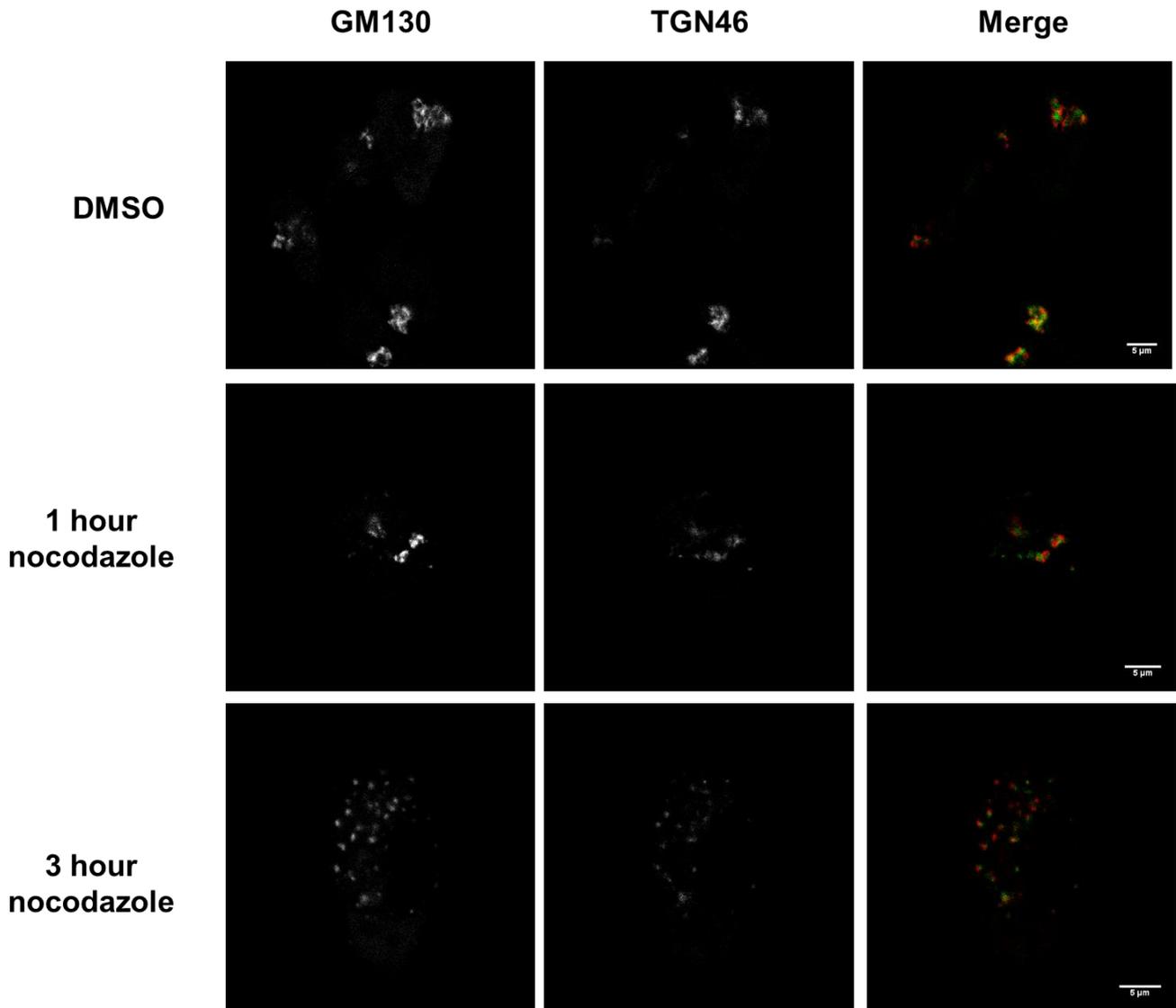
Figure 2.1 | Dispersal of Golgi ribbon following nocodazole treatment

Representative images of WT HEK293T cells treated with 5 µM DMSO for 3 hours or nocodazole for 1 or 3 hours and stained with anti-GM130 and anti-TGN46. Cells were then imaged using confocal microscopy. Scale bar is 5 µm.

2.2.5 Osteogenic differentiation

For osteogenic differentiation, immortalised MSCs were seeded at near confluency a 24-well tissue culture plate and media changed every 3-4 days with cell culture media containing L-ascorbic acid-2-phosphate (50 µg/mL), β-glycerophosphate (5 mM) and dexamethasone (10 nM) for 21 days.

2.2.6 Galectin inhibition with lactose

MSCs were seeded at near confluency in tissue culture wells and incubated with cell culture media containing 100 mM glycerol overnight. Following the pre-treatment with glycerol, the media was then changed to either cell culture media (basal) containing sucrose or lactose (100 mM) or the osteogenic media, described above, containing sucrose or lactose (100 mM). The medium was then refreshed each day for seven days. The sucrose/lactose media was then replaced with basal or osteogenic media only and the osteogenic procedure followed up to day 21.

2.3 Histochemistry

2.3.1 Alkaline phosphatase (ALP) and von Kossa stain

Cells were washed twice with phosphate buffered saline (PBS) before being incubated with napthol AS-MX (0.2 mg/mL) and Fast Red TR (1 mg/mL) in Tris (0.1 M, pH 9.2) containing N,N-dimethylformamide (0.1%) for two minutes. Cells were then washed again with PBS twice and fixed with formaldehyde (4%) (Thermo Scientific) for 5 minutes. Excess formaldehyde was removed with a PBS wash followed by a water (distilled) wash. The fixed cells were the incubated on a light box for 60 minutes in silver nitrate (1%, w/v). Cells were then washed three times with distilled water before being incubated with sodium thiosulphate (2.5%, w/v) for 5 minutes. Finally, cells were washed twice with distilled water and stored in glycerol (20% in PBS). Representative images of staining were acquired on a stereo microscope fitted with an Axiocam MrC5 (Ziess).

2.3.2 Alizarin Red S stain

Cells were first washed with PBS and fixed in formaldehyde (4%) (Kautex) for 20 minutes. Fixed cells were once again washed with PBS three times before incubation for 20 minutes in Alizarin Red S solution (40 mM, pH 4.3). Cells were then washed three time with PBS before being extensively washed with tap water to remove non-specific staining. Cells were then allowed to air dry before representative images of staining were acquired on a stereo microscope fitted with an Axiocam MrC5 (Ziess).

The Alizarin stain was then eluted through the addition of cetylpyridinium chloride (CPC) (10%, w/v). Following incubation with CPC for an hour at room temperature with shaking, the eluted stain was transferred into a fresh 96-well plate and quantified at 570 nm on a Multiskan FC microplate photometer (Thermo Scientific).

2.4 Mass spectrometry

2.4.1 Filter aided *N*-glycan separation (FANGS)

Cells were grown to near confluency in 10 cm cell culture dishes and the media changed 24 hours before being harvested to ensure nutrients were not limited. Cells were first washed five times with pre-warmed PBS. Cells were then harvested by scraping into a microcentrifuge tube and centrifuged at 14000×g for five minutes at 4 ˚C. The supernatant was removed and lysis buffer (4% (w/v) SDS, 100 mM Tris/HCl pH 7.6, 0.1 M dithiothreitol) was added at ten times the cell pellet volume. The sample was boiled at 97 ˚C for five minutes and then centrifuged at 14000×g for 10 minutes. Probe sonication was used if the lysate was too viscous to pipette. The supernatant was transferred into a fresh microcentrifuge tube and stored at -80 ˚C if required.

A urea solution (8 M urea in 100 mM Tris/HCl pH 8.5) was then used to dilute the cell lysate tenfold. The diluted cell lysate was then centrifuged through an ultrafiltration tube (Amicon Ultra-0.5, Ultracel-30 membrane, nominal mass cutoff 30 kDa, Millipore) in 400 µL increments at 15000×g for ten minutes until the complete sample solution passed through the filter. The membrane filter was washed with the urea solution three times, before the addition of iodoacetamide solution (40 mM in 300 µL urea solution) to prevent cysteine bridge formation. The sample was incubated in the dark for 15 minutes in the iodoacetamide solution and then centrifuged at 14000×g for 10 minutes. Following this the sample was washed with urea solution once followed by ammonium bicarbonate (300 µL, 50 mM pH 8) three times. The membrane filter was incubated at 37°C for 16 hours with 8 units of PNGase F in 100 µL ammonium bicarbonate (50 mM) solution. Released glycans were then eluted into a fresh collection tube with water (HPLC grade).

2.4.2 Permethylation

The released *N*-glycans were transferred into glass tubes and evaporated to dryness using a vacuum concentrator. The glycans were then dissolved in 20 drops of dimethyl-sulfoxide (DMSO) followed by the addition of two heaped micro spatulas of crushed sodium hydroxide pellets. Repeated addition of iodomethane to the sample solution was made as follows: 10 drops of iodomethane followed by a ten minute incubation; 10 drops of iodomethane followed by a ten minute incubation; 20 drops followed by a 20 minute incubation. The reaction was then quenched with the addition of sodium thiosulfate (1 mL, 100 mg/mL) and dichloromethane (1 mL). The organic phase which contains the permethylated glycans, was then subjected to repeated washes with water (HPLC grade). The remaining organic layer was then dried in a vacuum concentrator.

2.4.3 Fourier-Transform Ion Cyclotron mass spectrometry

The permethylated glycans were first dissolved in methanol (20 µL, HPLC grade). The sample solution (2 µL) was then mixed with 2,5-dihydroxybenzoic acid (2 µL, 20 mg/mL in 70% methanol) and sodium nitrate (1 µL, 0.5 M). The sample/matrix mixture (2 µL) was spotted onto a MALDI target plate. Mass spectra were acquired on a 9 T solariX FTICR mass spectrometer (Bruker Daltonics) recorded over an m/z range 400-4000 in positive ion mode. Eight scans each consisting of 500 laser shots were averaged and the laser power set between 40-60%. Spectra were calibrated using an external calibrant (Bruker peptide mix II).

2.4.4 Data analysis

Mass spectra were analysed using Data Analysis (Bruker Daltonics). A peak list was generated using the SNAP peak identification algorithm available on the Data Analysis software. For an ion signal to be added to the peak list the signal-to-noise ratio had to be above three. The generated mass list was exported as a .csv file along with the area of all isotopic peaks and the full width half maximum (FWHM) of each signal. A Python algorithm was written and used to identify masses from the mass list that corresponded to *N*-glycan structures. The algorithm also calculated the total peak intensity for each identified glycan from the FWHM and ion signal area. The automated glycan picking and quantification method gives a similar result to manual

quantification method as used in Abdul Rahman et al. 2014 (figure 2.2). When comparing spectra from different conditions, the total ion signal intensity for each glycan was normalized to the total ion signal for all glycans that were common to both conditions.
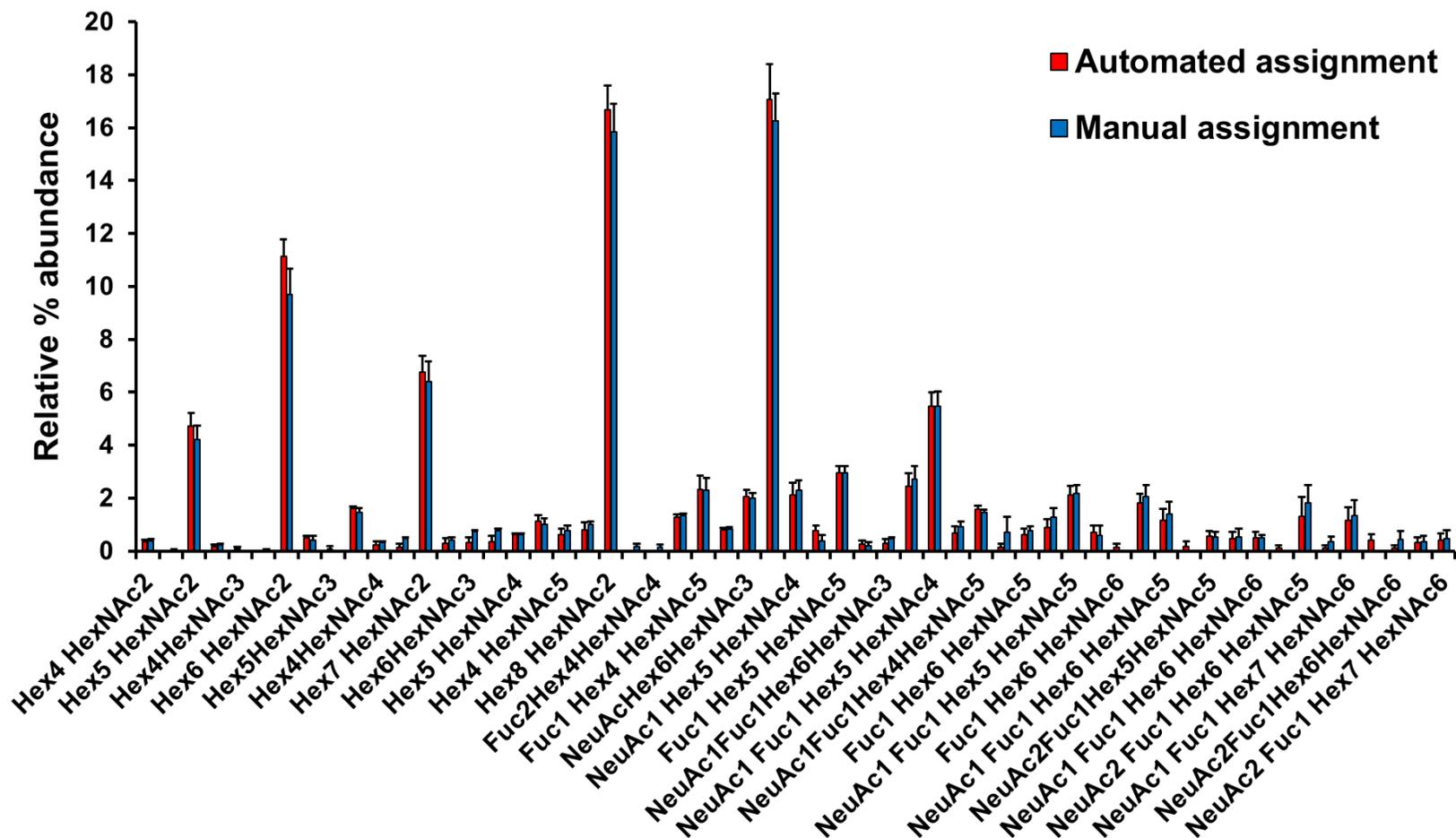
Figure 2.2 | Automation of *N*-glycan relative quantification

Normalized ion signal intensities for glycan identified in WT HEK293T cells. Quantification using automated assignment with Python algorithm to quantify glycan ion signals (red) compared with manual assignment (blue). Error bars are SEM, n = 3

2.4.5 Crude cell membrane proteomics

2.4.5.1 Digitonin treatment and urea lysis

Cells were grown to near confluency in a 10 cm tissue culture dish before being harvested by centrifugation at 800×g for three minutes at room temperature. The resulting pellet was then washed twice with PBS and then twice in ice cold permeabilization buffer (20mM HEPES pH7.4, 150 mM potassium acetate, 5 mM magnesium acetate, 5 mM DTT). The cell pellet was then resuspended and incubated in digitonin solution (0.1% digitonin in permeabilization buffer) at 4 ℃ with rotation for 30 minutes. Following permeabilization with digitonin, the total cell membrane fraction was pelleted by centrifugation at 800×g. The digitonin treatment was undertaken in order to enrich membrane proteins. The membrane pellet was then resuspended in urea lysis buffer (20 mM HEPES pH 8.0, 9 M urea, 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate and 1 mM β-glycerophosphate) and lysed with a sonic probe. The lysate was then cleared through centrifugation at 20,000×g for 15 minutes. The supernatant was then transferred into a LoBind microcentrifuge tube (Eppendorf).

2.4.5.2 Mass spectrometry and analysis

The following work was conducted by the staff (Adam Dowle) at the Proteomics facility at the University of York.

The protein fraction from the digitonin and urea lysate was reduced and alkylated before digestion with a combination of Lys-C and trypsin proteases. Peptides that resulted from the digestion were eluted into an Orbitrap Fusion mass spectrometer (Thermo Scientific) from a C18 EN PepMap column. Mass spectrometry was used to acquire tandem mass spectra of peptides which were searched against the human subset of the UniProt database using Mascot (Matrix Science). Peak areas were normalised to a subset of peak areas from ER proteins (table 2.1) allowing relative quantification of peptides. Relative fold differences and associated p-values for differential abundance were calculated in Progenesis QI (Waters).

Gene name	Description
DAD1	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1
SRPRB	Signal recognition particle receptor subunit beta
RRBP1	Ribosome-binding protein
STT3A	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3A
STT3B	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit STT3B
SEC61A1	Protein transport protein Sec61 subunit alpha isoform 1
SEC61B	Protein transport protein Sec61 subunit beta
CALR	Calreticulin
CANX	Calnexin
PDIA6	Protein disulfide-isomerase A6
PDIA3	Protein disulfide-isomerase A3

Table 2.1 | List of proteins used to normalise proteomic data for relative quantification

2.5 Western blotting

2.5.1 Whole cell protein lysis

Cells were cultured to confluency in 6-well cell culture plate until they reached confluency. Cells were then washed twice with PBS before incubation at room temperature with 250 µL of sample buffer (5% glycerol, 50 mM Tris, 50 mM dithiothreitol, 1% sodium dodecyl sulfate, 0.75 mM bromophenol blue) for five minutes. The sample solution was then transferred into a microcentrifuge tube and boiled at 97

˚C for five minutes. If required the sample solution was passed through a needle. Sample lysates were stored at -20 ˚C.

2.5.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Gels for SDS-PAGE separation of proteins consisted of a resolving gel (10% acrylamide, 375 mM Tris pH 8.0, 0.05% ammonium persulfate, 0.067% N, N, N', N'-tetramethylethylenediamine (TEMED)) and a stacking gel (4% acrylamide, 125 mM Tris pH 6.8, 0.1 % ammonium persulfate and 0.1% TEMED). Electrophoresis was conducted initially at 120 V until the sample lysates had reached the resolving gel and then at 180 V. A pre-stained protein ladder (All Blue Precision Plus, Bio-rad) was also loaded onto the gel.

2.5.3 Semi-dry transfer

Semi-dry transfer was used to transfer proteins from gels to a polyvinylidene fluoride (PVDF) membrane. The PVDF membrane was placed in methanol for 15 seconds before being transferred into semi-dry transfer buffer (48 mM Tris, 39 mM glycine, 20% methanol and 0.0375% SDS). Two pieces of Whatman paper (3 mm thickness) were soaked in semi-dry transfer buffer, one of which was placed on the semi-dry apparatus (Bio-rad) followed by, the activated PVDF membrane then the SDS-PAGE gel and finally the second Whatman paper. Semi-dry transfer was conducted at 20 V for 60 minutes.

2.5.4 Immunoblotting

Non-specific binding was blocked through incubation of the PVDF membranes in milk (5% milk powder (w/v) in PBS with 0.05% tween) for one hour at room temperature on a platform rocker. In the majority of cases the PVDF membranes were cut with a scalpel using the molecular weight markers from the pre-stained ladder as a guide. Primary antibodies (table 2.2) were diluted in the blocking solution and incubated with PVDF membranes overnight at 4℃ on a platform shaker. The membranes were then washed four times for ten minutes with the blocking solution while shaking. Secondary antibodies conjugated with horse radish peroxidase (HRP) were diluted (1/1000) in the blocking solution and incubated with the membranes for one hour at room temperature while shaking. PVDF membranes were then washed two times with the blocking

solution for ten minutes, followed by four washes with PBS containing 0.05% tween for ten minutes while shaking.

Antibody	Host	Dilution used	Supplier
MGAT1	Rabbit	1/500	Abcam (ab180578)
B4GalT1	Goat	1/500	R&D Systems (AF3609)
FUT8	Rabbit	1/1000	Abcam (ab191571)
LAMP1	Gift from Paul Pryor	1/100	Kind gift from Pryor group (University of York)
GAPDH	Mouse	1/500,000	Applied Biosystems (AM4300)

Table 2.2 | Primary antibodies used for western blotting

2.5.5 Imaging and image analysis

Protein bands were visualised using BM chemiluminescence western blotting substrate (Roche). The membranes were imaged on a Syngene GeneGenius system and analysis carried out with ImageJ.

2.6 Beta-hexosaminidase assay

Cells were seeded in a 24-well tissue culture plate and harvested in PBS. The cells were pelleted with centrifugation at 800×g and then resuspended in citrate buffer (100 µL, 40 mM sodium citrate, 60 mM citric acid) containing 0.15% triton-X 100. The lysate (25 µL) was added to the bottom of a LP3 tube (Thermo Scientific). The substrate solution (100 µL, 100 mM citrate buffer, pH 5.0, 0.5 mM mM 4-methylumbelliferyl-2-acetamido-2-deoxy-beta-D-glucopyranoside, 0.27 M sucrose) was then added to the sample solution. After exactly three minutes the reaction was stopped with the addition of sodium carbonate (1 mL, 1 M). Fluorometric quantification was achieved with an excitation wavelength of 360 nm and an emission wavelength of 445 nm.

2.7 Bicinchoninic acid assay (BCA)

BCA was conducted using a commercial kit (Thermo Scientific). A standard curve was created with bovine serum albumin ranging from 0-10 µg in sodium hydroxide (0.1 mM). The sample solution (5 µL) was added to sodium hydroxide (5 µL, 0.2 M) into a 96-well. To each 96-well, 196 µL of BCA reagent A and 4 µL copper sulfate (4% in water) were added. The 96-well plates were then incubated at 37℃ for 30 minutes. The optical density was measured at 560 nm on a Multiskan FC microplate photometer (Thermo Scientific).

2.8 Confocal microscopy

2.8.1 Immunofluorescence

48 hours post-transfection cells were treated with nocodazole (section 2.2.4.2), fixed, stained and mounted on microscope slides. Cells were first washed twice with PBS before being incubated in formaldehyde (4%) (Thermo Scientific) for 15 minutes at room temperature. The fixed cells were then washed with PBS twice then washed twice with glycine (20 mM glycine in PBS). Non-specific binding was then blocked by incubation of the fixed cells in blocking solution (2% BSA, 20 mM glycine in PBS) for 45 minutes.

Following blocking the coverslips containing the fixed cells were removed from the 24-well tissue culture plates and placed on a piece of blue roll, cell side up. Cells were stained with anti-GM130-Alexa647 (BD Laboratories, 558712) antibody. The antibodies were diluted (1/100) in the blocking solution and a 40 µL drop per coverslip pipetted onto parafilm. Each coverslip was then gently placed cell side down onto each drop and incubated for one hour. The cells were then washed four times in the blocking solution by moving the coverslip drop to drop. The cells were then stained with DAPI (1/2000) for two minutes before being dipped in PBS and allowed to dry cell side up on a piece of blue roll. Finally, coverslips were mounted onto microscope slides using mounting reagent (GeneTex) and stored at 4℃.

2.8.3 Microscopy

Slides were imaged on a Ziess LSM 880 operating in Airyscan mode with a 63× oil objective. All channels were visualised using a 488/561/633 nm beam splitter. GalT-YFP was imaged using a 488 nm laser and 420-480 nm band pass filter. Anti-GM130-Alexa647 was visualised using a 633 nm laser and 605 nm long pass filter. 16 bit images were acquired of 10 z-slices with a resolution of 488 × 488 pixels. Images were exported as .lsm files.

2.8.4 Data analysis

Pearson's colocalization coefficients were calculated for each individual Golgi stack and averaged for each cell analysed using the Coloc 2 add on for ImageJ. In total 151 and 113 Golgi stacks were analysed for WT and Cog4KO HEK293T cells respectively.

2.9 Computational modelling

The computation work was implemented using a custom written JAVA implementation, the code of which is included as e-materials with this thesis. Additional details of specific changes can be found in the relevant chapters where they are introduced.

2.9.1 Simulating glycosylation

The SSA used to simulate glycosylation reactions is based on the Gillespie algorithm as outlined above. Simulating the glycosylation reactions required the glycans to be represented in a form of linear notation, examples of which for different types of *N*-glycan are shown in table 2.3. The representation of *N*-glycans in this form allowed reactions that were simulated by the SSA to be implemented using string substitutions to add or remove monosaccharides to build new glycans. The linear notation includes different antennae, denoted within brackets, and linkage information. If required, the linear notation allows isobaric glycans produced by the model to be distinguished. The SSA utilised a set of 11 enzymes for HeLa cells or 12 enzymes for HEK293T cells and MSCs. These enzymes, the rules defining the reactions they instigate and their activities were saved as a .xls file (table 2.4). The activities of each enzyme were distributed into three (HeLa) or four (HEK293T and MSCs) compartments. The input glycans that are the initial substrates for the SSA were $GlcNAc_2Man_8$, $GlcNAc_2Man_9$,

GlcNAc$_2$Man$_9$Glc. In total the glycan processing of 100,000 input glycans were simulated, independently from each other.

N-glycan structure	Linear notation
	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1Man:)_m6.1Man(3.1Man:)_m6.1Man:@
	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man(3.1Man:)_m6.1Man:@
	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man2.1GlcNAc4.1Gal6.2Sia:@
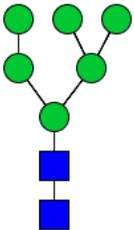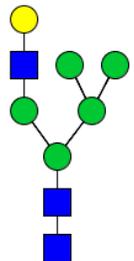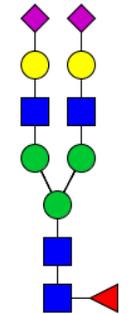
	GlcNAc(6.1Fuc:)4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man(2.1GlcNAc4.1Gal6.2Sia:)_m6.1GlcNAc4 .1Gal6.2Sia:@
	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man(2.1GlcNAc4.1Gal6.2Sia:)_m4.1GlcNAc4.1Gal6.2Sia:)_m6.1Man(2.1GlcN Ac4.1Gal6.2Sia:)_m6.1GlcNAc4.1Gal6.2Sia:@

Table 2.3 | Exemplar of linear notation used to represent glycans

Pictorial and linear representation of several classes of glycans. Numbers denote the linkage configuration of the glycans, while ":" denotes the terminal monosaccharide of an antenna, residues within "()" denote a separate antenna and "@" denotes the end of the glycan structure. The use of linear notation for the SSA allowed reactions to be represented as string substitutions.

Enzyme Name	Enzyme Target	Enzyme Result	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3
MAN1	1Man2.1Man:	1Man:	0.233	0.527	0.178
MAN1	1ManSS2.1Man:	1Man:	0.037	0.084	0.028
EndoMAN	Man3.1Gluc:	:	0.029	0.004	0.004
MAN2	Man(3.1Man2.1GlcNAc:)_m6.1Man(3.1Man:)_m6.1Man:	Man(3.1Man2.1GlcNAc:)_m6.1Man:	0.454	3.452	0.569
MGAT1	4.1Man(3.1Man:)_m6.1Man(3.1Man:)_m6.1Man:	4.1Man(3.1Man2.1GlcNAc:)_m6.1Man(3.1Man:)_m6.1Man:	0.014	0.305	0.009
MGAT2 (1)	GlcNAc:)_m6.1Man:	GlcNAc:)_m6.1Man2.1GlcNAc:	1.798	12.770	1.700
MGAT2 (2)	GlcNAc4.1Gal:)_m6.1Man:	GlcNAc4.1Gal:)_m6.1Man2.1GlcNAc:	1.798	12.770	1.700
MGAT2 (3)	GlcNAc4.1Gal6.2Sia:)_m6.1Man:	GlcNAc4.1Gal6.2Sia:)_m6.1Man2.1GlcNAc:	1.798	12.770	1.700
MGAT4	3.1Man2.1GlcNAc:	3.1Man(2.1GlcNAc:)_m4.1GlcNAc:	0.004	0.031	0.004
MGAT5	6.1Man2.1GlcNAc:	6.1Man(2.1GlcNAc:)_m6.1GlcNAc:	0.020	0.308	0.021
FUT8 (1)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man2.1GlcNAc	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man2.1GlcNAc	0.049	0.412	0.053

FUT8 (2)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man2.1GlcNAc	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man2.1GlcNAc	0.049	0.412	0.053
FUT8 (3)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man2.1GlcNAc	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man2.1GlcNAc	0.049	0.412	0.053
FUT8 (4)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man	0.049	0.412	0.053
FUT8 (5)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man	0.049	0.412	0.053
FUT8 (6)	GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal6.2Sia:)_m6.1Man	0.049	0.412	0.053
FUT8 (7)	GlcNAc4.1GlcNAc4.1Man(3.1Man:)_m6.1Man(3.1Man:)_m6.1Man:	GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man:)_m6.1Man(3.1Man:)_m6.1Man:	0.000	0.002	0.000
GalT (1)	GlcNAc:)_m6.1Man	GlcNAc4.1Gal:)_m6.1Man	0.452	0.449	3.149

GalT (2)	6.1Man2.1GlcNAc:	6.1Man2.1GlcNAc4.1Gal:	0.452	0.449	3.149
GalT (3)	(2.1GlcNAc:	(2.1GlcNAc4.1Gal:	0.452	0.449	3.149
GalT (4)	6.1GlcNAc:	6.1GlcNAc4.1Gal:	0.452	0.449	3.149
SiaT	.1Gal:	.1Gal6.2Sia:	0.006	0.006	0.128
GlcNAc-PhospoT	1Man2.1Man:	1Man2.1ManP:	0.126	0.015	0.014
			0.711^A	0.009^B	5.279^C

Table 2.4 | Table representing the enzyme .xls file

Example of the .xls spreadsheet used to store the enzyme names, enzyme rules and activities in each cisterna. Values under the heading's cisterna 1, 2 and 3 can be altered and used by the SSA to simulate a glycan profile. [A] proportion of Man9GlcNAc2, [B] proportion of Glc1Man9GlcNAc2, [C] transit time per cisterna.

Following the simulation of the glycan processing of 100,000 input glycans, the final simulated structures, how many times the structure was produced in the simulation, the relative abundance of the simulated glycan and the relative of the equivalent glycan from experimental data were outputted as a .csv file (table 2.5). The SSA was run three times and the average relative abundance of each glycan from the three simulations was used to plot a simulated glycan profile which is compared with the experimental glycan profiles. Simulated glycans at a relative abundance under 0.1% were removed. Although glycan isoforms can be distinguished within the model, isoforms were collated together to generate one value.

Simulated glycan structure	Mass	No. of times structure is simulated	Simulated relative % abundance	Observed relative % abundance	Summary statistic
GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man:@	1417	0	0	0.17	0.17
GlcNAc4.1GlcNAc4.1Man(3.1Man:)_m6.1Man(3.1Man:)_m6.1Man:@	1580	1042	10.42	11.34	3.35
GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man2.1GlcNAc:)_m6.1Man:@	1591	0	0	0.18	0.18
GlcNAc4.1GlcNAc4.1Man(3.1Man2.1GlcNAc4.1Gal:)_m6.1Man:@	1621	1	0.01	0.11	0.10
GlcNAc(6.1Fuc:)_g4.1GlcNAc4.1Man(3.1Man:)_m6.1Man(3.1Man:)_m6.1Man:@	1754	1	0.01	0.00	0.01
GlcNAc4.1GlcNAc4.1Man(3.1Man2.1Man:)_m6.1Man(3.1Man:)_m6.1Man:@	1784	2357	23.57	23.31	5.32

Table 2.5 | Representation of results .csv file as outputted from the SSA

The summary statistic is explained in section 2.9.2.

2.9.2 Approximate Bayesian Computation

Probability density functions (PDFs) for each parameter were constructed in Libre Open Office and used as the prior probability distributions. The mean μ, of each prior distribution was set as the activity of each corresponding enzyme obtained from literature (table 2.6). The variance for the prior distributions was set to μ^2. The prior distributions for each enzyme parameter were represented as log-normal distributions if the enzyme was predominantly in that cisterna or an exponential decay distribution if not (appendix 1). The variance of the prior distributions was increased in cases where more probability was required at the limits of the parameter values.

Enzyme	Rate (nmol mg^{-1} min^{-1})	Reference
MAN1	2880	(199)
MAN2	7000	(200)
MGAT1	101	(100)
MGAT2	27000	(201)
MGAT4	46.6	(103)
MGAT5	18800	(202)
FUT8	127.8	(203)
GalT	4000	(92)
SiaT	‡	N/A
GlcNAc-Phosphotransferase	166.7	(204)

Table 2.6 Table of enzymatic rates determined *in vitro*.

‡ the rate of sialylation was set to allow the correct level of total sialylation in WT HeLa cells. Literature rates for sialyation were found to be too low.

A summary statistic, referred to as the score, was set based on the simulation of the glycan processing utilising the prior distribution parameter means. The similarity between the simulated and experimental glycan abundances were quantified as a score.

$$score = \sum_{i=1}^{n} \left(sem_i - \left(|obs_i - sim_i|\right)\right)^2 \tag{6}$$

If:

$$|obs_i - sim_i| < sem_i$$
$$score = 0$$

Where sem_i is the standard error of the mean for the observed glycan i, obs_i and sim_i are the relative abundances of glycan i in the observed and simulated data set respectively.

Parameter values sampled from the prior distribution were used by the SSA to simulate the glycan processing reactions. Sampling is according to the prior distribution and is implicit in the ABC approach. The score for the simulated and experimental data was calculated for each set of parameter values sampled from the prior distributions. The computed score for each set of sampled parameters was accepted if this score fell below the threshold as defined by the starting score. The starting score was reduced by 10% if the acceptance rate of parameter values were greater than 7.001% (205). The starting similarity score is continually lowered until a user-defined threshold is reached. The algorithm samples in that region of the parameter space until 10,000 parameter values for each variable have been accepted. Accepted parameter values are used to construct the posterior distribution. This methodology is equivalent to the acceptance barriers technique (206).

The user-defined threshold was determined as the lowest similarity score that could be achieved in a reasonable amount of computational time (typically within 100 hours). Defining the threshold too high would result in the acceptance of parameter values at various local minima thus giving a wide range of parameter values from each individual fitting run. To avoid this possibility the convergence of each Markov chain is determined. To assess convergence of multiple Markov chains, the Gelman-Rubin R-Statistic was used (207) and the chains visually assessed. The Gelman-Rubin statistic compares the in-chain variance with the between-chain variance providing assurance that the posterior distribution has been explored to a satisfactory extent.

30 fitting procedures were run in parallel and the posterior probability distributions for each parameter were outputted for each individual run as .csv files. A custom written Python script was then used to calculate the posterior means of each parameter for each individual fitting run. The average parameter values of at least 10 completed individual fitting runs were used to generate simulated glycan profiles. To avoid large amounts of time evaluating space in the tails of prior distributions in a high dimensional parameter space and because of the high level of uncertainty in the numerical biological parameters, prior distributions where moved only when there was biological and numerical evidence to do so. A Mann-Whitney U test was used to assign significance in means between the prior and posterior distributions for each parameter. Only if the change in means was found to be significant was the prior distribution for that parameter altered, and the rejection algorithm run again to a lower threshold.

The parameter average value of each posterior from all completed individual fitting runs was then used to simulate a glycan profile using the SSA described above (section 2.9.1).

2.9.3 Glycan flux analysis

Flux maps were created by the SSA from 10,000 input glycans. The substrate, product, enzymes and how many times the reaction occurred were exported as .csv files. Flux maps were then generated in Cytoscape software (208).

2.9.4 York Advanced Research Computing Cluster (YARCC)

The fitting of glycan profiles using the ABC algorithm was performed in parallel on the YARCC server. YARCC is a tier 3' High Performance Computing (HPC) facility available to researchers at the University of York.

Chapter 3: Model development and simulating wild type mammalian cell glycan profiles

3.1 Rationale for computational modelling of mammalian *N*-glycosylation

Computational modelling of the glycan biosynthetic process has lagged behind the modelling of other metabolic processes, yet glycosylation has functional effects on proteins, including those used as therapeutics. The importance of glycosylation in the development of biologics has been the driving force behind previous models of glycosylation. The ability to model the glycosylation of individual proteins is undoubtedly useful to the pharmaceutical industry, however, such an approach does not capture glycosylation at the cellular level. In the case of genetic disruption to the Golgi trafficking machinery or in models of CDGs, modelling the glycosylation of a single protein produced by the cell provides very little information regarding the effect on the global glycosylation process, and therefore has limited usefulness in devising personalised therapies. Here, a computational model is developed and used to simulate the *N*-linked glycan profiles of two mammalian cell lines.

Previous models that simulate the glycosylation process have used deterministic methodology (190, 195). Such methodology is not ideal for the modelling of glycan biosynthesis, as the reactions that occur in glycosylation are confined to the Golgi apparatus. The approximate volume of a Golgi cisterna is $0.03\mu m^3$ (209) compared with that of the cytoplasm of HeLa cells ($940\mu m^3$). Within such a confined volume, substrate and enzyme numbers are small, meaning biological noise is significant in determining the glycan structures that are produced by the cell. The naturally high level of heterogeneity found in glycan biosynthesis (210, 211) is suggestive that a stochastic approach is the appropriate mathematical mechanism to adopt. Furthermore, glycan biosynthesis is a process with a large number of parameters, but by taking a probabilistic approach, such as that used in this work, several of these parameters can be combined into an effective parameter and fitted directly. For example, rather than parametrising the concentration of UDP-GlcNAc, the amount of

the UDP-GlcNAc transporter and the amount of the MGAT1 enzyme, a probabilistic approach can encompass all these factors into one parameter.

The broad aim of this research has been to develop a stochastic model of *N*-glycosylation and to use it to predict alterations in the *N*-glycosylation machinery, based on experimentally determined glycan profiles. This chapter sets out the model development process from the defining of the system parameters to the simulation of whole-cell mammalian glycan profiles. In order to ensure the validity of the modelling methodology's predictions, the glycan profile of cells treated with a glycosylation enzyme inhibitor have also been fitted.

3.2 Aims

As previously stated, the aim of the work described in this chapter was to develop a probabilistic model capable of simulating whole cell *N*-glycan profiles of wild type mammalian cell lines. To achieve this, the developed model was used to fit and simulate the glycan profiles of WT HeLa cells (60) and WT HEK293T cells (58). An additional aim was to model the glycan profile of cells treated with the MAN2 inhibitor, swainsonine, in order to evaluate the predictive power of the developed computational methodology.

3.3 Results

3.3.1 Model development and parameterization

Following *de novo* synthesis in the ER, the suite of *N*-glycan structures produced by a cell is vastly expanded through processing in the Golgi apparatus. The glycan profile generated by each cell is dependent on the levels of individual enzymes and their distribution between Golgi cisternae. In order to simulate the reactions that generate the diverse spectrum of *N*-glycans, a stochastic simulation algorithm (SSA) based on the Gillespie algorithm (180, 212) was developed and parameterized.

The SSA was parameterized with a set of enzymes and rules defining the reactions catalysed by each enzyme (table 3.1). The set of enzymes used in this work represented the smallest number of enzymes that could generate each glycan in the experimentally observed glycan profiles. The decision to choose the smallest set of

enzymes was taken to reduce the computational cost of the fitting process, this did however, exclude glycan structures that may contain, for example, bisecting GlcNAc or poly-lactosamine repeats. Table 3.1 shows the enzymes used in this work, the enzyme target (which can be considered the enzyme substrate) and the result of the reaction catalysed by that enzyme. Defining the reactions of each enzyme by more than one rule when called for, allowed control over substrate specificity. In other words, when driven by the fitting procedure and if described in the literature, enzymatic activities were allowed to differ for different substrates. This is described in more detail in the section describing the WT HEK293T cell data (section 3.3.3). It also proved impossible to define generic rules so that each enzyme was defined by a single rule, hence, the enzymes listed have several targets.

Enzyme	Enzyme target	Enzyme result	Cell line
MAN1 (1)			HL, HK, MC
MAN1 (2)			HL, HK, MC
Endo-Man			HL, HK, MC
MAN2			HL, HK, MC
MGAT1			HL, HK, MC
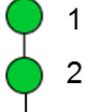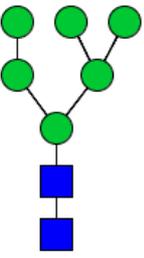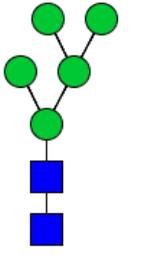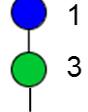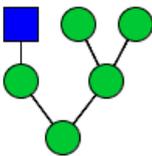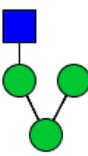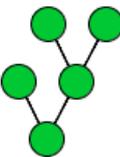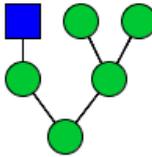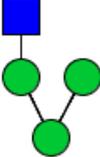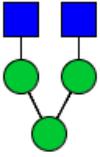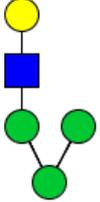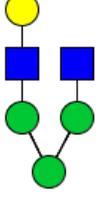
Enzyme	Substrate	Product	Tissue
MGAT2 (1)			HL, HK, MC
MGAT2 (2)			HL, HK, MC
MGAT2 (3)			HL, HK, MC
MGAT4			HL, HK, MC
MGAT5			HL, HK, MC
FUT8 (1)			HL, HK, MC
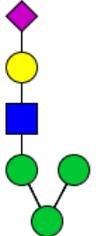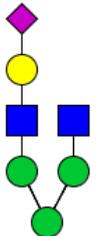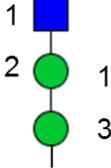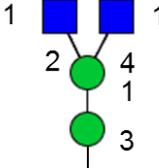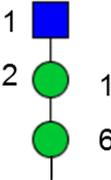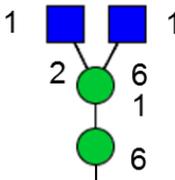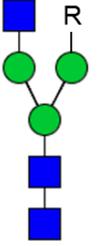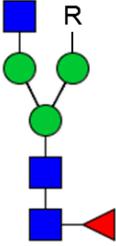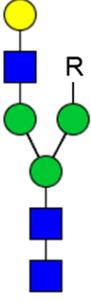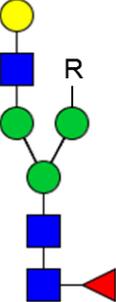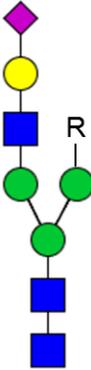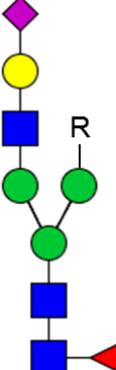
FUT8 (2)			HL, HK, MC
FUT8 (3)			HL, HK, MC
FUT8 (4)			HL, HK, MC
FUT8 (5)			HL, HK, MC
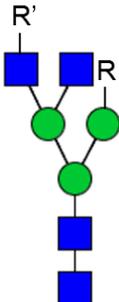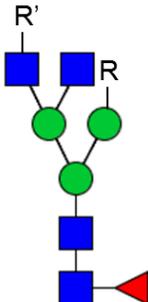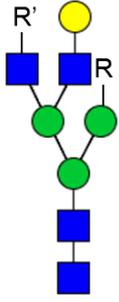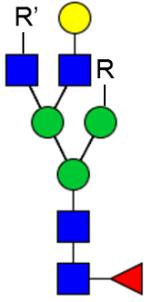
FUT8 (6)			HL, HK, MC
FUT8 (7)			HL, HK, MC
Antenna FUT			HK, MC
GalT (1)			HL, HK, MC
GalT (2)			HL, HK, MC
GalT (3)			HL, HK, MC
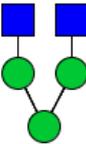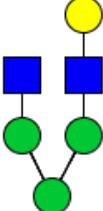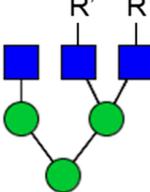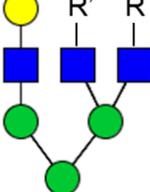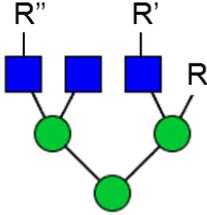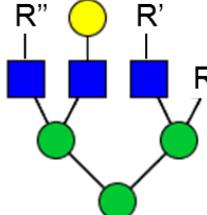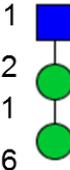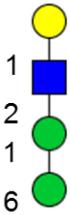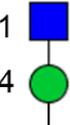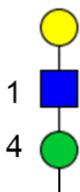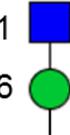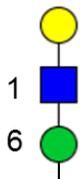
GalT (4)			HL, HK, MC
GalT (5)	R' R	R' R	HL, HK, MC
GalT (6)	R" R' R	R" R' R	HL, HK, MC
GalT (7)	1 2 1 6	1 2 1 6	HL, HK, MC
GalT (8)	1 4	1 4	HL, HK, MC
GalT (9)	1 6	1 6	HL, HK, MC
SiaT (1)			HL, HK, MC

SiaT (2)			HK, MC
SiaT (3)			HK, MC
GlcNAc-phosphotransferase			HL, HK, MC

Table 3.1 | Schematic representation of enzymatic rules

Glycosylation enzymes that are included in the SSA their substrates and the products they generate. HeLa (HL), HEK293T (HK) and mesenchymal stromal cells (MC). Where R and R' R" are further oligosaccharide chain extensions or branching as a result of other rules. Numbers represent linkage between residues.

To model the Golgi apparatus, the suite of enzymes defined in Table 3.1, were initially distributed in a non-linear fashion into three cisternae. Distributing the enzymes into separated compartments was required for several reasons. Firstly, much evidence suggests that Golgi-resident proteins are unevenly distributed throughout the Golgi stack (17, 22). Secondly, previous models have shown that simulating mammalian glycosylation reactions within a single compartment could not reproduce observed data (191, 195). Finally, the distribution of enzymes is an important factor in determining the glycan profiles of cells, as illustrated by COG-CDGs. The initial localization of enzymes was based on microscopy studies of enzyme Golgi localization, where available. For example, endo-mannosidase (213), GlcNAc-phosphotransferase (214), and MGAT1, MAN2, GalT and SiaT (22). Alternatively, the initial distribution of enzymes was determined based on the positions they act in the *N*-glycan pathway. The entire activity of each enzyme was not confined to just one cisterna (215) because of the nature of cisternal maturation. Cisternal maturation is likely to result in a degree of leakage of enzymes into adjacent cisternae.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	References
GlcNAc-phosphotransferase	80%	10%	10%	(214)
MAN1	80%	10%	10%	(216)
Endo-mannosidase	80%	10%	10%	(213)
MAN2	10%	80%	10%	(22)
MGAT1	10%	80%	10%	(22)
MGAT2	10%	80%	10%	(217)
MGAT4	10%	80%	10%	(218)
MGAT5	10%	80%	10%	(218)
FUT8	10%	80%	10%	(219)
GalT	10%	10%	80%	(22)
SiaT	10%	10%	80%	(22)

Table 3.2 | Initial distribution of enzymes in the three-cisterna model of WT HeLa cells

Total activity of each enzyme was distributed into three-cisterna within the Golgi apparatus. For example, SiaT is thought to be localized in the trans cisternae, so 80% of the activity was initially placed in cisterna 3. The distributions shown here are prior information and are subject to the fitting procedure.

To determine the relative activities of the glycan-modifying enzymes and their unique localizations within the Golgi stack, a Bayesian framework (189) was developed. The

unique advantage of using a Bayesian framework is that it allows the incorporation of existing knowledge, no matter how tentative or precise, to construct a series of prior distributions for our parameters. This results in a broad distribution for largely unknown parameters, or a sharp distribution centred around a single value for well-characterized parameters. The use of this Bayesian methodology, known as an approximate Bayesian computation without likelihoods, does not require exact well-defined enzymatic rates and localizations. The initial distribution of each enzyme in the model Golgi for WT HeLa cells is shown in table 3.2. The enzyme distributions fall into three categories: *cis*, *medial* and *trans*. While HeLa cells likely have at least four cisternae, the aim was to keep the model as simple as possible to minimize the parameter matrix. In the context of the Bayesian framework, the use of log normal probability distributions for the dominant cisterna (figure 3.1), or exponential decay probability distributions for the minor cisternae (figure 3.1) were used as the initial prior probability distributions.
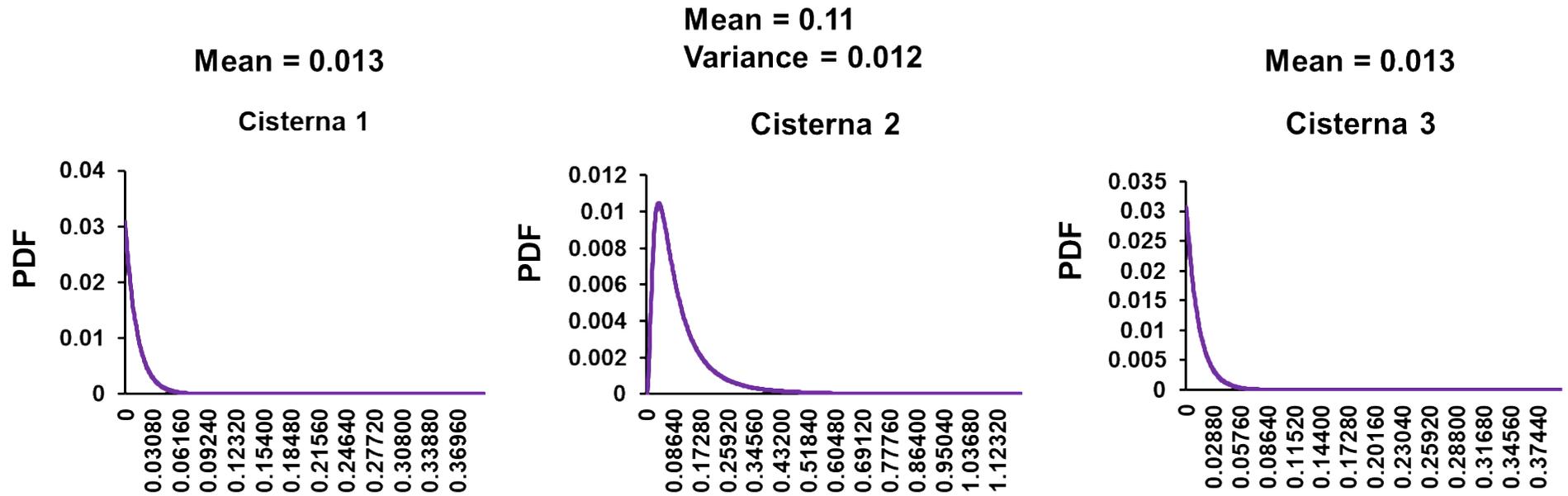
Figure 3.1 | Probability density functions (PDF) defining enzyme distributions in a three cisterna Golgi apparatus

Probability functions for the distribution of a medial localized enzyme in a three cisterna Golgi model. Exponential decay functions were used to represent the probability distributions of the enzyme in which the activity of the enzyme is low, in this case cisternae 1 and 3. A log normal probability distribution was used to represent the probability distribution of enzymatic activity in the cisterna with high enzymatic activity (cisterna 2). Values on the x-axis represent parameter values.
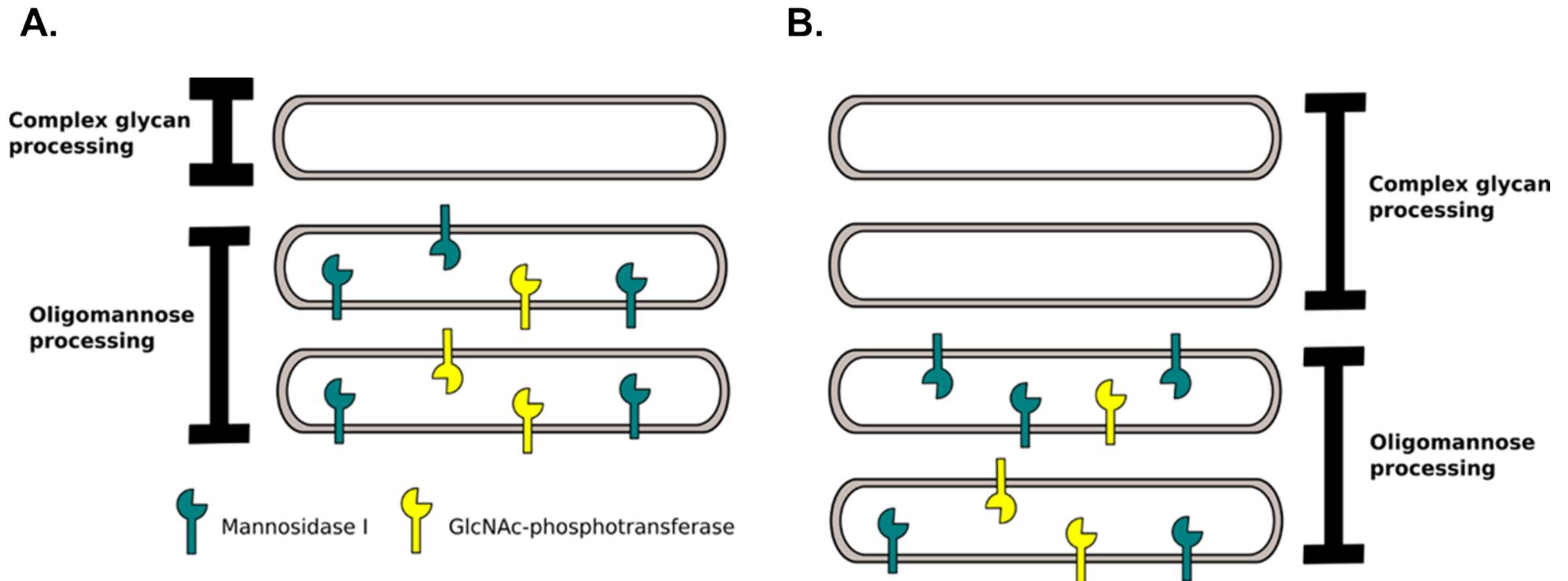
Figure 3.2 | Rationale for four-cisterna model in HEK293T cells

A) Cartoon representation of the three-cisterna model. B) Cartoon representation of the four-cisterna model. In both models of the Golgi apparatus, two cisternae are required to fit the distribution of the oligomannose glycans through the action of MAN1 and GlcNAc-phosphoT. The remaining cisternae being available for fitting the hybrid and complex glycan structures. In the case of the HeLa cells, three cisternae were adequate due to the low abundance of hybrid and complex glycans, however, in the HEK293T cells at least two cisternae are required for fitting the complex glycans.

The fitting of the HEK293T glycan profile was initially attempted with the Golgi model used for HeLa cells. However, due to the more elaborate glycan profile of HEK293T cells, the introduction into the model of a fourth cisterna was required to obtain a good fit with the observed data (figure 3.2). The addition of a fourth cisterna for the modelling of glycosylation in the HEK293T cells increased the glycan processing capacity of the model Golgi. This increased modelling capacity appeared necessary, as a minimum of two cisternae was required to model the oligomannose glycans adequately. In the model, the oligomannose distribution is determined by competition between MAN1 and GlcNAc-phosphoT. In the case of the HEK293T cells, a further two cisternae were required to model the complex glycans. The total activity for each enzyme was, therefore, distributed into four cisternae (table 3.3) for the fitting of the HEK293T glycan profile.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
GlcNAc-phosphotransferase	80%	10%	5%	5%
Endo-mannosidase	80%	10%	5%	5%
MAN1	10%	80%	5%	5%
MAN2	10%	40%	40%	10%
MGAT1	10%	40%	40%	10%
MGAT2	10%	40%	40%	10%
MGAT4	10%	40%	40%	10%
MGAT5	10%	40%	40%	10%
FUT8	10%	40%	40%	10%
GalT	10%	10%	40%	40%
SiaT	10%	10%	40%	40%

Table 3.3 | Initial distribution of enzymes in the four-cisterna model of WT HEK293T cells

Total activity of each enzyme was distributed into four-cisterna in line within the Golgi apparatus. For example, SiaT was placed equally in cisternae 3 and 4, with lower activity placed in cisternae 1 and 2. The distributions shown here are prior information and are subject to the fitting procedure.

3.3.2 Fitting and simulating the WT HeLa cell *N*-glycan profile

To generate an *in silico* glycan profile, the processing of 10,000 input glycans that enter the Golgi one at a time, was simulated. This allows explicit simulation of enzyme competition but not glycan competition. Each glycan is acted upon by a changing set of enzymes, determined by the available target sites; as new carbohydrate structures

are generated, possible target sites also change. This is modelled as a stochastic process and simulated using the standard Gillespie algorithm (180). A glycan structure that changes over time in the Golgi is the result of the simulation. The process is interrupted by the passage to a new cisterna and terminated when the time limit is reached, i.e. the glycan exits the Golgi.

Rather than first fitting the complete glycan profile of HeLa cells, a less computationally demanding experiment, the simulation of the oligomannose profile of HeLa cells was instead performed. Figure 3.3 shows the observed oligomannose profile of WT HeLa as previously published (60), alongside the simulated oligomannose glycan profile. The hybrid and complex glycans were pooled together, so the simulation of the processing of complex glycans was not included. Parameter values used for the simulation of the oligomannose profile shown in figure 3.3 were based on the enzymatic turnover numbers of Golgi endo-mannosidase (endoMan), MAN1, MGAT1 and N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphoT), but without fitting (table 2.6). The enzyme complex of GlcNAc-PhosphoT, which is responsible for adding GlcNAc-phosphate to mannose residues, was included in the enzyme suite. The Man-6-phosphate (Man6P) tag is used to target glycoproteins to the lysosome, where the phosphate is removed. At the same time, the presence of Man6P prevents conversion of glycans to hybrid and complex forms in the Golgi apparatus. Given that glycan profiles from whole cell lysates are used in this work, it is likely that they contain a significant proportion of lysosomal oligomannose glycans. It was therefore important to include this modification route. The presence of $Glc_1Man_9GlcNAc_2$ in N-glycan mass spectra indicated that Golgi endoMan should be included in the modelling framework. Including this enzyme allowed for the possibility that the major N-glycan exiting the ER was $Glc_1Man_9GlcNAc_2$, which could then be processed further in the Golgi apparatus.

As the simulated glycan profile shown in figure 3.3 was generated using parameter values that had not undergone the fitting procedure, the simulated glycan abundances do not match well with the observed glycan abundances. The simulated distribution of the oligomannose glycans does not create the doubly peaked distribution of the observed oligomannose distribution (section 1.1.3.1). Rather, the simulated oligomannose profile creates a singly peaked distribution, peaking at $Hex_6HexNAc_2$. The competition between two enzymes, in this case MAN1 and GlcNAc-phosphoT,

cannot generate a double peaked distribution. In addition to this discrepancy between the observed and simulated data, the simulated profile also generates too high a level of hybrid/complex glycan compared with the observed.
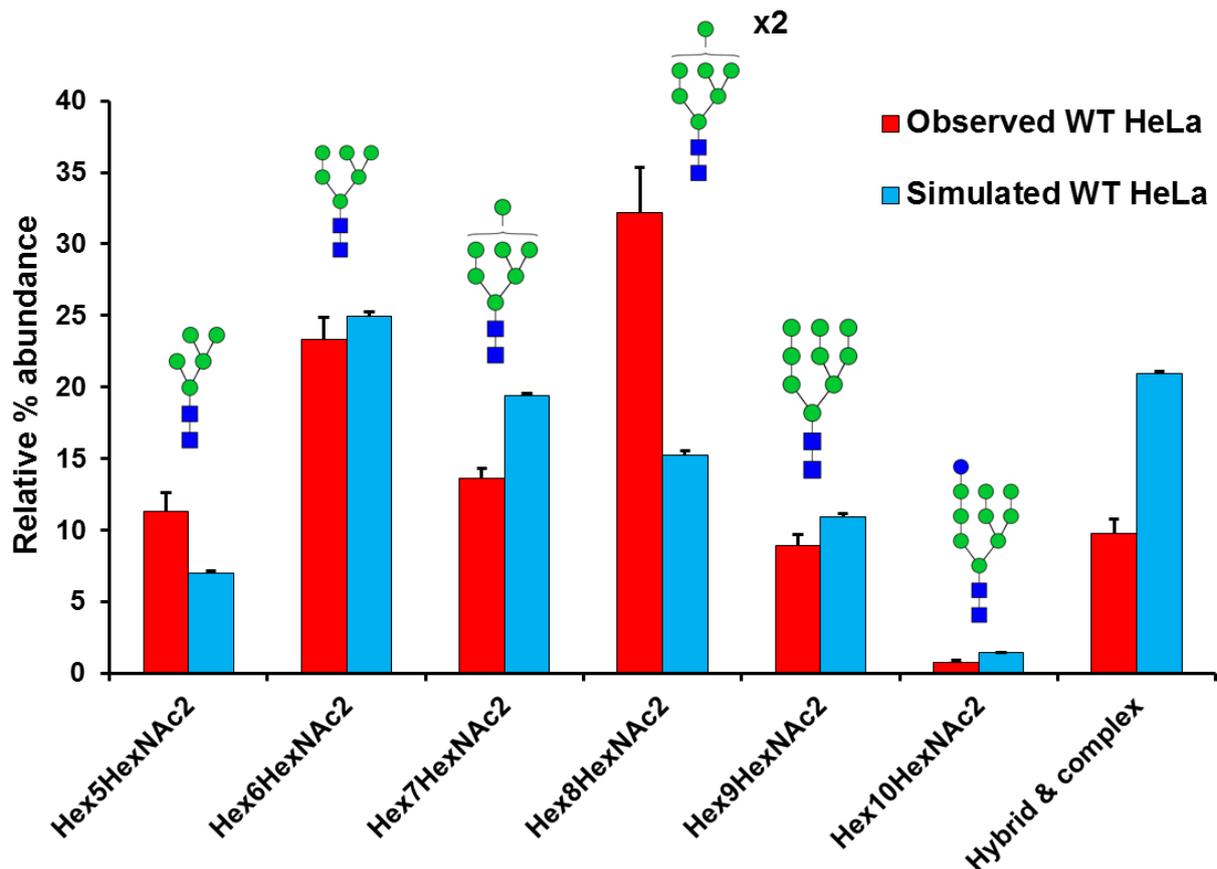
Figure 3.3 | Initial parameter estimation for modelling oligomannose glycans in HeLa cells

Observed (red) and simulated (blue) oligomannose glycan profile of WT HeLa cells. Observed glycan abundances have been previously published (Abdul Rahman et al. 2014). The glycan profile was simulated three times using the SSA and the endo-mannosidase, MAN1, GlcNAc-phosphoT and MGAT1 enzymes. The processing of hybrid and complex glycans were excluded from the simulation. The parameter values used to simulate the glycan profile were the initial values used to parameterize the model. Error bars are SEM, n = 3.

The inability of the model after fitting to reproduce this characteristic doubly peaked oligomannose distribution led to the addition of a $Man_6GlcNAc_2$ trimming scale-factor. In other words, the trimming of $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$ was allowed to proceed at a different rate to the trimming of all other oligomannose species by MAN1, an enzymatic property that has been noted in the literature (99). This scale-factor was

implemented as an internal property of the MAN1 enzyme and is defined as the ratio of the enzymatic activity for MAN1 trimming of $Man_{9-7}GlcNAc_2$ divided by the enzymatic activity of MAN1 trimming of $Man_6GlcNAc_2$.
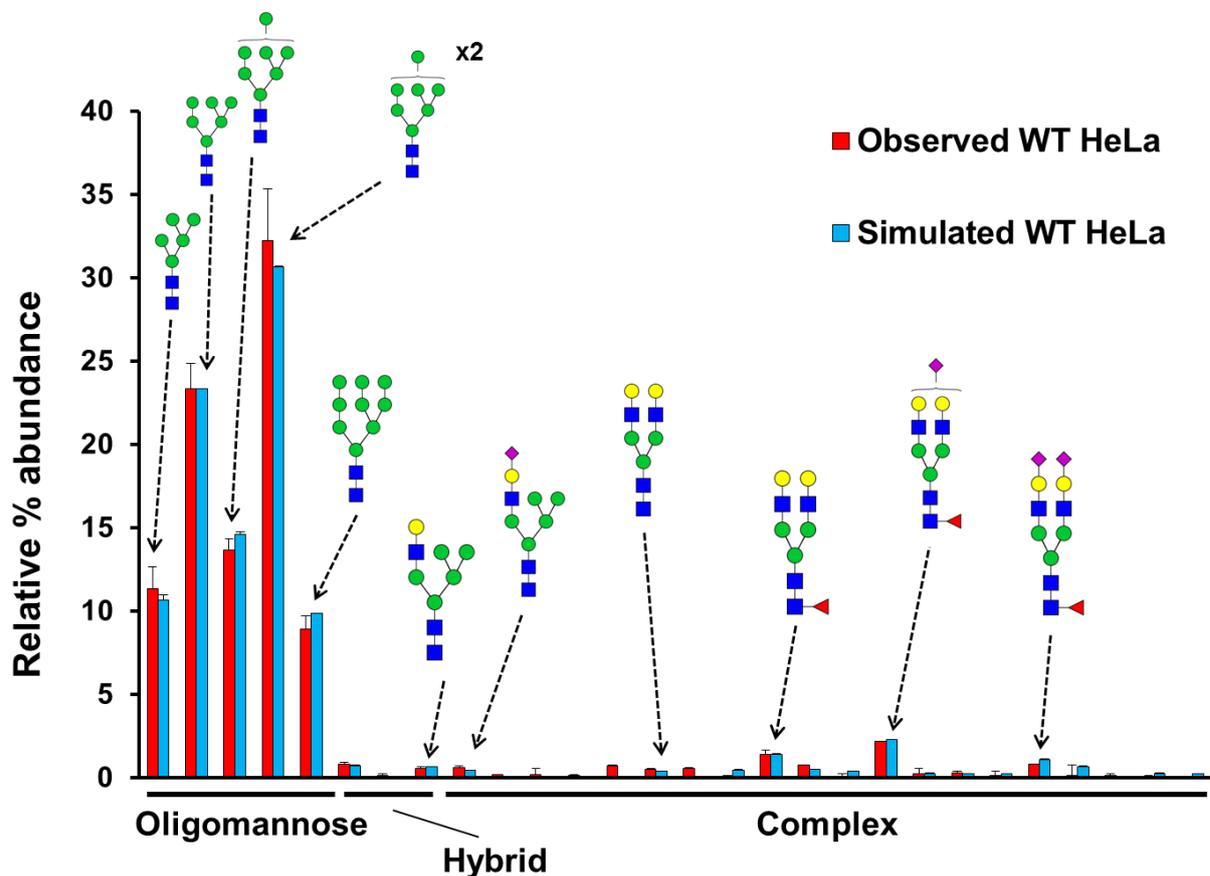
Figure 3.4 | Modelling of whole cell *N*-glycome from HeLa cells

Observed (red) and simulated (blue) glycan profiles of WT HeLa cells. The observed glycan profile of whole cell WT HeLa cells has been previously published (Abdul Rahman et al. 2014). The glycan profile is simulated three times using the whole suit of HeLa enzymes as shown in table 3.1 following the fitting procedure. The scale-factor for the trimming of Man6GlcNAc2 is included. Error bars are SEM, n = 3. Note all simulated glycan abundances under 0.1% are not shown.

Having determined the need for a $Man_6GlcNAc_2$ trimming scale-factor, the whole cellular glycan profile, including the hybrid and complex glycans, was then fitted (figure 3.4). Unlike in figure 3.3, the simulated distribution of oligomannose glycans matches the experimentally observed abundances. Indeed, the whole cell HeLa glycan profile could be reproduced *in silico* with a model distributing enzymes into three cisternae. Due to the high abundance of oligomannose glycans in the WT HeLa cells, the

strongest predictions can be made for the oligomannose-processing enzymes such as MAN1. An initial assumption based on the literature was that MAN1 was predominately located in the *cis* Golgi (17, 220). However, after fitting the observed glycan profile of HeLa cells the model predicted that the enzyme was spread over the whole stack, with a weak *medial*-Golgi preference (table 3.4). Although often considered a marker of the *cis*-Golgi, MAN1 localization has been shown to vary across cell types (221), consistent with the *in silico* prediction. It should also be noted that the MAN1 enzyme remains in the *cis* relative to the majority of the other enzymes in the system.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3
EndoMAN	79±14%	11±3%	11±2%
GlcNAc-PhosphoT	81±4%	9±2%	9±2%
MAN1	25±2%	56±5%	19±2%
MAN2	10±2%	77±12%	13±2%
MGAT1	4±1%	93±15%	3±0.003%
MGAT2	11±2%	78±11%	10±2%
MGAT4	11±2%	79±15%	11±3%
MGAT5	6±1%	88±17%	6±1%
FUT8	10±2%	80±8%	10±1%
omFUT8	8±2%	84±19%	8±2%
GalT	11±1%	11±1%	78±11%
SiaT	4±1%	4±1%	91±6%

Table 3.4 | Optimised distribution of enzymes for fitted HeLa cell glycome

The predicted distribution of enzymes in WT HeLa Golgi apparatus following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 20 individual fitting runs and confidence interval is standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

GlcNAc-PhosphoT was allowed to phosphorylate $Man_9GlcNAc_2$, $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$, as this was found to be the most appropriate target set (figure 3.5). As shown in figure 3.5, the best fit of the HeLa oligomannose distribution could be achieved with $Man_9GlcNAc_2$, $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ (Man9-7) being the substrates for GlcNAc-PhosphoT. Expanding the substrates for GlcNAc-PhosphoT to include $Man_6GlcNAc_2$ (Man9-6) resulted in an increase in $Man_7GlcNAc_2$ and a concomitant decrease in the abundance of $Man_8GlcNAc_2$. Following this result, all

other glycan profiles have been simulated with the Man9-7 substrate set for the GlcNAc-PhosphoT enzyme.
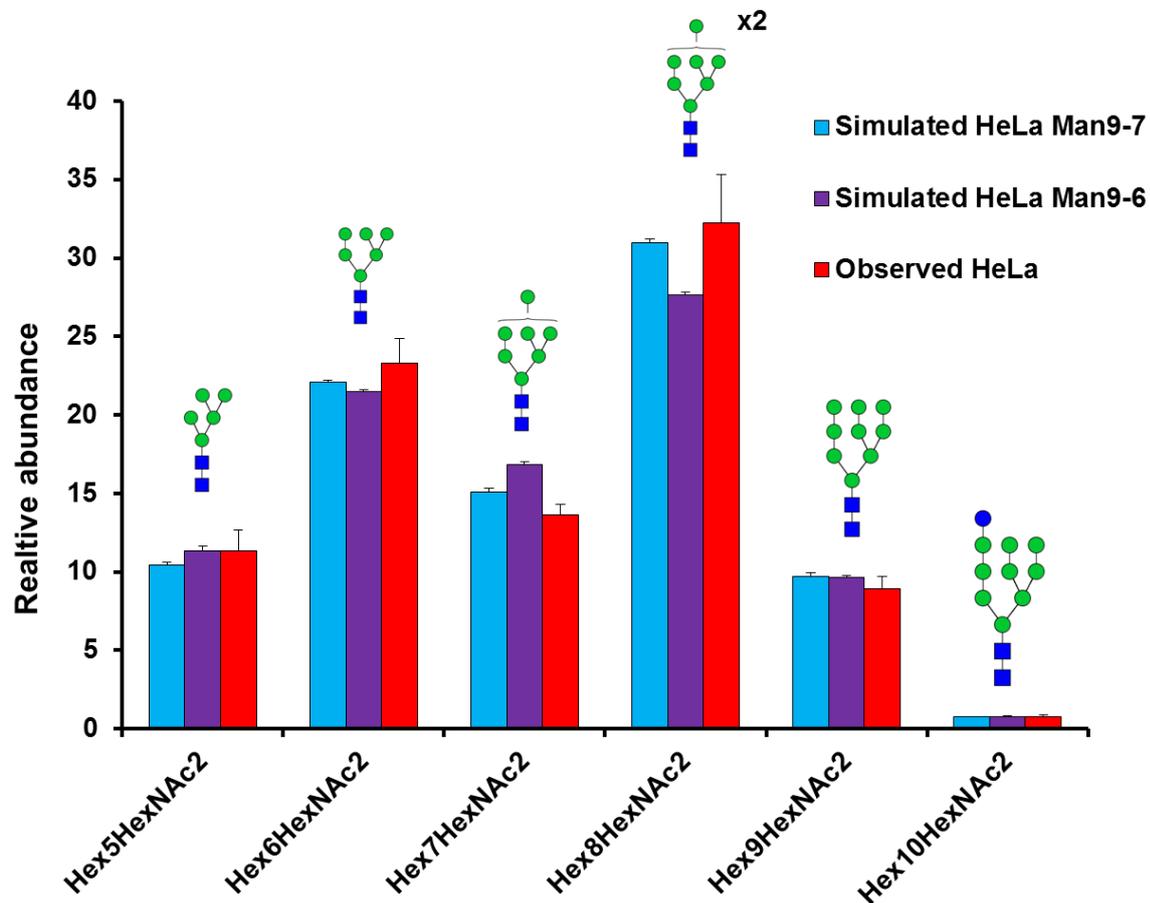
Figure 3.5 | Rationale for GlcNAc-phosphotransferase rule set

Observed (blue) and simulated oligomannose glycan profile. The observed glycan profile has been previously published (Abdul Rahman et al. 2014). The glycan profile was either simulated using enzymatic rules that allowed GlcNAc-phosphoT to act on Man9-7GlcNAc2 (blue) or Man9 6GlcNAc2 (purple). The glycan profile is simulated three times using the SSA. Error bars are SEM, n = 3.

Importantly, the modelling methodology developed was able to predict enzyme parameter values that were used to accurately simulate the whole cell glycan profile of HeLa cells. To test that the modelling methodology could be applied to other mammalian cell lines with more elaborate glycan profiles, the modelling methodology was next used for glycan profiles derived from HEK293T cells.

3.3.3 Fitting and simulating the WT HEK293T cell *N*-glycan profile and comparison with WT HeLa cell predictions

Before modelling the glycan processing of HEK293T cells, a whole cell glycan profile for this cell line needed to be obtained experimentally. To generate a glycan profile for HEK293T cells, the *N*-glycans were harvested, released, permethylated and analysed by MALDI mass spectrometry. The *N*-glycans were identified by their accurate mass, the isotope pattern of their ionic signals, and knowledge of the biosynthetic pathway. 53 *N*-glycans were detected and structures proposed using this method, including oligomannose, hybrid and complex types. A selection of the most abundant glycans detected is shown annotated in a representative mass spectrum from the HEK293T glycan profile (figure 3.6). It is possible that the time of cell line harvesting can affect the glycan profile, influenced by the availability of nutrients in the culture medium or reactions of the cell to its environment (e.g. confluency). To minimise the effect of harvest time on the glycan distribution the cell lines were harvested at the same confluency throughout. This work was done in collaboration with Vladimir Lupashin (University of Arkansas for Medical Sciences) to investigate the role of different COG complex subunits on cellular phenotypes (58).

Comparison of the relative abundance of *N*-glycans from WT HeLa and HEK293T cells showed significant differences. The most striking difference between the two cell lines was the abundance of oligomannose glycans, which made up 90% of the total glycan abundance in HeLa cells (figure 3.7). This compared with 58% oligomannose glycans in the HEK293T glycan profile (figure 3.7). Features of complex glycans, such as fucosylation, sialylation and antenna number were also significantly reduced in the HeLa cell line when compared with HEK293T cells (figure 3.7).
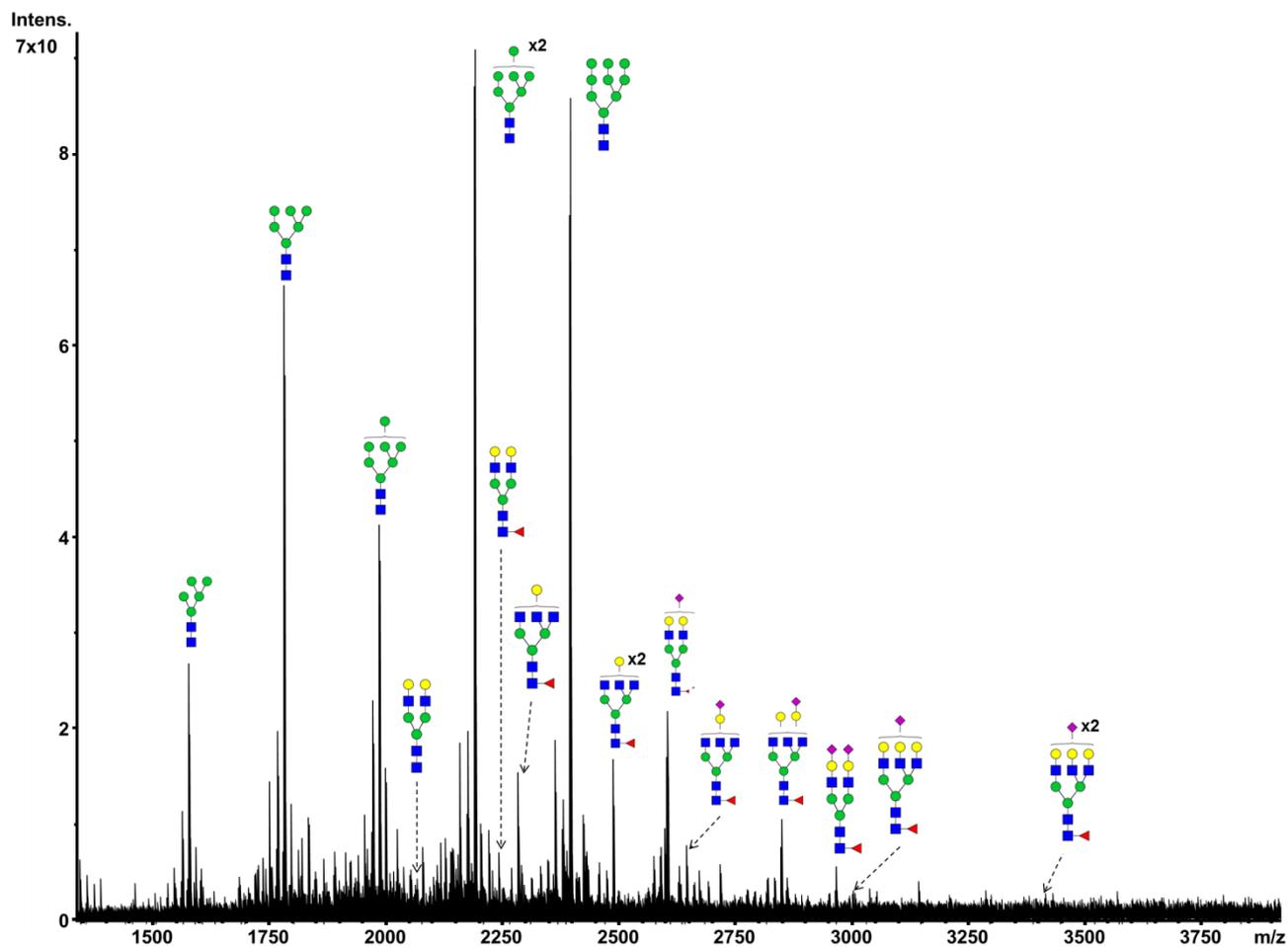
Figure 3.6 | *N*-glycan mass spectrum of wild type HEK293T cells

A representative mass spectrum of *N*-glycans released from WT HEK293T cells. *N*-glycans were released with PNGaseF and permethylated. Mass spectra were acquired using FT-ICR MS. Eight scans consisting of 500 laser shots were averaged to generate a mass spectrum. The signal to noise ratio had to be >3 for the ion to be included. Glycan assignments for a selection of the more abundant signals are shown.
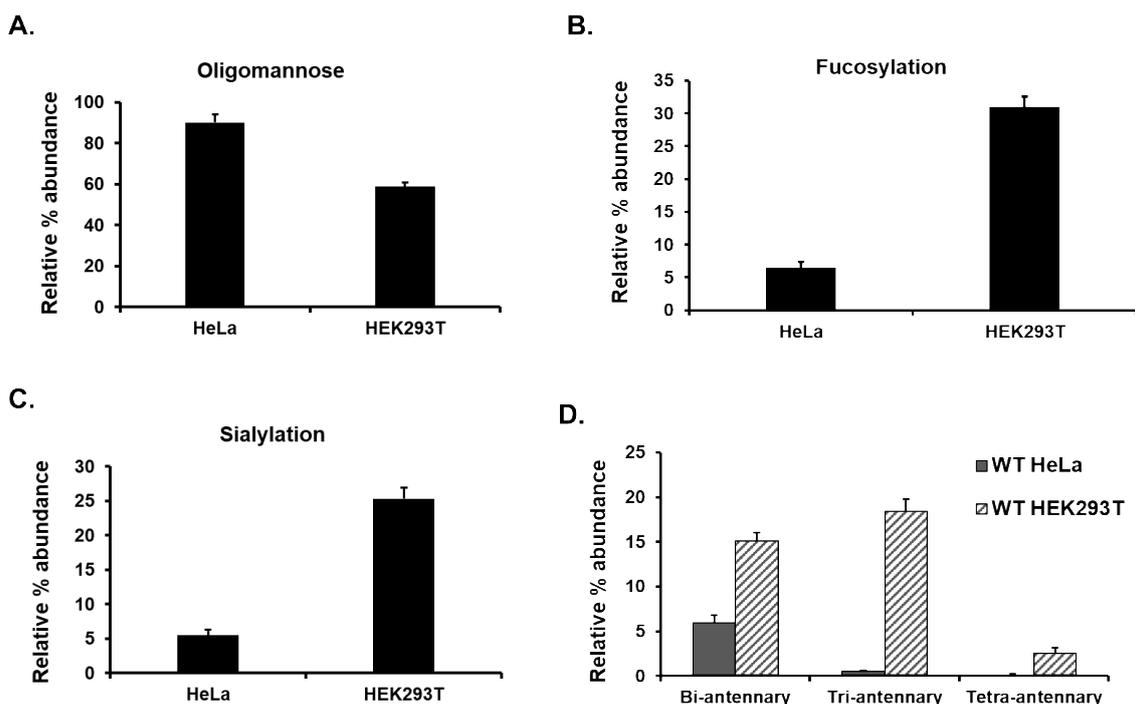
Figure 3.7 | Comparison of WT HeLa and WT HEK293T *N*-glycans

Glycan peaks were identified from mass spectra and quantified as outlined in the material and methods. Quantification of glycans from WT HeLa cells was achieved using previously published mass spectra (Abdul Rahman et al. 2014). Relative % abundance of: oligomannose glycans (A), fucosylated glycans (B), sialylated glycans (C) in WT HeLa and HEK293T cells. Relative % abundance of bi-, tri and tetra-antennary glycans (D). Error bars are SEM, n = 3.

For fitting the HEK293T glycan profile, the optimised HeLa cell parameter values were used as a starting point, allowing a direct comparison of the predictions for the two cell lines after fitting. A satisfactory fit between the observed and simulated glycan profiles could be achieved (figure 3.8). The significant proportion of complex glycans in the HEK293T glycan profile allows more confident predictions to made about the level and localization of glycosylation enzymes involved in the processing of complex glycans. The predicted percentage activities of each enzyme in the four-cisterna system are shown in table 3.5. Similar to the predictions for HeLa cells, in HEK293T cells endo-mannosidase and GlcNAc-phosphoT are predicted to be the enzymes acting earliest in the Golgi, followed by MAN1 and MGAT1. The *medial* enzymes, such as MGAT2, 4 and 5, are evenly split between cisternae 2 and 3. This even split between cisternae 2 and 3 suggests that both oligomannose and complex glycan processing is occurring in cisterna 2. GalT and SiaT are the only two enzymes that are predicted by the model to be localized in the *trans* Golgi.
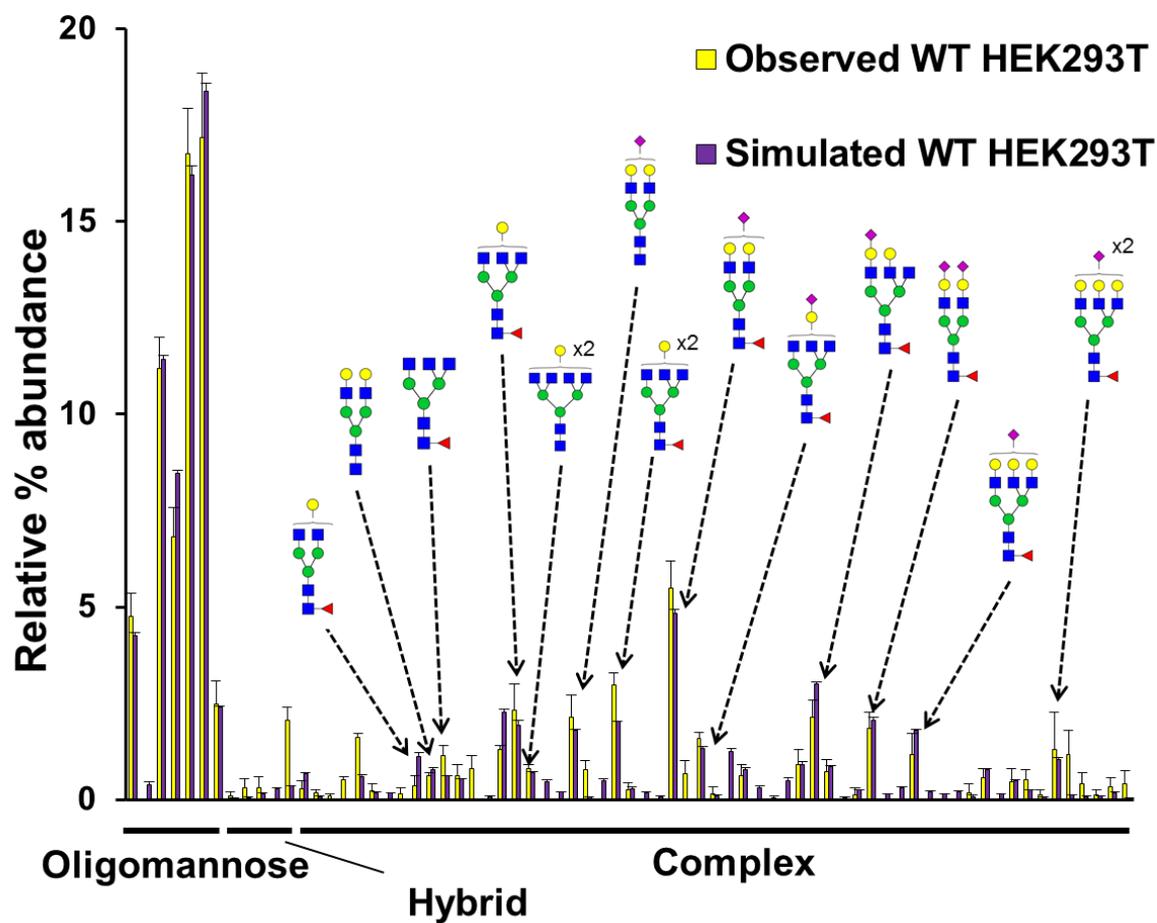
Figure 3.8 | Modelling of whole cell *N*-glycome from HEK293T cells

Observed (yellow) and simulated (purple) glycan profile of whole cell WT HEK293T cells. Glycan peaks were identified from mass spectra. The total peak area intensities were used to quantify each identified glycan structure to generate an observed glycan profile. The glycan profile was simulated three times using the SSA with the mean parameter values from n = 15 individual fitting runs. Error bars are SEM, n = 3.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
EndoMAN	76±2%	13±2%	10±1%	1±0.2%
GlcNAc-PhosphoT	66±2%	15±1%	18±2%	2±0.3%
MAN1	9±0.2%	80±3%	10±1%	1±0.2%
MAN2	9±1%	50±10%	32±9%	9±2%
MGAT1	1±0.05%	85±5%	13±1%	1±0.1%
MGAT2	11±0.5%	41±3%	37±4%	11±2%
MGAT4	9±1%	41±13%	40±11%	10±2%
MGAT5	4±0.4%	36±6%	53±8%	7±3%
FUT8	14±1%	52±6%	21±2%	13±3%
ant FUT	4±0.4%	52±9%	41±8%	4±0.6%
GalT	2±0.1%	13±2%	38±4%	47±5%
SiaT	3±0.5%	3±0.1%	42±16%	51±9%

Table 3.5 | Optimised distribution of enzymes for fitted HEK293T cell glycome

The predicted distribution of enzymes in WT HEK293T Golgi apparatus following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 15 individual fitting runs and confidence interval is standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

Over the course of fitting whole cell glycan profiles, a number of additional parameters were incorporated into the enzymes' rules, all of which have precedent in the literature and improve our biological understanding of the process of glycosylation. These parameters included a different MAN1 rate for the conversion of $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$ (95, 99) than for all other mannose-cleaving reactions by the same enzyme (as previously discussed); a separate rate of sialylation for the galactose residing on the 3.1Man and 6.1Man antennae (111, 222) and separate rates of galactosylation for bi-antennary versus tri- and tetra-antennary glycans (92) (table 3.9).

The scale-factors for galactosylation and sialylation were introduced during the WT HEK293T glycan profile fitting process. These additions were likely not required for the modelling of the HeLa cell glycan profile, due to the low abundance of complex glycans. The GalT scale factor is defined as the ratio of the enzymatic activity of GalT acting on a bi-antennary glycan over the enzymatic activity of GalT acting on tri- or tetra-antennary glycans. Similarly, the sialylation scale-factor is defined as the enzymatic activity of SiaT acting on the 3.1Man antenna divided by the activity of SiaT acting on the 6.1Man antenna.

As the fitted HeLa cell parameters were used as the prior distributions for the modelling of the HEK293T glycan profile, the fitted parameter values of the cell lines may be directly compared (figure 3.9). A comparison of the enzyme localization reveals a strong similarity between the two cell lines. As would be expected, the introduction of a fourth cisterna for the HEK293T cell line allows the elucidation of finer glycan processing than for the HeLa cells. For example, it is clear that MAN1 and MGAT5 are predicted to be spatially separated in the HEK293T cells, while this is less clear for the predicted localizations for the HeLa cells.
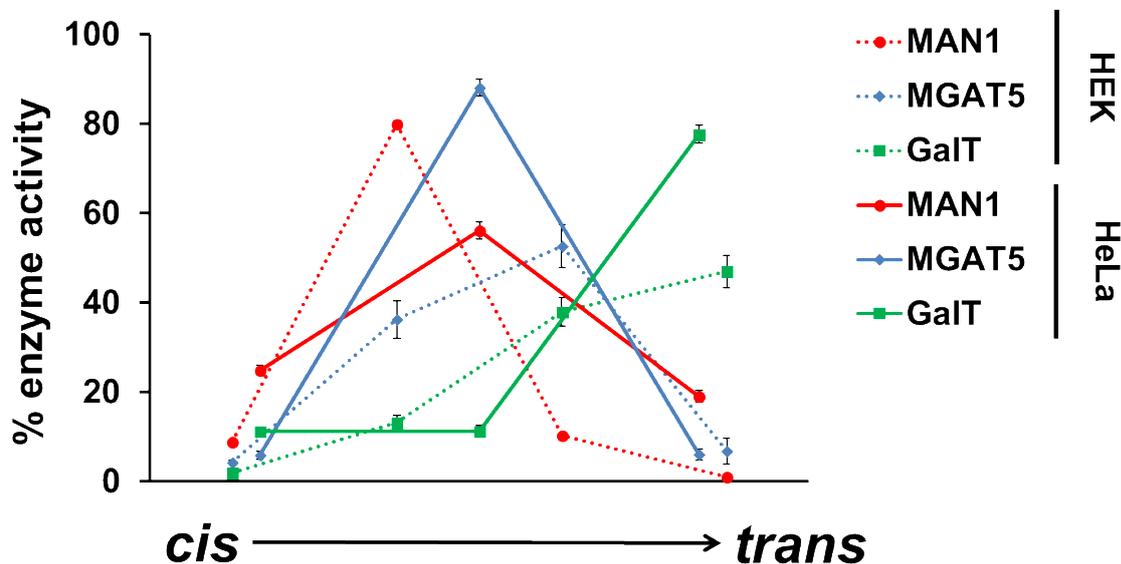
Figure 3.9 | Comparison of selected parameter distributions for fitted HeLa and HEK293T cells

Predicted enzymatic distributions of MAN1, MGAT5 and GalT in HeLa and HEK293T cells following the fitting procedure. Error bars are standard deviation for n = 20 (HeLa) and n = 15 (HEK293T) individual fitting procedures.

The ability to directly compare cell lines is an important feature of the modelling framework developed in this work. In addition to the fitting of enzymatic parameters, the proportion of each precursor glycan entering the Golgi apparatus and the time taken to transit through each cisterna are also fitted (figure 3.10). The transit time through each cisterna is predicted to be longer in the HEK293T cells compared with the HeLa cells (figure 3.10 C). The compositions of precursor glycans entering the

Golgi apparatus are also different between the two cell lines. The major precursor glycan in the HeLa cells was predicted to be $Man_8GlcNAc_2$, as opposed to $Man_9GlcNAc_2$ in the HEK293T cells (figure 3.10 A, B).
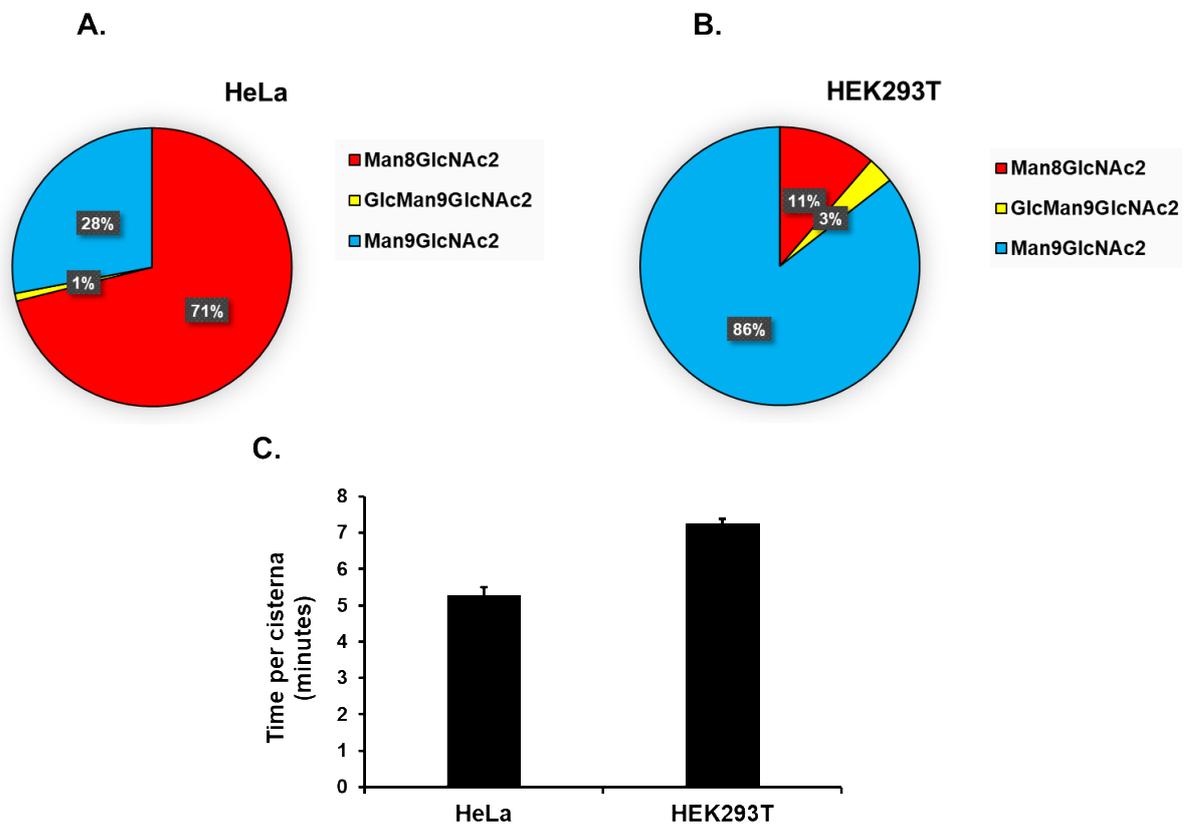
Figure 3.10 | Comparison of input glycans and transit time per cisterna in HeLa and HEK293T cells

Mean fitted values for the proportion of the input glycans for HeLa (A) and HEK293T (B) cells. Transit time per cisterna following fitting of glycan profiles for WT HeLa and HEK293T cell lines (C). Error bars are standard deviation for n = 20 (HeLa) and n = 15 (HEK293T).

The modelling framework that has been developed, namely an SSA capable of simulating the glycosylation reactions that occur within the Golgi apparatus, and an ABC without likelihood algorithm to optimise the parameters, was able to simulate accurately the elaborate HEK293T whole cell glycan profile. Fitting of the HEK293T glycan profile required introduction of several additional parameters. It is important to emphasize that the addition of these substrate-dependent parameters was driven by the desire for a well fitted model and then justified based on information in the literature, suggesting the model is physiologically relevant.

3.3.4 Fitting and simulation of swainsonine-treated HeLa and HEK293T cells

The power of the model developed is to make predictions relative to the prior distribution, hence the ability to compare the organization of the glycosylation machinery in HeLa and HEK293T cell lines. To demonstrate that the modelling methodology can make relevant and correct predictions, the model was used to investigate the glycan profile of swainsonine-treated HeLa (60) and HEK293T cells. Swainsonine is an alkaloid drug that inhibits MAN2 (106), preventing the conversion of hybrid to complex-type glycans. HEK293T cells were treated with swainsonine (see material and methods), the *N*-glycans were harvested, released, permethylated and analysed by MALDI mass spectrometry. As before, the accurate mass, isotope pattern and knowledge of the biosynthetic pathway was used to propose structures for the *N*-glycans. Due to the inhibition of MAN2 and the elimination of detectable complex glycans, only 13 glycans were identified. The most abundant hybrid glycans detected are shown in figure 3.11 and are the same as those reported in the swainsonine-treated HeLa cells.
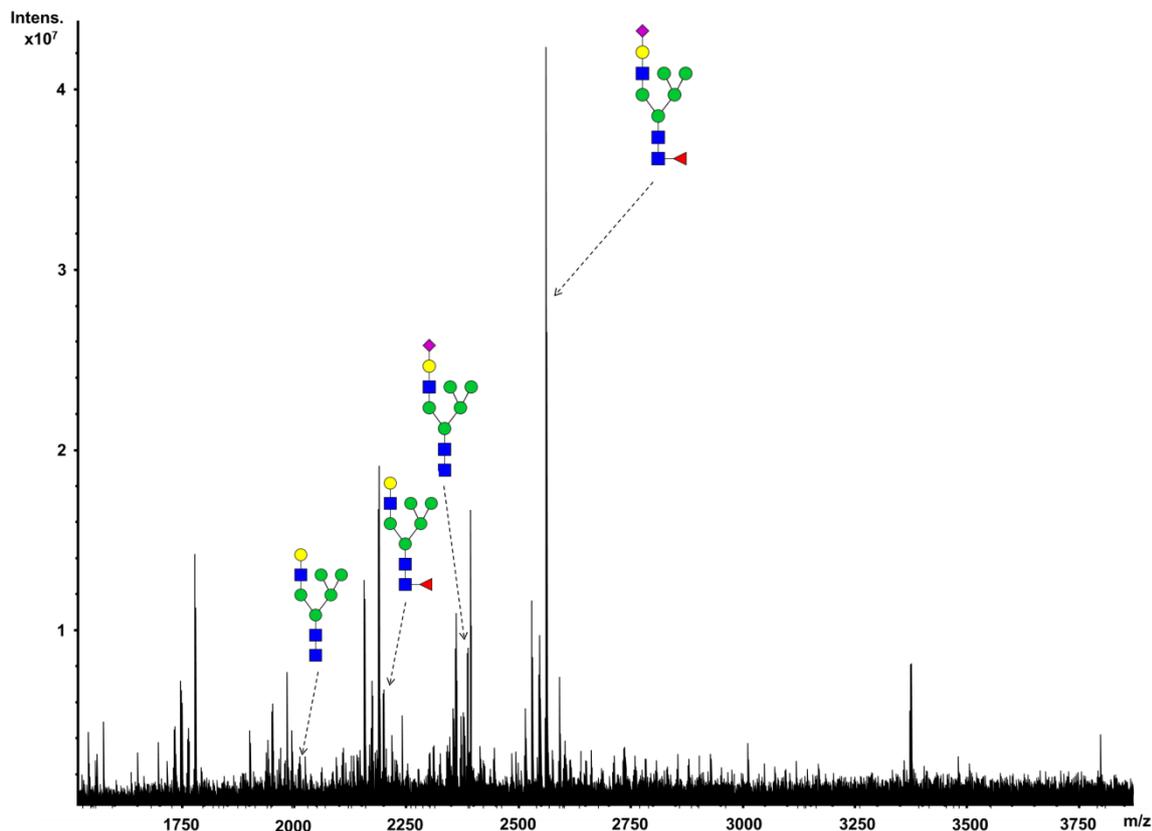
Figure 3.11 | Representative *N*-glycan mass spectrum of swainsonine-treated HEK293T cells

A representative mass spectrum of *N*-glycans released from swainsonine-treated HEK293T cells. *N*-glycans were released with PNGaseF and permethylated. Mass spectra were acquired using FT-ICR MS. Eight scans consisting of 500 laser shots were averaged to generate a mass spectrum. The signal to noise ratio had to be >3 for the ion to be included. Glycan assignments for a selection of the more abundant signals corresponding to hybrid structures are shown.

The glycan profiling data from swainsonine-treated cells were fitted in two ways. In the first instance, the fitted WT HeLa parameters, excluding MAN2 activity completely, were used to simulate a glycan profile without further fitting (figure 3.12). The removal of MAN2 activity accounts for the known effects of swainsonine, however the simulated glycan profile only partially matches that of the observed glycan profile (figure 3.12). The exclusion of MAN2 from the simulation did increase the relative abundance of hybrid type glycans, as expected, but could not reproduce the different types of hybrid glycans present in the observed glycan profile. For example, the simulated glycan profile produced higher levels of asialo-hybrid glycans ($Gal_1Man_5GlcNAc_3$ and $Fuc_1Gal_1Man_5GlcNAc_3$) than in the observed data (figure

3.12). Concurrently, the high abundance of the predominant sialylated hybrid glycan $NeuAc_1Fuc_1Gal_1Man_5GlcNAc_3$ in the observed profile, was not achieved in the simulation. Furthermore, the relative abundance of $Fuc_1Man_5GlcNAc_2$ could not be reproduced in this simulation.
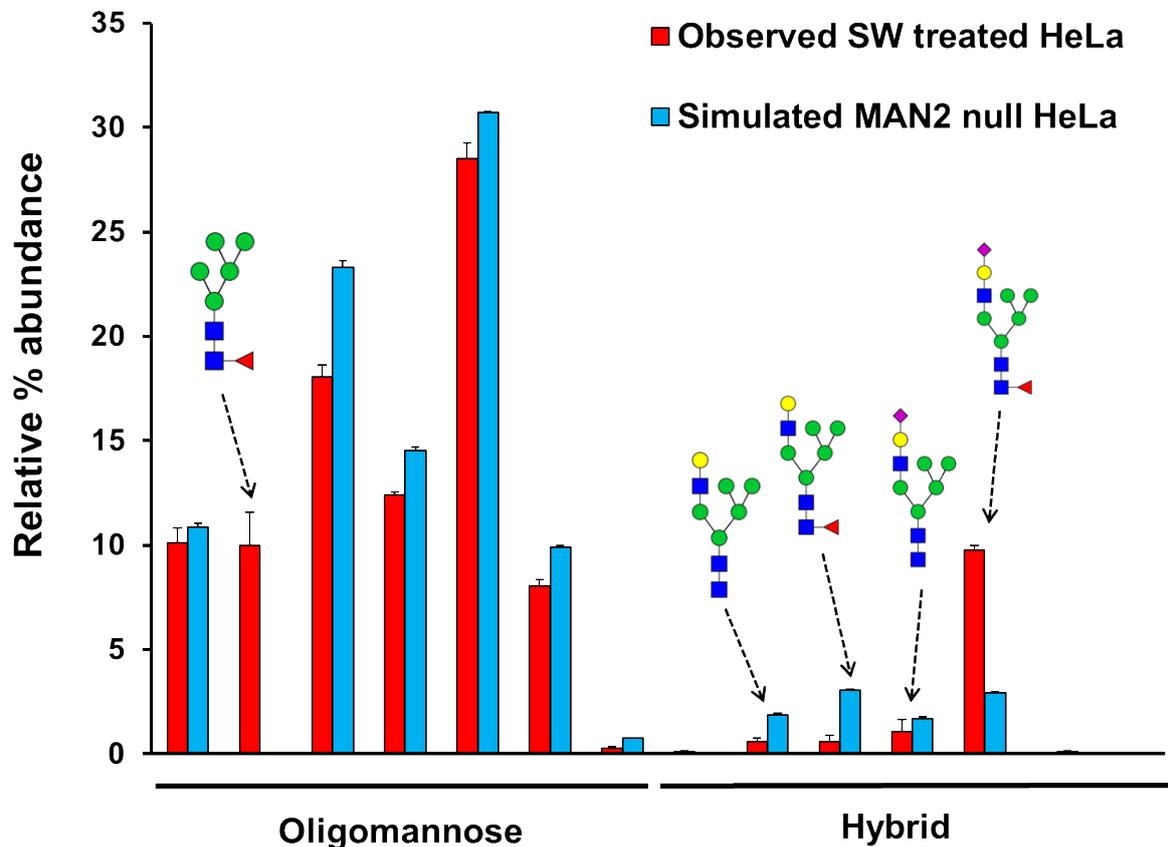
Figure 3.12 | HeLa MAN2 null observed and simulated glycan profiles

Observed (red) and simulated (blue) glycan profile of whole cell swainsonine-treated HeLa cells. The observed glycan profile was acquired by Chris Watson, a former member of the Ungar and Wood groups. The glycan profile was simulated three times using the SSA. The simulated glycan profile was generated using the fitted parameters obtained from the untreated HeLa glycan profile minus the activity of MAN2. Error bars are SEM, n = 3.

Similar to the swainsonine-treated HeLa cell work, the optimised WT HEK293T parameters with MAN2 activity removed completely, were used to simulate a glycan profile (figure 3.13). As with the HeLa cells, the resulting simulated glycan profile gave a qualitative match to that of the observed but did not correctly simulate the level of the $NeuAc_1Fuc_1Gal_1Man_5GlcNAc_3$ or $Fuc_1Man_5GlcNAc_2$. There are several

explanations for why the removal of MAN2 activity from the HeLa and HEK293T cell parameters did not result in a satisfactory fit between the simulated and observed data.
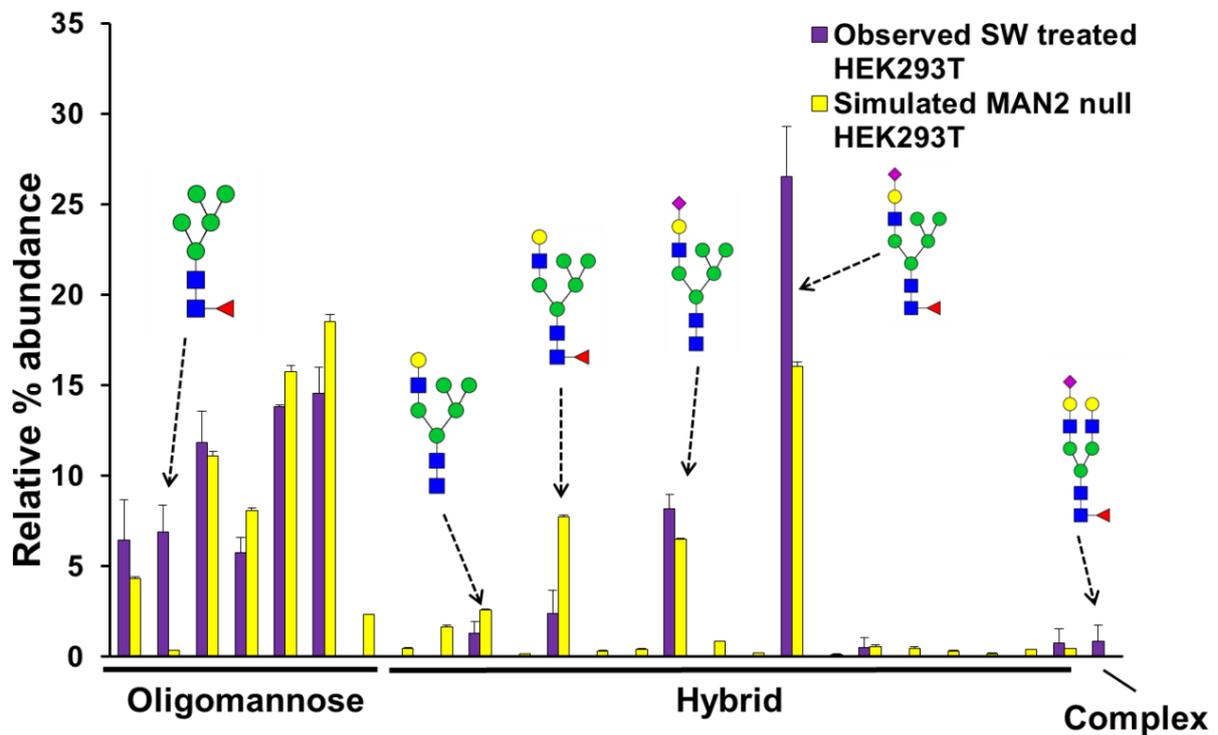
Figure 3.13 | HEK293T mannosidase II null observed and simulated glycan profile

Observed (purple) and simulated (yellow) glycan profile of whole cell swainsonine-treated HEK293T cells. Glycan peaks were identified from mass spectra. The total peak area intensities were used to quantify each identified glycan structure to generate an observed glycan profile. The glycan profile was simulated three times using the SSA and the mean of the three simulations is shown. The simulated glycan profile was generated using the fitted parameters obtained from the untreated HEK293T glycan profile minus the activity of MAN2. Error bars are SEM, n = 3.

First, it is possible that hybrid glycans behave differently as substrates for glycosylation enzymes to complex glycans, meaning enzymatic rates are different between the two glycan types (92, 93). As the abundance of complex glycans in the WT glycan profiles exceeds that of the hybrid type, the fitted enzymatic parameters for the WT profiles are strongly biased towards the enzymatic rates with complex glycans as substrates. Second, the addition of swainsonine potentially alters the glycosylation machinery, either by having effects on enzyme expression levels or by changing the physical properties of the glycosylation enzymes, many of which are themselves glycosylated (table 3.6). The presence of $Fuc_1Man_5GlcNAc_2$ necessitated FUT8 to act on

Man$_5$GlcNAc$_2$ (114, 115, 223, 224), albeit at a lower rate than that at which FUT8 modifies other substrates. *In vitro* studies had previously shown that FUT8 was incapable of acting on oligomannose glycans (225). A scale-factor for the fucosylation of Man$_5$GlcNAc$_2$ was therefore introduced (table 3.9), this was defined as the ratio of the enzymatic activity for FUT8 acting of any hybrid or complex-type glycan divided by the enzymatic activity of FUT8 acting on of Man$_5$GlcNAc$_2$.

Enzyme	Confirmed glycosylation site	Predicted glycosylation site	*N*- or O-glycosylation
MAN1	No	Yes	*N*-
MAN2	Yes (226)	Yes	*N*-
Endo-Man	No	No	N/A
MGAT1	No	No	N/A
MGAT2	No	Yes	*N*-
MGAT3	No	Yes	*N*-
MGAT4	No	Yes	*N*-
MGAT5	No	Yes	*N*-
FUT8	No	No	N/A
GalT	No	Yes	*N*-
ST3Gal2	No	Yes	*N*-
ST6Gal2	Yes (112)	Yes	*N*- and O-
GlcNAc-PhosphoT	Yes (227, 228)	Yes	*N*-

Table 3.6 | Glycosylation of glycosylation enzymes

Table of *N*-glycosylation enzymes used in the modelling and their glycosylation status. The majority of enzymes are predicted to contain a glycosylation site by the presence of the consensus sequence. Three enzymes have been identified as glycoproteins experimentally: MAN2, ST6Gal2 and GlcNAc-PhosphoT. N/A denotes not applicable.
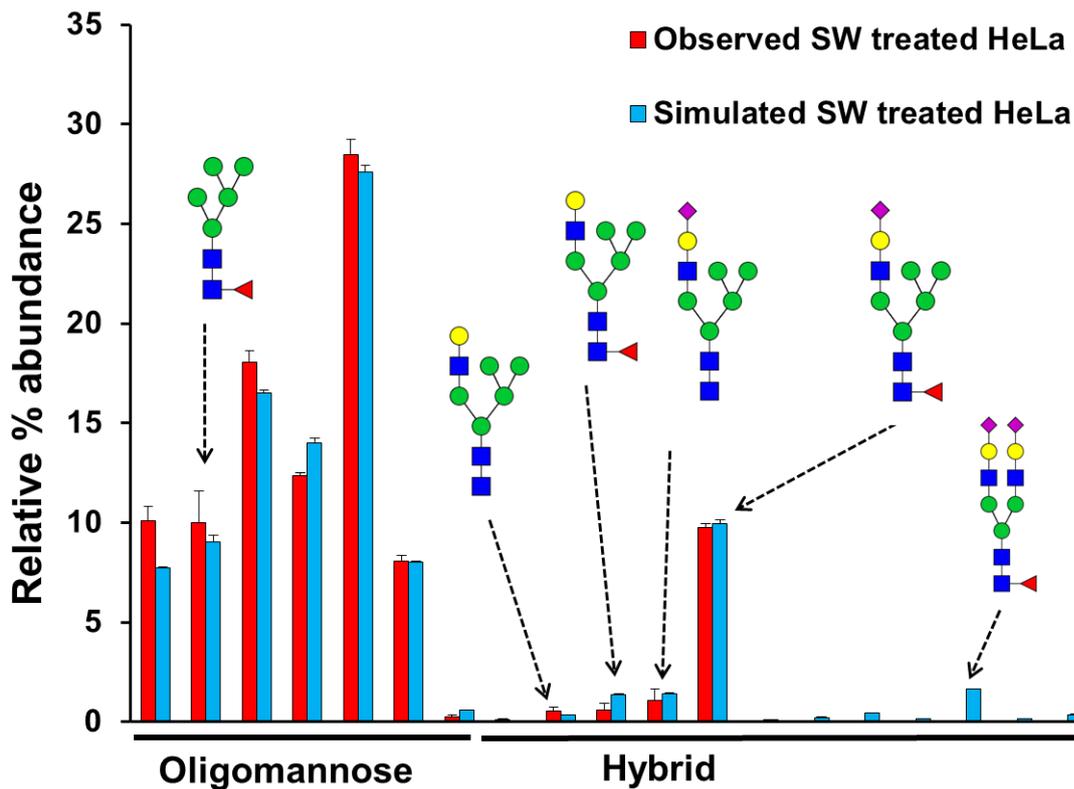
Figure 3.14 | Observed and simulated glycan profiles of swainsonine-treated HeLa cells after fitting

Observed (red) and simulated (blue) glycan profile of whole cell swainsonine-treated HeLa cells. The observed glycan profile was acquired by Chris Watson, a former member of the Ungar and Wood groups. The glycan profile was simulated three times using the SSA with the mean parameter values from n = 13 individual fitting runs. Error bars are SEM, n = 3.

Beginning with the fitted HeLa cell parameters that simulate the glycan profile of the WT cells accurately, the Bayesian fitting methodology was used to optimise the parameters for the swainsonine-treated HeLa glycan profile (figure 3.14). In contrast to the simple removal of MAN2, the fitting procedure was able to optimise the parameters, and to simulate a glycan profile that accurately matches the observed profile. Likewise, the optimised WT HEK293T parameters were used as the prior distributions for fitting the swainsonine treated HEK293T glycan profile (figure 3.15). The quality of the simulated glycan profiles produced by the fitted parameters suggests that the fitting procedure is absolutely necessary and demonstrates the power of the Bayesian methodology when compared to reasoning based on elimination of specific enzymes only.
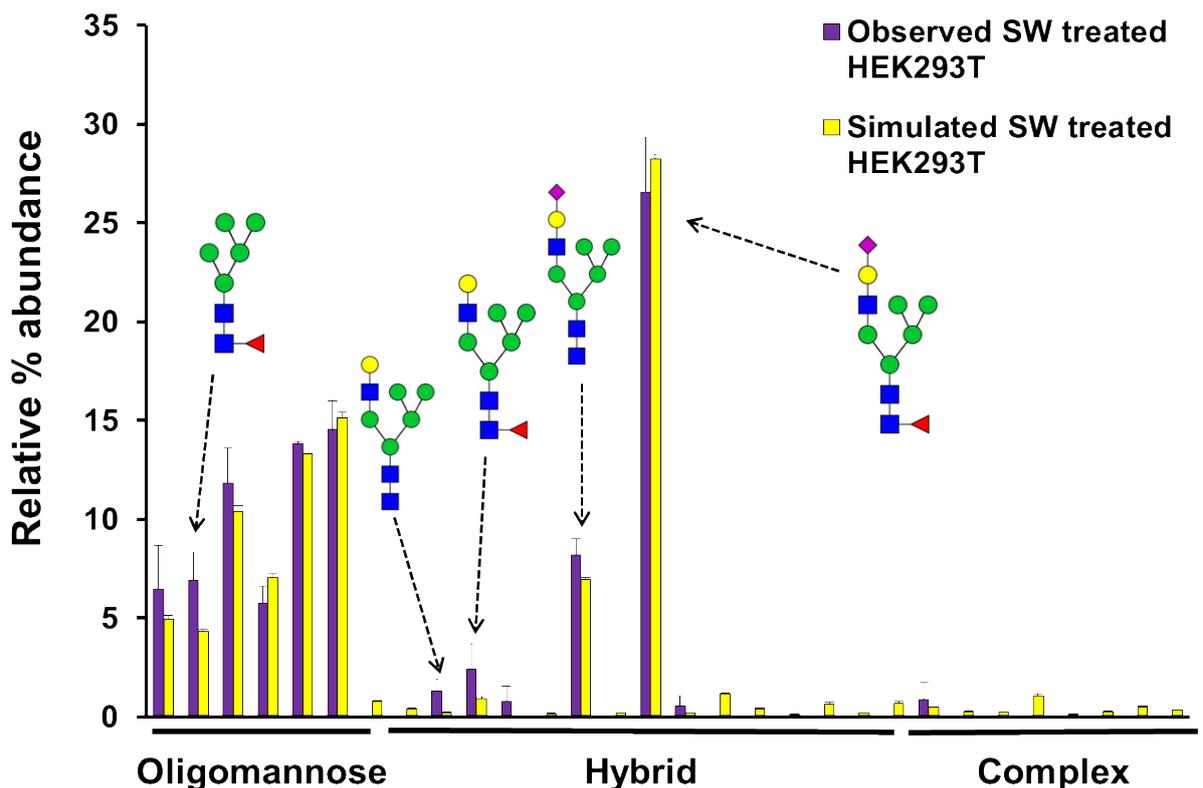
Figure 3.15 | Simulated glycan profile of swainsonine treated HEK293T cells after fitting

Observed (purple) and simulated (yellow) glycan profile of swainsonine-treated HEK293T cells. Glycan peaks were identified from mass spectra. The total peak area intensities were used to quantify each identified glycan structure and to generate an observed glycan profile. The glycan profile was simulated three times using the SSA with the mean parameter values from n = 16 individual fitting runs. Error bars are SEM, n = 3.

As expected, the fitting of both swainsonine-treated cell lines showed large decreases in the total activity of MAN2 when compared with the untreated activity (figure 3.16). The predicted decrease in MAN2 activity is greater for the HEK293T cell line than the HeLa cells. This is likely due to the higher abundance in the untreated HEK293T cells of complex glycans, the abundance of which needs to be reduced in the swainsonine-treated condition. There were, however, no changes in the relative distribution of each enzyme in the Golgi stack for the swainsonine-treated HeLa (table 3.7) and HEK293T (table 3.8) cells. Interestingly, the model also predicts an unexpected increase in sialylation and galactosylation activity (figure 3.16). The increases in sialylation and galactosylation activity are the reasons why the removal of MAN2 activity was not adequate for simulating the swainsonine-treated glycan profiles, thus, underlining the requirement for a robust fitting procedure.
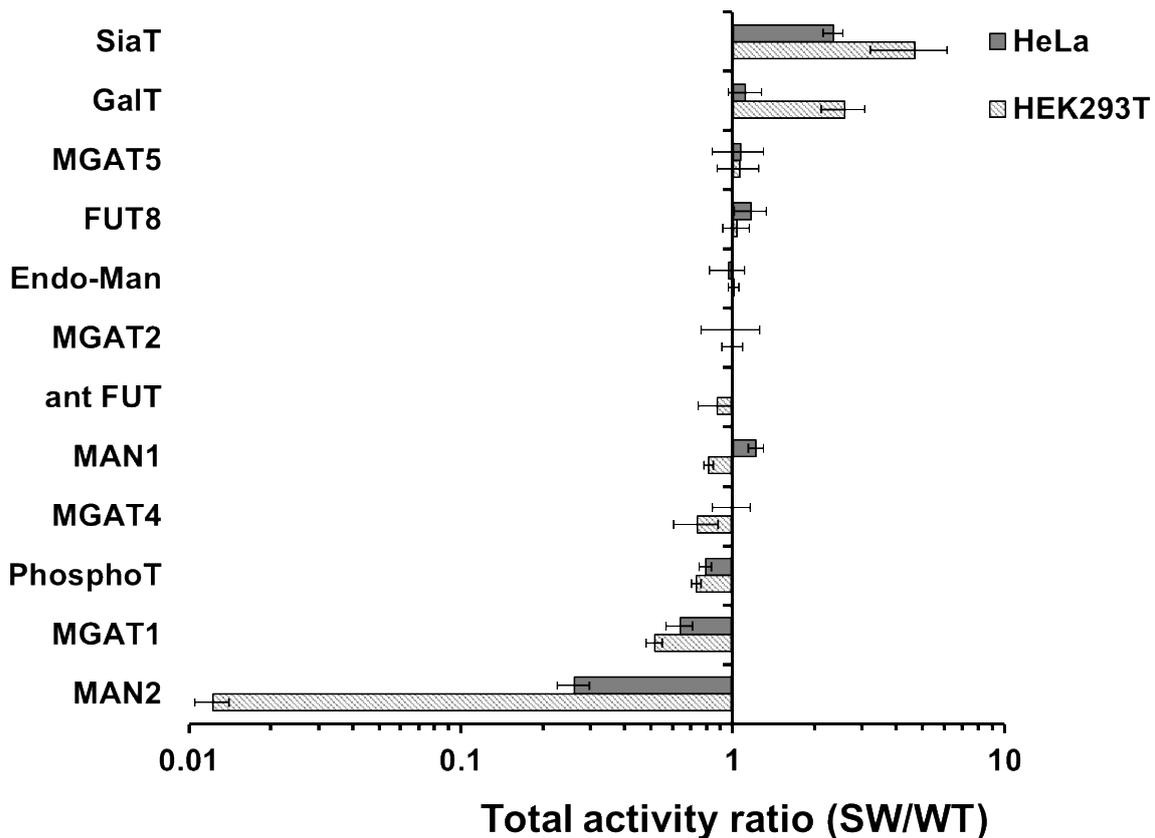
Figure 3.16 | Comparison of swainsonine-treated HeLa and HEK293T cell optimised parameters

The predicted ratio of total activity levels of the glycosylation enzymes in HeLa and HEK293T cells when treated with swainsonine. The predicted total activity for swainsonine-treated HeLa (n = 13) and HEK293T (n = 16) is divided by the predicted total activity of untreated HeLa (n = 20) and HEK293T (n = 15), respectively. Error bars are the propagated standard deviation. SW denotes swainsonine-treated.

There are several potential causes for the increase in sialyation activity. Firstly, the low relative abundance of hybrid glycans in the WT cell lines means the parameters pertaining to hybrid glycan synthesis are poorly defined and could well be different from parameters for complex type glycans. In addition, it may be linked to the different sialyation rates of different glycan antennae (111, 222). In that case, the swainsonine model would merely be representing the altered sialylation rates of different antennae. Secondly, sialyltransferases are themselves glycosylated and their enzymatic activity can be modulated by the glycans that reside on the transferase itself (112). Swainsonine treatment will almost certainly alter the glycosylation status of the sialyltransferases from complex to hybrid (table 3.6), which could lead to increased

119

activity. Thirdly, the protein levels of sialyltransferases are different between the two conditions. This suggestion is in good agreement with unpublished data showing that the protein levels of a galactosyltransferase and sialyltransferase are increased in MGAT1 KO cells (Vladimir Lupashin, personal communication).

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3
GlcNAc-PhosphoT	75±4%	13±2%	11±2%
EndoMAN	77±12%	12±2%	12±2%
MAN1	18±1%	66±4%	16±1%
MAN2	N/A	N/A	N/A
MGAT1	3±0.003%	93±3%	4±1%
MGAT2	12±3%	78±28%	11±3%
MGAT4	12±2%	77±12%	11±2%
MGAT5	6±1%	89±23%	6±1%
FUT8	9±2%	91±16%	0%
omFUT8	0%	100±13%	0%
GalT	11±2%	17±2%	73±13%
SiaT	2±0.002%	2±0.002%	96±10%

Table 3.7 | Optimised distribution of enzymes for fitted swainsonine-treated HeLa cell glycome

The predicted distribution of enzymes in swainsonine-treated HeLa Golgi apparatus following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 13 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
Endo-Man	75±3%	14±3%	10±1%	1±0.2%
PhosphoT	52±2%	21±2%	24±3%	3±0.6%
MAN1	8±0.4%	77±4%	14±2%	1±0.2%
MAN2	N/A	N/A	N/A	N/A
MGAT1	2±0.1%	73±7%	23±2%	2±0.4%
MGAT2	11±0.7%	41±3%	36±6%	12±3%
MGAT4	11±0.7%	40±6%	35±4%	13±2%
MGAT5	4±0.5%	37±13%	52±10%	7±2%
FUT8	14±1%	54±8%	19±2%	13±2%
ant FUT	4±0.4%	57±10%	35±5%	4±0.6%
GalT	1±0.2%	10±1%	31±5%	58±17%
SiaT	1±0.1%	1±0.3%	27±5%	70±8%

Table 3.8 | Optimised distribution of enzymes for fitted swainsonine-treated HEK293T cell glycome

The predicted distribution of enzymes in swainsonine-treated HEK293T Golgi apparatus following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 16 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

3.4 Summary of scale-factors

The scale-factors for: MAN1 trimming of $Man_6GlcNAc_2$, antenna sialylation, antenna number galactosylation and fucosylation of $Man_5GlcNAc_2$ were all parameters that were fitted. Table 3.9 shows the fitted scale-factors for HeLa and HEK293T cells. Previous models of glycosylation have included K_m values for different enzymatic substrates (192, 195) and *in vitro* studies on enzymatic specificity has demonstrated varying enzymatic preferences for substrates (93). The inclusion of K_m values is not valid for the approach used in this work as substrate competition is not modelled, however the variability in enzymatic specificity supports the approach taken in this work. The biological significance of differential rates of reactions for varying substrates is an added level of control. In this work the addition of scale-factors results in altered competition between enzymes for particular substrates therefore directing the *N*-glycosylation pathway down particular routes. The evolution of these enzymes to incorporate substrate specificities potential adds a layer of control to reduce heterogeneity and produce desired glycan structures.

In both HeLa and HEK293T cell lines MAN1 trimming of $Man_6GlcNAc_2$ was predicted to be slower than MAN1 acting on other oligomannose glycans. In HeLa cells the trimming of $Man_6GlcNAc_2$ was predicted to be 6.27 times slower compared with other oligomannose glycans (table 3.9). For HEK293T cells the scale-factor for MAN1 trimming was predicted to be 8.87 (table 3.9). The scale-factors for SiaT and GalT were only included in the fitting of HEK293T cells. In the case of GalT, the model predicted that the galactosylation rate of tri- and tetra-antennary glycans proceeds 2.81 times slower than the galactosylation of bi-antennary glycans (table 3.9). For sialylation, the activity of SiaT acting on the 6.1Man antenna was predicted to be 2.18 times slower compared with 3.1Man antenna (table 3.9). Finally, the fucosylation of $Man_5GlcNAc_2$ became a relevant factor in the swainsonine-treated conditions. Fucosylation of $Man_5GlcNAc_2$ occurred at only very low levels in the untreated conditions.

Internal enzymatic property	Prior distribution mean	Fitted HeLa	Fitted HEK293T
Scale-factor for $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$	10	6.27	8.87
Scale-factor for sialylation of separate branches	1	-	2.18
Scale-factor for galactosylation of bi- vs tri- and tetra-antennary glycans	1	-	2.81
Scale-factor for fucosylation of $Man_5GlcNAc_2$	20	1.24*	2.90*

Table 3.9 | Summary of fitted parameter values for additional internal enzymatic parameters in WT and swainsonine-treated cell lines

Mean predicted internal enzymatic parameter values for n = 20 (HeLa) and n = 15 (HEK293T) individual fitting runs. The fitted values represent scale-factors for the specified reaction. For example, MAN1 removes mannose from Man6GlcNAc2 6.27 times slower than removing mannose from any of the other oligomannose glycans. * fitted scale-factor is for swainsonine-treated condition n = 13 (HeLa) and n = 15 (HEK293T).

3.5 Conclusion

In order to make predictions regarding changes in the organization of the mammalian glycosylation machinery, a computational model incorporating a large suite of glycosylation enzymes was developed. The modelling framework developed consisted of two algorithms, an SSA, which simulated the glycosylation reactions, and a Bayesian fitting algorithm. The use of a probabilistic model to simulate the glycosylation reactions explicitly includes biological noise, an inherent property of reaction systems with smaller numbers of reactants. A second advantage of using an SSA is that additional factors that need parameterizing are subsumed into the probabilities, thus reducing the parameter matrix of an already complex system. The Bayesian fitting approach adopted in this work allows for the integration of experimental data and knowledge, which are characterised by high uncertainty.

To determine the ability of the modelling framework to fit mammalian glycan profiles, WT HeLa and HEK293T whole cell glycan profiles were determined using mass spectrometric methods and then fitted *in silico*. The modelling methodology developed here was capable of simulating the glycan profiles of the two cell lines accurately. The optimised enzyme distributions within the Golgi stack predicted by the model were consistent with information in the literature. The more elaborate glycan profile of HEK293T cells necessitated an additional model cisterna and several substrate-specific parameters. The glycan profiles of the two swainsonine-treated cell lines were also fitted in order to show that the developed model made rational and accurate predictions. The fitting of the glycan profiles of swainsonine-treated cells revealed a decrease in MAN2 total activity, as expected, and an increase in sialylation. The fitting of the swainsonine-treated glycan profile suggested that the modelling methodology was able to capture novel features of the glycan processing machinery.

In conclusion, a modelling framework was developed that could fit mammalian glycan profiles and make predictions regarding the activity of enzymes and their respective locations in a Golgi stack. The model was validated through the fitting of swainsonine-treated cell glycan profiles, which showed the expected decrease in MAN2 activity as well as an increase in sialylation. The following chapter describes use of the model to

investigate the effect of perturbing intra-Golgi trafficking on glycosylation and the organization of the glycosylation machinery.

Chapter 4: Predicting the results of Golgi trafficking defects in mammalian cell lines

4.1 Rationale for predicting alterations in the glycosylation machinery in trafficking-defective cells

A modelling framework has been developed in this work that is capable of making predictions regarding the organization of the *N*-glycosylation machinery. In the previous chapter, this modelling framework was applied to WT and swainsonine-treated cellular glycan profiles of both HeLa and HEK293T cells. The glycan structures that are produced by the cell are only partially the result of enzyme levels (118, 119). A complementary factor in regulating glycan structure is the organization of these enzymes within the Golgi apparatus. However, the organisation of glycosylation enzymes, including levels and localizations within the Golgi apparatus has only been accessible with advanced microscopy techniques and the information is limited to a small number of enzymes.

The COG complex has been implicated in ensuring the organizational fidelity of the glycosylation machinery (56, 62). Furthermore, mutations in the COG complex have been shown to result in CDGs. To investigate the role of the COG complex in regulating the glycosylation machinery of mammalian cells, the modelling framework has been applied to Cog4-deficient cells lines. Predicted relative changes in the enzyme activities and localizations can then be validated experimentally to provide confidence in the modelling methodology. The ability to model changes in glycosylation allows other conditions in which glycosylation is altered to be assessed; CDGs, cancers and cellular differentiation can all alter the levels and localizations of glycosylation enzymes

4.2 Aims

The aim of this chapter is to describe work demonstrating the ability of the modelling methodology to predict alterations in enzyme levels and localizations in a system

where Golgi trafficking has been disrupted. To achieve this, the glycan profile of HeLa cells in which Cog4 has been knocked down (KD), has been modelled. Furthermore, a Cog4 knock out (KO) HEK293T cell line (created by Jessica Blackburn-Bailey, University of Arkansas for Medical Sciences), has been glycan profiled by the author and then fitted using the modelling framework. A further aim of this work was to validate predicted changes in enzyme levels and localizations using experimental methods, such as western blotting and confocal microscopy.

4.3 Results

4.3.1 Modelling the Cog4KD HeLa cell glycan profile

Cog4KD HeLa cells have been previously glycan profiled (60). The glycan profile of Cog4KD HeLa cells is shown in figure 4.1 juxtaposed with the WT profile for convenience. Limited changes are observed between the glycan profile of the two cell lines (figure 4.1). The main changes that occur as a result of Cog4KD are in the oligomannose type glycans. For example, in the glycan profile of Cog4KD, the abundances of $Man_6GlcNAc_2$ and $Man_9GlcNAc_2$ are increased, while the level of $Man_8GlcNAc_2$ is decreased. Due to the low abundance of complex glycans detected in HeLa cells, no significant changes were observed in the hybrid and complex glycan structures. The limited changes in glycan profile that are observed are likely due to the nature of the knockdown, which resulted in only a partial depletion of Cog4.
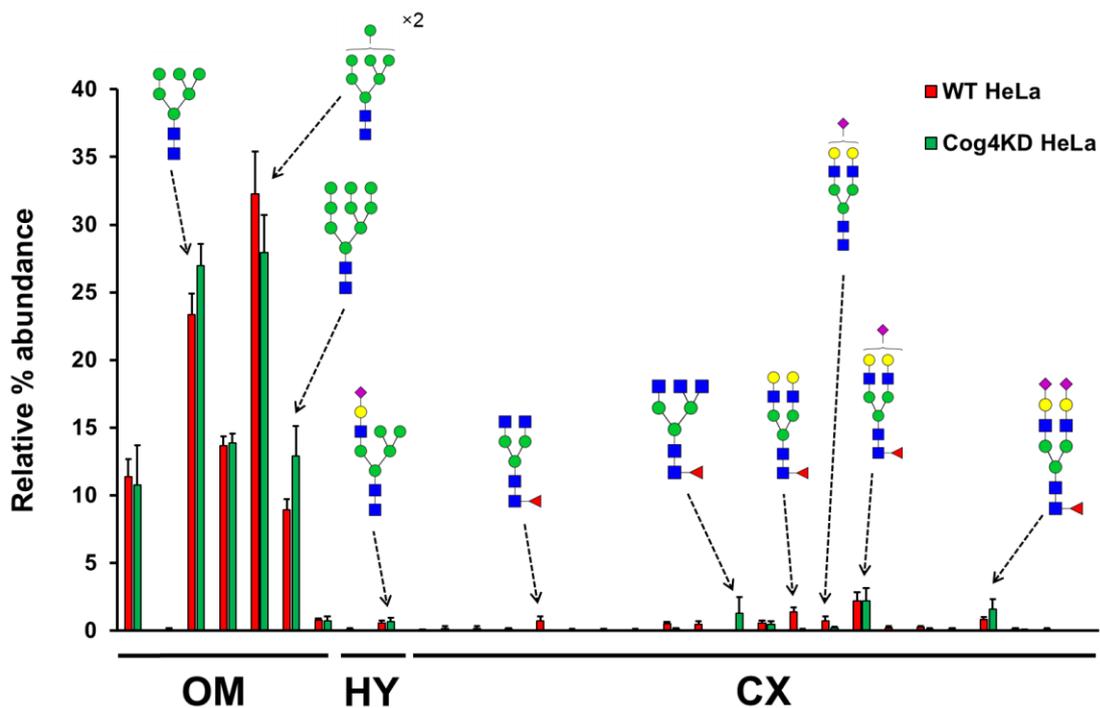
Figure 4.1 | Comparison of WT and Cog4KD HeLa cell glycan profiles

Observed glycan profiles of whole cell WT and Cog4KD HeLa cells. The observed glycan profiles are those published (Abdul Rahman et al. 2014). The quantification of the two profiles are shown together for the readers' convenience. OM = oligomannose, HY = hybrid, CX = complex.

In order to model the Cog4KD HeLa cell glycan profile, the prior distributions for the Bayesian fitting were set to the optimised WT HeLa cell parameters. Setting the prior distributions to the WT values allows a direct comparison between the glycan profiles of the two cell lines. Following fitting, the changes in the oligomannose glycans due to knocking down Cog4 could be reproduced *in silico* (figure 4.2). Overall, there were no changes to the localizations and total activities of the majority of enzymes (figure 4.3, table 4.1). Interestingly, the alteration to the oligomannose abundances in the glycan profile did lead the model to predict the MAN1 enzyme distribution to flatten out and shift to a more *trans*-Golgi localization (figure 4.4). Moreover, the total enzymatic activity for the MAN1 enzyme was predicted to be decreased in the Cog4KD HeLa cells compared with WT (figure 4.3).
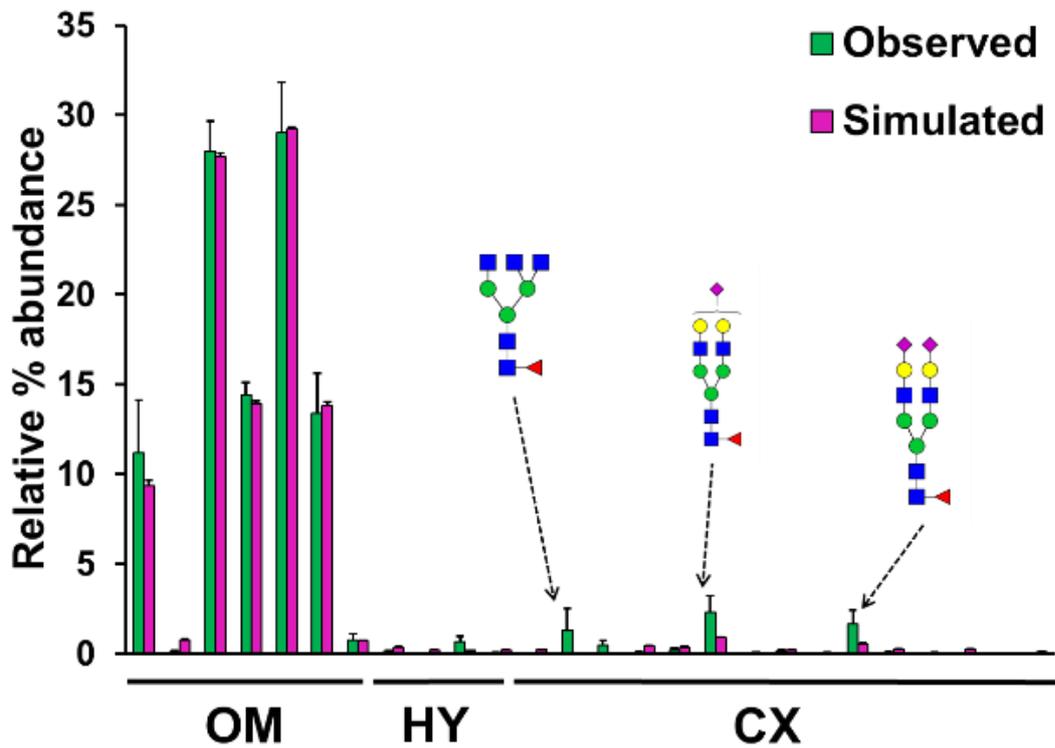
Figure 4.2 | Modelling of whole cell *N*-glycome from Cog4KD HeLa cells

Observed (green) and simulated (magenta) glycan profile of whole cell Cog4KD HeLa cells. The observed glycan profiles are those published (Abdul Rahman et al. 2014). To generate the simulated glycan profile the SSA was run three times using the mean parameter values from n = 20 individual fitting runs. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.
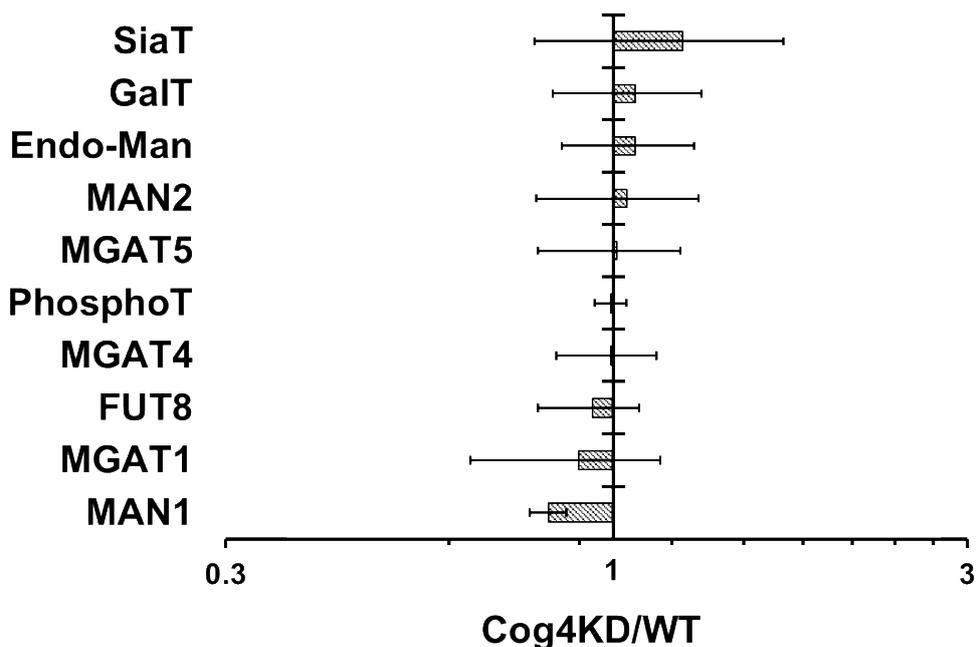
Figure 4.3 | Predicted total activity of MAN1 is reduced in Cog4KD HeLa cells

The predicted ratio of total activity levels of the glycosylation enzymes in Cog4KD HeLa cells in comparison with WT HeLa cells. The predicted total activity in Cog4KD HeLa cells (n = 20) is divided by the predicted total activity of WT HeLa cells (n = 20), following fitting. Error bars are the propagated standard deviation.

The predicted changes in MAN1, but not enzymes involved in complex glycan processing, is not surprising. Based on the dominance of oligomannose glycans in the HeLa cell glycan profiles, the strongest predictions can be made regarding the enzymes that process oligomannose glycans. The shift in the MAN1 distribution in the *trans* direction is consistent with the COG complexes' role in retrograde Golgi trafficking; the shift in enzyme distribution suggests that the efficiency of retrograde recycling of MAN1 has been reduced as a result of disruption to the COG complex. This prediction of MAN1 shifting in the *trans* direction has been made by the model despite not having the *a priori* information about the role of the COG complex. Having accomplished the aim of modelling the Cog4KD HeLa cell glycan profile, the more elaborate glycan profile of Cog4KO HEK293Ts was modelled next.

131
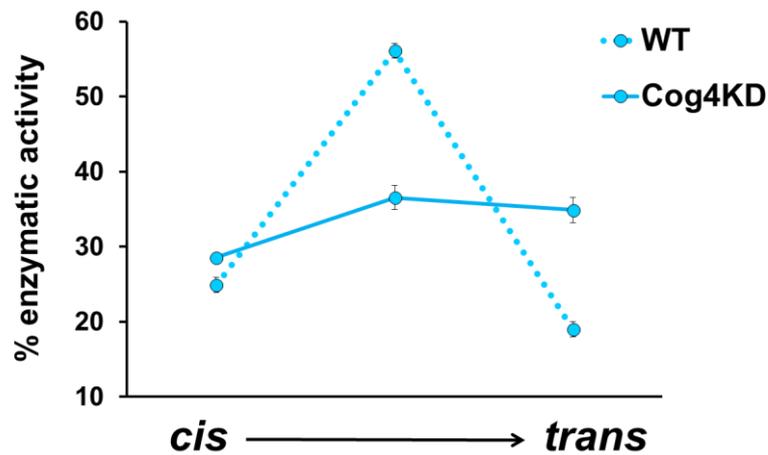
Figure 4.4 | Cog4KD results in a flattening of the MAN1 distribution in the Golgi stack

Predicted distribution of MAN1 in WT (dotted) and Cog4KD (solid) HeLa cells following fitting. Error bars are standard deviation for n = 20 (WT HeLa) and n = 20 (Cog4KD HeLa) individual fitting procedures.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3
GlcNAc-PhosphoT	80±3%	10±1%	10±2%
EndoMAN	79±21%	10±2%	11±2%
MAN1	29±1%	37±2%	35±2%
MAN2	10±3%	77±29%	13±4%
MGAT1	2±1%	95±36%	3±1%
MGAT2	11±2%	78±19%	11±2%
MGAT4	11±2%	78±12%	11±1%
MGAT5	6±1%	88±23%	6±1%
FUT8	12±3%	88±19%	0%
omFUT8	0%	100±10%	0%
GalT	10±2%	11±2%	79±26%
SiaT	4±1%	4±1%	93±50%

Table 4.1 | Predicted distribution of enzymes in the Golgi of Cog4KD HeLa cells

The predicted distribution of enzymes in the Golgi apparatus of Cog4KD HeLa cells following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 20 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

4.3.2 Modelling the Cog4KO HEK293T cell glycan profile

In contrast to the Cog4KD HeLa cells, in which the changes to oligomannose glycans predominate, the knockout of Cog4 in HEK293T cells produces marked changes in complex glycans (58). This is potentially due to the targeted CRISPR-Cas9-based gene editing used to knockout Cog4 in HEK293T cells , that results in a complete elimination of Cog4 protein levels (58). Guide RNA sequences for Cog4 were provided by Horizon Discovery and targeted CGAATCAAGGTC CGCCATCT within the Cog4 gene. For further details please see Blackburn-Bailey *et al.* (58). A representative mass spectrum of the *N*-glycans is shown in figure 4.5, indicating the dominant glycans detected. A comparison between the glycan profiles of WT HEK293T and Cog4KO HEK293T cells reveals striking alterations in the levels of a variety of glycans (figure 4.6). A total of 40 glycans were detected in the Cog4KO glycan profile compared with 58 in that of WT HEK293T cells. Highly processed glycans containing sialic acids, fucoses and multiple antennae are either absent or were detected at much lower levels in the Cog4KO HEK293T cells than the WT. The distribution of oligomannose glycans is also altered as a result of Cog4KO (figure 4.6). Cog4KO resulted in an increase in the levels of $Man_5GlcNAc_2$, $Man_7GlcNAc_2$ and $Man_8GlcNAc_2$ and a decrease in $Man_6GlcNAc_2$.
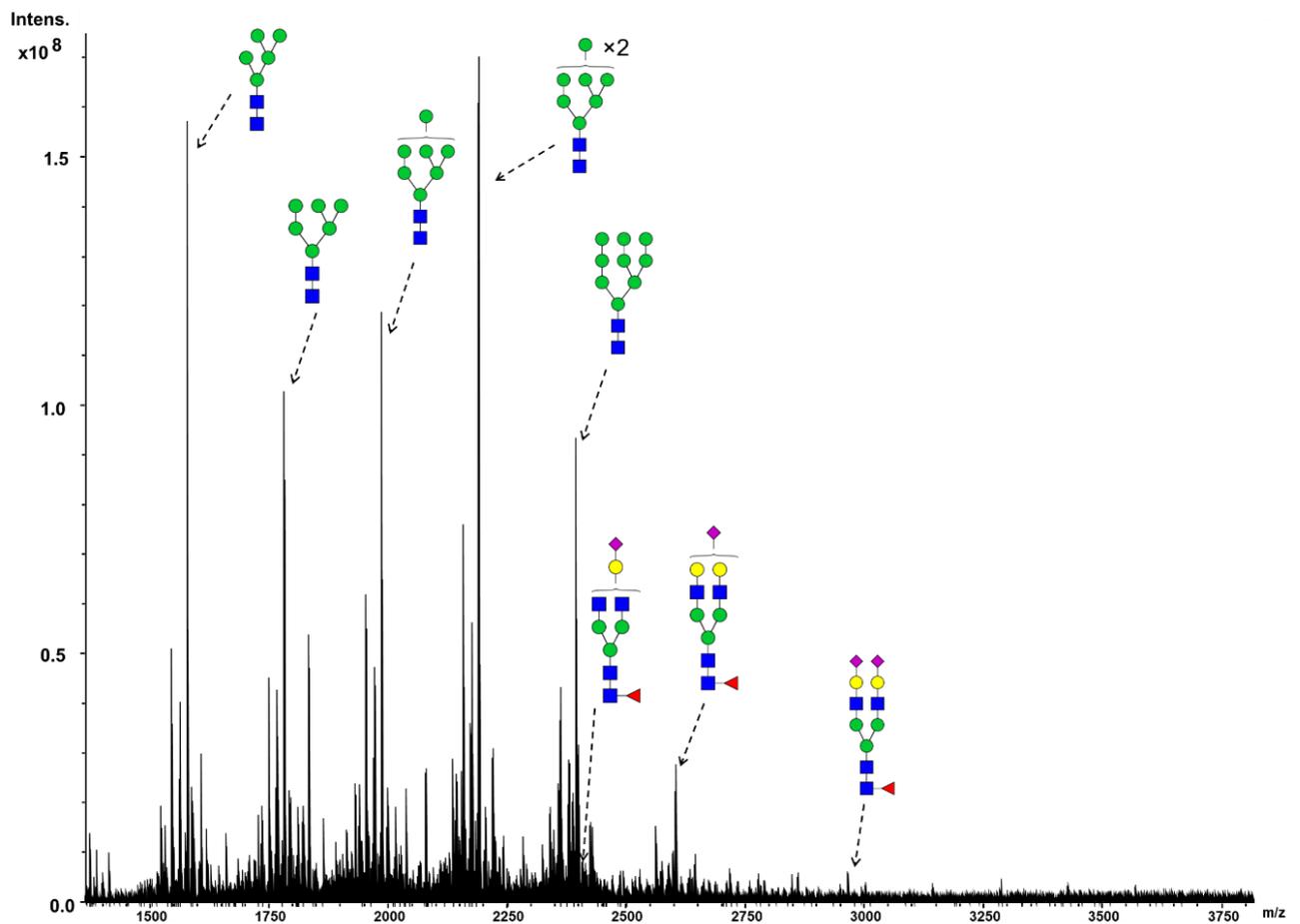
Figure 4.5 | Representative mass spectrum of Cog4KO HEK293T cell glycans

A representative mass spectrum of *N*-glycans released from Cog4KO HEK293T cells. *N*-glycans were released with PNGaseF and permethylated. Mass spectra were acquired using FT-ICR MS. Eight scans consisting of 500 laser shots were averaged to generate a mass spectrum. The signal to noise ratio had to be >3 for the ion to be included. Glycan structural proposals for a selection of the more abundant signals are shown.
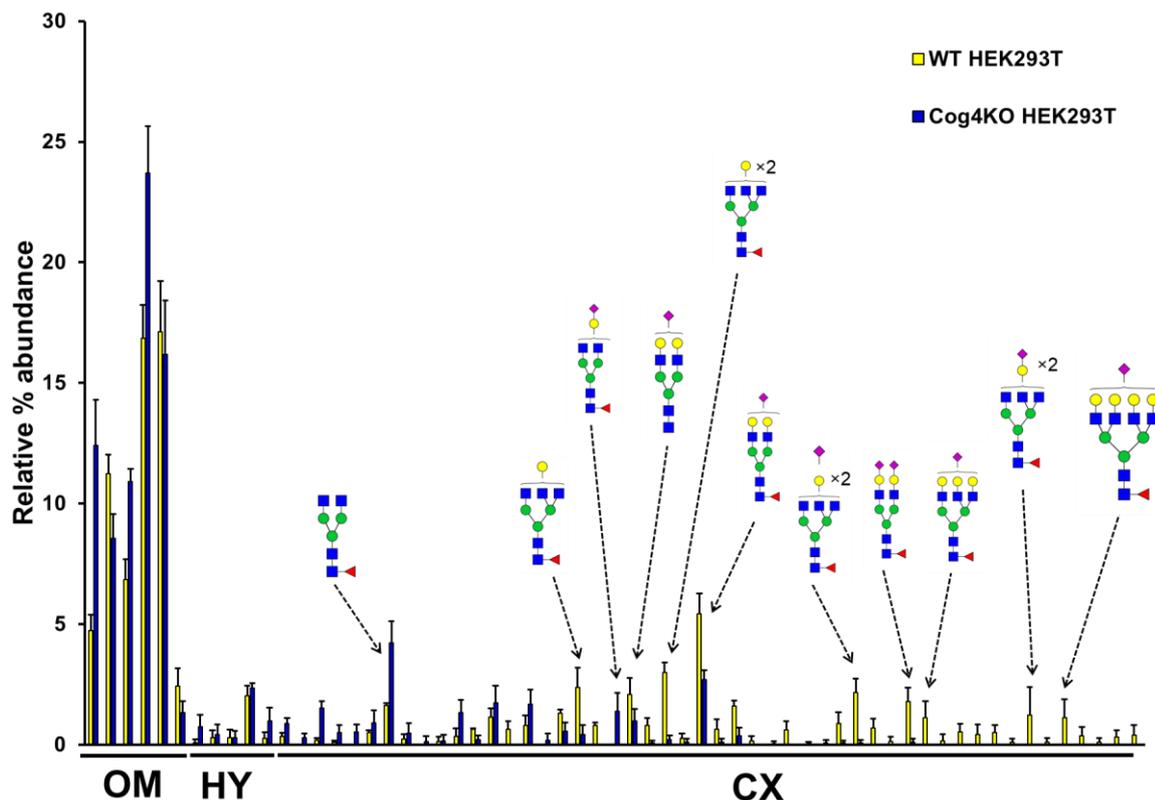
Figure 4.6 | Comparison of WT and Cog4KO HEK293T glycan profiles

Glycan profiles of WT (yellow) and Cog4KO (purple) HEK293T cells. Glycans were identified from mass spectral signals. The total peak area intensities were used to quantify each identified glycan structure to generate an observed glycan profile, normalised to the total intensity of signals present in each cell line. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.

In addition to the changes in the distribution of oligomannose glycans, Cog4KO in HEK293T cells resulted in an increase in the overall levels of oligomannose glycans and a decrease in total complex glycans (figure 4.7A). The total levels of fucosylation and sialylation were also decreased (figure 4.7B, C). Interestingly, the levels of bi-antennary glycans did not change between the two cell lines, suggesting that MGAT2 remained functional in the Cog4KO cell line, or the loss of enzymes downstream resulted in reduced competition, allowing MGAT2 to act. However, the levels of tri- and tetra-antennary glycans were reduced in the Cog4KO HEK293T cell line (figure 4.7D).
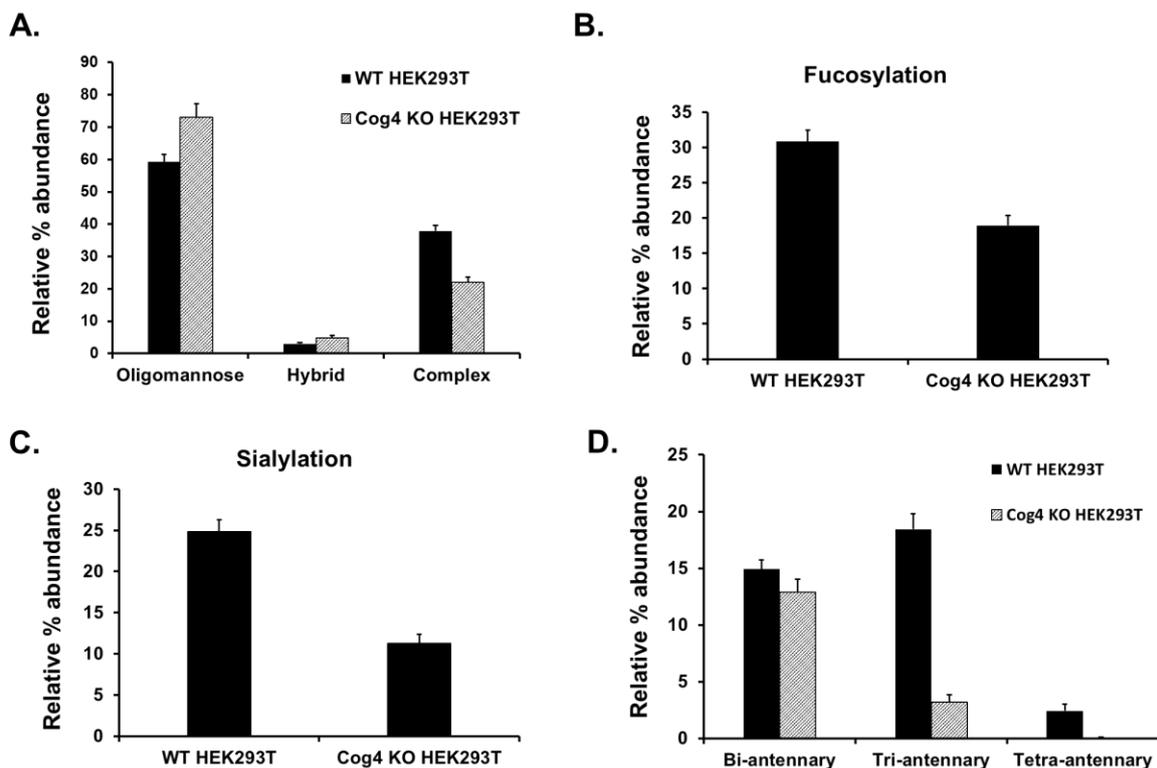
Figure 4.7 | Summary comparison of glycans produced by WT and Cog4KO HEK293T cells

A) Quantification of *N*-glycans grouped into oligomannose, hybrid and complex from WT and Cog4KO HEK293T cells. Quantification of *N*-glycans identified from mass spectra that are fucosylated B) or sialylated C). Quantification of bi, tri and tetra-antennary glycans identified by mass spectrometry D). Error bars are SEM, n = 3.

Changes in the proteome can also result in changes in the glycan profile. Disruption to the COG complex has previously been shown to reduce the levels of some cell surface proteins, for example the low density lipoprotein receptor (63). To eliminate the possibility that changes in the whole cell glycan profile of Cog4KO HEK293T cells were the result of a reduction in a selection of cell surface glycoproteins and/or changes in site-specific glycosylation (229), a crude total membrane proteomic experiment was conducted. Cells were permeabilized with digitonin and then centrifuged in order to remove cytosolic proteins and enrich membrane proteins. The membrane protein fraction was then digested using a combination of proteases before the resulting peptides were identified using liquid chromatography (LC)-MS. The chromatorgaphic peak areas for each pepetide were normalised to a subset of ER

proteins (table 2.1) allowing relative quantification between the two cell lines. A threshold of two unique peptides was required for identification.

In order to identify transmembrane proteins the unique protein accession numbers for the identified proteins were compared with a list of transmembrane proteins from www.geneontology.org.(figure 4.8). In total, 2488 proteins were observed of which 220 were identified as transmembrane proteins. The levels of the transmembrane proteins: Complement 3 (C3), Transmembrane 9 superfamily member 2 (TM9SF2), Transmembrane 9 superfamily member 3 (TM9SF3) and Cleft lip and palate transmembrane protein 1-like protein (CLPTM1L) were found to be significantly decreased. Golgi proteins were excluded from the list of transmembrane proteins because the permeabilization of the cells with digitonin and subsequent centrifguation may have resulted in a loss of vesicles, and therefore vesicular cargo, from the sample prior to tryptic digestion. A previous study of a Cog4 KD HeLa cell line has demonstrated that the overall level of exogenously expressed MGAT1 is unaltered when compared with WT HeLa cells but rather, MGAT1 accumulated in Golgi vesicles in the Cog4KD HeLa cells while in WT cells it remained in the Golgi apparatus (57). Overall the crude transmembrane proteomics results suggest that alterations in the proteome between WT and Cog4KO HEK293T cells are unlikely to account for the large changes between the glycan profiles.
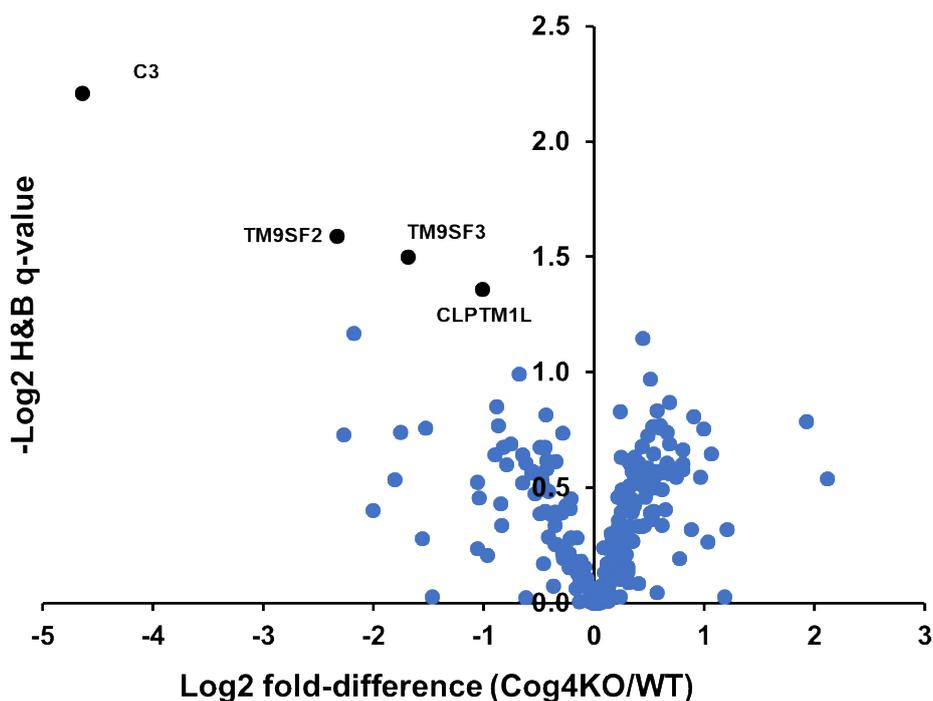
Figure 4.8 | Alterations in glycosylation are not the result of changes in the transmembrane proteome

Volcano plot of transmembrane proteins identified using GeneOntology.org. Normalisation was performed between acquisitions against total precursor intensities for a subset of known ER proteins (see material and methods). ANOVA-derived p-values were converted to multiple test-corrected q-values using the Hochberg and Benjamini (H&B q-values) approach. Excluding Golgi apparatus-associated proteins, Complement 3 (C3), Transmembrane 9 superfamily member 2 (TM9SF2), Transmembrane 9 superfamily member 3 (TM9SF3) and Cleft lip and palate transmembrane protein 1-like protein (CLPTM1L) were found to be significantly decreased in Cog4KO HEK293T cells compared with WT.

After excluding the possibility that alterations in the glycan profile of Cog4KO HEK293T cells were due to changes in protein expression, it is likely that changes in the glycan profile the result of changes to the glycosylation machinery. The modelling methodology was then used to fit the glycan profile of Cog4KO HEK293T cells (figure 4.9). Upon fitting of the Cog4KO HEK293T profile, three enzymes exhibited localization changes when compared with WT (figure 4.10, table 4.2). As in the HeLa cell example, the distribution of MAN1 was shifted in the *trans* direction, although not to the same degree as in the Cog4KD HeLa cells. The proportion of the MGAT5 activity falling in the third cisterna, and the GalT activity in the fourth cisterna were both reduced upon Cog4KO, indicating a shift of these enzymes towards the *cis*-side of the Golgi.
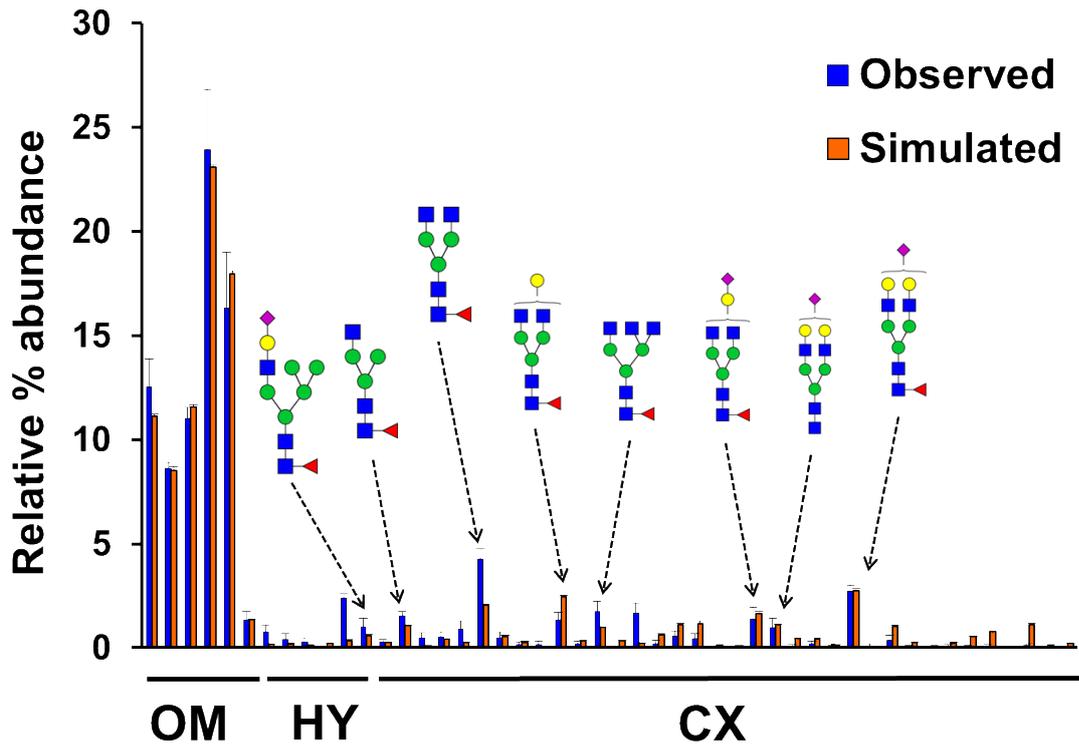
Figure 4.9 | Modelling of whole cell *N*-glycome from Cog4KO HEK293T cells

Observed (blue) and simulated (orange) *N*-glycan profiles of Cog4KO HEK293T cells. Glycans were identified from mass spectral signals. The total peak area intensities were used to quantify each identified glycan structure to generate an observed glycan profile. The glycan profile was simulated three times using the SSA with the mean parameter values from n = 21 individual fitting runs. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.
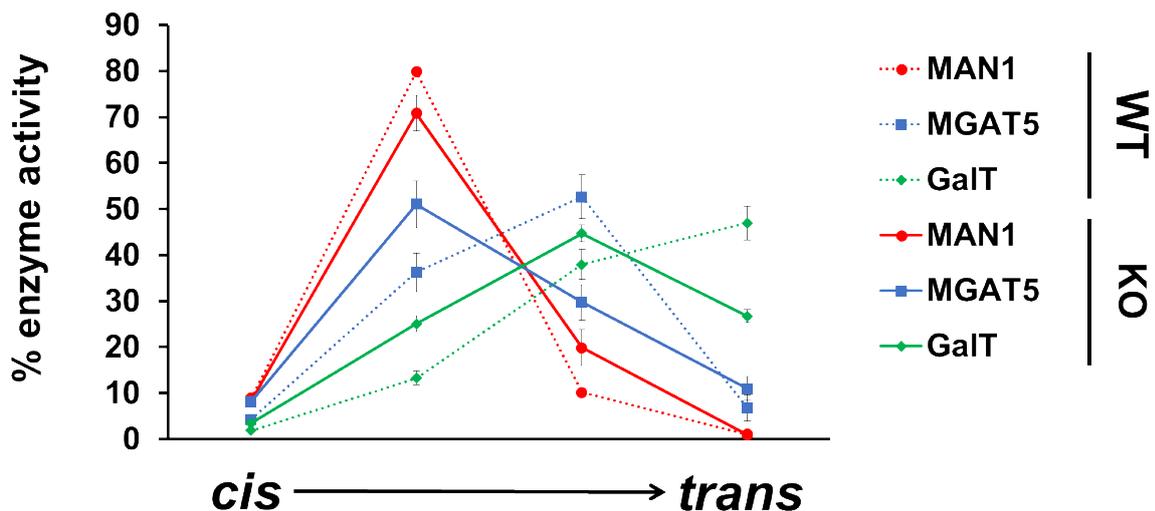
Figure 4.10 | Alterations in a subset of glycosylation enzymes as a result of Cog4KO

Predicted enzymatic distributions of enzymes in WT (dotted) and Cog4KO (solid) HEK293T cells. The distributions of MAN1 (red), MGAT5 (blue) and GalT (green) in both cell lines following the fitting procedure are shown. Error bars are standard deviation for n = 15 (WT HEK293T) and n = 21 (Cog4KO HEK293T) individual fitting procedures.

In order to validate the predicted shift in GalT localization, GalT-YFP was exogenously expressed in WT and Cog4KO HEK293T cells and its localization with respect to the *cis*-Golgi marker GM130, was compared in the two cell lines. Cells were treated with nocodazole to disperse the Golgi ribbon, producing Golgi mini-stacks. In WT HEK293T cells, GalT-YFP and GM130 showed minimal co-localization (figure 4.11A, B). In contrast, GalT-YFP and GM130 showed partial co-localization in Cog4KO HEK293T cells (figure 4.11A, B). The co-localization in Cog4KO HEK293T cells is not complete, indicating that GalT is not located in the *cis*-Golgi, which is in good agreement with the model prediction (figure 4.10). This shift in the *cis* direction of GalT could be explained in a number of different ways. Firstly, Cog4KO could have resulted in the delivery of COPI vesicles containing GalT to the wrong cisterna; however, this explanation would mean that disrupting the COG complex increases the rate of retrograde trafficking. This is not consistent with known role of the COG complex in retrograde trafficking. Secondly, GalT was specifically depleted from *trans* cisternae, or thirdly, the Golgi architecture was altered as a result of Cog4KO.

A.

B.
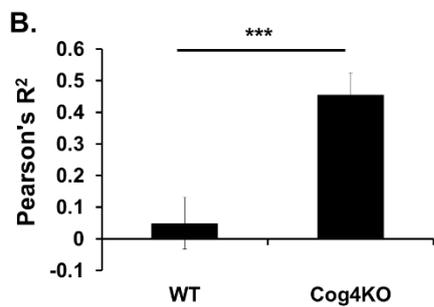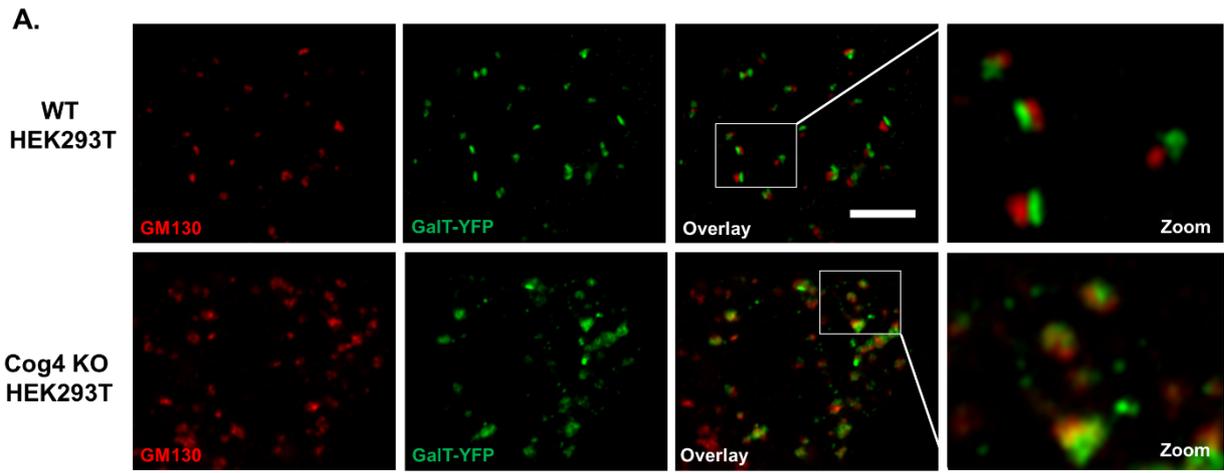
Figure 4.11 | GalT-YFP localizes with the cis marker, GM130, in Cog4KO HEK293T cells

A) Representative images acquired using Airyscan confocal microscopy of WT and Cog4KO HEK293T cells treated with nocodazole. Cells were transiently transfected with GalT-YFP (green) and Golgi mini-stacks were stained with anti-GM130-Alexa647 (red). B) Pearson's colocalization coefficients for GM130 and GalT-YFP in WT and Cog4KO HEK293T cells. Pearson's colocalization coefficients were calculated for each individual Golgi stack and error bars are standard deviation for n = 4 cells with 151 (WT) and 131 (Cog4KO) Golgi stacks. Scale bar is 5 μm. *** denotes P ≤ 0.001 for a student's t-test.

141

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
GlcNAc-phosphoT	76±2%	10±1%	12±1%	1±0.2%
Endo-Man	75±3%	14±2%	11±2%	1±0.2%
MAN1	8±0.2%	71±2%	20±1%	1±0.4%
MAN2	10±3%	51±22%	28±7%	11±3%
MGAT1	2±0.2%	81±15%	14±2%	3±0.4%
MGAT2	10±2%	43±15%	35±20%	12±3%
MGAT4	9±1%	40±8%	39±11%	12±3%
MGAT5	8±1%	52±13%	29±4%	11±2%
FUT8	12±2%	49±13%	25±5%	13±3%
ant FUT	3±1%	49±12%	44±21%	4±1%
GalT	3±0.2%	25±3%	45±5%	27±4%
SiaT	2±0.3%	3±1%	46±12%	49±7%

Table 4.2 | Predicted distribution of enzymes in the Golgi of Cog4KO HEK293T cells

The predicted distribution of enzymes in the Golgi apparatus of Cog4KO HEK293T cells following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 21 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna, red denotes lower activity, green denotes higher activity.

In addition to the changes in enzyme localization, in order to simulate the glycan profile of these Cog4KO HEK293T cells, the model predicted a relative decrease in the total activity of MGAT1, MGAT5, MAN2 and GalT (figure 4.12). It was assumed that the enzymatic rate constants do not change between the cell lines compared in figure 4.12, instead the Cog4KO alters enzyme sorting, not the physical state of the enzyme itself. Therefore, predictions of the changes in enzyme activity parameters represent changes in the amount of enzyme available within specific cisternae or the cell as a whole. Indeed, predicted decreases in GalT and MGAT1 levels were validated through western blotting (figure 4.13A, B, D). Although it cannot be ruled out that changes to other enzymatic attributes have also been altered, the experimental validation makes the assumption, that the physical state of the enzyme is unaltered, credible. Perhaps unexpectedly, the activity of FUT8 was predicted not to be decrease in the Cog4KO cell line, despite the total levels of glycan fucosylation being reduced in the observed Cog4KO profile compared with WT HEK293T cells (figure 4.7B, 118). This prediction

of the model could also be validated through western blot showing no significant change in FUT8 levels (figure 4.13C, D).
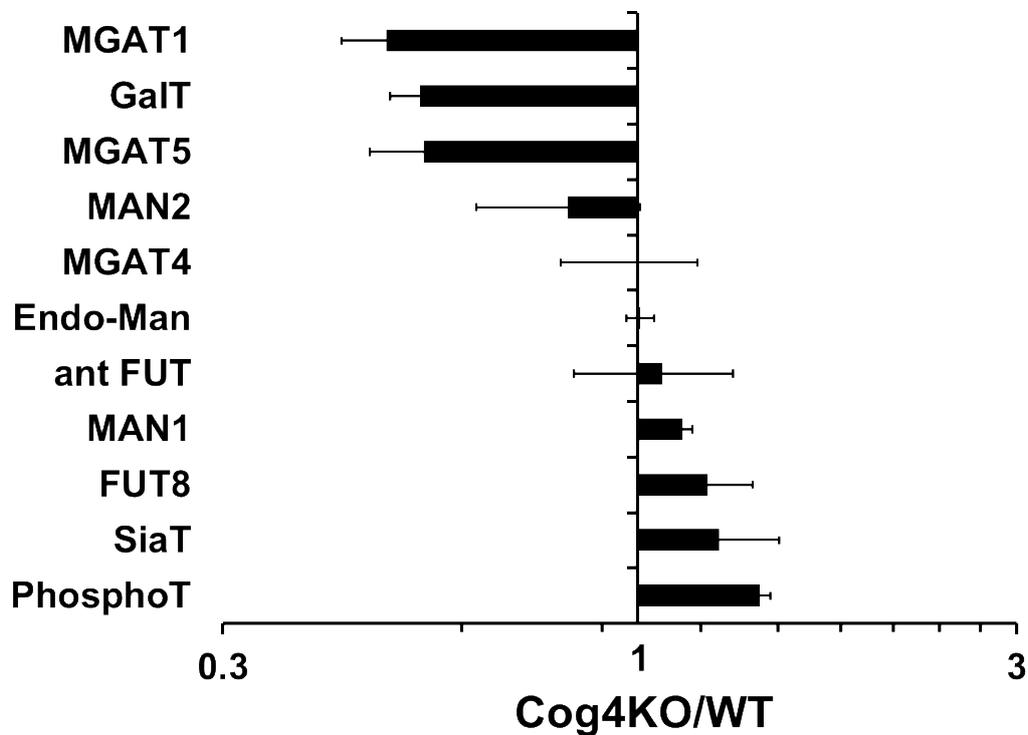
Figure 4.12 | Predicted total activity changes in Cog4KO HEK293T cells

Predicted total enzymatic activity changes upon Cog4KO in HEK293T cells following fitting. The comparison is between WT HEK293T cells and Cog4KO HEK293T cells. The predicted total activity for each enzyme in Cog4KO HEK293T (n = 21) is divided by the predicted total activity of each enzyme in WT HEK293T cells (n = 15). Error bars are the propagated standard deviation.
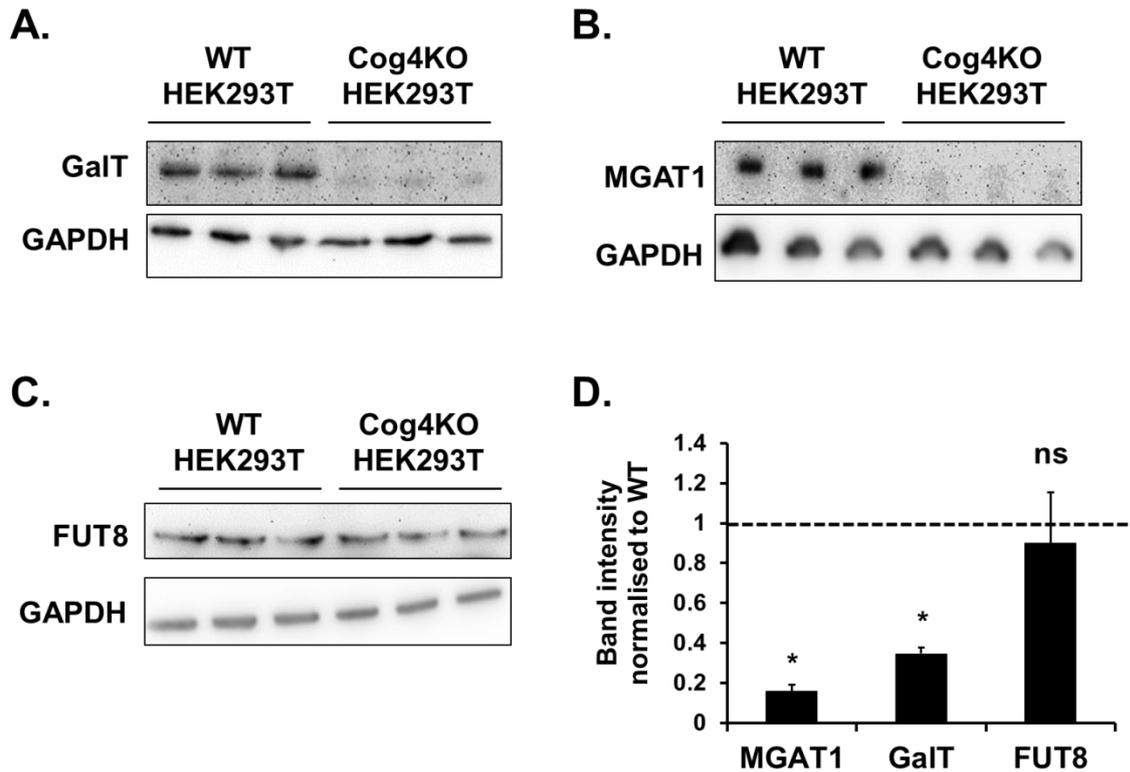
Figure 4.13 | Predicted relative changes in enzymes levels could be validated with western blot analysis

Western blot analysis of endogenous GalT (A), MGAT1 (B) and FUT8 (C) from WT and Cog4KO HEK293T cells. D) Quantification of enzyme band intensities from Cog4KO HEK293T cells normalised to WT HEK293T levels. Error bars are standard deviation, n = 3. * denotes $P \leq 0.05$, ns denotes not significant for a student's t-test.

The fitting of the glycan profile of WT HEK293T cells was not sensitive to changes in MGAT2 activity due to the high relative activity used in the initial parameterization, which was based on literature rates. However, the fitting of the Cog4KO HEK293T glycan profile necessitated a reduction in the MGAT2 activity parameters. The fitting of the Cog4KO HEK293T glycan profile was found to be more sensitive to the MGAT2 activities. MGAT2 activities were therefore manually tuned in all cisternae prior to fitting of the Cog4KO profile. It should be noted that the low MGAT2 activities resulting from this Cog4KO fitting did not change the simulated glycan profile of WT cells, suggesting that the levels of MGAT2 are not altered as a result of Cog4KO. This could be validated experimentally (figure 4.14). The result suggests that the total protein levels of MGAT2 our unchanged between WT and Cog4KO HEK293T and that the prior activity parameter value for MGAT2 was set too high, although changes to the intrinsic properties of MGAT2 due to Cog4KO, such as its rate constant, cannot be ruled out.

144
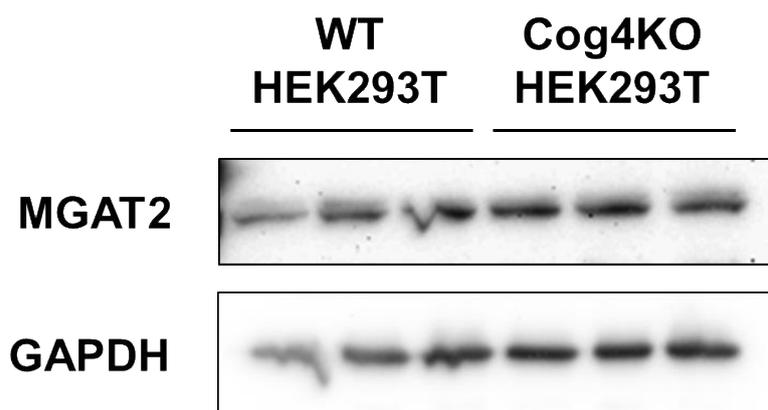
Figure 4.14 MGAT2 protein levels are unaltered in Cog4KO HEK293T cells

Western blot analysis of endogenous MGAT2 in WT and Cog4KO HEK293T cells

4.3.3 Flux map analysis of WT and Cog4KO HEK293T glycosylation

It was puzzling that neither the level or the localization of FUT8 was predicted to be altered in the Cog4KO HEK293T cells, despite the clear reduction in the overall level of fucosylation compared to WT (figure 4.7). Thus, to further investigate the action of FUT8, flux maps for all the reactions that generate the predicted glycans of WT and Cog4KO HEK293T were constructed (figure 4.15, 4.16). Interestingly, for both WT and Cog4KO cells, the fucosylation reactions occur early on in the glycan processing pathway, just after MGAT1 has acted. This early action of FUT8 was despite the most abundant fucosylated glycans observed being mature complex glycans, often capped with sialic acids and having multiple antennae. The preferred substrates of FUT8 (93, 230, 231) are in good agreement with the prediction of FUT8 acting early in the Golgi glycosylation pathway.
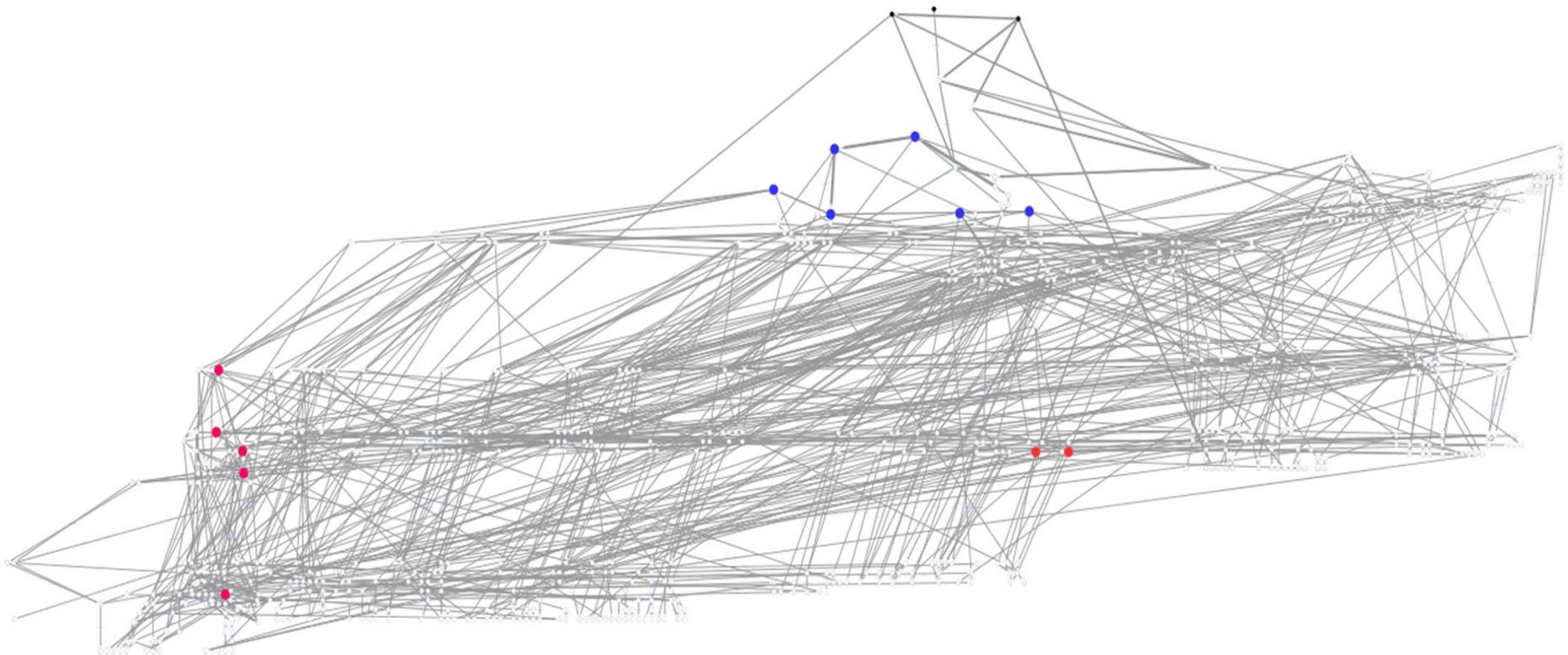
Figure 4.15 | Total flux map of in silico glycosylation in WT HEK293T

Simulation of all WT HEK293T enzymatic glycan processing reactions for 10,000 input glycans plotted using Cytoscape software and shown on a flux map, following fitting. Black dots at the top of the map represent the input glycans GlcMan9GlcNAc2, Man9GlcNAc2, Man8GlcNAc2. Blue dots represent the six substrates with the highest FUT8 flux. Red dots represent the most abundant fucosylated glycans in WT HEK293T cells.
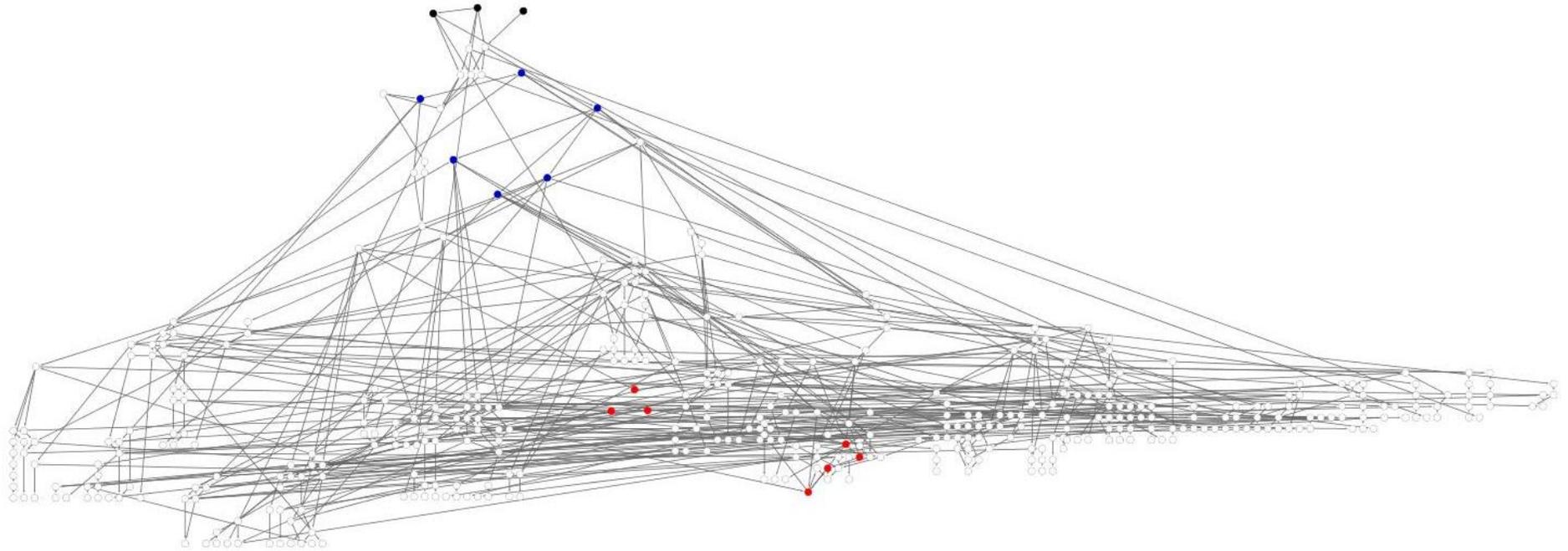
Figure 4.16 | Total flux map of in silico glycosylation in Cog4KO HEK293T

Simulation of all Cog4KO HEK293T enzymatic glycan processing reactions for 10,000 input glycans plotted using Cytoscape software and shown on a flux map, following fitting. Black dots represent the input glycans GlcMan9GlcNAc2, Man9GlcNAc2, Man8GlcNAc2. Blue dots represent the six substrates with the highest FUT8 flux. Red dots represent the most abundant fucosylated glycans in Cog4KO HEK293T cells.
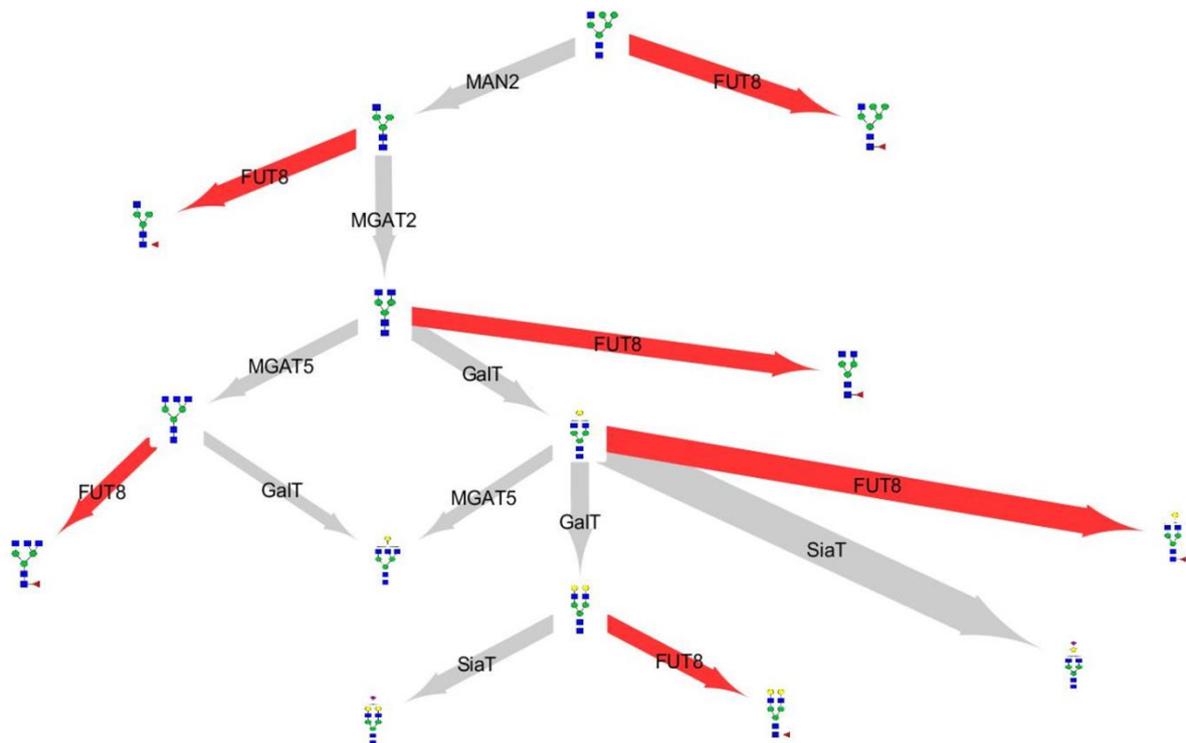
Figure 4.17 | Normalized ratio of the Cog4KO/WT flux for the top six substrates for FUT8.

Normalized ratio of Cog4KO/WT fluxes for the top six substrates for FUT8 and competing reactions. *N*-glycan structures of substrates and products are shown. The top FUT8 fluxes were identified from the total flux maps of WT and Cog4KO HEK293T cells (figure 4.14 and 4.15). Normalized flux map was plotted using Cytoscape software. Red arrows highlight the fluxes of the FUT8 enzyme.

To further investigate the core fucosylation reactions occurring in WT and Cog4KO HEK293T cells, the reactions of the FUT8 enzyme competing for the top six substrates were compared (figure 4.17, table 4.3). The flux for each reaction was first normalized to the total fucosylation flux in each cell line, to compensate for the fact that there is less total complex glycan in the Cog4KO HEK293T cell line than in the WT. The total FUT8 flux for WT HEK293T cells was 2875 ± 59 (n = 3) reactions compared with 1971 ± 46 (n = 3) reactions in the Cog4KO HEK293T cell line. Indeed, this analysis suggests that the observed decrease in overall fucosylation in the glycan profile of Cog4KO HEK293T cells is the result of there being a smaller total amount of complex glycans.

Despite the overall reduction in fucosylated *N*-glycans being the result of lower levels of total complex glycans in the Cog4KO HEK293T cells, changes were predicted in the FUT8 fluxes of the two cell lines. For example, due to the reduced competition with

MGAT5 and GalT, FUT8 had a higher preference for $Hex_4HexNAc_4$ in the Cog4KO HEK293T cells than in WT HEK293T cells. Furthermore, $Hex_3HexNAc_4$ is more galactosylated in Cog4KO cells at the expense of branching by MGAT5, whereas galactosylation of $Hex_4HexNAc_4$ is lost in favour of sialylation and fucosylation in the Cog4KO cells. The ability to generate and probe glycosylation flux maps offers the potential to identify key reactions that determine glycan profiles of cells or proteins.

Glycan substrate	Normalized KO/WT	Glycan product	Enzyme
$Hex_3HexNAc_3$	0.83	$Fuc_1Hex_3HexNAc_3$	FUT8
$Hex_3HexNAc_3$	0.67	$Hex_3HexNAc_4$	MGAT2
$Hex_3HexNAc_4$	0.79	$Hex_4HexNAc_4$	GalT
$Hex_3HexNAc_4$	0.81	$Fuc_1Hex_3HexNAc_4$	FUT8
$Hex_3HexNAc_4$	0.37	$Hex_3HexNAc_5$	MGAT5
$Hex_3HexNAc_5$	0.49	$Fuc_1Hex_3HexNAc_5$	FUT8
$Hex_3HexNAc_5$	0.17	$Hex_4HexNAc_5$	GalT
$Hex_4HexNAc_4$	0.82	$Hex_5HexNAc_4$	GalT
$Hex_4HexNAc_4$	1.53	$Fuc_1Hex_4HexNAc_4$	FUT8
$Hex_4HexNAc_4$	2.64	$NeuAc_1Hex_4HexNAc_4$	SiaT
$Hex_4HexNAc_4$	0.31	$Hex_4HexNAc_5$	MGAT5
$Hex_5HexNAc_3$	0.97	$Fuc_1Hex_5HexNAc_3$	FUT8
$Hex_5HexNAc_3$	0.71	$Hex_3HexNAc_3$	MAN2
$Hex_5HexNAc_4$	0.60	$Fuc_1Hex_5HexNAc_4$	FUT8
$Hex_5HexNAc_4$	0.61	$NeuAc_1Hex_5HexNAc_4$	SiaT

Table 4.3 Normalized flux for top 6 substrates of FUT8 and competing reactions

The ratio of fluxes through each reaction for WT and Cog4KO HEK293T was normalized to the ratio of total core fucosylation for each cell line to account for the differences in overall levels of complex glycans in the two cell lines. Values > 1.0 indicate a dominance of the reaction in the Cog4KO cells compared with WT cells. Values < 1.0 indicate a dominance of the reaction in the WT cells compared with Cog4KO cells. Values were used to plot the normalized flux map in figure 5.17.

4.4 Conclusion

Defects in the COG complex result in aberration in the cellular glycan profiles (58, 60). Due to the complex alterations to enzyme homeostasis, it is very difficult to pinpoint which enzymes and/or reactions are most critically altered in these COG-defective cell lines. As demonstrated in the previous chapter, a computational model has been developed that allows the organization of glycosylation enzymes to be predicted. This modelling methodology was here used to investigate the effects of Cog4KD in HeLa cells. The modelling of glycosylation in Cog4KD HeLa cells led to the prediction of a shift in the *trans* direction of the MAN1 enzyme.

As complex glycans are detected at low abundance in HeLa cells, only predictions regarding the enzymes acting on oligomannose glycans can made confidently. In contrast, modelling the glycan profile of Cog4KO HEK293T cells predicted shifts in the levels and/or localizations of several enzymes acting on complex glycans, including GalT and MGAT5. The total activity of both enzymes was predicted to be decreased and both also shifted toward the *cis*-face of the Golgi as a result of Cog4KO. Other enzyme levels such as those of FUT8 were predicted to be less sensitive to Cog4KO and several of these predictions, including the altered GalT localization, could indeed be validated experimentally.

Importantly, the model allowed the generation of flux maps that can be used to investigate which enzymatic reactions are altered as a consequence of the mutation. This information could be used in the future to highlight which reactions should be targeted to correct glycosylation in CDG patients. For example, by identifying the dominant substrates for the FUT8 enzyme in both WT and Cog4KO cells, $Hex_4HexNAc_4$ was shown to be a more favourable substrate for FUT8 in the Cog4KO cells than it is in WT cells. Less competition from GalT and MGAT5 for $Hex_4HexNAc_4$ resulted in increased FUT8 activity for this substrate. The consequence of this finding is that, although the overall reduction in total fucosylation for Cog4KO cells was the result of a reduction in MGAT1 levels, simply overexpressing MGAT1 in Cog4KO cells would result in a glycan profile containing fucosylated glycans that are different from those in the WT. The model can therefore be a powerful tool in the development of

COG-CDG treatments and in the glyco-engineering of recombinant therapeutic proteins.

Chapter 5: Glycosylation enzyme homeostasis and osteogenesis

5.1 Rationale for investigating glycosylation enzyme homeostasis in mesenchymal stromal cells and osteoblasts

In chapter 3, the development is described of a computational model that utilises a stochastic algorithm to simulate Golgi glycosylation reactions and a Bayesian-based framework to fit observed glycan profiles. This modelling methodology has been used to fit the glycan profiles of two mammalian cell lines, HeLa and HEK293T cells. In chapter 4, the glycan profiles of HeLa and HEK293T cell lines that were defective in Golgi trafficking were modelled. HeLa and HEK293T cells are commonly used by cell biologists but have do not have any specialized function. In order to apply the developed modelling methodology to a biological process (i.e. differentiation), the glycan profile of mesenchymal stromal cells (MSCs) and osteoblasts were modelled.

A cellular model that has previously been used to study glycan function is MSCs. The ability of MSCs to differentiate down a number of lineages, including into bone, opens up the possibility that MSCs can be used in regenerative medicine. For example, large bone defects could be repaired following implantation of MSCs on microporous hydroxyapatite scaffolds (232). MSCs can be isolated from bone marrow and have previously been immortalized into cell lines amenable to tissue culture (148). James et al. isolated and immortalized four different MSC lines, termed Y101, Y201, Y102 and Y202. In this chapter the Y101 and Y201 MSC lines are used (148). Importantly, the Y101 and Y201 cell lines were capable of undergoing osteogenesis and have indistinguishable *N*-glycan profiles (107).

The process of osteogenesis is complex, with numerous levels of control. The precise roles of glycosylation during the osteogenesis process are not yet known, although glycosylation has been shown to modulate the ability of MSCs to differentiate (147) and the glycan profiles of osteoblasts differ from those of the MSCs from which they are derived (107). The shift in glycan profile between stem cells and osteoblasts suggests changes either in enzyme levels and/or their localization. More information

about the changes at the level of enzyme organization is needed to establish molecular details of the functional link between glycan processing and differentiation (147).

To investigate the changes in glycosylation that occur during osteogenesis, the developed modelling methodology was used to simulate the glycan profile of undifferentiated Y101 MSCs. Similar to the HeLa and HEK293T cell lines, the glycan profile of the related Y201 MSC line treated with swainsonine was also fitted. Finally, the glycan profile of osteoblasts derived from the Y101 MSC line was fitted, giving rise to predicted changes in the glycosylation machinery, which have been further investigated.

5.2 Aims

The broad aim of the work described in this chapter was to demonstrate the ability of the modelling methodology to make predictions regarding the glycosylation machinery in the context of a cellular event such as differentiation. Specifically, the aims were to simulate the *N*-glycan profile of undifferentiated MSCs and to confirm the validity of the modelling by fitting the glycan profile of swainsonine-treated MSCs. Following this, the aim was to simulate the glycan profile of osteoblasts derived from MSCs in order to predict potential alterations in the glycosylation machinery. The final aim was to investigate and independently validate predictions that arise from the modelling of the osteoblast glycan profile.

5.3 Results

5.3.1 Modelling the glycan profile of an immortalized mesenchymal stromal cell line

Glycosylation is known to play an important role in MSC differentiation potential (147). MSCs from bone marrow are made up of a heterogenous population of cells with differing levels of potency. The four MSC clonal lines isolated by James *et al.* were positioned at different extremes of the differentiation competency scale (148). The Y101 and Y201 cell lines were found to be tri-potent, while the Y102 and Y202 cell lines were relatively impotent (148). The Y101 and Y201 cell lines can undergo osteogenesis, adipogenesis and chondrogenesis. Glycan profiles of the different MSC

clonal lines could not be used to predict differentiation capacity (107) and did not differ significantly.

In order to investigate the organization of the glycosylation machinery in one of the immortalized MSC clones (designated Y101), the *N*-glycan profile was fitted (figure 5.1). The fitted WT HEK293T cell parameters were used as the prior probability distributions for the modelling of the Y101 MSCs. The WT HEK293T parameters were chosen as the prior distributions, as the elaborate glycan profile of the HEK293T cells more closely resembled that of the MSCs, compared with the relatively simple HeLa cell glycan profile. Having demonstrated the validity of the fitting methodology by modelling HeLa and HEK293T glycan profiles, following Bayesian based fitting the Y101 MSC parameters were manually fine-tuned. Manually fine-tuning parameters in this case is acceptable because the fitting methodology has already been validated extensively in chapters 3 and 4. The organization of the glycosylation machinery in MSCs is also not being compared to that of the WT HEK293T cells. The glycan profile of Y101 MSCs contains the highest proportion of complex glycans of all the cell lines examined in this work. Therefore, predictions made by the model are equally strong for the oligomannose and complex glycan processing enzymes.
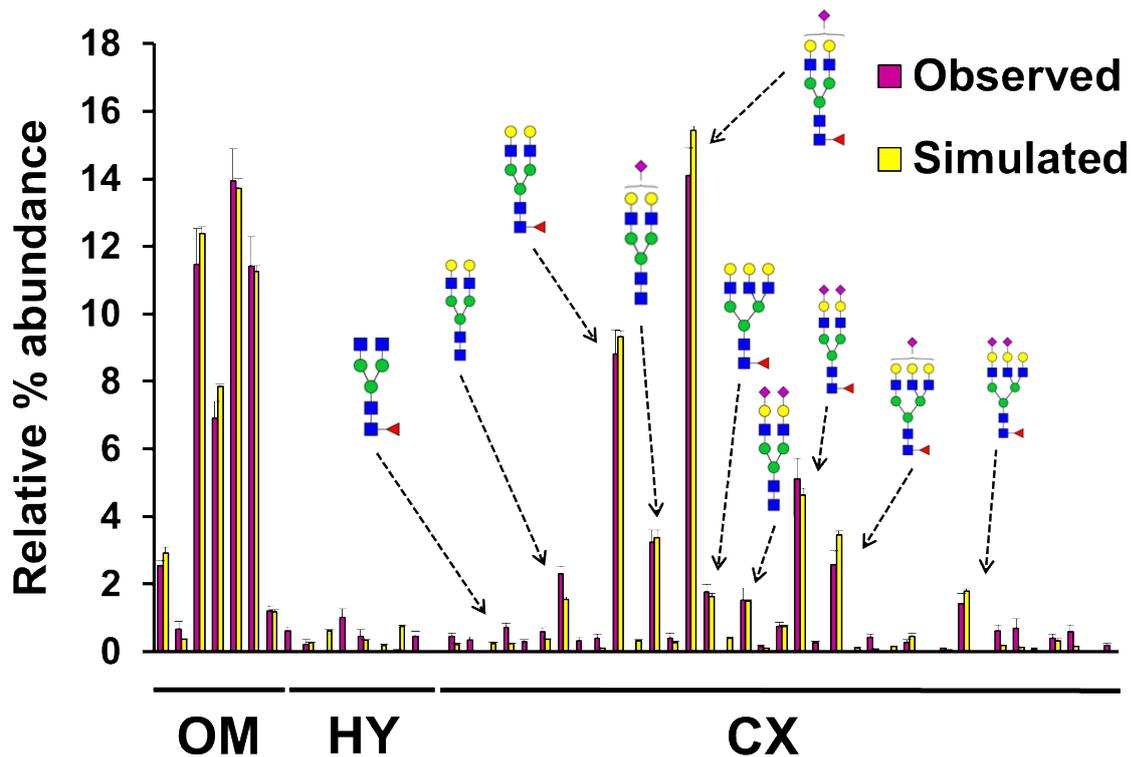
Figure 5.1 | Modelling of whole cell *N*-glycome of MSCs Y101

Observed (magenta) and simulated (yellow) glycan profile of whole cell Y101 MSCs. The observed glycan profile has been previously published (Wilson et al. 2016). To generate the simulated glycan profile the SSA was run three times using the mean parameter values from n = 14 individual fitting runs followed by manual fine-tuning of the MAN1, MGAT1, ant FUT and GlcNAc-phosphoT parameters. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.

The predicted relative distribution of enzymes in the Golgi of Y101 MSCs is shown in table 5.1. Similar to the predicted distribution in WT HEK293T cells, endo-mannosidase and GlcNAc-phosphotransferase are located in the first cisterna, followed by MAN1 and MGAT1. MAN2 is predicted to be located in the early *medial* Golgi. MGAT2, 4 and 5 have similar localizations in the late *medial* Golgi, while FUT8 has a distinct localization peaking in the second cisterna with the remainder evenly spread across the remaining cisternae. GalT and SiaT are the only enzymes predominately localized in the *trans* Golgi.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
MAN1	11±1%	81±9%	7±2%	1±0.2%
Endo-Man	75±17%	14±4%	10±2%	1±0.2%
MAN2	6±1%	57±15%	31±5%	6±1%
MGAT1	1±0.1%	88±21%	10±2%	1±0.2%
MGAT2	10±1%	33±5%	42±5%	14±2%
MGAT4	10±1%	34±6%	43±14%	13±3%
MGAT5	11±1%	31±2%	41±3%	17±3%
FUT8	15±1%	48±8%	18±3%	18±4%
ant FUT	5±1%	71±11%	19±4%	5±1%
GalT	1±0.1%	5±1%	58±10%	36±9%
SiaT	6±1%	7±1%	50±7%	37±4%
GlcNAc-PhosphoT	60±4%	18±3%	20±1%	2±0.3%

Table 5.1 | Predicted distribution of enzymes in Golgi apparatus of MSCs

The predicted distribution of enzymes in Y101 MSC Golgi apparatus following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 14 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna.

In order to ensure that the model would make rational predictions in the context of glycosylation in MSCs, the glycan profile of a different MSC clonal line (Y201), treated with swainsonine, was also modelled. Both the Y101 and Y201 immortalized cell lines have the ability to differentiate into osteoblasts (148) and have very similar glycan profiles (107). Starting with the fitted parameters for the undifferentiated Y101 MSCs, the N-glycan profile of swainsonine-treated Y201 MSCs could be simulated (figure 5.2). As shown in figure 5.2, swainsonine treatment of Y201 MSCs resulted in a large increase in the abundance of the fucosylated oligomannose glycan, Fuc_1Hex_5-$HexNAc_2$. In addition to an increase in the levels of those hybrid glycans observed in the HeLa and HEK293T cell lines, swainsonine treatment of the Y201 MSCs also resulted in an increase in several multi-antennary hybrid glycans (figure 5.2), such as $NeuAc_1Fuc_1Hex_7HexNAc_4$ and $NeuAc_2Fuc_1Hex_7HexNAc_4$. As with modelling the glycan profiles of swainsonine-treated HeLa and HEK293T cells, a large decrease in the total activity of MAN2 was expected, as swainsonine is known to inhibit MAN2.
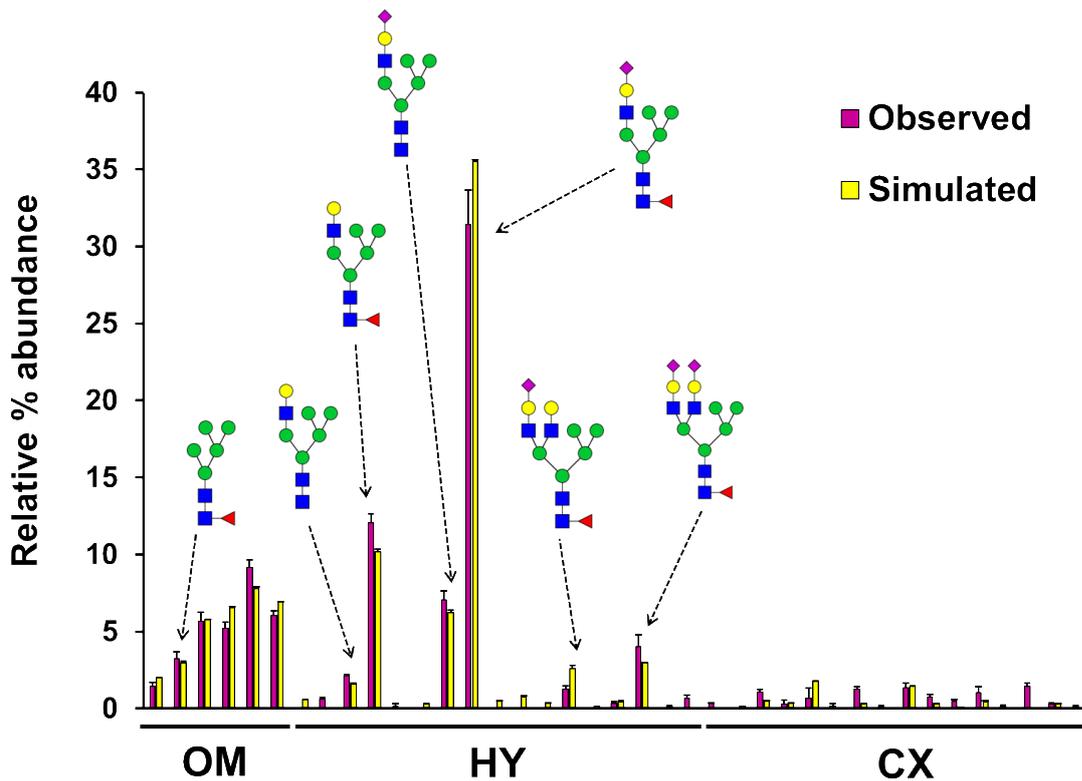
Figure 5.2 | Modelling of the glycan profile derived from swainsonine-treated Y201 MSCs

Observed (magenta) and simulated (yellow) glycan profile of whole cell Y201 MSCs treated with swainsonine. The observed glycan profile was acquired by Katie Wilson, a former member of the Ungar group. To generate the simulated glycan profile the SSA was run three times using the mean parameter values from n = 11 individual fitting runs followed by manual fine-tuning of the MGAT1 enzyme and the fucosylation of Man5GlcNAc2 parameters. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.

Indeed, in order to simulate the glycan profile of swainsonine-treated Y201 MSCs, the model predicted a large decrease in MAN2 (figure 5.3A), as expected. To account for the increase in the branched hybrid glycans, the model predicted an increase in the activity of MGAT4, suggesting the model can identify novel features of the glycosylation process. An increase in the rate of sialylation in the swainsonine-treated HeLa and HEK293T cell lines was a prediction previously made (section 3.3.4). However, for the fitting of swainsonine-treated Y201 MSCs, the total activity of sialylation decreased relative to that in untreated cells. Rather than increasing the total activity levels of SiaT, the fitting procedure elected to decrease the SiaT scale-factor in the swainsonine-treated Y201 MSCs (figure 5.3B). Decreasing the SiaT scale-factor by 0.42-fold has the effect of increasing the overall rate of sialylation for hybrid glycans.

This prediction is consistent with the previous predictions regarding HeLa and HEK293T cells (section 3.3.4) and suggests that the increased SiaT rate in swainsonine-treated HeLa and HEK293T cells is due to the altered rate of sialylation for different antenna.
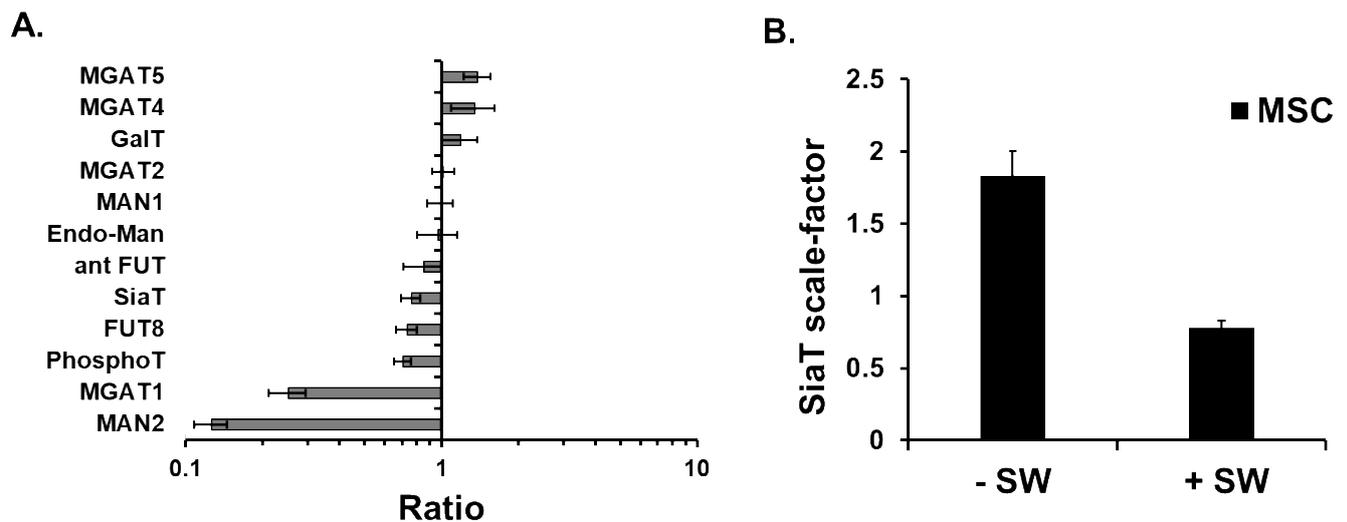
Figure 5.3 | Predicted total activity changes in swainsonine-treated MSCs

A) The predicted ratio of total activity levels of the glycosylation enzymes in MSCs when treated with swainsonine. The comparison is between untreated Y101 MSCs and swainsonine-treated Y201 MSCs. The predicted total activity of swainsonine-treated Y201s (n = 11) is divided by the predicted total activity untreated Y101s (n = 14). B) Predicted SiaT scale-factor in untreated Y101 MSCs and swainsonine-treated Y201 MSCs. Error bars are the propagated standard deviation.

5.3.2 Predicting alteration in glycosylation enzyme homeostasis in osteogenesis

Although the *N*-glycan profile of osteoblasts contained the same set of glycans as the undifferentiated Y101 MSCs, there were significant alterations in the relative abundances of individual glycans (figure 5.4). The glycan profiles of the undifferentiated MSCs (Y101) and osteoblasts derived from them are reproduced from data obtained by K. Wilson and previously published (107). Figure 5.4 shows a comparison of the glycan profiles for the readers' convenience. The most obvious difference between the Y101 MSCs and osteoblasts is the prevalence of

oligomannose glycans, which is significantly reduced in the osteoblasts, while the abundance of complex glycans is increased (figure 5.4). Importantly, the differences in the *N*-glycan profiles of Y101 MSCs and osteoblasts, were not a result of time in tissue culture (107).
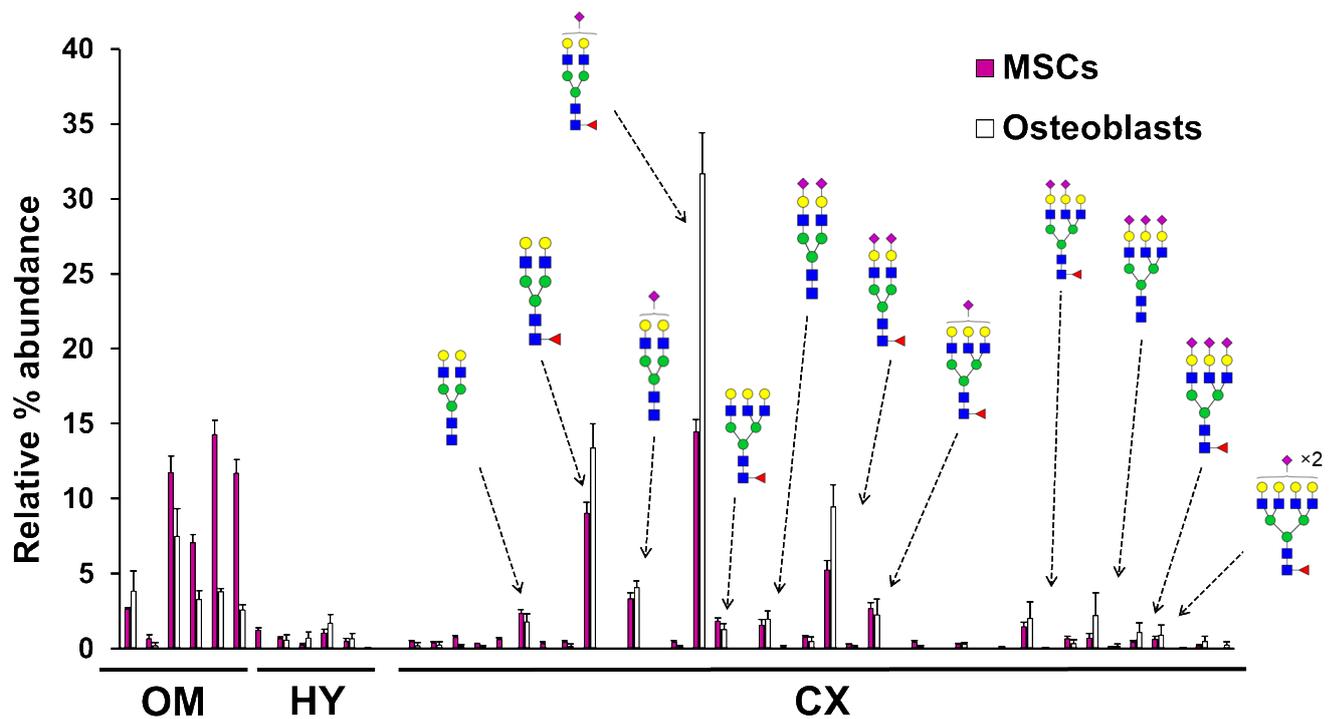
Figure 5.4 | Comparison of Y101 MSC and osteoblast glycan profiles

N-glycan profile of Y101 MSCs (magenta) and osteoblasts derived from Y101s (white). The glycan profiles have been previously published (Wilson et al. 2016) and are shown here for the readers' convenience. OM = oligomannose, HY = hybrid, CX = complex.

Having demonstrated that the modelling methodology could be applied to the elaborate *N*-glycan profile of undifferentiated MSC cells, the glycan profile of the osteoblasts obtained from Y101 MSCs following a 21 day differentiation protocol (148) was then fitted (figure 5.5). The Y101 MSC model parameters were used as the prior distributions in order to directly compare differentiated with undifferentiated, and therefore reveal changes to the glycosylation machinery that occur during osteogenic differentiation. The major alterations to the glycan profile as a result of the Y101 MSCs

undergoing osteogenesis (figure 5.1), could be fitted (figure 5.5), notably the shift from oligomannose to complex glycans.
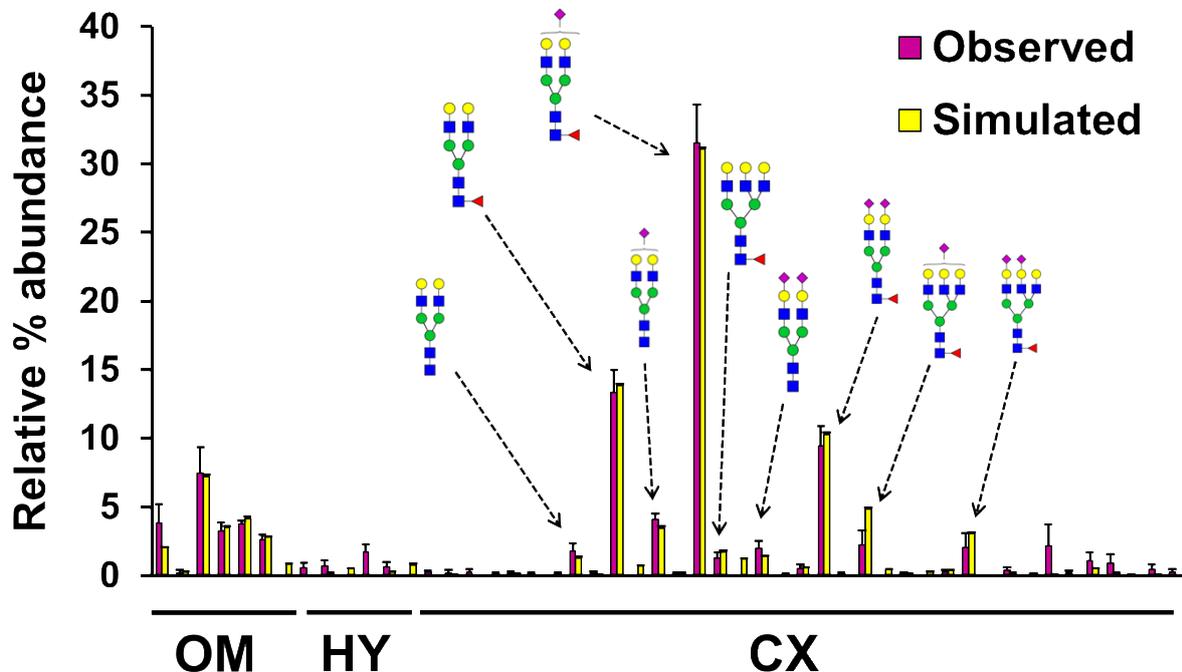
Figure 5.5 | Modelling of whole cell *N*-glycome of osteoblasts derived from Y101 MSCs
Observed (magenta) and simulated (yellow) glycan profile of osteoblasts derived from Y101 MSCs following a 21-day osteogenic differentiation. The observed osteoblast glycan profile has been previously published (Wilson et al. 2016). The glycan profile was simulated three times using the SSA with the mean parameter values from n = 20 individual fitting runs. Error bars are SEM, n = 3. OM = oligomannose, HY = hybrid, CX = complex.

In order to fit the overall decrease in oligomannose glycans on differentiation, the model predicted a relative decrease in the enzymatic activity of the GlcNAc-phosphotransferase (figure 5.6). GlcNAc-phosphotransferase is responsible for initiating the targeting of hydrolases to the lysosomes, hence a decrease in the activity of this enzyme suggests a decrease in the overall lysosomal content in the osteoblasts relative to that in the Y101 MSCs. It follows that a lower abundance of oligomannose glycans would indicate a lower abundance of lysosomal proteins which could indicate either less or smaller lysosomes. To test this hypothesis, the number/size of lysosomes of MSCs and osteoblasts was measured indirectly by the levels of protein LAMP1 (a marker of the lysosome) (figure 5.7A, B) and the activity of β-hexosaminidase

(lysosomal glycosidase) (figure 5.7C). In contrast to the proposed hypothesis, a relative increase in lysosomal content of the cell as a result of osteogenesis was observed. Both the protein levels of LAMP1 (figure 5.7B) and the activity of β-hexosaminidase (figure 5.7C) increased in osteoblasts relative to undifferentiated MSCs. This increase in lysosomal capacity is in agreement with the literature (233, 234). The reason for the shift from oligomannose to complex glycans in the osteoblasts is discussed below.
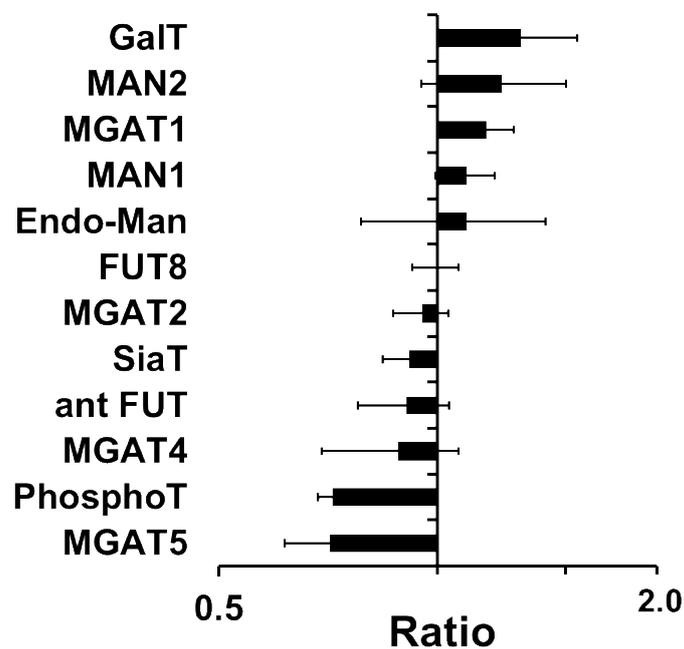
Figure 5.6 | Predicted total enzymatic activity changes in osteoblasts with respect to Y101 MSCs

Predicted total enzymatic activity changes upon differentiation of Y101 MSCs into osteoblasts following fitting of the osteoblast glycan profile. The comparison is between undifferentiated Y101 MSCs and osteoblasts derived from Y101 MSCs. The predicted total activity for MSCs (n = 14) is divided by the predicted total activity of osteoblasts (n = 20). Error bars are the propagated standard deviation.

A.

B.

C.
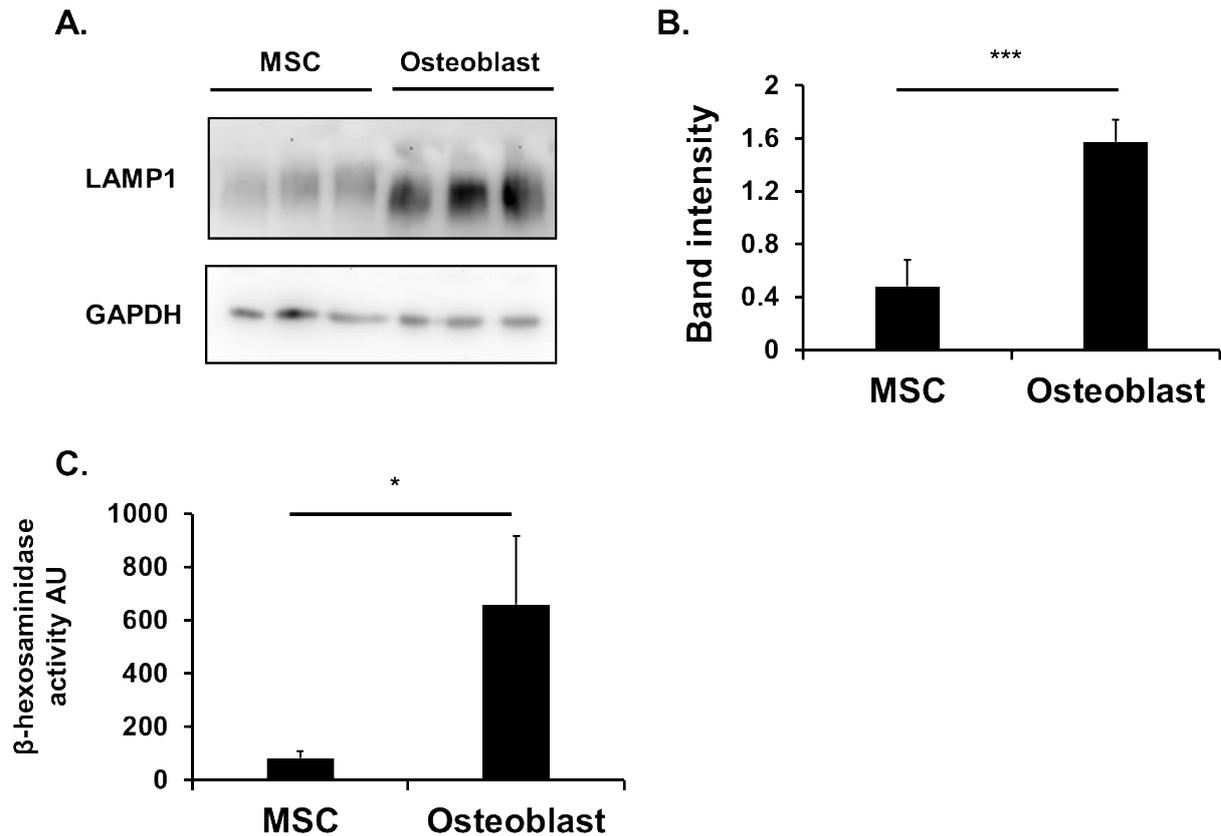
Figure 5.7 | Lysosomal content of MSCs and osteoblasts

A) Western blot analysis of the lysosomal marker, LAMP1 in undifferentiated Y101 MSCs and osteoblasts derived from Y101 MSCs following a 21-day differentiation. B) Quantification of LAMP1 band intensities normalised to GAPDH band intensities. C) Quantification of β-hexosaminidase activity in Y101 MSCs and osteoblasts derived from Y101 MSCs, normalised to total protein content as determined through BCA assay. * denotes $P \leq 0.05$ and *** denotes $P \leq 0.001$ for students' t-test. Error bars are standard deviation, n = 3. Osteogenic differentiation was confirmed by ALP and von Kossa staining.

Of potential interest is that the model predicted a decrease in the total enzymatic activity of MGAT5 (figure 5.6). Furthermore, the predicted distribution of MGAT5 in osteoblasts was flattened out across the Golgi cisternae (figure 5.8, table 5.2), compared with that in the MSCs (table 5.1). The reasons for the changes in MGAT5 are not clear from directly comparing the glycan profiles of osteoblasts and Y101 MSCs (figure 5.4). The predicted decrease in MGAT5 suggests there is a relative decrease in the levels of multi-antennary glycans. By excluding the oligomannose glycans and considering just the complex glycans, the large differences between the two cell lines are no longer apparent (figure 5.9A). The similarity in the complex glycan profiles over the two cells lines is likely the reason why the model does not predict

164

more changes in the organization of glycosylation machinery. Importantly, the relative abundance of the bi-antennary glycan $NeuAc_1Fuc_1Hex_5HexNAc_4$ is elevated in the osteoblasts compared with the MSCs. The overall level of tri-antennary glycans is consequently reduced in the osteoblasts (figure 5.9B), which is the reason the model predicted a decrease in MGAT5 activity.
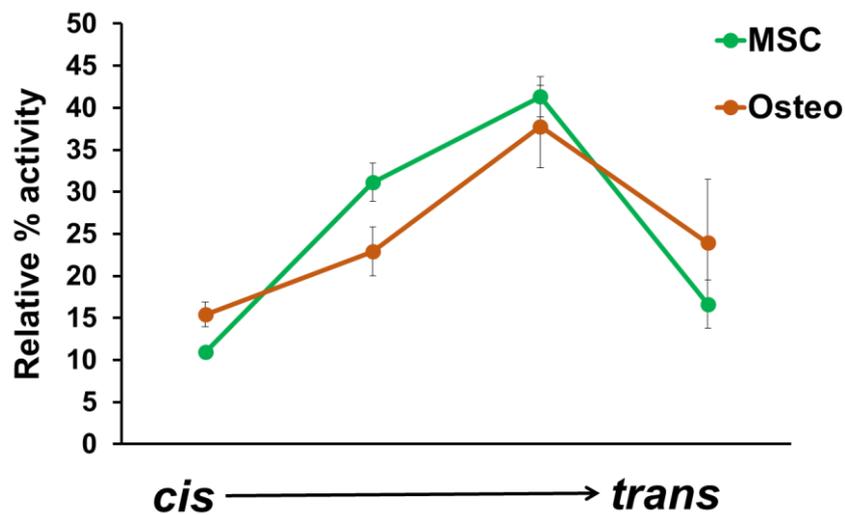
Figure 5.8 | Predicted distribution of MGAT5 is shifted in osteoblasts

Predicted enzymatic distributions of MGAT5 in Y101 MSCs (green) and osteoblasts derived from Y101 MSCs (orange) following the fitting procedure. Error bars are standard deviation for n = 14 (Y101 MSCs) and n = 20 (osteoblasts) individual fitting procedures.

Enzyme	Cisterna 1	Cisterna 2	Cisterna 3	Cisterna 4
MAN1	13±2%	78±11%	8±2%	1±0.2%
Endo-Man	77±30%	13±4%	9±3%	1±0.3%
MAN2	5±1%	58±21%	32±9%	6±2%
MGAT1	1±0.1%	87±11%	11±2%	1±0.2%
MGAT2	11±1%	28±5%	44±6%	16±2%
MGAT4	11±2%	26±5%	48±20%	15±4%
MGAT5	15±2%	23±3%	37±6%	25±14%
FUT8	15±1%	49±7%	17±1%	19±2%
ant FUT	7±1%	85±19%	5±1%	4±1%
GalT	1±0.1%	3±1%	66±17%	31±7%
SiaT	7±0.3%	8±2%	48±4%	37±3%
GlcNAc-PhosphoT	32±2%	30±4%	35±2%	3±1%

Table 5.2 | Predicted distribution of enzymes in Golgi apparatus of osteoblasts

The predicted distribution of enzymes in the Golgi apparatus of osteoblasts derived from Y101 MSC following fitting of the whole cell glycan profile. Each predicted % activity is the mean of n = 20 individual fitting runs and confidence intervals are standard deviation. Colour denotes % activity in each cisterna.
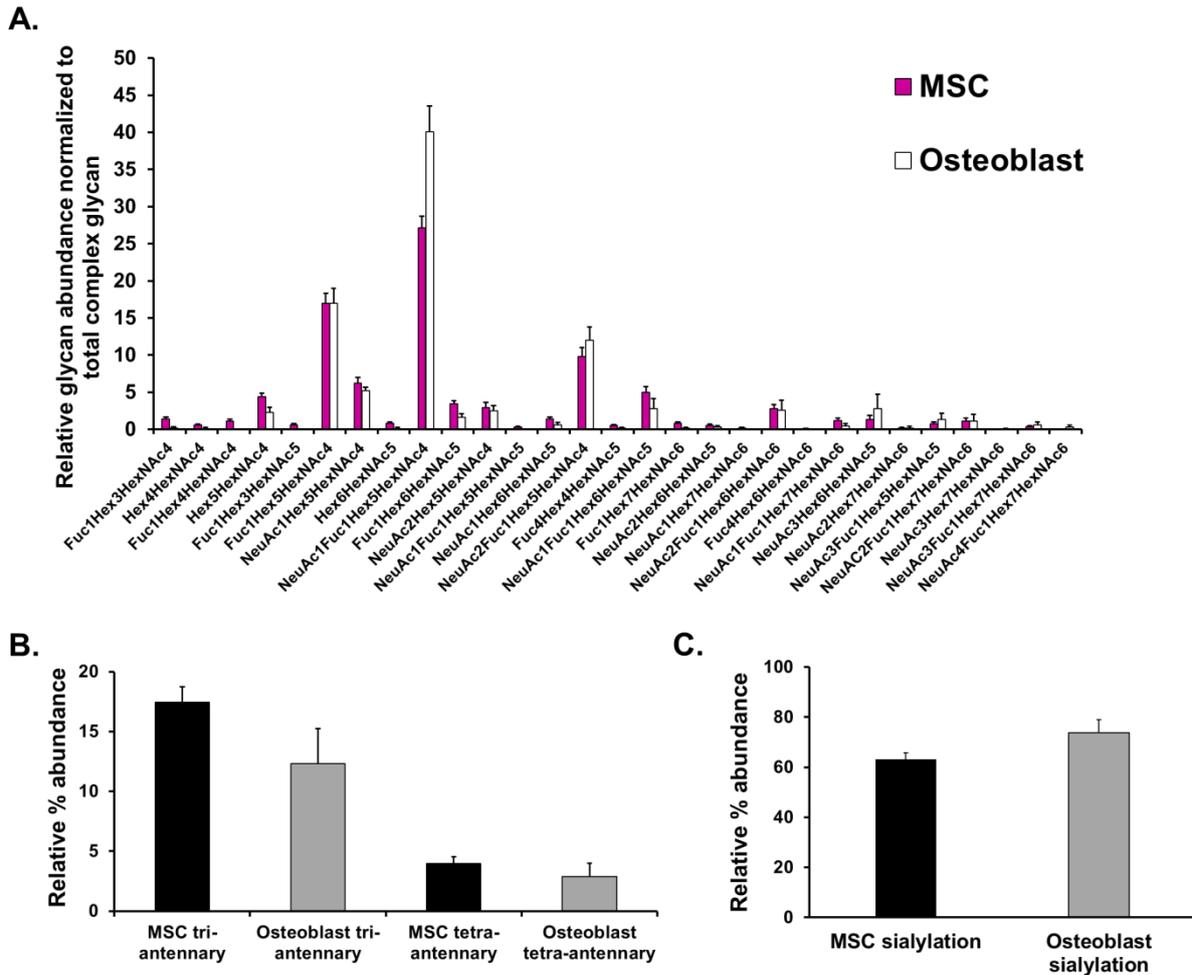
Figure 5.9 | Comparison of complex glycans only in Y101 MSCs and osteoblasts

A) Observed glycan profile of Y101 MSCs and osteoblasts derived from Y101 MSCs. Glycan ion intensities for the two condition were normalised to the total ion intensities of complex and hybrid glycans, excluding the intensities of oligomannose glycans. B) Relative abundances of tri- and tetra-antennary glycans in MSCs and osteoblasts. C) Relative abundance of sialylated *N*-glycans in MSCs and osteoblasts. Error bars are SEM, n = 3.

Increased branching in *N*-glycans has been linked to galectin binding (105). Galectins are a family of proteins containing carbohydrate recognition domains that recognise and bind to oligosaccharides containing galactose, for example the Gal1,4GlcNAc moiety of *N*-glycans. Galectins interact weakly with individual glycan moieties, but galectins can oligomerise and form multiple interactions with multiple galactose-containing moieties. Therefore, the avidity between galectins and glycans can be increased through increasing the antenna number in *N*-glycans. The degree of galectin binding in turn has been shown to modulate cell proliferation and differentiation (133), as well as osteogenic differentiation of pre-osteoblasts through Notch signalling (235).

In addition to lower numbers of antennae, osteoblasts also exhibit increased relative levels of sialylation when compared with their MSCs (figure 5.9C), which can also inhibit galectin binding (236).

A.
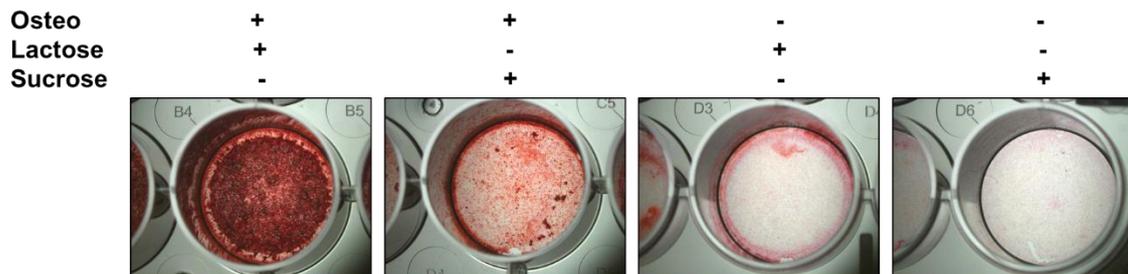
B.
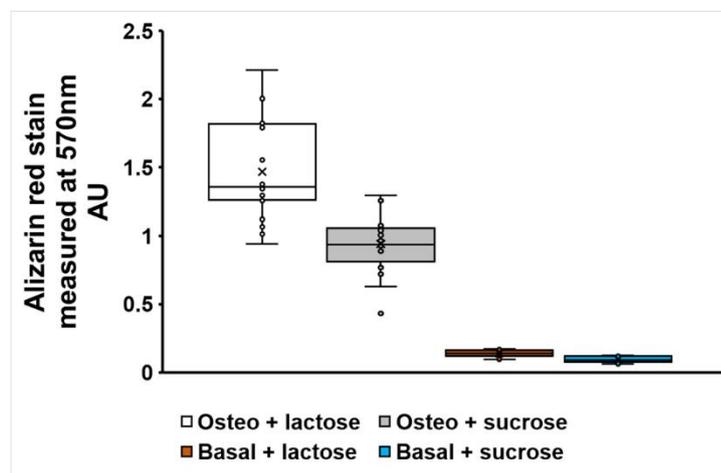
Figure 5.10 | Galectin inhibition promotes osteogenesis

A) Representative images of alizarin red staining of Y101 MSCs after a 21-day osteogenic differentiation. B) Quantification of eluted alizarin red stain as measured at 570nm for n = 20 (osteogenic plus lactose or sucrose treatments) and n = 11 (osteogenic treatment and basal) from two independent experiments.

Previous experiments have demonstrated the role of glycosylation early on in the osteogenesis process (147). Wilson *et al.* found that when the immortalized Y101 MSC line was treated with the drug kifuensine (an inhibitor of MAN1), the osteogenic potential was increased. Kifuensine treatment prevents the biosynthesis of hybrid and complex glycans, thus indirectly preventing galectin binding. In order to further investigate the role of *N*-glycan antenna number, galectin binding in osteogenesis, the

Y101 MSC cell line was treated with either lactose or sucrose for the first seven days of a 21-day osteogenesis protocol. Lactose inhibits galectin binding while sucrose does not. As shown in figure 5.10, inhibition of galectin binding by lactose during the first 7 days of osteogenesis increased osteogenic potential over sucrose treatment, as assayed by Alizarin Red staining for calcium. Osteoblasts deposit calcium extracellularly in a process known as mineralization. The increase in Alizarin Red staining for the lactose treatment over that of the sucrose treatment implies that the reduction in *N*-glycan branching observed in the osteoblast glycan profile is a contributing factor in osteogenic differentiation.

5.4 Conclusions

An SSA and Bayesian fitting procedure was used to simulate the glycan profile of the immortalized Y101 hTERT MSC line (148). As for the other cell lines used in this work, the glycan profile of swainsonine-treated MSCs was also fitted. Fitting of the swainsonine-treated MSC glycan profile revealed an expected decrease in MAN2, demonstrating that the modelling methodology is applicable to modelling the elaborate glycan profiles of MSCs.

The reason for validating the model for MSC was to be able to predict adaptions of the glycosylation machinery that occur when the Y101 hTERT MSC line undergoes differentiation into osteoblasts. Previously, inhibition of MAN1 with kifunensine in Y101 MSCs has been shown to promote osteogenesis (147), highlighting the important role of *N*-linked glycans in this process. Due to the insensitivity of the model to glycan degradation, meaningful conclusions about the oligomannose/complex glycan ratio could not be made. However, by looking at just complex *N*-glycans, a notable switch from tri-antennary to bi-antennary glycans upon differentiation was observed. The model predicts this switch to be controlled through a decrease in the levels of the branching enzyme MGAT5. The degree of MGAT5-initiated *N*-glycan branching has been intimately linked to galectin binding (105), with a consequence of altered receptor endocytosis that impacts on cell proliferation and differentiation (133, 235, 237). Lactose inhibition of galectin binding was effective in the first week of the three-week differentiation process, suggesting that the switch from tri- to bi-antennary glycans releases an inhibitory effect on differentiation.

Chapter 6: Discussion

Glycosylation is an important aspect of biology, the role of which has, until recently, been underappreciated. The glycosylation of proteins challenges the central dogma of biology, as specific glycan structures can often alter the structure and properties of particular proteins (112, 238). It is therefore of great interest and importance to understand the biosynthetic pathway of glycans. Despite glycosylation being inherently heterogenous, the heterogeneity is not random. Mass spectrometry and other techniques have demonstrated reproducible glycan profiles for individual cell lines and proteins. At the enzymatic level much is now known regarding the reactions that generate and modify *N*-glycans. The glycan structures that are produced by the cell are influenced by the localization and quantitative amount of enzymes within the Golgi apparatus. While glycan profiles are relatively easy to acquire, eliciting the organization of all relevant glycosylation enzymes within the Golgi apparatus is difficult and time consuming.

Computational modelling of complex biological systems and processes has aided understanding of cellular biological processes. Therefore, in order to understand the contribution of enzyme organization in the Golgi apparatus, a computational workflow capable of predicting enzymatic activities and localizations, based on experimentally determined *N*-glycan profiles has been developed. The modelling methodology developed in this work is complementary to different *in silico* models of glycosylation and has wide ranging applications, for example: predicting strategies for tailoring glycosylation of biologics; identifying disrupted glycosylation reactions in CDG patients; or as a process of discovery in cellular events such as differentiation, or diseases such as cancer.

6.1 Developing an SSA to model mammalian *N*-glycosylation

A computational model capable of simulating Golgi *N*-glycan processing and fitting the glycan profiles of mammalian cells has been developed. This model has the ability to generate testable hypotheses regarding the *N*-glycosylation machinery. The model was based on two algorithms, an SSA to simulate glycosylation reactions and an ABC algorithm to fit the model to observed data. For several reasons an SSA was chosen and developed to simulate the glycosylation reactions occurring in the Golgi

apparatus. Firstly, the process of glycosylation is inherently heterogenous, due to the non-templated basis of glycan biosynthesis. Heterogeneity of glycosylation between two different glycoproteins can be the result of the quaternary structure of each protein, the degree of biological noise in the system and differences in the trafficking of each protein through the Golgi apparatus. Furthermore, different glycan structures may be found at the same glycosylation site of two separate molecules of the same protein (210, 238): this phenomenon is known as microheterogeneity. Glycan microheterogeneity is the result of enzyme competition, substrate competition and biological noise. The existence of microheterogeneity as a consequence of biological noise warrants a stochastic model of glycosylation. The inclusion of intrinsic noise is an important aspect of stochastic simulations and has been recognised as a contributing factor in the field of biology, for example in genetics (175). In the context of glycan biosynthesis in the Golgi apparatus, biological noise is likely a contributing factor to glycan heterogeneity. This is because the number of molecules residing in the confined volume of the Golgi apparatus is not large enough to average out the fluctuations that are the result of biological noise. The contribution of biological noise to heterogeneity increases as the number of molecules decreases because noise behaves as $1/\sqrt{X}$, where X is the number of molecules (239).

The second reason for developing a stochastic model of mammalian glycosylation was to simplify an already complex system. Glycan biosynthesis is dependent on a range of factors, such as the availability of monosaccharide nucleotides, the levels of the monosaccharide-nucleotide transporters, enzyme levels and secretory protein levels. The probabilistic nature of an SSA allows these factors to be subsumed within the probabilities, thus eliminating the need for estimating other kinetic parameters (181). For example, the concentration of UDP-GlcNAc, the rate of the UDP-GlcNAc transporter and the rate of MGAT1 can all be encapsulated into the probability of the MGAT1 catalysed reaction occurring. Although such an approach allows the simplification of the reaction system, information is discarded in this approach. This means that the predictions that are made in this work, based on the stochastic simulation algorithm, requires further experimentation to elucidate the true mechanism. This is the ideal situation, in which *in silico* model and experiment are intertwined to further understanding of the biological system being studied.

The modelling framework developed in this work simulates *N*-glycosylation in a stochastic fashion thus accounting for the heterogenous nature of glycan biosynthesis. The biological noise of the system is not included in previous deterministic models of glycosylation (190, 195). Furthermore, the distribution of enzymes in these models are not fitted. Taking a stochastic approach to simulating glycosylation also simplifies the process. Reaction rates and concentrations are combined into one parameter eliminating the requirement of excessive parametrization. By combing a stochastic approach to simulation glycosylation with an ABC-based fitting methodology allows uncertain parameters to be sampled and optimised for a complex data set, such as a whole cell glycan profile.

6.2 Enzyme rules and specificities

During the development of an SSA that could simulate glycosylation reactions, a set of enzymes was chosen and rules regarding each enzyme defined. For HeLa cells, a total of 11 enzymes was used, for HEK293T cells and MSCs a total of 12 enzymes was used. In the four-cisterna model used for HEK293T cells and MSCs, 12 enzymes equated to 56 parameters, including input glycans, transit time and scale-factors. The enzymes chosen represented the lowest number of enzymes that could explain all the *N*-glycan species identified from mass spectra. The lowest number of enzymes and the number of cisternae were chosen in order to keep the parameter matrix low, and therefore reduce the computational cost of fitting. The inclusion of additional enzymes such as MGAT3 has the potential to improve the modelling of the mammalian *N*-glycosylation pathway, but would increase the computational cost of fitting glycan profiles.

To further avoid an unacceptably large increase in the number of required parameters, a phenomena often termed parameter explosion, enzyme efficiency was initially assumed to be independent of the glycan substrate. This is not strictly correct as there are several examples in which the *N*-glycosylation enzymes have known substrate specificities (92, 93, 224). However, introducing substrate independence for each enzyme would have increased the total number of parameters in the four-cisterna model from 56 to at least 130 parameters. The model development process was used

to identify which substrate-specific parameters were most critical to simulating a whole cell *N*-glycan profile. The use of an ABC-based fitting methodology was advantageous in this respect, as it allows refinement of the model based on prior knowledge of the system. It is important to note that the addition of substrate specificities was driven by the requirement for fitting the glycan profile and evidence then found in the literature. The fact that the modelling highlighted the need for such parameters and evidence for these parameters can be found in the literature (92, 99, 112) suggests the model is physiologically relevant. Furthermore, each additional substrate specificity was modelled as a scale-factor that itself was fitted. If the substrate specificity was not required, the scale-factor would converge to 1.0 during the fitting process. This did not occur for any of the scale factors.

MAN1 is active early on in the Golgi glycosylation pathway and is responsible for the trimming of mannose residues from oligomannose-type glycans. A feature of the glycan profiles from the three cell lines investigated in this work is that of a doubly-peaked oligomannose distribution. Indeed, a doubly peaked distribution cannot be fitted using only two competing enzymes (MAN1 and GlcNAc-phosphoT). The oligomannose distribution could not be simulated without accounting for MAN1 substrate specificity. Therefore, MAN1 was allowed to act on $Man_6GlcNAc_2$ at a different rate than on the other oligomannose glycans. MAN1 has been shown to trim $Man_6GlcNAc_2$ to $Man_5GlcNAc_2$ at a slower rate than the trimming of other oligomannose glycans (99). The predicted scale-factors for trimming of $Man_6GlcNAc_2$ were 6.26, 8.88 and 11.39 for WT HeLa, HEK293T cells and MSCs while the MAN1 enzyme from pig has been shown to act 40 times slower on $Man_6GlcNAc_2$ compared with $Man_9GlcNAc_2$ (99). The discrepancy between the predicted scale-factors and that found by Bause *et al*, could be due to different species or the *in vitro* nature of the experiment. The addition of MAN1 substrate specificity allowed the satisfactory fitting of the oligomannose distribution in all cell lines investigated in this work.

During the fitting of WT HEK293T cells glycan profile, the requirement for enzyme specificity for the GalT and SiaT enzymes was highlighted. Substrate specificity was introduced using scale-factors for each enzyme. In the case of GalT, the rate of galactosylation for bi-antennary glycans was allowed to differ from that of tri- and tetra-antennary glycans (92). Following the fitting of WT HEK293T and MSC glycan profiles

the GalT scale factor did not converge to 1.0, indicating that to fit these glycan profiles, substrate specificities for GalT were required. Indeed, differences in the activity of GalT towards bi-, tri and tetra-antennary glycans have been shown (92). For sialylation, SiaT was allowed to act on each antenna of a bi-antennary glycan at different rates. Similar to the GalT scale-factor, the SiaT scale-factor was fitted and converged to a value greater than 1.0 in all instances. Evidence for antenna specificity for SiaT enzyme could also be found in the literature (111, 222). For example, the majority of sialylation (95%) occurs on the 1-3Man branch of a bi-antennary glycan attached to an IgG (111).

More evidence for differences in SiaT substrate specificity is shown in the modelling of glycosylation in swainsonine-treated cell lines. Accounting for the known effects of swainsonine (inhibition of MAN2), by completely removing MAN2 activity from the WT cell line models without further fitting, could not reproduce the experimentally observed glycan profiles of swainsonine-treated HeLa (figure 3.12) or HEK293T (figure 3.13) cells. This was characterized by a sialylation rate that was too low and required increasing to fit the glycan profiles of swainsonine-treated cell lines. The inability of the simulation with MAN2 removed to fully reproduce the glycan profiles of swainsonine-treated cells can be explained in two ways. Firstly, hybrid and complex-type glycans could act differently as substrates. Due to the higher abundance of complex glycans compared with hybrid glycans, the fitting of WT glycan profiles is strongly biased towards enzymes acting on complex glycans. This explanation is supported by examining the simulated and experimental glycan profile of the WT HEK293T cells, in which the fit for the hybrid-type glycans is relatively poor. As hybrid glycans are in relatively low abundance in the cell lines studied in this work and because of the desire to avoid increasing parameter numbers and simulation times, scale-factors for enzymatic activity on hybrid substrates were not used in this work. In the context of sialylation, hybrid glycans may only be sialylated on the 1-3Man branch, which is the preferential branch for SiaT (111, 222) further supporting the argument that hybrid glycans act differently as substrates compared with complex glycans.

A second explanation is that the glycosylation machinery has been fundamentally altered as a result of inhibiting MAN2, either by altering enzyme expression levels or by changing the physical properties of the glycosylation enzymes. The glycan profiles

of swainsonine-treated HeLa (figure 3.14) and HEK293T (figure 3.15) cells could be reproduced accurately following fitting, with the model predicting an increase in sialylation (HeLa and HEK293T) and galactosylation (HEK293T). The predicted increase in the activities of SiaT and GalT could be the result of changes in expression levels. Alternatively, many of the glycosylation enzymes are themselves glycosylated or predicted to be glycosylated (table 3.7). The enzymatic rate of SiaT in particular has been shown to be affected by the structure of the glycan that resides on it (112). The absence of galactose but not sialic acid was found to reduce the activity of SiaT by 60% (112).

6.3 Fitting whole cell glycan profiles of mammalian cell lines

In chapter 3, the glycan profiles of WT HeLa and WT HEK293T cells were modelled. As discussed in section 6.2, the fitting process necessitated the inclusion of substrate specificities for MAN1, GalT and SiaT. The developed modelling methodology was able to reproduce quantitative fits for the glycan profiles of mammalian cell lines. However, there were discrepancies between the fitted and experimental glycan profiles. It was not expected that the modelling framework would be able to reproduce perfectly the experimental glycan profiles. The discrepancies between the simulated and experimental glycan profiles can arise from a number of different sources, technical and biological, including enzyme specificities (section 6.2).

The whole cell glycan profile is a global average readout of the glycosylation process. The influence of protein structure in determining the glycan structures is not considered in the modelling framework. This is likely a major factor in the discrepancies between simulated and experimental glycan profiles. However, by modelling the whole cell glycan profile rather than the glycan profile of an individual protein, predictions can be made regarding the whole Golgi glycan processing pathway. This is potentially an important feature of the model, for example in investigating the effects of COG complex mutations on the cellular glycosylation pathway. A second likely source of discrepancies between the simulated and experimental glycan profiles is the absence of glycan competition in the model. In the model, glycans are processed one at a time meaning that differences in protein flux

through the secretory pathway between cell lines or conditions is not accounted for. The introduction of glycan competition is potentially a future addition to the SSA.

Despite the presence of some discrepancies, the whole cell glycan profile of WT HeLa and HEK293T cells could be simulated following the Bayesian fitting procedure described earlier. The more elaborate glycan profile of WT HEK293T cells required the Golgi model to consist of four compartments compared with the three required for modelling the HeLa cell glycan profile (figure 3.2). Furthermore, the transit time per cisterna was predicted to be longer in WT HEK293T cells than in HeLa cells (figure 3.10C), supporting the notion that a model of WT HEK293T glycosylation requires increased glycan processing capacity. The predicted transit time per cisterna would result in a total Golgi residence time of 15.81 minutes and 29.04 minutes for WT HeLa and HEK293T cells, respectively. This is in good agreement with the characteristic half-times of secretory proteins in the Golgi, which range from 5 to 20 minutes (240).

For both the WT HeLa and HEK293T cells, the MAN1 distribution was predicted to be *cis*/*medial* and *cis*/early-*medial* respectively (figure 3.9). Often considered a marker of the *cis*-Golgi, MAN1 has been shown to have a variable distribution throughout the Golgi apparatus (221). Indeed in a variety of mammalian cell lines, the distribution of MAN1 was shown to overlap considerably with that of MAN2 (221), a feature which is predicted in both HeLa (table 3.5) and HEK293T (table 3.6) cells by the model.

To validate the computational approach that has been developed in this work and to demonstrate that predictions made by the model were rational and verifiable, the *N*-glycan profiles of swainsonine-treated HeLa and HEK293T cells were modelled (section 3.3.4). Unlike in the case of SiaT, in which evidence existed for potential differences in substrate specificities for hybrid and complex glycans, the *in vitro* activity of FUT8 for hybrid glycans was shown to be similar to that of complex glycans (93). For example, the conversion percentage of $Man_5GlcNAc_3$ was 75% compared with 90% for $Man_3GlcNAc_4$ (93). This similar conversion rate of FUT8 towards hybrid and complex glycans explains why no change in FUT8 activity was predicted by the model for the swainsonine-treated cell lines.

Despite this finding FUT8 does have a wide range of substrate specificities (93, 113, 224). To include all of these into the modelling methodology as parameters subject to the fitting process would have been unfeasible, because of the large number of extra parameters that would have to be introduced and the difficulty in constructing compatible rules in linear notation. More extensive substrate specificities are included in other computational models of glycosylation as K_m values (190, 192) but are not subject to fitting. An unavoidable feature of FUT8 that needed to be included was the ability of FUT8 to fucosylate $Man_5GlcNAc_2$ (figure 3.12, figure 3.13). This reaction was allowed to occur in the fitting of WT cell glycan profiles but only became relevant in fitting the swainsonine-treated cells. Fucosylation of $Man_5GlcNAc_2$ has been observed in MGAT1-deficient cell lines (115) and FUT8 has been identified as the sole enzyme involved in the synthesis of $Fuc_1Man_5GlcNAc_2$ (114). $Fuc_1Man_5GlcNAc_2$ was detected in swainsonine-treated cells (figure 3.11, 3.12), and in cells deficient in MGAT1 (115).

The modelling of swainsonine-treated cell lines described in this thesis does offer an explanation for why $Fuc_1Man_5GlcNAc_2$ is observed. In the fitting of glycan profiles from swainsonine-treated cells, the predicted total activity of MGAT1 in swainsonine-treated cells is reduced along with the activity of MAN2. Therefore, the modelling suggests that $Man_5GlcNAc_2$ is a substrate for FUT8 but that FUT8 is out-competed by MGAT1 and MAN2 in normal cases. Indeed, FUT8 has recently been shown to act on Man_5-$GlcNAc_2$ *in vitro*, given the correct polypeptide context. In the case where MGAT1 activity is perturbed, such as in MGAT1-deficient cells, FUT8 can now compete and act on $Man_5GlcNAc_2$. Why MGAT1 activities are predicted to be reduced in swainsonine-treated cells requires further investigation.

6.4 Semi-quantitative *N*-glycan analysis of Cog4KO HEK293T cells

In order to investigate the role of the COG complex in dictating the *N*-glycosylation of secretory proteins, mass spectrometry was used to glycan profile Cog4KO HEK293T cells (figure 4.5). The glycan profiling of Cog4KO HEK293T cells was undertaken by the author as part of a larger collaboration with Vladimir Lupashin (University of Arkansas for Medical Sciences) and is published (58). As expected, the complete depletion of Cog4 resulted in large changes in the glycan profile compared with that of WT HEK293T cells. The distribution of oligomannose glycans is altered as a result

of Cog4KO; for example, the level of Man$_5$GlcNAc$_2$ is markedly increased in the Cog4KO HEK293T cell line (figure 4.6). This increase in the relative abundance of Man$_5$GlcNAc$_2$ appears to be a common consequence of disrupting the COG complex (58, 60).

In addition to the changes in oligomannose glycan distribution, Cog4KO also resulted in an increase in the total levels of oligomannose glycans and a concomitant decrease in complex-type glycans. Decreases in the functionally important modifications of fucosylation and sialylation are also observed (figure 4.7B, C), suggesting that Cog4 depletion affects both early and later stages of glycosylation. This finding is in contrast to previous suggestions that lobe B, not lobe A subunits are the effectors of the final stages of glycan processing (86). Lobe A and B may have different functions but depletion of either lobe may result in the same outcome (64). Cog3KD HeLa cells displayed abnormal Golgi structure but no change in the levels of GalT and SiaT, while Cog7KD HeLa cells had reduced levels of GalT and SiaT but normal Golgi structure (64). However, in both cell lines terminal Golgi glycosylation was not as efficient as in WT cells. Analysis of the observed glycan profiles revealed a decrease in tri- and tetra-antennary glycans but not bi-antennary glycans suggesting that MGAT2 is unaffected by Cog4KO in HEK293T cells (figure 4.7D).

6.5 Alterations to Golgi glycosylation homeostasis in Golgi trafficking perturbed cell lines

Mutations in the COG complex have been implicated in a number of CDGs (section 1.2.5). As none of the Cog subunits have any known enzymatic activity but have been implicated in the trafficking of glycosylation enzymes (reviewed in Fisher and Ungar, 2016), alterations to glycosylation are believed to be the result of disrupted enzyme homeostasis. As perturbations to the COG complex are likely to affect numerous enzymes (34, 62, 65, 241), it is difficult to identify which reactions are critically impacted. The model developed in this work has been shown to be capable of predicting changes in glycosylation enzyme activities and localizations as a result of Cog4KD in HeLa cells (figure 4.2) and Cog4KO in HEK293T cells (figure 4.9).

Cog4KD in HeLa cells resulted in minor changes to the glycan profile (60). Notable alterations that did occur as a result of Cog4KD centred around the oligomannose glycan distribution (figure 4.1) and led the model to predict a flattening out of the MAN1 distribution (figure 4.4). The shift towards the *trans*-Golgi of the MAN1 enzyme distribution was able to account for the relative increases in $Man_9GlcNAc_2$ and $Man_6GlcNAc_2$. Similarly to the prediction in Cog4KD HeLa cells, MAN1 was shifted in the *trans* direction in the Cog4KO HEK293T cells, albeit to a lesser extent than in the Cog4KD HeLa cells. In the context of cisternal maturation, the flattening out of an enzyme distribution would suggest a defect in retrograde trafficking (figure 6.1). The inability to recycle MAN1 in a retrograde fashion to more immature cisternae would result in a diffusion of the enzyme across the Golgi stack.
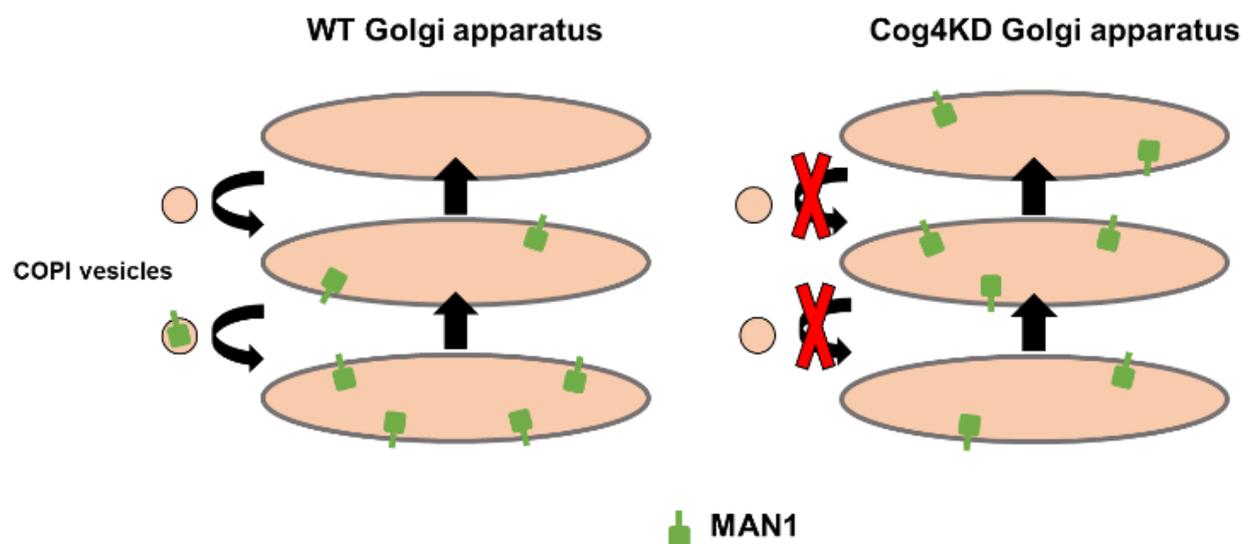
Figure 6.1 | Defective retrograde trafficking in Cog4KD HeLa cells

Proposed model of Golgi cisternal maturation and COPI-vesicle retrograde trafficking of MAN1 in WT and Cog4KD HeLa cells.

MAN1 acts early on in the Golgi glycosylation pathway; it therefore follows that without correct oligomannose glycan processing, the downstream modifications to form complex glycans could potentially be faulty. If individual secreted or cell surface proteins had been modelled instead of the whole cell glycan profile, this alteration to the MAN1 distribution would have been missed because only a subset of glycans end

up on cell surface or secreted proteins and these tend not to be oligomannose glycans. Despite the challenges in modelling whole cell glycan profiles, for making predictions relevant to the mechanistic nature of COG-CDGs for example, modelling whole-cell glycosylation is necessary.

The modelling of Cog4KO HEK293T cells also led to predicted shifts in the localizations of MGAT5 and GalT (figure 4.10). These predicted shifts were not predicted to occur in the Cog4KD HeLa cell line most likely due to the small amount of complex glycans in the HeLa cells. Interestingly, the distribution of GalT and MGAT5 was predicted to be shifted towards the *cis*-Golgi in Cog4KO HEK293T cells. This shift in distribution was validated using confocal microscopy to visualise the localization of GalT-YFP relative to the *cis*-Golgi marker, GM130 (figure 4.11). To visualise the localization of GalT-YFP in the Golgi apparatus, the cells were treated with nocodazole in order to disperse the Golgi ribbon. Nocodazole treatment simplifies the Golgi architecture by disrupting microtubules and generating functional Golgi ministacks. Assessing the localization of glycosylation enzymes within the Golgi apparatus necessitates the use of nocodazole, as cisternae are intertwined and overlap in the Golgi ribbon making analysis difficult.

The shift of GalT towards the *cis*-Golgi as a result of Cog4KO can be explained in a number of ways. Firstly, the shift could be caused by the mis-sorting of GalT into earlier cisternae (figure 6.2). In this scenario, GalT would be trafficked more efficiently from the later, more mature cisternae into early cisternae in the Cog4KO HEK293T cells. This explanation is not consistent with the known role of the COG complex in retrograde trafficking at the Golgi apparatus (242). Defects in the COG complex would be expected to result in disruption to the retrograde trafficking of Golgi-resident proteins, such as GalT, not increased retrograde trafficking.

A second explanation for a *cis* shift in the localization of GalT in Cog4KO cells relative to WT HEK293T cells, is that GalT has been depleted in the *trans* cisternae to a greater extend compared with other cisternae (figure 6.2). In this case, GalT would be packaged into COPI-vesicles but these vesicles would not be delivered to adjacent cisternae due to the depletion of the Cog4. This would implicate lobe A subunits in the trafficking of *trans*-Golgi proteins as previously proposed (58). The overall level of

endogenous GalT is also reduced in Cog4KO HEK293T cells (figure 4.13), suggesting that the GalT in vesicles originating from the *trans*-Golgi are ultimately degraded or secreted by Cog4KO HEK293T cells. In lobe A-deficient CHO cells, the Golgi SNARE protein, GOS-28, has been shown to be rapidly degraded by the proteasome following mis-localization to the ER (62), suggesting a potential outcome of GalT in Cog4KO HEK293T cells.

Finally, the shift towards the *cis*-Golgi in Cog4KO HEK293T cells could be a result of disrupted architecture of the Golgi apparatus (figure 6.3). The Golgi morphology has been shown to be altered in COG-CDG patient cells (52) and in cell lines deficient in COG complex subunits (34, 58). A dilated and fragmented Golgi apparatus would potentially result in the loss of separation between the *trans*- and *cis*-Golgi. While the model does not predict changes in Golgi morphology, a convergence of the Golgi cisternae as a result of Cog4KO is equivalent to a shift from *trans* to *cis*. This explanation would mean that all *trans*-Golgi enzymes should be shifted in the *cis* direction, this is not a prediction made by the model.

Modelling of Cog4KO HEK293T cells predicted reductions in the total activities of MGAT1, MGAT5 and GalT (figure 4.12). The predicted decreases in the activities of these enzymes can be explained by reductions in protein level, and the levels of MGAT1 and GalT were validated with western blotting (figure 4.13). Although not investigated in this work, MAN2 has been shown to be sensitive to the depletion of other lobe A COG subunits (62). The predicted decrease in MGAT5 activity explains the lack of tri- and tetra-antennary glycans observed in the glycan profile of Cog4KO HEK293T cells (figure 4.7D). The decreases in protein amount for these enzymes are most likely explained by trafficking defects, due to the known role of the COG complex. The enzymes are no longer retained within the Golgi apparatus and are degraded as a result of disruption to the COG complex.

The predicted total activity of MGAT1 was decreased in Cog4KO HEK293T compared with WT HEK293T cells (figure 4.12). It was important to validate this prediction (figure 4.13B, D) as MGAT1 catalyses the reaction that converts oligomannose to hybrid and then complex glycans. Oligomannose glycans are likely to be attached to lysosomal hydrolases, while hybrid and complex glycans dominate cell surface glycoproteins

(243). Due to the differing fates of oligomannose glycans in relation to hybrid/complex glycans, the model could effectively use the MGAT1 activity parameter to account for differences in degradation between oligomannose and hybrid/complex glycans in the two cell lines. In other words, if the steady-state level of cell surface proteins was reduced in the Cog4KO HEK293T cells due to alterations in glycosylation, this would manifest itself as a decrease in the relative abundance of complex glycans in a whole cell glycan profile. A total cell membrane proteomics experiment comparing protein levels between the WT and Cog4KO HEK293T cell lines was used as an additional control experiment and to ensure that changes in glycosylation were not the result in changes to the proteome. There was no global reduction in cell surface transmembrane proteins (figure 4.8). For example, the fold-change between Cog4KO and WT cells for the transferrin receptor (four glycosylation sites) was 0.93, while for the insulin receptor (18 glycosylation sites) the fold-change was 0.71.
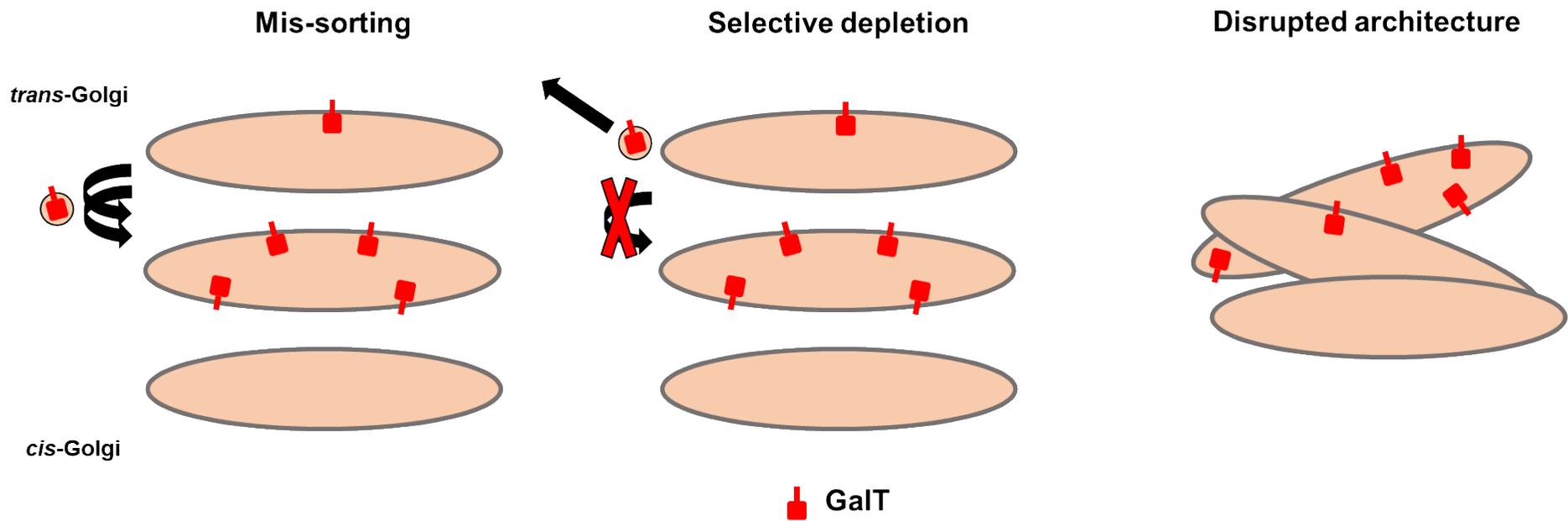
Figure 6.2 | Mis-localization of GalT in Cog4KO HEK293T cells

Potential explanations for the shift towards the cis-Golgi of GalT as a result of Cog4KO in HEK293T cells.

6.6 FUT8 is insensitive to Cog4KO

Despite the reduction in total fucosylation as a result of Cog4KO (figure 4.7), the model did not predict a decrease in the total activity of FUT8 following fitting of the Cog4KO HEK293T cell line (figure 4.12). The overall protein level of FUT8 was shown to be unaltered between WT and Cog4KO HEK293T cells through western blotting (figure 4.13C, D). To investigate the action of FUT8 in Cog4KO HEK293T cells, flux maps were generated *in silico* and the competing reactions for the top six substrates of FUT8 analysed (figure 4.17). The activity of FUT8 was not restricted to a subset of substrates in the model but the predicted top six *N*-glycans acted on by FUT8 (table 4.3) were in line with the preferred substrates, $Man_5GlcNAc_3$ and $Man_3GlcNAc_4$, of FUT8 (93, 244).

The flux map normalized to the total fucosylation flux for each cell line (figure 4.17, table 4.3) demonstrates fucosylation pattern differences between the cell lines. The differences are due to the decrease in enzymatic competition in the Cog4KO HEK293T cell line. The loss of enzymatic competition from MGAT5 and GalT in Cog4KO HEK293T cells results in a relative increase in the fucosylation of $Hex_4HexNAc_4$, compared with WT. Therefore, even though the reduction in total fucosylation was due to a decrease in the levels of complex glycans, restoring the levels of total complex glycan in Cog4KO HEK293T cells would not lead to the same fucosylated glycans observed in the WT HEK293T glycan profile. The ability to identify glycosylation reactions that are affected by mutations in the trafficking machinery has the potential to provide targeted solutions to correct glycosylation in CDG patients. Furthermore, proposed treatments for CDGs can be replicated *in silico* using the developed model and the outcomes assessed, potentially saving time and money.

6.7 Modelling glycan alterations during osteogenesis

Having applied and then demonstrated that a stochastic model of glycosylation with a Bayesian methodology to fit the model to the observed data can be used to predict alterations in the Golgi glycosylation machinery, the model was applied to a cell biological process (osteogenesis). Furthermore, changes to *N*-glycosylation have been shown to influence osteogenesis. For example, inhibition of MAN1 in undifferentiated MSCs promoted osteogenesis, while knockdown of Cog4 in MSCs

prevented osteogenesis (147), although it should be noted that depletion of COG subunits would likely affect both *N-* and O-glycosylation (61).

Comparison of the glycan profiles from Y101 MSCs and osteoblasts derived from them showed a marked decrease in the levels of oligomannose glycans as a result of differentiation (107). Oligomannose glycans are often found on lysosomal hydrolases as a result of the Man6P targeting tag. This pathway is accounted for by the model with the inclusion of the GlcNAc-phosphoT enzyme, the activity of which is predicted to be depressed in the osteoblasts relative to the Y101 MSCs (figure 5.6). Therefore, the levels of oligomannose glycans can be an indirect measure of the lysosomal volume of the cell. Fewer and/or smaller lysosomes would suggest lower levels of lysosomal hydrolases, which in turn would result in lower levels of oligomannose glycans in an observed glycan profile. The levels of LAMP1 (lysosomal marker) and the activity of β-hexosaminidase (lysosomal glycosidase) were increased in osteoblasts relative to MSCs indicating the lysosomal content of osteoblasts was in fact increased above that of MSCs (figure 5.7). This finding is in good agreement with proteomic (233) and microscopic data (234). These findings suggest that the observed decrease in oligomannose glycan in osteoblasts is not related to changes in the lysosomal content of the cell. By considering just the complex glycans in MSCs and osteoblasts (figure 5.9), the significant differences in glycan profiles of the two cell lines are largely diminished. The small changes in complex glycans between MSCs and osteoblasts means there are few changes in the glycosylation machinery as predicted by the model. The differing fates of oligomannose and hybrid/complex glycans is a likely explanation for the changes to the oligomannose/complex glycan ratios between osteoblasts and MSCs. In the case of osteoblasts, the cells produce extracellular matrix (ECM) (containing glycoproteins) which accumulates over the course of the 21-day differentiation protocol, increasing the relative amount of complex glycans compared with oligomannose glycans. Meaningful conclusions regarding the oligomannose/complex glycan ratio cannot therefore be made without additional information, such as the absolute amount of *N*-glycan in each sample.

Despite the abundance of complex glycans increasing in osteoblasts relative to MSCs, the predicted total activity of MGAT5 was reduced in osteoblasts (figure 5.6). Excluding the oligomannose glycans, a switch from tri-antennary to bi-antennary glycans upon

osteogenesis can be observed (figure 5.9B), a feature that may not have been noticed had the model not predicted a decrease in the activity of MGAT5. MGAT5-mediated *N*-glycan antenna number has been intimately linked to galectin binding (105). Galectins are a family of small soluble proteins capable of binding galactose-containing moieties and can aggregate with ligands resulting in galectin oligomerization (245), which is important for inducing signalling pathways. Once in the extracellular space, galectins may either promote endocytosis of cell-surface proteins (246) or inhibit endocytic uptake of cell-surface proteins (105, 247). Altered receptor endocytosis has been shown to impact cell proliferation and differentiation (133). This is due to an intricate balance between proliferation and differentiation that is maintained by UDP-GlcNAc levels influencing antenna number (133).

Inhibiting galectin interactions using lactose in the first week of a three-week osteogenesis experiment promoted osteogenesis when compared with a control (sucrose) treatment (figure 5.10). This finding is consistent with previous work demonstrating that galectin-3 downregulates the expression of osteogenic markers and inhibits osteogenesis (235). Furthermore, MSCs pre-treated with kifunensine, which completely abrogates MGAT5-mediated branching, promoted osteogenesis (147). The increase in osteogenic potential due to kifunensine pre-treatment of MSCs agrees with the results of the inhibition of galectin binding in the first week of the 21-day osteogenesis protocol and suggests an existence of an early switch from proliferation to differentiation, influenced by glycosylation. These findings suggest that by switching from tri- to bi-antennary glycans, an inhibitory effect on osteogenesis is released.

6.8 Future work

In terms of the modelling framework, the introduction of glycan competition would be a logical addition. The inclusion of glycan competition in the modelling framework would help identify changes in secretory protein flux, which is of particular importance for overexpression experiments. This is an important consideration when producing biotherapeutics, the glycosylation of which could also be modelled in the future. As a goal of cellular engineering is to increase production of biologics from cells, the effect this increased flux through the secretory pathway has on the glycosylation process

could be predicted using a modelling framework incorporating glycan competition. Previous work has already shown the importance of secretory protein flux on the glycosylation process (248). Alongside modelling the glycan profiles of biologics, hypothetical glycan profiles expressing the desired glycoforms for specific biologics could also be modelled, in order to predict manipulations to the glycosylation machinery that could achieve the desired glycosylation.

The developed model could be used to gain much deeper insight into COG-dependent Golgi trafficking. Instead of depleting COG subunits, by studying COG mutants with disrupted interactions with other Golgi trafficking players the role in glycosylation of individual interactions may be assessed. By modelling the glycosylation of these COG-mutant cell lines, binding partners can be identified that are primarily responsible for the localization of specific enzymes. Additionally, the glycosylation of cells expressing CDG-causing COG mutations could also be modelled. Predicting alterations to the glycosylation machinery in COG-CDG mutation-expressing cells has the ability to propose targeted therapies, made even more accurate through the analysis of flux maps.

In regard to chapter 5, further work is required to fully understand the role of glycosylation in determining cell fate. The development and characterization of a MGAT5KO MSC would demonstrate the role of MGAT5 in dictating osteogenic potential. Identifying the key receptor(s) that may be affected by *N*-glycan branching and galectin binding is difficult, although it has been suggested that inhibition of osteogenesis by galectins occurs through the Notch signally pathway (235). Investigating the glycosylation of the Notch receptors would be a good starting point. MSCs also have the ability to differentiate into adipocytes; glycan profiling and the modelling of the glycosylation into adipocytes could therefore also provide interesting avenues of investigation. The modelling of whole-cell glycan profiles can be seen as a process of discovery, generating hypotheses that wait to be tested and explored.

Abbreviations

ABC = approximate Bayesian computation

ADCC = antibody-dependent cell-mediated cytotoxicity

ALG = asparagine-linked glycosylation

ALP = alkaline phosphatase

Ant FUT = antenna fucosylation

BCA = bicinchoninic acid

CATCHR = complexes associated with tethering containing helical rods

CDG = congenital disorders of glycosylation

CHO = Chinese hamster ovary

COG = conserved oligomeric Golgi

COPI = coatomer protein complex-I

DHB = 2, 6-dihydroxybenzonic acid

DMSO = dimethyl sulfoxide

DTT = dithiothreitol

ECM = extracellular matrix

EDTA = ethylenediaminetetraacetic acid

Endo-MAN = glycoprotein endo-alpha-1,2-mannosidase

ER = endoplasmic reticulum

ERGIC = endoplasmic reticulum Golgi intermediate complex

FTICR = Fourier-transform ion cyclotron resonance

Fuc = fucose

FUT8 = fucosyltransferase 8

GalT = galactosyltransferase

GlcNAc = *N*-acetylglucosamine

GlcNAc-PhosphoT = *N*-acetylglucosamine-1-phosphate transferase

Glc = glucose

Gal = galactose

GRASP = Golgi reassembly stacking protein

HEK = human embryonic kidney

HPLC = high performance liquid chromatography

hTERT = human telomerase reverse transcriptase

KD = knockdown

KO = knockout

LDL = low-density lipoprotein

MALDI = matrix-assisted laser desorption

Man = mannose

MAN1 = mannosyl-oligosaccharide 1,2-alpha-mannosidase 1

MAN2 = alpha-mannosidase 2

MCMC = Markov chain Monte Carlo

MGAT1 = alpha-1,3-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase

MGAT2 = alpha-1,6-mannosyl-glycoprotein 2-beta-*N*-acetylglucosaminyltransferase

MGAT4 = alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase

MGAT5 = alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase

MLE = maximum likelihood estimation

MS = mass spectrometry

MSC = mesenchymal stromal cells

NeuAc = *N*-acetylneuraminic acid

PBS = phosphate buffered saline

PNGase F = peptide *N*-glycosidase F

PVDF = polyvinylidene fluoride

RF = radiofrequency

SDS = sodium dodecyl sulfate

SEM = standard error of the mean

SiaT = sialyltransferase

SNARE = soluble *N*-ethylmaleimide sensitive factor attachment protein receptor

SSA = stochastic simulation algorithm

TGN = *trans* Golgi network

VSV-G = vesicular stomatitis virus G protein

WT = wild-type

List of references

1.   Klumperman J (2011) Architecture of the mammalian Golgi. *Cold Spring Harb Perspect Biol* 3(7). doi:10.1101/cshperspect.a005181.

2.   Xiang Y, Wang Y (2010) GRASP55 and GRASP65 play complementary and essential roles in Golgi cisternal stacking. *J Cell Biol* 188(2).

3.   Derganc J, Mironov AA, Svetina S (2006) Physical Factors that Affect the Number and Size of Golgi Cisternae. *Traffic* 7(1):85–96.

4.   Trucco A, et al. (2004) Secretory traffic triggers the formation of tubular continuities across Golgi sub-compartments. *Nat Cell Biol* 6(11):1071–1081.

5.   Wei J-H, Seemann J (2010) Unraveling the Golgi ribbon. *Traffic* 11(11):1391–400.

6.   Lucocq JM, Warren G (1987) Fragmentation and partitioning of the Golgi apparatus during mitosis in HeLa cells. *EMBO J* 6(11):3239–46.

7.   Minin AA (1997) Dispersal of Golgi apparatus in nocodazole-treated fibroblasts is a kinesin-driven process. *J Cell Sci* 110 (Pt 19):2495–505.

8.   Losev E, et al. (2006) Golgi maturation visualized in living yeast. *Nature* 441(7096):1002–6.

9.   Glick BS, Nakano A (2009) Membrane traffic within the Golgi apparatus. *Annu Rev Cell Dev Biol* 25:113–32.

10.   Tang D, et al. (2015) Mena-GRASP65 interaction couples actin polymerization to Golgi ribbon linking. *Mol Biol Cell*. doi:10.1091/mbc.E15-09-0650.

11.   Franzusoff A, Redding K, Crosby J, Fuller RS, Schekman R (1991) Localization of components involved in protein transport and processing through the yeast Golgi apparatus. *J Cell Biol* 112(1):27–37.

12.   Kondylis V, Rabouille C (2009) The Golgi apparatus: Lessons from *Drosophila*. *FEBS Lett* 583(23):3827–3838.

13.   Dupree P, Sherrier DJ (1998) The plant Golgi apparatus. *Biochim Biophys Acta - Mol Cell Res* 1404(1–2):259–270.

14.   Mironov AA, Sesorova IS, Seliverstova E V., Beznoussenko G V. (2017) Different Golgi ultrastructure across species and tissues: Implications under functional and pathological conditions, and an attempt at classification. *Tissue Cell* 49(2):186–201.

15. Joshi G, Chi Y, Huang Z, Wang Y (2014) Aβ-induced Golgi fragmentation in Alzheimer's disease enhances Aβ production. *Proc Natl Acad Sci U S A* 111(13):E1230-9.

16. Dunlop MH, et al. (2017) Land-locked mammalian Golgi reveals cargo transport between stable cisternae. *Nat Commun* 8(1):432.

17. Dunphy WG, Rothman JE (1983) Compartmentation of asparagine-linked oligosaccharide processing in the Golgi apparatus. *J Cell Biol* 97(1). Available at: http://jcb.rupress.org/content/97/1/270?ijkey=6f33e23928167e1000c7f3f3563d 42ebc466660a&keytype2=tf_ipsecsha [Accessed September 14, 2017].

18. Morré DJ, Ovtracht L (1977) Dynamics of the Golgi apparatus: membrane differentiation and membrane flow. *Int Rev Cytol Suppl* (5):61–188.

19. Beznoussenko G V, et al. (2014) Transport of soluble proteins through the Golgi occurs by diffusion via continuities across cisternae. *Elife* 3. doi:10.7554/eLife.02009.

20. Patterson GH, et al. (2008) Transport through the Golgi apparatus by rapid partitioning within a two-phase membrane system. *Cell* 133(6):1055–67.

21. Chen Y, Gershlick DC, Park SY, Bonifacino JS (2017) Segregation in the Golgi complex precedes export of endolysosomal proteins in distinct transport carriers. *J Cell Biol* 216(12):4141–4151.

22. Rabouille C, et al. (1995) Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *J Cell Sci* 108 (Pt 4:1617–27.

23. Martinez-Menárguez JA, et al. (2001) Peri-Golgi vesicles contain retrograde but not anterograde proteins consistent with the cisternal progression model of intra-Golgi transport. *J Cell Biol* 155(7):1213–24.

24. Bonfanti L, et al. (1998) Procollagen traverses the Golgi stack without leaving the lumen of cisternae: evidence for cisternal maturation. *Cell* 95(7):993–1003.

25. Glick BS, Elston T, Oster G (1997) A cisternal maturation mechanism can explain the asymmetry of the Golgi stack. *FEBS Lett* 414(2):177–81.

26. Fisher P, Ungar D (2016) Bridging the Gap between Glycosylation and Vesicle Traffic. *Front Cell Dev Biol* 4. doi:10.3389/fcell.2016.00015.

27. Whyte JR, Munro S (2001) The Sec34/35 Golgi transport complex is related to the exocyst, defining a family of complexes involved in multiple steps of membrane traffic. *Dev Cell* 1(4):527–37.

28. Yu I-M, Hughson FM (2010) Tethering factors as organizers of intracellular vesicular traffic. *Annu Rev Cell Dev Biol* 26:137–56.

29. Ungar D, et al. (2002) Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. *J Cell Biol* 157(3):405–15.

30. Ungar D, Oka T, Vasile E, Krieger M, Hughson FM (2005) Subunit architecture of the conserved oligomeric Golgi complex. *J Biol Chem* 280(38):32729–35.

31. Lees JA, Yip CK, Walz T, Hughson FM (2010) Molecular organization of the COG vesicle tethering complex. *Nat Struct Mol Biol* 17(11):1292–7.

32. Ha JY, et al. (2016) Molecular architecture of the complete COG tethering complex. *Nat Struct Mol Biol*. doi:10.1038/nsmb.3263.

33. Ha JY, et al. (2014) Cog5-Cog7 crystal structure reveals interactions essential for the function of a multisubunit tethering complex. *Proc Natl Acad Sci U S A* 111(44):15762–7.

34. Zolov SN, Lupashin V V (2005) Cog3p depletion blocks vesicle-mediated Golgi retrograde trafficking in HeLa cells. *J Cell Biol* 168(5):747–59.

35. Willett R, et al. (2013) COG complexes form spatial landmarks for distinct SNARE complexes. *Nat Commun* 4:1553.

36. Malsam J, Satoh A, Pelletier L, Warren G (2005) Golgin tethers define subpopulations of COPI vesicles. *Science* 307(5712):1095–8.

37. Wong M, Munro S (2014) Membrane trafficking. The specificity of vesicle traffic to the Golgi is encoded in the golgin coiled-coil proteins. *Science* 346(6209):1256898.

38. Munro S (2005) The Golgi apparatus: defining the identity of Golgi membranes. *Curr Opin Cell Biol* 17(4):395–401.

39. Barr FA (2009) Rab GTPase function in Golgi trafficking. *Semin Cell Dev Biol* 20(7):780–783.

40. Bonifacino JS, Glick BS (2004) The mechanisms of vesicle budding and fusion. *Cell* 116(2):153–66.

41. Söllner T, et al. (1993) SNAP receptors implicated in vesicle targeting and fusion. *Nature* 362(6418):318–324.

42. Pobbati A V., Stein A, Fasshauer D (2006) N- to C-Terminal SNARE Complex Assembly Promotes Rapid Membrane Fusion. *Science (80-)* 313(5787):673–676.

43.   Melia TJ, et al. (2002) Regulation of membrane fusion by the membrane-proximal coil of the t-SNARE during zippering of SNAREpins. *J Cell Biol* 158(5):929–940.

44.   Willett R, et al. (2016) COG lobe B sub-complex engages v-SNARE GS15 and functions via regulated interaction with lobe A sub-complex. *Sci Rep* 6:29139.

45.   Miller VJ, et al. (2013) Molecular insights into vesicle tethering at the Golgi by the conserved oligomeric Golgi (COG) complex and the golgin TATA element modulatory factor (TMF). *J Biol Chem* 288(6):4229–40.

46.   Sohda M, et al. (2007) The interaction of two tethering factors, p115 and COG complex, is required for Golgi integrity. *Traffic* 8(3):270–84.

47.   Sohda M, et al. (2010) Interaction of Golgin-84 with the COG complex mediates the intra-Golgi retrograde transport. *Traffic* 11(12):1552–66.

48.   Shestakova A, Suvorova E, Pavliv O, Khaidakova G, Lupashin V (2007) Interaction of the conserved oligomeric Golgi complex with t-SNARE Syntaxin5a/Sed5 enhances intra-Golgi SNARE complex stability. *J Cell Biol* 179(6):1179–1192.

49.   Laufman O, Hong W, Lev S (2013) The COG complex interacts with multiple Golgi SNAREs and enhances fusogenic assembly of SNARE complexes. *J Cell Sci* 126(Pt 6):1506–16.

50.   Laufman O, Kedan A, Hong W, Lev S (2009) Direct interaction between the COG complex and the SM protein, Sly1, is required for Golgi SNARE pairing. *EMBO J* 28(14):2006–17.

51.   Fukuda M, Kanno E, Ishibashi K, Itoh T (2008) Large scale screening for novel rab effectors reveals unexpected broad Rab binding specificity. *Mol Cell Proteomics* 7(6):1031–42.

52.   Kudlyk T, Willett R, Pokrovskaya ID, Lupashin V (2013) COG6 interacts with a subset of the Golgi SNAREs and is important for the Golgi complex integrity. *Traffic* 14(2):194–204.

53.   Farkas RM, Giansanti MG, Gatti M, Fuller MT (2003) The Drosophila Cog5 homologue is required for cytokinesis, cell elongation, and assembly of specialized Golgi architecture during spermatogenesis. *Mol Biol Cell* 14(1):190–200.

54.   Schnorrer F, et al. (2010) Systematic genetic analysis of muscle morphogenesis and function in Drosophila. *Nature* 464(7286):287–91.

55. Tan X, et al. (2016) Arabidopsis COG Complex Subunits COG3 and COG8 Modulate Golgi Morphology, Vesicle Trafficking Homeostasis and Are Essential for Pollen Tube Growth. *PLOS Genet* 12(7):e1006140.

56. Shestakova A, Zolov S, Lupashin V (2006) COG complex-mediated recycling of Golgi glycosyltransferases is essential for normal protein glycosylation. *Traffic* 7(2):191–204.

57. Pokrovskaya ID, et al. (2011) Conserved oligomeric Golgi complex specifically regulates the maintenance of Golgi glycosylation machinery. *Glycobiology* 21(12):1554–69.

58. Bailey Blackburn J, Pokrovskaya I, Fisher P, Ungar D, Lupashin V V (2016) COG Complex Complexities: Detailed Characterization of a Complete Set of HEK293T Cells Lacking Individual COG Subunits. *Front cell Dev Biol* 4:23.

59. Richardson BC, et al. (2009) Structural basis for a human glycosylation disorder caused by mutation of the COG4 gene. *Proc Natl Acad Sci U S A* 106(32):13329–34.

60. Abdul Rahman S, et al. (2014) Filter-aided N-glycan separation (FANGS): a convenient sample preparation method for mass spectrometric N-glycan profiling. *J Proteome Res* 13(3):1167–76.

61. Skeene K, et al. (2017) One Filter, One Sample, and the *N* - and O-Glyco(proteo)me: Toward a System to Study Disorders of Protein Glycosylation. *Anal Chem* 89(11):5840–5849.

62. Oka T, Ungar D, Hughson FM, Krieger M (2004) The COG and COPI complexes interact to control the abundance of GEARs, a subset of Golgi integral membrane proteins. *Mol Biol Cell* 15(5):2423–35.

63. Kingsley DM, Kozarsky KF, Segal M, Krieger M (1986) Three types of low density lipoprotein receptor-deficient mutant have pleiotropic defects in the synthesis of N-linked, O-linked, and lipid-linked carbohydrate chains. *J Cell Biol* 102(5):1576–85.

64. Peanne R, et al. (2010) Differential effects of lobe A and lobe B of the Conserved Oligomeric Golgi complex on the stability of 1,4-galactosyltransferase 1 and 2,6-sialyltransferase 1. *Glycobiology* 21(7):864–876.

65. Blackburn JB, Kudlyk T, Pokrovskaya I, Lupashin V V. (2018) More than just sugars: COG complex deficiency causes glycosylation-independent cellular

defects. *Traffic*. doi:10.1111/tra.12564.

66. Hansen L, et al. (2015) A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology* 25(2):211–24.

67. Hiraoka S, et al. (2007) Nucleotide-sugar transporter SLC35D1 is critical to chondroitin sulfate synthesis in cartilage and skeletal development in mouse and human. *Nat Med* 13(11):1363–1367.

68. Mohamed M, et al. (2013) Intellectual disability and bleeding diathesis due to deficient CMP--sialic acid transport. *Neurology* 81(7):681–7.

69. Charuk JH, et al. (1995) Carbohydrate-deficient glycoprotein syndrome type II. An autosomal recessive N-acetylglucosaminyltransferase II deficiency different from typical hereditary erythroblastic multinuclearity, with a positive acidified-serum lysis test (HEMPAS). *Eur J Biochem* 230(2):797–805.

70. Foulquier F, et al. (2006) Conserved oligomeric Golgi complex subunit 1 deficiency reveals a previously uncharacterized congenital disorder of glycosylation type II. *Proc Natl Acad Sci U S A* 103(10):3764–9.

71. Kodera H, et al. (2015) Mutations in COG2 encoding a subunit of the conserved oligomeric golgi complex cause a congenital disorder of glycosylation. *Clin Genet* 87(5):455–60.

72. Reynders E, et al. (2009) Golgi function and dysfunction in the first COG4-deficient CDG type II patient. *Hum Mol Genet* 18(17):3244–56.

73. Ng BG, et al. (2011) Identification of the first COG-CDG patient of Indian origin. *Mol Genet Metab* 102(3):364–7.

74. Paesold-Burda P, et al. (2009) Deficiency in COG5 causes a moderate form of congenital disorders of glycosylation. *Hum Mol Genet* 18(22):4350–6.

75. Lübbehusen J, et al. (2010) Fatal outcome due to deficiency of subunit 6 of the conserved oligomeric Golgi complex leading to a new type of congenital disorders of glycosylation. *Hum Mol Genet* 19(18):3623–33.

76. Huybrechts S, et al. (2012) Deficiency of Subunit 6 of the Conserved Oligomeric Golgi Complex (COG6-CDG): Second Patient, Different Phenotype. *JIMD Rep* 4:103–8.

77. Rymen D, et al. (2015) Key features and clinical variability of COG6-CDG. *Mol Genet Metab* 116(3):163–70.

78. Wu X, et al. (2004) Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nat Med* 10(5):518–23.

79. Morava E, et al. (2007) A common mutation in the COG7 gene with a consistent phenotype including microcephaly, adducted thumbs, growth retardation, VSD and episodes of hyperthermia. *Eur J Hum Genet* 15(6):638–645.

80. Zeevaert R, et al. (2009) A new mutation in COG7 extends the spectrum of COG subunit deficiencies. *Eur J Med Genet* 52(5):303–5.

81. Kranz C, et al. (2007) COG8 deficiency causes new congenital disorder of glycosylation type IIh. *Hum Mol Genet* 16(7):731–41.

82. Foulquier F, et al. (2007) A new inborn error of glycosylation due to a Cog8 deficiency reveals a critical role for the Cog1-Cog8 interaction in COG complex formation. *Hum Mol Genet* 16(7):717–30.

83. Oka T, et al. (2005) Genetic analysis of the subunit organization and function of the conserved oligomeric golgi (COG) complex: studies of COG5- and COG7-deficient mammalian cells. *J Biol Chem* 280(38):32736–45.

84. Rymen D, et al. (2012) COG5-CDG: expanding the clinical spectrum. *Orphanet J Rare Dis* 7:94.

85. Steet R, Kornfeld S (2006) COG-7-deficient Human Fibroblasts Exhibit Altered Recycling of Golgi Proteins. *Mol Biol Cell* 17(5):2312–21.

86. Laufman O, Freeze HH, Hong W, Lev S (2013) Deficiency of the Cog8 subunit in normal and CDG-derived cells impairs the assembly of the COG and Golgi SNARE complexes. *Traffic* 14(10):1065–77.

87. Apweiler R (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta - Gen Subj* 1473(1):4–8.

88. Waetzig GH, et al. (2010) N-linked glycosylation is essential for the stability but not the signaling function of the interleukin-6 signal transducer glycoprotein 130. *J Biol Chem* 285(3):1781–9.

89. Shields RL, et al. (2002) Lack of Fucose on Human IgG1 *N* -Linked Oligosaccharide Improves Binding to Human FcγRIII and Antibody-dependent Cellular Toxicity. *J Biol Chem* 277(30):26733–26740.

90. Kornfeld S, Li E, Tabas I (1978) The synthesis of complex-type oligosaccharides. II. Characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein. *J Biol Chem* 253(21):7771–8.

91. Spahn PN, et al. (2016) A Markov chain model for N-linked protein glycosylation - towards a low-parameter tool for model-driven glycoengineering. *Metab Eng* 33:52–66.

92. Ramasamy V, et al. (2005) Oligosaccharide Preferences of β1,4-Galactosyltransferase-I: Crystal Structures of Met340His Mutant of Human β1,4-Galactosyltransferase-I with a Pentasaccharide and Trisaccharides of the N-Glycan Moiety. *J Mol Biol* 353(1):53–67.

93. Calderon AD, et al. (2016) Substrate specificity of FUT8 and chemoenzymatic synthesis of core-fucosylated asymmetric N-glycans. *Org Biomol Chem* 14(17):4027–4031.

94. Gonzalez DS, Karaveg K, Vandersall-Nairn AS, Lal A, Moremen KW (1999) Identification, expression, and characterization of a cDNA encoding human endoplasmic reticulum mannosidase I, the enzyme that catalyzes the first mannose trimming step in mammalian Asn-linked oligosaccharide biosynthesis. *J Biol Chem* 274(30):21375–86.

95. Lal A, et al. (1998) Substrate specificities of recombinant murine Golgi alpha1, 2-mannosidases IA and IB and comparison with endoplasmic reticulum and Golgi processing alpha1,2-mannosidases. *Glycobiology* 8(10):981–95.

96. Herscovics A, Schneikert J, Athanassiadis A, Moremen KW (1994) Isolation of a mouse Golgi mannosidase cDNA, a member of a gene family conserved from yeast to mammals. *J Biol Chem* 269(13):9864–71.

97. Tremblay LO, Herscovics A (2000) Characterization of a cDNA Encoding a Novel Human Golgi α1,2-Mannosidase (IC) Involved in *N* -Glycan Biosynthesis. *J Biol Chem* 275(41):31655–31660.

98. Xiang Y, Karaveg K, Moremen KW (2016) Substrate recognition and catalysis by GH47 α-mannosidases involved in Asn-linked glycan maturation in the mammalian secretory pathway. *Proc Natl Acad Sci U S A* 113(49):E7890–E7899.

99. Bause E, Breuer W, Schweden J, Roeser R, Geyer R (1992) Effect of substrate structure on the activity of Man9-mannosidase from pig liver involved in N-linked oligosaccharide processing. *Eur J Biochem* 208(2):451–7.

100. Fujiyama K, et al. (2001) Human N-acetylglucosaminyltransferase I. Expression in Escherichia coli as a soluble enzyme, and application as an immobilized enzyme for the chemoenzymatic synthesis of N-linked

oligosaccharides. *J Biosci Bioeng* 92(6):569–74.

101. Ioffe E, Stanley P (1994) Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc Natl Acad Sci U S A* 91(2):728–32.

102. Bendiak B, Schachter H (1987) THE JOURNAL OF BIOLOGICAL CHEMISTRY Control of Glycoprotein Synthesis PURIFICATION OF UDP-N-ACETYLGLUCOSAMINECX-D-MANNOSIDE p1-2 N-ACETYLGLUCOSAMINYLTRANSFERASE I1 FROM RAT LIVER*. 262(12):5175–5783.

103. Oguri S, Yoshida A, Minowa MT, Takeuchi M (2006) Kinetic properties and substrate specificities of two recombinant human N-acetylglucosaminyltransferase-IV isozymes. *Glycoconj J* 23(7–8):473–480.

104. Nihon Seikagakkai. J, et al. (1993) *The Journal of biochemistry.* (Japanese Biochemical Society) Available at: https://www.jstage.jst.go.jp/article/biochemistry1922/113/5/113_5_614/_article [Accessed May 25, 2017].

105. Demetriou M, Granovsky M, Quaggin S, Dennis JW (2001) Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature* 409(6821):733–739.

106. Elbein AD, Solf R, Dorling PR, Vosbeck K (1981) Swainsonine: an inhibitor of glycoprotein processing. *Proc Natl Acad Sci U S A* 78(12):7393–7.

107. Wilson KM, Thomas-Oates JE, Genever PG, Ungar D (2016) Glycan Profiling Shows Unvaried N-Glycomes in MSC Clones with Distinct Differentiation Potentials. *Front Cell Dev Biol* 4. doi:10.3389/fcell.2016.00052.

108. Powell JT (1980) Purification and properties of lung lectin. Rat lung and human lung beta-galactoside-binding proteins. *Biochem J* 187(1):123–9.

109. Harduin-Lepers A, et al. (2001) The human sialyltransferase family. *Biochimie* 83(8):727–737.

110. Oates JA, Wood AJJ, Dwyer JM (1992) Manipulating the Immune System with Immune Globulin. *N Engl J Med* 326(2):107–116.

111. Barb AW, Brady EK, Prestegard JH (2009) Branch-specific sialylation of IgG-Fc glycans by ST6Gal-I. *Biochemistry* 48(41):9705–7.

112. Breen KC (2002) The role of protein glycosylation in the control of cellular *N* - sialyltransferase activity. *FEBS Lett* 517(1–3):215–218.

113. Tseng T-H, et al. (2017) Substrate Preference and Interplay of Fucosyltransferase 8 and *N* -Acetylglucosaminyltransferases. *J Am Chem Soc* 139(28):9431–9434.

114. Yang Q, Wang L-X (2016) Mammalian α1,6-fucosyltransferase (FUT8) is the sole enzyme responsible for the N-acetylglucosaminyltransferase I-independent core fucosylation of high-mannose N-glycans. *J Biol Chem*. doi:10.1074/jbc.M116.720789.

115. Crispin M, et al. (2006) Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology* 16(8):748–56.

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

117. Chen X, Flynn GC (2007) Analysis of N-glycans from recombinant immunoglobulin G by on-line reversed-phase high-performance liquid chromatography/mass spectrometry. *Anal Biochem* 370(2):147–161.

118. Nairn A V, et al. (2008) Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J Biol Chem* 283(25):17298–313.

119. Nairn A V, et al. (2012) Regulation of glycan structures in murine embryonic stem cells: combined transcript profiling of glycan-related genes and glycan structural analysis. *J Biol Chem* 287(45):37835–56.

120. Ohtsubo K, Marth JD (2006) Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126(5):855–867.

121. Varki Ajit (2017) Biological roles of glycans. *Glycobiology* 27(1):3,49.

122. Gibson R, Schlesinger S, Kornfeld S (1979) The nonglycosylated glycoprotein of vesicular stomatitis virus is temperature-sensitive and undergoes intracellular aggregation at elevated temperatures. *J Biol Chem* 254(9):3600–7.

123. Sousa M, Parodi AJ (1995) The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *EMBO J* 14(17):4196–203.

124. Hammond C, Braakman I, Helenius A (1994) Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc Natl Acad Sci U S A* 91(3):913–7.

125. Rajagopalan S, Xu Y, Brenner MB (1994) Retention of unassembled components of integral membrane proteins by calnexin. *Science* 263(5145):387–90.

126. Wang X, et al. (2005) Dysregulation of TGF-beta1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice. *Proc Natl Acad Sci U S A* 102(44):15791–6.

127. Liu Y-C, et al. (2011) Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells. *Proc Natl Acad Sci* 108(28):11332–11337.

128. Dube DH, Bertozzi CR (2005) Glycans in cancer and inflammation — potential for therapeutics and diagnostics. *Nat Rev Drug Discov* 4(6):477–488.

129. Altevogt P, et al. (1983) Different patterns of lectin binding and cell surface sialylation detected on related high- and low-metastatic tumor lines. *Cancer Res* 43(11):5138–44.

130. Yogeeswaran G, Salk PL (1981) Metastatic potential is positively correlated with cell surface sialylation of cultured murine tumor cell lines. *Science* 212(4502):1514–6.

131. Rabinovich GA, Toscano MA, Jackson SS, Vasta GR (2007) Functions of cell surface galectin-glycoprotein lattices. *Curr Opin Struct Biol* 17(5):513–520.

132. Lagana A, et al. (2006) Galectin binding to Mgat5-modified N-glycans regulates fibronectin matrix remodeling in tumor cells. *Mol Cell Biol* 26(8):3181–93.

133. Lau KS, et al. (2007) Complex N-Glycan Number and Degree of Branching Cooperate to Regulate Cell Proliferation and Differentiation. *Cell* 129(1):123–134.

134. Skropeta D (2009) The effect of individual N-glycans on enzyme activity. *Bioorg Med Chem* 17(7):2645–2653.

135. Berg-Fussman A, Grace ME, Ioannou Y, Grabowski GA (1993) Human acid beta-glucosidase. N-glycosylation site occupancy and the effect of glycosylation on enzymatic activity. *J Biol Chem* 268(20):14861–6.

136. Kato T, Suzuki M, Murata T, Park EY (2005) The effects of N-glycosylation sites and the N-terminal region on the biological function of β1,3-N-acetylglucosaminyltransferase 2 and its secretion. *Biochem Biophys Res Commun* 329(2):699–705.

137. Bieberich E, Tencomnao T, Kapitonov D, Yu RK (2000) Effect of N-glycosylation on turnover and subcellular distribution of N-acetylgalactosaminyltransferase I and sialyltransferase II in neuroblastoma cells. *J Neurochem* 74(6):2359–64.

138. Liwosz A, Lei T, Kukuruzinska MA (2006) *N* -Glycosylation Affects the Molecular Organization and Stability of E-cadherin Junctions. *J Biol Chem* 281(32):23138–23149.

139. Pinho SS, et al. (2013) E-cadherin and adherens-junctions stability in gastric carcinoma: Functional implications of glycosyltransferases involving N-glycan branching biosynthesis, N-acetylglucosaminyltransferases III and V. *Biochim Biophys Acta - Gen Subj* 1830(3):2690–2700.

140. Granovsky M, et al. (2000) Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nat Med* 6(3):306–312.

141. Hall M, et al. (2016) Predominant Expression of Hybrid N-Glycans Has Distinct Cellular Roles Relative to Complex and Oligomannose N-Glycans. *Int J Mol Sci* 17(6):925.

142. Hamouda H, et al. (2013) N-Glycosylation Profile of Undifferentiated and Adipogenically Differentiated Human Bone Marrow Mesenchymal Stem Cells: Towards a Next Generation of Stem Cell Markers. *Stem Cells Dev* 22(23):3100–3113.

143. Hasehira K, Hirabayashi J, Tateno H (2017) Structural and quantitative evidence of α2-6-sialylated N-glycans as markers of the differentiation potential of human mesenchymal stem cells. *Glycoconj J* 34(6):797–806.

144. Tateno H, et al. (2016) α2–6 sialylation is a marker of the differentiation potential of human mesenchymal stem cells. *Glycobiology* 26(12):cww039.

145. Komekado H, Yamamoto H, Chiba T, Kikuchi A (2007) Glycosylation and palmitoylation of Wnt-3a are coupled to produce an active form of Wnt-3a. *Genes to Cells* 12(4):521–534.

146. Saremba S, et al. (2008) Type I receptor binding of bone morphogenetic protein   6 is dependent on N-glycosylation of the ligand. *FEBS J* 275(1):172–183.

147. Wilson KM, et al. (2018) Glycans modify mesenchymal stem cell differentiation to impact the function of resulting osteoblasts. Available at: http://jcs.biologists.org/content/joces/early/2018/01/09/jcs.209452.full.pdf

[Accessed January 22, 2018].

148. James S, et al. (2015) Multiparameter Analysis of Human Bone Marrow Stromal Cells Identifies Distinct Immunomodulatory and Differentiation-Competent Subtypes. *Stem cell reports* 4(6):1004–15.

149. Planinc A, Bones J, Dejaegher B, Van Antwerpen P, Delporte C (2016) Glycan characterization of biopharmaceuticals: Updates and perspectives. *Anal Chim Acta* 921:13–27.

150. Yamane-Ohnuki N, et al. (2004) Establishment ofFUT8 knockout Chinese hamster ovary cells: An ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol Bioeng* 87(5):614–622.

151. Shinkawa T, et al. (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. *J Biol Chem* 278(5):3466–3473.

152. Kanda Y, et al. (2007) Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: the high-mannose, hybrid, and complex types. *Glycobiology* 17(1):104–118.

153. Karas M, Bachmann D, Hillenkamp F (1985) Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal Chem* 57(14):2935–2939.

154. Karas M, Bachmann D, Bahr U, Hillenkamp F (1987) Matrix-assisted ultraviolet laser desorption of non-volatile compounds. *Int J Mass Spectrom Ion Process* 78:53–68.

155. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60(20):2299–301.

156. Tanaka K, et al. (1988) Protein and polymer analyses up tom/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 2(8):151–153.

157. Strupat K, Karas M, Hillenkamp F (1991) 2,5-Dihydroxybenzoic acid: a new matrix for laser desorption—ionization mass spectrometry. *Int J Mass Spectrom Ion Process* 111:89–102.

158. Zenobi R, Knochenmuss R (1998) Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev* 17(5):337–366.

159. Karas M, Gl�ckmann M, Sch�fer J (2000) Ionization in matrix-assisted laser desorption/ionization: singly charged molecular ions are the lucky survivors. *J Mass Spectrom* 35(1):1–12.

160. Ehring H, Karas M, Hillenkamp F (1992) Role of photoionization and photochemistry in ionization processes of organic molecules and relevance for matrix-assisted laser desorption Ionization mass spectrometry. *Org Mass Spectrom* 27(4):472–480.

161. Jaskolla TW, Karas M (2011) Compelling Evidence for Lucky Survivor and Gas Phase Protonation: The Unified MALDI Analyte Protonation Mechanism. *J Am Soc Mass Spectrom* 22(6):976–988.

162. Mock KK, Davey M, Cottrell JS (1991) The analysis of underivatized oligosaccharides by matrix-assisted laser desorption mass spectrometry. *Biochem Biophys Res Commun* 177(2):644–51.

163. Wada Y, et al. (2007) Comparison of the methods for profiling glycoprotein glycans--HUPO Human Disease Glycomics/Proteome Initiative multi-institutional study. *Glycobiology* 17(4):411–22.

164. Zaia J (2010) Mass Spectrometry and Glycomics. *Omi A J Integr Biol* 14(4):401–418.

165. Cancilla MT, Penn SG, Carroll JA, Lebrilla CB (1996) Coordination of Alkali Metals to Oligosaccharides Dictates Fragmentation Behavior in Matrix Assisted Laser Desorption Ionization/Fourier Transform Mass Spectrometry. *J Am Chem Soc* 118(28):6736–6745.

166. Treuheit MJ, Costello CE, Kirley TL (1993) Structures of the complex glycans found on the beta-subunit of (Na,K)-ATPase. *J Biol Chem* 268(19):13914–9.

167. Stahl B, et al. (1994) The oligosaccharides of the Fe(III)-Zn(II) purple acid phosphatase of the red kidney bean. Determination of the structure by a combination of matrix-assisted laser desorption/ionization mass spectrometry and selective enzymic degradation. *Eur J Biochem* 220(2):321–30.

168. North SJ, et al. (2010) Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals *N*-Glycans of a Novel Size and Complexity. *J Biol Chem* 285(8):5759–5775.

169. Patnaik SK, Stanley P (2006) Lectin-Resistant CHO Glycosylation Mutants. *Methods Enzymol* 416:159–182.

170. Xiang Y, et al. (2013) Regulation of protein glycosylation and sorting by the Golgi matrix proteins GRASP55/65. *Nat Commun* 4:1659.

171. Kam Y, Rejniak KA, Anderson ARA (2012) Cellular modeling of cancer invasion: Integration of in silico and in vitro approaches. *J Cell Physiol* 227(2):431–438.

172. Hughey JJ, Lee TK, Covert MW (2010) Computational modeling of mammalian signaling networks. *Wiley Interdiscip Rev Syst Biol Med* 2(2):194–209.

173. Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. *Nature* 438(7067):443–448.

174. Umaña P, Bailey JE (1997) A mathematical model of N-linked glycoform biosynthesis. *Biotechnol Bioeng* 55(6):890–908.

175. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467(7312):167–173.

176. Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* 309(5743):2010–3.

177. Spudich JL, Koshland DE (1976) Non-genetic individuality: chance in the single cell. *Nature* 262(5568):467–471.

178. Lehner B (2008) Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* 4:170.

179. Forger DB, Peskin CS (2005) Stochastic simulation of the mammalian circadian clock. *Proc Natl Acad Sci U S A* 102(2):321–4.

180. Gillespie DT (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22(4):403–434.

181. Spahn PN, Hansen AH, Kol S, Voldborg BG, Lewis NE (2017) Predictive glycoengineering of biosimilars using a Markov chain glycosylation model. *Biotechnol J* 12(2):1600489.

182. Weber P, Hornjik M, Olayioye MA, Hausser A, Radde NE (2015) A computational model of PKD and CERT interactions at the trans-Golgi network of mammalian cells. *BMC Syst Biol* 9(1):9.

183. Vanlier J, Tiemann CA, Hilbers PAJ, van Riel NAW (2012) A Bayesian approach to targeted experiment design. *Bioinformatics* 28(8):1136–1142.

184. Fagundes NJR, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci* 104(45):17614–17619.

185. Woods ML, Barnes CP (2016) Mechanistic Modelling and Bayesian Inference

Elucidates the Variable Dynamics of Double-Strand Break Repair. *PLOS Comput Biol* 12(10):e1005131.

186. Tanaka MM, Francis AR, Luciani F, Sisson SA (2006) Using approximate Bayesian computation to estimate tuberculosis transmission parameters from genotype data. *Genetics* 173(3):1511–20.

187. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface* 6(31):187–202.

188. Wilkinson DJ (2007) Bayesian methods in bioinformatics and computational systems biology. *Brief Bioinform* 8(2):109–16.

189. Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A* 100(26):15324–8.

190. Krambeck FJ, Betenbaugh MJ (2005) A mathematical model of N-linked glycosylation. *Biotechnol Bioeng* 92(6):711–28.

191. Hossler P, Mulukutla BC, Hu W-S (2007) Systems analysis of N-glycan processing in mammalian cells. *PLoS One* 2(8):e713.

192. Krambeck FJ, et al. (2009) A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology* 19(11):1163–1175.

193. Bennun S V, Yarema KJ, Betenbaugh MJ, Krambeck FJ (2013) Integration of the transcriptome and glycome for identification of glycan cell signatures. *PLoS Comput Biol* 9(1):e1002813.

194. McDonald AG, et al. (2014) Galactosyltransferase 4 is a major control point for glycan branching in N-linked glycosylation. *J Cell Sci* 127(Pt 23):5014–26.

195. Krambeck FJ, Bennun S V., Andersen MR, Betenbaugh MJ (2017) Model-based analysis of N-glycosylation in Chinese hamster ovary cells. *PLoS One* 12(5):e0175376.

196. Jimenez del Val I, Nagy JM, Kontoravdi C (2011) A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnol Prog* 27(6):1730–1743.

197. Sou SN, et al. (2017) Model-based investigation of intracellular processes determining antibody Fc-glycosylation under mild hypothermia. *Biotechnol Bioeng* 114(7):1570–1582.

198. Suga A, Nagae M, Yamaguchi Y (2018) Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. *Glycobiology* 28(10):774–785.

199. Ichishima E, et al. (1999) Molecular and enzymic properties of recombinant 1, 2-alpha-mannosidase from Aspergillus saitoi overexpressed in Aspergillus oryzae cells. *Biochem J* 339 (Pt 3):589–97.

200. Zhu Y, et al. (2010) Mechanistic insights into a Ca2+-dependent family of ?-mannosidases in a human gut symbiont. *Nat Chem Biol* 6(2):125–132.

201. Bendiak B, Schachter H (1987) Control of glycoprotein synthesis. Kinetic mechanism, substrate specificity, and inhibition characteristics of UDP-N-acetylglucosamine:alpha-D-mannoside beta 1-2 N-acetylglucosaminyltransferase II from rat liver. *J Biol Chem* 262(12):5784–90.

202. Shoreibah MG, Hindsgaul O, Pierce M (1992) Purification and characterization of rat kidney UDP-N-acetylglucosamine: alpha-6-D-mannoside beta-1,6-N-acetylglucosaminyltransferase. *J Biol Chem* 267(5):2920–7.

203. Yanagidani S, et al. (1997) Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 FucT) from human gastric cancer MKN45 cells. *J Biochem* 121(3):626–32.

204. Bao M, Elmendorf BJ, Booth JL, Drake RR, Canfield WM Bovine UDP-N-acetylglucosamine:Lysosomal-enzyme N-Acetylglucosamine-1-phosphotransferase. Available at: http://www.jbc.org/content/271/49/31446.full.pdf [Accessed June 13, 2017].

205. Sherlock C, Thiery AH, Roberts GO, Rosenthal JS (2015) On the efficiency of pseudo-marginal random walk Metropolis algorithms. *Ann Stat* 43(1):238–275.

206. Toni T, Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26(1):104–110.

207. Gelman A, Rubin DB (1992) Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci* 7:457–472.

208. Shannon P, et al. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13(11):2498–2504.

209. Ladinsky MS, Mastronarde DN, McIntosh JR, Howell KE, Staehelin LA (1999) Golgi structure in three dimensions: functional insights from the normal rat kidney cell. *J Cell Biol* 144(6):1135–49.

210. Stanley P, Sudo T (1981) Microheterogeneity among carbohydrate structures at the cell surface may be important in recognition phenomena. *Cell* 23(3):763–9.

211. An HJ, et al. (2012) Extensive determination of glycan heterogeneity reveals an unusual abundance of high mannose glycans in enriched plasma membranes of human embryonic stem cells. *Mol Cell Proteomics* 11(4):M111.010660.

212. Doob JL (1945) Markoff Chains--Denumerable Case. *Trans Am Math Soc* 58(3):455.

213. Zuber C, Spiro MJ, Guhl B, Spiro RG, Roth J (2000) Golgi apparatus immunolocalization of endomannosidase suggests post-endoplasmic reticulum glucose trimming: implications for quality control. *Mol Biol Cell* 11(12):4227–40.

214. Franke M, Braulke T, Storch S (2013) Transport of the GlcNAc-1-phosphotransferase α/β-subunit precursor protein to the Golgi apparatus requires a combinatorial sorting motif. *J Biol Chem* 288(2):1238–49.

215. Roth J, Berger EG (1982) Immunocytochemical localization of galactosyltransferase in HeLa cells: codistribution with thiamine pyrophosphatase in trans-Golgi cisternae. *J Cell Biol* 93(1):223–9.

216. Tabas I, Kornfeld S (1979) Purification and characterization of a rat liver Golgi alpha-mannosidase capable of processing asparagine-linked oligosaccharides. *J Biol Chem* 254(22):11655–63.

217. Harpaz N, Schachter H (1980) Control of glycoprotein synthesis. Processing of asparagine-linked oligosaccharides by one or more rat liver Golgi alpha-D-mannosidases dependent on the prior action of UDP-N-acetylglucosamine: alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase I. *J Biol Chem* 255(10):4894–902.

218. Schachter H, Narasimhan S, Gleeson P, Vella G (1983) Control of branching during the biosynthesis of asparagine-linked oligosaccharides. *Can J Biochem Cell Biol* 61(9):1049–66.

219. Longmore GD, Schachter H (1982) Product-identification and substrate-specificity studies of the GDP-L-fucose:2-acetamido-2-deoxy-beta-D-glucoside (FUC goes to Asn-linked GlcNAc) 6-alpha-L-fucosyltransferase in a Golgi-rich fraction from porcine liver. *Carbohydr Res* 100:365–92.

220. Marra P, et al. (2001) The GM130 and GRASP65 Golgi proteins cycle through

and define a subdomain of the intermediate compartment. *Nat Cell Biol* 3(12):1101–13.

221. Velasco A, et al. (1993) Cell type-dependent variations in the subcellular distribution of alpha-mannosidase I and II. *J Cell Biol* 122(1):39–51.

222. Joziasse DH, et al. (1987) Branch specificity of bovine colostrum CMP-sialic acid: Gal beta 1----4GlcNAc-R alpha 2----6-sialyltransferase. Sialylation of bi-, tri-, and tetraantennary oligosaccharides and glycopeptides of the N-acetyllactosamine type. *J Biol Chem* 262(5):2025–33.

223. Lin AI, Philipsberg GA, Haltiwanger RS (1994) Core fucosylation of high-mannose-type oligosaccharides in GlcNAc transferase I-deficient (Lec1) CHO cells. *Glycobiology* 4(6):895–901.

224. Yang Q, Zhang R, Cai H, Wang L-X (2017) Revisiting the substrate specificity of mammalian α1,6-fucosyltransferase (FUT8) reveals that it catalyzes core fucosylation of N-glycans lacking α1,3-arm GlcNAc. *J Biol Chem*:jbc.M117.804070.

225. Shao MC, Sokolik CW, Wold F (1994) Specificity studies of the GDP-[L]-fucose: 2-acetamido-2-deoxy-beta-[D]-glucoside (Fuc-->Asn-linked GlcNAc) 6-alpha-[L]-fucosyltransferase from rat-liver Golgi membranes. *Carbohydr Res* 251:163–73.

226. Liu T, et al. (2005) Human Plasma *N* -Glycoproteome Analysis by Immunoaffinity Subtraction, Hydrazide Chemistry, and Mass Spectrometry. *J Proteome Res* 4(6):2070–2080.

227. Chen R, et al. (2009) Glycoproteomics Analysis of Human Liver Tissue by Combination of Multiple Enzyme Digestion and Hydrazide Chemistry. *J Proteome Res* 8(2):651–661.

228. Tiede S, et al. (2004) A novel mutation in UDP- *N* -acetylglucosamine-1-phosphotransferase gamma subunit (GNPTAG) in two siblings with mucolipidosis type III alters a used glycosylation site. *Hum Mutat* 24(6):535–535.

229. Losfeld M-E, et al. (2017) Influence of protein/glycan interaction on site-specific glycan heterogeneity. *FASEB J* 31(10):4623–4635.

230. Li L, et al. (2015) Efficient Chemoenzymatic Synthesis of an N-glycan Isomer Library. *Chem Sci* 6(10):5652–5661.

231. Takahashi M, Kuroki Y, Ohtsubo K, Taniguchi N (2009) Core fucose and

bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydr Res* 344(12):1387–1390.

232. Quarto R, et al. (2001) Repair of Large Bone Defects with the Use of Autologous Bone Marrow Stromal Cells. *N Engl J Med* 344(5):385–386.

233. Taniguchi T, et al. (2010) Induction of endosomal/lysosomal pathways in differentiating osteoblasts as revealed by combined proteomic and transcriptomic analyses. *FEBS Lett* 584(18):3969–3974.

234. Nabavi N, Urukova Y, Cardelli M, Aubin JE, Harrison RE (2008) Lysosome Dispersion in Osteoblasts Accommodates Enhanced Collagen Production during Differentiation. *J Biol Chem* 283(28):19678–19690.

235. Nakajima K, et al. (2014) Galectin-3 Inhibits Osteoblast Differentiation through Notch Signaling. *Neoplasia* 16(11):939–949.

236. Stowell SR, et al. (2008) Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. *J Biol Chem* 283(15):10109–23.

237. Weilner S, et al. (2016) Vesicular Galectin-3 levels decrease with donor age and contribute to the reduced osteo-inductive potential of human plasma derived extracellular vesicles. *Aging (Albany NY)* 8(1):16–30.

238. Wu D, Struwe WB, Harvey DJ, Ferguson MAJ, Robinson C V. (2018) N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc Natl Acad Sci*:201807439.

239. Székely T, Burrage K (2014) Stochastic simulation in systems biology. *Comput Struct Biotechnol J* 12(20–21):14–25.

240. Bibila TA, Flickinger MC (1991) A model of interorganelle monoclonal antibody transport and secretion in mouse hybridoma cells. *Biotechnol Bioeng* 38(7):767–780.

241. Climer LK, Pokrovskaya ID, Blackburn JB, Lupashin V V. (2018) Membrane detachment is not essential for COG complex function. *Mol Biol Cell*:mbc.E17-11-0694.

242. Smith RD, et al. (2009) The COG complex, Rab6 and COPI define a novel Golgi retrograde trafficking pathway that is exploited by SubAB toxin. *Traffic* 10(10):1502–17.

243. Hamouda H, et al. (2014) Rapid Analysis of Cell Surface N-Glycosylation from Living Cells Using Mass Spectrometry. *J Proteome Res* 13(12):6144–6151.

244. Takahashi M, Kuroki Y, Ohtsubo K, Taniguchi N (2009) Core fucose and

bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins. *Carbohydr Res* 344(12):1387–1390.

245. Hsu DK, Zuberi RI, Liu FT (1992) Biochemical and biophysical characterization of human recombinant IgE-binding protein, an S-type animal lectin. *J Biol Chem* 267(20):14167–74.

246. Lakshminarayan R, et al. (2014) Galectin-3 drives glycosphingolipid-dependent biogenesis of clathrin-independent carriers. *Nat Cell Biol* 16(6):592–603.

247. James C. Sacchettini ‡, Linda G. Baum § and, C. Fred Brewer* ‖ (2001) Multivalent Protein−Carbohydrate Interactions. A New Paradigm for Supermolecular Assembly and Signal Transduction. doi:10.1021/BI002544J.

248. Jimenez del Val I, Fan Y, Weilguny D (2016) Dynamics of immature mAb glycoform secretion during CHO cell culture: An integrated modelling framework. *Biotechnol J* 11(5):610–623.

Appendix 1

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3
MAN1	Log-normal	Log-normal	Exponential decay
EndoMAN	Log-normal	Exponential decay	Exponential decay
MAN2	Exponential decay	Log-normal	Exponential decay
MGAT1	Exponential decay	Log-normal	Exponential decay
MGAT2 (1)	Exponential decay	Log-normal	Exponential decay
MGAT2 (2)	Exponential decay	Log-normal	Exponential decay
MGAT2 (3)	Exponential decay	Log-normal	Exponential decay
MGAT4	Exponential decay	Log-normal	Exponential decay
MGAT5	Exponential decay	Log-normal	Exponential decay
FUT8 (1)	Exponential decay	Log-normal	Exponential decay
FUT8 (2)	Exponential decay	Log-normal	Exponential decay
FUT8 (3)	Exponential decay	Log-normal	Exponential decay
FUT8 (4)	Exponential decay	Log-normal	Exponential decay
FUT8 (5)	Exponential decay	Log-normal	Exponential decay
FUT8 (6)	Exponential decay	Log-normal	Exponential decay
FUT8 (7)	Exponential decay	Log-normal	Exponential decay
GalT (1)	Exponential decay	Exponential decay	Log-normal
GalT (2)	Exponential decay	Exponential decay	Log-normal
GalT (3)	Exponential decay	Exponential decay	Log-normal
GalT (4)	Exponential decay	Exponential decay	Log-normal
SiaT	Exponential decay	Exponential decay	Log-normal
GlcNAc-PhospoT	Log-normal	Exponential decay	Exponential decay

Appendix 1 | Table representing assignment of PDFs for enzymatic parameters in HeLa cells

Appendix 2

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3
MAN1	0.233496705	0.527139663	0.178111358
MAN1(2)	0.037217926	0.084022791	0.028389845
EndoMAN	0.029461455	0.003994726	0.004010006
MAN2	0.453718821	3.451994737	0.568873221
MGAT1	0.01393476	0.304813621	0.009163032
MGAT2(1)	1.798471895	12.77035395	1.700470263
MGAT2(2)	1.798471895	12.77035395	1.700470263
MGAT2(3)	1.798471895	12.77035395	1.700470263
MGAT4	0.004194392	0.031407158	0.004211912
MGAT5	0.020214963	0.307998276	0.020739824
FUT8(1)	0.049015259	0.412099132	0.053306307
FUT8(2)	0.049015259	0.412099132	0.053306307
FUT8(3)	0.049015259	0.412099132	0.053306307
FUT8(4)	0.049015259	0.412099132	0.053306307
FUT8(5)	0.049015259	0.412099132	0.053306307
FUT8(6)	0.049015259	0.412099132	0.053306307
FUT8(7)	0.000147627	0.001514319	0.000146666
GalT(1)	0.451799263	0.448824895	3.148618158
GalT(2)	0.451799263	0.448824895	3.148618158
GalT(3)	0.451799263	0.448824895	3.148618158
GalT(4)	0.451799263	0.448824895	3.148618158
SiaT	0.006008948	0.006231213	0.127969208
GlcNAc-phosphoT	0.125711779	0.014646036	0.014101402
	0.710849474	0.009048474	5.279178421

Appendix 2 | Table representing the fitted parameter values for WT HeLa cells

Appendix 3

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3
MAN1	0.217140421	0.716860211	0.174680053
MAN1(2)	0.039337033	0.12986598	0.031644937
EndoMAN	0.047610013	0.004292252	0.004242789
MAN2	0.020611184	0.082378211	0.020556605
MGAT1	0.005638958	0.173814332	0.010241469
MGAT2(1)	1.869815789	13.12817263	1.741911842
MGAT2(2)	1.869815789	13.12817263	1.741911842
MGAT2(3)	1.869815789	13.12817263	1.741911842
MGAT4	0.004275932	0.030062337	0.004408021
MGAT5	0.021468105	0.3103385	0.021851513
FUT8(1)	0.050996627	0.547423579	0.001383333
FUT8(2)	0.050996627	0.547423579	0.001383333
FUT8(3)	0.050996627	0.547423579	0.001383333
FUT8(4)	0.050996627	0.547423579	0.001383333
FUT8(5)	0.050996627	0.547423579	0.001383333
FUT8(6)	0.050996627	0.547423579	0.001383333
FUT8(7)	0.000149171	0.537967974	6.40199E-05
GalT(1)	0.466929947	0.549856632	2.952280789
GalT(2)	0.466929947	0.549856632	2.952280789
GalT(3)	0.466929947	0.549856632	2.952280789
GalT(4)	0.466929947	0.549856632	2.952280789
SiaT	0.006608056	0.006707846	0.713332926
GlcNAc-phosphoT	0.098124853	0.014375818	0.014271601
	0.725984295	0.008530408	6.358872105

Appendix 3 | Table representing the fitted parameter values for swainsonine-treated HeLa cells

Appendix 4

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3
MAN1	0.233496705	0.527139663	0.178111358
MAN1(2)	0.037217926	0.084022791	0.028389845
EndoMAN	0.029461455	0.003994726	0.004010006
MAN2	0.453718821	3.451994737	0.568873221
MGAT1	0.01393476	0.304813621	0.009163032
MGAT2(1)	1.798471895	12.77035395	1.700470263
MGAT2(2)	1.798471895	12.77035395	1.700470263
MGAT2(3)	1.798471895	12.77035395	1.700470263
MGAT4	0.004194392	0.031407158	0.004211912
MGAT5	0.020214963	0.307998276	0.020739824
FUT8(1)	0.049015259	0.412099132	0.053306307
FUT8(2)	0.049015259	0.412099132	0.053306307
FUT8(3)	0.049015259	0.412099132	0.053306307
FUT8(4)	0.049015259	0.412099132	0.053306307
FUT8(5)	0.049015259	0.412099132	0.053306307
FUT8(6)	0.049015259	0.412099132	0.053306307
FUT8(7)	0.000147627	0.001514319	0.000146666
GalT(1)	0.451799263	0.448824895	3.148618158
GalT(2)	0.451799263	0.448824895	3.148618158
GalT(3)	0.451799263	0.448824895	3.148618158
GalT(4)	0.451799263	0.448824895	3.148618158
SiaT	0.006008948	0.006231213	0.127969208
GlcNAc-phosphoT	0.125711779	0.014646036	0.014101402
	0.710849474	0.009048474	5.279178421

Appendix 4 | Table representing the fitted parameter values for Cog4KD HeLa cells

Appendix 5

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.194485087	1.7516927	0.22349944	0.022924432
MAN1(2)	0.021928699	0.197507902	0.025200142	0.002584789
EndoMAN	0.030639145	0.005333927	0.004076729	0.000423947
MAN2	0.50131336	2.9467012	1.914476833	0.547513267
MGAT1	0.007119538	0.636804333	0.098816213	0.009268861
MGAT2(1)	1.796796067	6.5077812	5.896882133	1.7318214
MGAT2(2)	1.796796067	6.5077812	5.896882133	1.7318214
MGAT2(3)	1.796796067	6.5077812	5.896882133	1.7318214
MGAT4	0.004176106	0.02006288	0.019514897	0.004981294
MGAT5	0.020047691	0.168627987	0.24528994	0.031690672
FUT8(1)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(2)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(3)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(4)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(5)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(6)	0.049176632	0.177289367	0.071655067	0.04276546
FUT8(7)	0.002713389	0.009782188	0.003953668	0.002359644
FUT8OFF	0.099446587	0.099446587	0.099446587	0.099446587
Ant FUT	0.000999622	0.014081887	0.011049582	0.001026963
GalT(1)	0.010020635	0.072018267	0.2061617	0.25531468
GalT(2)	0.010020635	0.072018267	0.2061617	0.25531468
GalT(3)	0.010020635	0.072018267	0.2061617	0.25531468
GalT(4)	0.010020635	0.072018267	0.2061617	0.25531468

GalT(5)	0.003562637	0.025604659	0.073296684	0.090772046
GalT(6)	0.003562637	0.025604659	0.073296684	0.090772046
GalT(7)	0.003562637	0.025604659	0.073296684	0.090772046
GalT(8)	0.003562637	0.025604659	0.073296684	0.090772046
GalT(9)	0.003562637	0.025604659	0.073296684	0.090772046
SiaT(1)	0.006000835	0.006713236	0.084122005	0.09841716
SiaT(2)	0.006000835	0.006713236	0.084122005	0.09841716
SiaT(3)	0.002754859	0.003081907	0.038618664	0.045181272
GlcNAc-phosphoT	0.052862328	0.011772589	0.014043215	0.001410821
	0.1132855	0.030996753	7.2580224	

Appendix 5 | Table representing the fitted parameter values for WT HEK293T cells

Appendix 6

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.148468438	1.382024	0.242616881	0.023389545
MAN1(2)	0.018257739	0.169952844	0.029835538	0.002876303
EndoMAN	0.030787209	0.005775171	0.003909426	0.000452394
MAN2	0.025405775	0.025219844	0.015022541	0.024132106
MGAT1	0.007097191	0.285030219	0.091399238	0.009116399
MGAT2(1)	1.794842375	6.60331125	5.829263	1.94838375
MGAT2(2)	1.794842375	6.60331125	5.829263	1.94838375
MGAT2(3)	1.794842375	6.60331125	5.829263	1.94838375
MGAT4	0.004211199	0.01504282	0.013251575	0.004850105
MGAT5	0.020005323	0.183632925	0.257015981	0.033191759
FUT8(1)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(2)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(3)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(4)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(5)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(6)	0.049294611	0.192566797	0.066660625	0.046202119
FUT8(7)	0.017024658	0.06650593	0.023022281	0.015956618
FUT8OFF	0.099503951	0.099503951	0.099503951	0.099503951
Ant FUT	0.001002962	0.013080549	0.007934707	0.000886347
GalT(1)	0.009985707	0.071914269	0.229779956	0.432610181
GalT(2)	0.009985707	0.071914269	0.229779956	0.432610181
GalT(3)	0.009985707	0.071914269	0.229779956	0.432610181
GalT(4)	0.009985707	0.071914269	0.229779956	0.432610181
GalT(5)	0.003531843	0.025435346	0.081270836	0.153009826
GalT(6)	0.003531843	0.025435346	0.081270836	0.153009826
GalT(7)	0.003531843	0.025435346	0.081270836	0.153009826
GalT(8)	0.003531843	0.025435346	0.081270836	0.153009826
GalT(9)	0.003531843	0.025435346	0.081270836	0.153009826
SiaT(1)	0.005991593	0.006270562	0.122704588	0.31972855

SiaT(2)	0.005991593	0.006270562	0.122704588	0.31972855
SiaT(3)	0.002851303	0.002984059	0.058393141	0.152153678
GlcNAc-phosphoT	0.030139624	0.012389008	0.013977596	0.001502368
	0.11671605	0.012219881	9.275291	

Appendix 6 | Table representing the fitted parameter values for swainsonine-treated HEK293T cells

Appendix 7

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.207270829	1.769651714	0.497360333	0.024201713
MAN1(2)	0.023412045	0.199889032	0.056178781	0.002733677
EndoMAN	0.03048921	0.005530166	0.004349325	0.00041441
MAN2	0.497703229	2.447501333	1.352916762	0.522109143
MGAT1	0.007135823	0.294988483	0.051401381	0.009110397
MGAT2(1)	0.178693757	0.737605371	0.593994476	0.197388386
MGAT2(2)	0.178693757	0.737605371	0.593994476	0.197388386
MGAT2(3)	0.178693757	0.737605371	0.593994476	0.197388386
MGAT4	0.004205056	0.019599488	0.019078714	0.005590132
MGAT5	0.020243149	0.128970705	0.073751829	0.027164443
FUT8(1)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(2)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(3)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(4)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(5)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(6)	0.049662796	0.206351155	0.105898033	0.055317005
FUT8(7)	4.96674E-05	0.00020637	0.000105908	5.53221E-05
FUT8OFF	0.072470219	0.072470219	0.072470219	0.072470219
Ant FUT	0.001002334	0.014333406	0.012777042	0.001043961
GalT(1)	0.009959058	0.07245599	0.15254361	0.10177224
GalT(2)	0.009959058	0.07245599	0.15254361	0.10177224
GalT(3)	0.009959058	0.07245599	0.15254361	0.10177224
GalT(4)	0.009959058	0.07245599	0.15254361	0.10177224
GalT(5)	0.003525369	0.025648424	0.05410513	0.03609722
GalT(6)	0.003525369	0.025648424	0.05410513	0.03609722
GalT(7)	0.003525369	0.025648424	0.05410513	0.03609722
GalT(8)	0.003525369	0.025648424	0.05410513	0.03609722
GalT(9)	0.003525369	0.025648424	0.05410513	0.03609722
SiaT(1)	0.006027384	0.006607882	0.113236252	0.121332895

SiaT(2)	0.006027384	0.006607882	0.113236252	0.121332895
SiaT(3)	0.002794113	0.003063214	0.052492891	0.056246249
GlcNAc-phosphoT	0.086969672	0.011764207	0.014015796	0.001537617
	0.390088557	0.018145524	6.693468952	

Appendix 7 | Table representing the fitted parameter values for Cog4KO HEK293T cells

Appendix 8

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.15	1.757304643	0.154082229	0.024309597
MAN1(2)	0.013169447	0.154202283	0.01352061	0.002133151
EndoMAN	0.030250348	0.005568547	0.003948526	0.00041535
MAN2	0.506190971	4.903260857	2.659008286	0.545422786
MGAT1	0.00693554	0.5	0.1	0.009975619
MGAT2(1)	0.179587836	0.564087143	0.727762607	0.248050429
MGAT2(2)	0.179587836	0.564087143	0.727762607	0.248050429
MGAT2(3)	0.179587836	0.564087143	0.727762607	0.248050429
MGAT4	0.004190141	0.014352356	0.018409193	0.005405872
MGAT5	0.020134409	0.057334779	0.076209429	0.030761723
FUT8(1)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(2)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(3)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(4)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(5)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(6)	0.048188871	0.15043975	0.057443964	0.057784225
FUT8(7)	0.002654511	0.008287059	0.003164333	0.003183077
FUT8OFF	0.082556143	0.082556143	0.082556143	0.082556143
Ant FUT	0.000998262	0.013721026	0.0007	0.0007
GalT(1)	0.010078676	0.065822614	0.784380357	0.483927857
GalT(2)	0.010078676	0.065822614	0.784380357	0.483927857
GalT(3)	0.010078676	0.065822614	0.784380357	0.483927857
GalT(4)	0.010078676	0.065822614	0.784380357	0.483927857
GalT(5)	0.005056857	0.033025725	0.393553645	0.242805127
GalT(6)	0.005056857	0.033025725	0.393553645	0.242805127
GalT(7)	0.005056857	0.033025725	0.393553645	0.242805127
GalT(8)	0.005056857	0.033025725	0.393553645	0.242805127
GalT(9)	0.005056857	0.033025725	0.393553645	0.242805127
SiaT(1)	0.006077692	0.006680604	0.04679148	0.034442945

SiaT(2)	0.006077692	0.006680604	0.04679148	0.034442945
SiaT(3)	0.003315519	0.003644421	0.025525814	0.018789408
GlcNAc-phosphoT	0.03	0.01252273	0.013914026	0.001323202
	0.3	0.017735158	10.03370857	

Appendix 8 | Table representing the fitted parameter values for hTERT-MSCs (Y101s)

Appendix 9

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.224048482	1.693398	0.211269164	0.029133404
MAN1(2)	0.02097832	0.158557845	0.019781754	0.002727846
EndoMAN	0.028452668	0.005762128	0.004382028	0.000465687
MAN2	0.507186291	0.015138	0.042572204	0.524604818
MGAT1	0.007002892	0.2	0.1	0.011384847
MGAT2(1)	0.177453336	0.566492882	0.736911818	0.269098955
MGAT2(2)	0.177453336	0.566492882	0.736911818	0.269098955
MGAT2(3)	0.177453336	0.566492882	0.736911818	0.269098955
MGAT4	0.004201773	0.026264944	0.0210087	0.005631788
MGAT5	0.020347337	0.076022202	0.125928314	0.032938956
FUT8(1)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(2)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(3)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(4)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(5)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(6)	0.049140775	0.077683573	0.045423322	0.058024591
FUT8(7)	0.01965631	0.031073429	0.018169329	0.023209836
FUT8OFF	0.073580291	0.073580291	0.073580291	0.073580291
Ant FUT	0.000999117	0.014057616	0.00080452	0.000637988
GalT(1)	0.010049891	0.064834567	0.983165636	0.539144564
GalT(2)	0.010049891	0.064834567	0.983165636	0.539144564
GalT(3)	0.010049891	0.064834567	0.983165636	0.539144564
GalT(4)	0.010049891	0.064834567	0.983165636	0.539144564
GalT(5)	0.009464088	0.061055394	0.925857419	0.507718105
GalT(6)	0.009464088	0.061055394	0.925857419	0.507718105
GalT(7)	0.009464088	0.061055394	0.925857419	0.507718105
GalT(8)	0.009464088	0.061055394	0.925857419	0.507718105
GalT(9)	0.009464088	0.061055394	0.925857419	0.507718105
SiaT(1)	0.006091524	0.006510832	0.027940549	0.030893268

SiaT(2)	0.006091524	0.006510832	0.027940549	0.030893268
SiaT(3)	0.007820717	0.008359052	0.035871994	0.039662897
GlcNAc-phosphoT	0.0211944	0.012634842	0.014064311	0.001281484
	0.209372209	0.009511833	13.81438091	

Appendix 9 | Table representing the fitted parameter values for swainsonine-treated hTERT-MSCs (Y201s)

Appendix 10

Enzyme Name	Rate Cisterna 1	Rate Cisterna 2	Rate Cisterna 3	Rate Cisterna 4
MAN1	0.307558695	1.7744673	0.17476228	0.026185174
MAN1(2)	0.030061662	0.173441485	0.017081763	0.002559413
EndoMAN	0.033662145	0.00588473	0.004029459	0.000402867
MAN2	0.5079001	6.093339	3.377208	0.587262
MGAT1	0.007114036	0.6225316	0.079531065	0.009841085
MGAT2(1)	0.18019222	0.46439071	0.7227887	0.26702385
MGAT2(2)	0.18019222	0.46439071	0.7227887	0.26702385
MGAT2(3)	0.18019222	0.46439071	0.7227887	0.26702385
MGAT4	0.004192536	0.009751601	0.017818555	0.005511443
MGAT5	0.019955005	0.029678331	0.04895692	0.032459256
FUT8(1)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(2)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(3)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(4)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(5)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(6)	0.047849834	0.154176175	0.0523304	0.05802216
FUT8(7)	0.002452617	0.007902537	0.002682275	0.002974015
FUT8OFF	0.092797335	0.092797335	0.092797335	0.092797335
Ant FUT	0.000995563	0.01234311	0.000676167	0.000579693
GalT(1)	0.010083057	0.05723082	1.14757235	0.53553684
GalT(2)	0.010083057	0.05723082	1.14757235	0.53553684
GalT(3)	0.010083057	0.05723082	1.14757235	0.53553684
GalT(4)	0.010083057	0.05723082	1.14757235	0.53553684
GalT(5)	0.005173889	0.029366679	0.588850366	0.274798416
GalT(6)	0.005173889	0.029366679	0.588850366	0.274798416
GalT(7)	0.005173889	0.029366679	0.588850366	0.274798416
GalT(8)	0.005173889	0.029366679	0.588850366	0.274798416
GalT(9)	0.005173889	0.029366679	0.588850366	0.274798416
SiaT(1)	0.006091488	0.007073644	0.041135007	0.03149241

SiaT(2)	0.006091488	0.007073644	0.041135007	0.03149241
SiaT(3)	0.002920068	0.003390883	0.019718831	0.015096473
GlcNAc-phosphoT	0.013152243	0.012452489	0.014328323	0.001424251
	0.301190675	0.012826073	12.091441	

Appendix 10 | Table representing the fitted parameter values for osteoblasts derived from hTERT-MSCs (Y101s)