Investigating the rules governing N-linked glycosylation of Biologics

Josephine Elizabeth Ayre

Master of Science (by research)

University of York

Biology

December 2023

Abstract

Biologics, drugs produced in living organisms, such as monoclonal antibodies (mAbs) can treat various medical conditions, so improving their production and efficacy is important. Large scale mAb manufacturing involves expressing recombinant DNA in mammalian cells, for example Chinese hamster ovary (CHO) cells, which are cultured in bioreactors. mAb glycosylation is monitored for quality control, as glycans dictate mAb stability, functionality, and efficacy. Glycans are processed in the Golgi and glycosylation enzyme localisation within Golgi cisternae is determined by intra Golgi vesicle transport. The conserved oligomeric Golgi (COG) complex participates in retrograde trafficking of glycosylation enzymes within the Golgi facilitated through Rab interactions. Understanding the glycosylation machinery and how alterations influence glycosylation patterns aids in improving mAb production and efficacy. This thesis addresses the role mutations in COG complex subunit Cog4 play in mAb glycosylation. Cog4 interactions with Rabs and stability of Cog4 mutants were examined using a computationally generated Cog4 structure. Rab30 and Rab41 structures were computationally docked to the Cog4 structure, revealing a preferential binding to the Cog4 N-terminus. Stability assessment of Cog4 upon mutation predicted Cog4-L36P is unstable, suggesting L36 is indirectly involved in Rab30 binding. Whereas R34C, Q67R and F137S had negligible effect on Cog4 stability suggesting these positions are directly involved in Rab41 binding. To determine how Cog4 mutations affect mAb glycosylation CHO cell lines Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R were transiently transfected with Herceptin. Glycan profiling of Herceptin products unveiled potential decreases in *trans*-Golgi to *medial*-Golgi glycosylation enzyme recycling in Cog4-KO and Cog4-L36P cell lines. Comparing the Herceptin glycan profiles of Cog4-L773R and Cog4-WT revealed an increase in galactosylation, suggesting increased recycling of galactosylation enzymes to the *trans*-Golgi. Furthering the understanding of Cog4 functions and how Cog4 mutants influence glycosylation can facilitate the development of CHO cell lines with specific mAb glycosylation.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank the project supervisors Daniel Ungar and Nia Bryant for their guidance and advice. I would also like to thank members of the Ungar and Bryant lab groups for their advice and support. I am grateful to the project industry supervisors Jessica Kane-Fidgeon and Rachel Richer from Fujifilm Diosynth Biotechnologies for their support and guidance. Thanks, are also due to Fujifilm Diosynth Biotechnologies for funding this work. Lastly, I am thankful for my partner, family, and friends for providing their continuous support.

Declaration

I declare that this thesis is a presentation of original work, and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Introduction

Drugs produced in living organisms, known as Biologics, are used to treat a variety of medical conditions. In the field of biologics, monoclonal antibodies (mAbs) have emerged as effective therapeutic agents, providing targeted treatment for cancers, autoimmune conditions, and a spectrum of diseases (Lu et al., 2020). On the pharmaceutical market 121 mAbs have been approved for use by the EU, of which 13 were approved in 2022 (Reichert et al., 2023). As secreted proteins, mAbs undergo post-translational modifications, for example glycosylation which attaches carbohydrate components called glycans. Glycans dictate mAb stability, functionality, and efficacy through influencing interactions with mAb biological targets and the immune system (Nose and Wigzell, 1983; Mimura et al., 2000; Wu et al., 2010). Understanding the factors that govern glycosylation and its machinery, is important for improving mAb production and efficacy.

N-linked Glycosylation

Glycans are found on secreted proteins and are involved in many protein functions. The process of protein glycosylation lacks a template, therefore within a cell a heterogeneous population of glycans can be found on different molecules of the same protein (Higel et al., 2016). Macro-heterogeneity describes the degree of occupation of glycosylation sites on a protein, this is determined in the endoplasmic reticulum (ER) where glycan precursors are added. Whereas micro-heterogeneity refers to the variety and amount of glycans found at one glycosylation site. This means glycan processing can be determined by glycosylation site used, as well as cell and protein type. Glycan heterogeneity makes glycosylation analysis challenging; therefore, a variety of analytical approaches are implemented to determine glycan patterns on glycoproteins. Glycosylation analysis can be displayed as glycan profiles, various methods can be used to analyse a glycoprotein including HPLC-MS (High-performance liquid chromatography coupled with mass-spectrometry), capillary gel electrophoresis, isoelectric focusing, and lectin-based approaches (Mariño et al., 2010; Stavenhagen et al., 2013; Planinc et al., 2016; Zhang, Luo and Zhang, 2016). These methods can be employed independently or in combination to evaluate protein glycosylation.

N-glycans are usually linked to the nitrogen atom of an asparagine side chain found in an Asparagine-X-Serine/Threonine consensus sequence, where X can be any amino acid except

proline. All *N*-glycans share a common core sequence consisting of two N-acetylglucosamine (GlcNAc) and three mannose monosaccharides, beyond the core they can be one of three types; oligomannose, complex or hybrid. Oligomannose *N*-glycans contain only mannose antennae, complex *N*-glycan antennae have a variety of monosaccharides and hybrid *N*-glycans have both antenna types (Figure 1A) (Kornfeld and Kornfeld, 1985). Protein glycosylation begins in the ER, all glycans exit the ER with the same structure thus acting as quality control for degradation or further processing (Hebert, Garman, and Molinari, 2005). Glycoproteins are transported to the Golgi for glycan processing, here glycan antennae number and length are established. In the *cis*-Golgi the MAN1 enzyme is found, which removes mannose residues from oligomannose glycans to form $GlcNAc_2Man_5$. Oligomannose glycans larger than $GlcNAc_2Man_5$ may leave the Golgi without undergoing further processing. The *medial*-Golgi contains the N-acetylglucosamine transferase MGAT1 enzyme that add a GlcNAc sugar to $GlcNAc_2Man_5$. $GlcNAc_3Man_5$ glycans can become hybrid or complex glycans by following one of two paths, keeping the mannose antennae, and becoming a hybrid glycan or MAN2 can remove two mannose residues and the N-acetylglucosamine transferase MGAT2 can add GlcNAc to the other antennae resulting in $GlcNAc_4Man_3$. Complex glycans have various capping arrangements that are added in the *trans*-Golgi. Additional glycosyltransferases are found in the *trans*-Golgi, these add sugars such as galactose and sialic acid to the glycan chain. Galactose is added by galactosyltransferase (GalT) and sialic acid by sialyltransferase (SiaT). *N*-glycans can also undergo core modifications such as addition of fucose, by fucosyltransferase (FUT8) (Kornfeld and Kornfeld, 1985) (Figure 1B). Some glycosylation enzymes compete to modify glycoproteins, competition can be influenced by Golgi organisation controlled by membrane trafficking components (Ohtsubo and Marth, 2006).

Localisation of glycosylation enzymes in the Golgi cisternae is controlled by intra Golgi vesicle transport. The Golgi maturation model suggests that cisternae within the Golgi stack mature from the *cis*-Golgi to the *trans*-Golgi. In this model glycosylation enzymes are recycled to their cisternae via retrograde transport (Glick, Elston and Oster, 1997; Matsuura-Tokita et al., 2006; Rizzo et al., 2013). Intra Golgi vesicle transport machinery directs vesicle

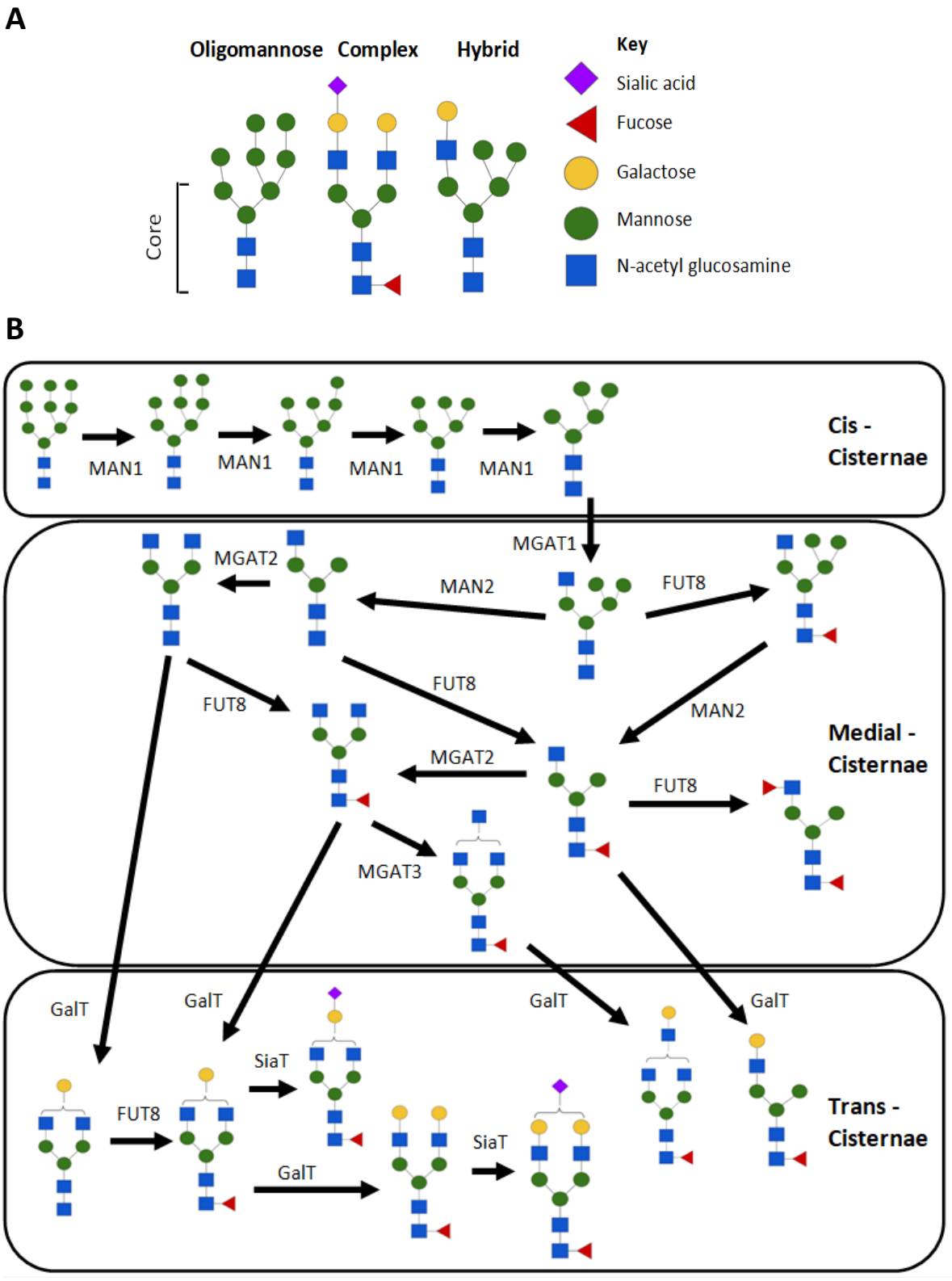

Figure 1 - Glycan processing in the Golgi. A) Glycan structures have a common core structure of two N-Acetyl glucosamine (GlcNAc) and three mannose residues. Beyond the core glycans can be one of three types, oligomannose, complex or hybrid containing only mannose, a variety of monosaccharides or antennae of both types respectively. B) Glycan processing occurs in the Golgi. In the *cis*-Golgi MAN1 removes mannose residues. In the *medial*-Golgi mannose is cleaved by MAN2, GlcNAc residues are added by MGAT enzymes, and fucose residues are added by FUT8. Finally in the *trans*-Golgi galactose is added by GalT and sialic acid by SiaT.

formation, budding, and fusion. The coat protein COPI plays a key role in vesicle budding, selecting cargo, and indicating destination (Rabouille and Klumperman, 2005). Membrane tethering is facilitated by golgins, the conserved oligomeric Golgi (COG) complex, and Rabs. Tethering is followed by membrane fusion which is mediated by Sec1/Munc18 (SM) proteins and Soluble N-ethylmaleimide-Sensitive Factor Attachment Proteins receptors (SNAREproteins). Component interaction specificity is essential for intra Golgi vesicle trafficking. Thus, glycosylation enzyme organisation within the Golgi is maintained through the interactions of tethering and fusion mediators (Willett, Ungar and Lupashin, 2013).

Conserved Oligomeric Golgi (COG) complex

The COG complex is a critical component of the glycosylation machinery, due to its involvement in the retrograde trafficking of glycosylation enzymes within the Golgi. All eight COG subunits are essential in maintaining the integrity of the complex and ensuring proper function of the glycosylation process, however, the roles of individual subunits are not fully understood (Podos et al., 1994; Shestakova, Zolov and Lupashin, 2006; Pokrovskaya et al., 2011; Blackburn et al., 2016). COG subunits are organised into two subcomplexes, lobe A contains subunits one to four and lobe B contains subunits five to eight. These lobes interact via subunits one and eight (Fotso et al., 2005; Ungar et al., 2005). In Human embryonic kidney HEK cells knocking out any of the COG subunits results in glycosylation enzyme recycling defects, mainly sialylation and fucosylation enzymes, and defects in Golgi morphology. The severity of defects was varied based on which subunit was targeted, with Lobe A knockouts having greater Golgi morphology defects than lobe B knockouts and Cog2, 3, 4, 5, and 7 knockouts displaying more defective glycosylation than other subunit knockouts (Blackburn et al., 2016). Glycosylation enzyme sorting to cisternae by retrograde transport involves the COG complex interacting with multiple tethering components (D'Souza, Taher and Lupashin, 2020). The COG complex subunit Cog4 is a unique subunit in its interactions, Cog4 interacts with several SNARE and Rab proteins and is the only lobe A subunit to do so (Shestakova et al., 2007; Laufman, Hong and Lev, 2013; Miller et al., 2013). Cog4 also has a unique interaction with clathrin heavy chain and GEA2, an Arf1 Guanidine exchange factor (Chen et al., 2011). To investigate the interaction between Cog4 and the SNARE STX5, truncated Cog4 was used to determine potential binding sites. Residues 1 to 222 of Cog4 were found to interact with STX5 (Shestakova et al., 2007). This interaction site

for STX5 was later refined to residues 84 to 153 of Cog4, and the SM protein Sly1 was found to interact with the first 83 residues of Cog4. This Cog4-Sly1-STX5 interaction is required for SNARE complex assembly and retrograde intra Golgi trafficking (Laufman et al., 2009). Pathological mutation of Cog4 have been associated with impaired glycosylation enzyme sorting, leading to defects in glycosylation processes and the resulting glycan structures (Reynders et al., 2009).

Rabs play a key role in coordinating membrane tethering and fusion, thus delivering glycosylation enzymes to the correct compartments. Rabs are GTPases that function as molecular switches by cycling between active GTP (Guanosine triphosphate) bound and inactive GDP (Guanosine diphosphate) bound forms. Active Rabs associate with Golgi and vesicle membranes and recruit specific effector proteins to regulate membrane trafficking steps. The yeast Rab Sec4 was found to recruit the exocyst tethering complex which in turn interacts with Rho GTPases, thus connecting vesicle membranes to their target membrane (Guo et al., 1999; Guo, Tamanoi and Novick, 2001). Rab5 participates in membrane docking and fusion through its interactions with SM proteins and SNAREs (McBride et al., 1999; Simonsen et al., 1999; Morrison et al., 2008). Rab30 is almost exclusively localised to the Golgi, and Rab41 is involved in *cis*-Golgi organisation and structure (Goud, Liu and Storrie, 2018). Rab30 interacts with *cis*-golgins, GM130 and GMAP210 and the *trans*-golgins, golgin-97, golgin-245/p230, GCC88 and GCC185 (Sinka et al., 2008; Hayes et al., 2009). Rab41 interacts with dynein motor complex subunit dynactin 6 and STX8 (Liu et al., 2016). Cog4 interacts with the active forms of Rab1a and Rab30 (Miller et al., 2013). Previously, the Ungar lab used a yeast two hybrid screen to identify a mutation in Cog4 that disrupts its interaction with Rab30 (unpublished). In this same screen a set of mutations in Cog4, known as mutant22, were found to cause an enhanced interaction with Rab41.

Congenital disorders of glycosylation (CDGs)

Mutations in the COG complex have been identified in patients with congenital disorders of glycosylation (CDGs), a group of rare genetic diseases characterized by defects in glycosylation machinery components (Zeevaert et al., 2008). Diagnostic methods for CDGs include high-performance liquid chromatography, capillary zone electrophoresis, and mass spectrometry (Carchon et al., 2004; Quintana et al., 2009; Casetta et al., 2020). CDGs have

been linked to mutations in genes encoding enzymes, transporters, and chaperone proteins involved in glycosylation, resulting in diverse clinical presentations (Ondruskova et al., 2021). Patients suffer with a wide range of clinical symptoms often involving multiple organ systems, including the central nervous system, liver, skin, and musculoskeletal system (Lipiński and Tylki-Szymańska, 2021). The management of CDGs is based on supportive care to alleviate specific symptoms and improve the quality of life for patients, causative treatments are available however only for a few CDG types (Lipiński and Tylki-Szymańska, 2021). CDGs linked to *N*-glycosylation defects are divided into two CDG types (CDG-I and CDG-II). CDG-I is characterised by defects in glycan synthesis and attachment to proteins, resulting in proteins with unoccupied glycan sites. For CDG-II defects are found in the machinery involved in the modification of glycans in the Golgi, causing defective glycoproteins (Freeze, 2007; Piedade et al., 2022). Mutations in COG complex members result in CDG-II (Wu et al., 2004; Foulquier et al., 2006; Reynders et al., 2009). While the precise mechanisms vary between CDG types, typically these mutations lead to incomplete or abnormal glycan structures, that can alter glycoprotein stability, localization, and functionality.

Cog4 mutations identified in CDG patients, R729W and L773R cause altered Cog4 levels that lead to glycosylation defects. L773R Cog4 mutation is known to cause a 66% reduction in Cog4 levels without affecting the levels of the other subunits (Ng et al., 2011). Whereas the R729W Cog4 mutation causes 80% reduction in levels of Cog4 along with lower levels of Cog1 (25%), Cog2 (40%), Cog3 (25%), and Cog5 (40%). The R729W mutation was associated with defects in sialylation (Reynders et al., 2009). Defects in sialylation were also identified in patient cells with the L773R mutation, which were linked to defects in galactosylation (Miura et al., 2005).

mAb production and glycosylation

The field of cancer therapy was transformed by targeted mAbs, for example Herceptin which treats HER2-positive breast cancer. The majority of marketed mAbs, including Herceptin, belong to the immunoglobulin G (IgG) class. IgG mAbs are glycosylated in their fragment crystallisable (Fc) region, which binds various receptors influencing how the antibody interacts with immune system components. For example, the Fc region binds to

Fcγ receptors (FcγR) and the C1 complex which participate in antibody-dependent cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC) respectively (Hughes-Jones and Gardner, 1979; Sarmay et al., 1992). Glycans are found between the two heavy chain C_H2 domains of IgG mAbs and are thought to influence FcγR interactions (Voynov et al., 2009; Hanson and Barb, 2015). Glycosylation is known to improve mAb stability, solubility and reduce aggregation tendencies (Kayser et al., 2011).

Manufacturing mAbs on a large scale involves expressing recombinant DNA in mammalian cells, preferably Chinese Hamster Ovary (CHO) cells as these can be grown on a large scale whilst achieving human-like glycosylation of mAb products. CHO cells are cultured in bioreactors, where conditions such as pH, ammonium concentration and temperature have been shown to influence mAb glycosylation (Goochee and Monica, 1990; Borys, Linzer and Papoutsakis, 1993; Monica, Goochee and Maiorella, 1993; Borys, Linzer and Papoutsakis, 1994; Andersen et al., 2000). The production of non-glycosylated mAbs increases during a bioreactor culture run, which has been linked to altered glycan processing rather than degradation (Curling et al., 1990). Throughout the bioreactor run nutrients will be depleted and metabolic by-products will accumulate, these have been linked to the production of non-glycosylated mAbs (Hayter et al., 1992; Jenkins and Curling, 1994; Nyberg et al., 2000). As glycosylation can be influenced by many factors in the mAb production process, mAbs can have batch to batch variability in glycosylation.

The effect of glycosylation features on mAb functionality has been widely investigated. In one study a non-fucosylated form of Herceptin was shown to enhance ADCC in whole blood cell assays using breast cancer tumour cells (Suzuki et al., 2007). mAbs with non-fucosylated glycans have improved FcγR binding which enhances ADCC (Chiu et al., 2019). Oligomannose glycans have also been linked to increased ADCC activity (Boune et al., 2020). However, oligomannose glycans cannot be fucosylated therefore, the lack of fucosylation may cause the enhanced ADCC activity. Cell lines have been engineered to study the effects of mAb glycosylation on functionality. For example, a FUT8 KO (knock out) CHO cell line with or without glycosidase inhibitors swainsonine or kifunensine were generated to produce each of the three types of glycan, oligomannose, hybrid and complex, exclusively. This revealed mAbs with non-naturally occurring *N*-glycans, such as oligomannose and hybrid types, had reduced receptor affinity (Kanda et al., 2007).

Due to the therapeutic importance of correct mAb glycosylation and the potential variability in glycosylation from cell culturing conditions, the monitoring of mAb glycosylation is part of the quality control of mAbs (Federici et al., 2013). Monitoring is often conducted using small samples from the culturing vessel. From these samples mAbs are purified using protein A columns. Following this the glycans are released from the purified mAbs and glycans are labelled with RapiFluor for analysis by LC-MS (Liquid chromatography coupled with mass-spectrometry) (Figure 2).

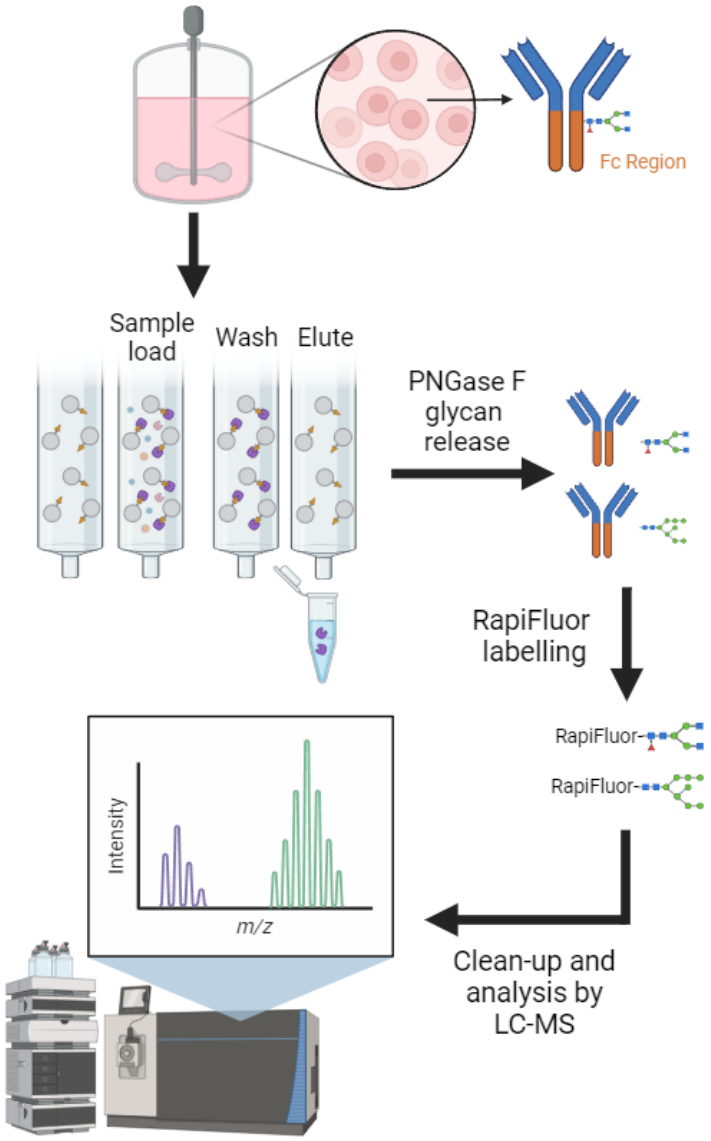

Figure 2 - mAb purification and glycan profiling workflow. Samples are taken from bioreactor cultures and mAbs are purified using protein A columns. Glycans are released from mAbs by PNGaseF and labelled with RapiFluor. The RapiFluor glycans are then detected by LC-MS. Diagram created with BioRender.

Aims

Glycosylation is a critical factor in the therapeutic efficacy of mAbs. Therefore, a greater understanding of mAb glycosylation will allow enhancing of mAb production and efficacy. Specifically, investigating how mutating glycosylation machinery influences the glycosylation of mAbs. This research aims to explore the impact of Cog4 mutations on glycosylation, investigate Cog4 interactions with Rab proteins, and examine how Cog4 mutations affect mAb production and glycosylation.

Cog4 mutations, L773R and R729W, were identified in patients with CDGs. These mutations are known to cause reduced levels of Cog4, and the R729W mutation reduces levels of other COG subunits (Reynders et al., 2009; Ng et al., 2011). The COG complex function in membrane tethering is affected by reductions in the levels of any COG subunit (Blackburn et al., 2016). Therefore, to hypothesise why L773R and R729W mutations have different effects on levels of other subunits the stability of Cog4 upon mutation was assessed. These results provide insight into the effects of these mutations on Cog4 function and thus glycosylation.

Cog4 mutations that alter their ability to interact with Rabs have been identified by the Ungar lab. Cog4-L36P has a reduced interaction with GTP-Rab30 (Spencer, 2022). R34C, Q67R and F137S have an enhanced interaction with Rab41 (unpublished data). COG complex interactions with Rabs are important for glycosylation enzyme sorting. Therefore, the effect of these mutations on Cog4 stability was investigated to determine how the mutations could be altering Rab binding. Rab30 and Rab41 were also computationally docked to Cog4 to determine the location of Rab binding. These results provide a greater understanding of the potential modes of action of Rab binding to Cog4. This data also supplies potential targets for further Rab binding experiments.

L773R and R729W mutations of Cog4 identified in CDG patients can cause defects in glycosylation. These mutations reduce levels of Cog4 which can alter COG complex membrane tethering function thus glycosylation enzyme trafficking (Shestakova et al., 2006; Richardson et al., 2009; Ng et al., 2011). Whole cell glycan profiling of the Cog4-L36P stable CHO cell line revealed the L36P mutation affects glycosylation. Previous Herceptin transfections of the Cog4-L36P stable CHO cell line suggest this mutation increases

production of Herceptin (Spencer, 2022). Therefore, to investigate how Cog4 mutations L36P, R729W and L773R influence glycosylation of mAbs, Herceptin transient transfections were performed, and Herceptin was glycan profiled. These results provide an insight into the effects of these mutation on glycosylation and their potential to be used in mAb production.

Chapter 1: Investigating Cog4 mutants using computationally generated structures.

Introduction

Cog4 mutations, L773R and R729W, were identified in patients with congenital disorders of glycosylation (CDGs). The L773R Cog4 mutation is known to cause reduced levels of Cog4 without affecting the levels of the other subunits (Ng et al., 2011). Whereas the R729W Cog4 mutation causes reduced levels of Cog4 along with Cogs1, 2, 3 and 5 (Reynders et al., 2009). Reductions in the levels of the Cog4 subunit have been shown to alter COG complex function in aiding membrane tethering (Blackburn et al., 2016). Previous studies of Cog4 by the Ungar lab identified Cog4 mutations that alter the ability of Cog4 to interact with Rabs. Cog4-L36P was found to have a reduced interaction with GTP-Rab30 (Spencer, 2022). Three Cog4 mutations were found to have an enhanced interaction with Rab41, it remains unknown whether this is due to an individual mutation or a combination. The three mutations were R34C, Q67R and F137S (unpublished data). In this chapter a computationally generated Cog4 structure is used to determine the stability of Cog4 when mutated and to assess the potential binding sites on Cog4 for Rabs.

Protein modelling methods

Protein structures

Cog4 structures were collected from AlphaFold (AF-Q9H9E3-F1) and RCSB protein data bank (PDB) (3HR0) and aligned using PyMOL (Richardson et al. 2009; Schrödinger 2015; Jumper et al. 2021). AlphaFold provided a confidence score for the prediction of Cog4 based on the lDDT-Cα metric and a predicted aligned error (PAE) plot for the Cog4 structure (Jumper et al., 2021). GTP bound Rab30 and Rab41 structures were collected from the RCSB PDB (2EW1) and AlphaFold (AF-Q5JT25-F1), respectively (Wang et al., 2005; Jumper et al., 2021). Cog4 and Rab structures were viewed in PyMOL, where residues and regions were highlighted, and the structure surface was viewed (Schrödinger, 2015).

Aligning sequences

Cog4 sequences were collected from the UniProtKB database for *H. sapiens* (Q9H9E3)*, G. gallus* (A0A8V0ZU03)*, X. tropicalis* (B1WAP4)*, D.* rerio (Q29RB1)*, D. melanogaster* (Q95TN4)*, A. thaliana* (Q8L838)*, C. elegans* (Q95XZ0)*, C. griseus* (A0A8C2LM94)*, S. cerevisiae* (Q06096), and *A. carolinensis* (H9G614) (Bateman et al., 2023). The *H. sapiens* sequence was used as the query sequence to align the subject sequences using protein BLAST with the blastp algorithm, BLOSUM62 scoring matrix, an expect threshold of 0.05, word size of 3, max matches in a query range of 0, Gap costs existence 11 and extension 1, and conditional compositional score matrix adjustment (Madden, Tatusov and Zhang, 1996). The multiple sequence alignment (MSA) was viewed in Jalview, where residues were highlighted in boxes and the MSA was coloured based on conservation (Waterhouse et al., 2009). The conservation score measures the number of conserved physico-chemical properties of the column of the alignment. The calculation of the conservation score is based in the AMAS method of multiple sequence alignment analysis (Livingstone and Barton, 1993).

Stability analysis using DDMut

The stability of Cog4 mutants was determined using DDMut. The Cog4 structure file from AlphaFold was uploaded to DDMut and the mutations were inputted. The results showed a predicted folding stability change, ΔΔG, measured in kcal/mol, and whether this value was deemed stabilising (ΔΔG ≥ 0 kcal/mol) or destabilising (ΔΔG < 0 kcal/mol). ΔΔG is the difference in ΔG between wild type and mutant protein forms, and ΔG is the unfolding free energy of the protein. Contact diagrams were also produced, displaying the amino acid interactions in both the wild-type and mutant proteins (Zhou et al., 2023).

STRING analysis

A *H. Sapiens* Cog4 protein full STRING diagram of experimentally determined interactions between Cog4 and Golgi proteins was generated using Cytoscape. The STRING diagram was created using significant protein interaction datasets and protein-protein interaction databases. A default medium confidence interaction score of 0.45 was specified and a Golgi location confidence score of 0.5 was used to ensure only Golgi proteins were included. Edge thickness was set to indicate the confidence of interaction, and edges showing interactions with Cog4 were coloured green. Nodes were coloured by group, GARP (Golgi-associated

retrograde protein) complex members, SNAREs, Rabs, and an SM protein (Shannon et al., 2003).

Protein property analysis

The *H. Sapiens* Cog4 structure collected from AlphaFold was viewed using PyMOL for property analysis. The electrostatic potentials of the Cog4, Rab30, and Rab41 structures were determined using the APBS server. For all structures PDB2PQR were generated by inputting the Cog4 structure PDB files, a pKa value of 7 was used, it was selected to use PROPKA to assign protonation states at provided pH and the forcefield option selected was PARSE. Additional options were selected for new atoms not to be rebuilt too close to existing atoms, an optimised hydrogen bonding network, removal of waters from the output file, and creation of an APBS input file. The results were used to calculate APBS, the calculation type selected was mg-auto (Holst and Saied, 1993, 1995; Jurrus et al., 2018). The APBS output files were downloaded for viewing using PyMOL, where residues commonly involved in protein-protein interactions were highlighted on the Cog4 structure surface (Schrödinger, 2015). For N-terminal analysis of Cog4, the N-terminus was defined as residues 1-165. N-terminal area and feature counts were calculated using PyMOL (Schrödinger, 2015).

Protein docking using Hdock

The hydrophobic binding pocket and triad of aromatic amino acids involved in all Rab interactions were identified on Rab8 (6ZSH) (Rai et al., 2016). To determine the location of these sites on Rab30 and Rab41, they were aligned to Rab8 in PyMOL (Itzen and Goody, 2011; Schrödinger, 2015). Rab30 and Rab41 were separately docked to the Cog4 AlphaFold structure using Hdock by specifying the hydrophobic binding pocket and triad of aromatic amino acid sites. A total of 100 potential models were generated from the Cog4 and Rab structures (Yan et al., 2020; Jumper et al., 2021). Structural interactions were examined and counted using PyMOL (Schrödinger, 2015). To calculate the probability of Rabs binding to the N-terminus randomly, the surface area of this region was calculated as a percentage of the total surface area of Cog4. This percentage was taken as the random chance for Rab binding to this region.

Results

Stability of Cog4 with CDG mutations may explain the Cog4 levels seen in patients.

Cog4 mutations identified in patients with CDG have been linked to altered levels of Cog4 and other COG complex subunits. R729W patient cells had 80% less Cog4 than wild-type cells and lower levels of Cog1 (25%), Cog2 (40%), Cog3 (25%), and Cog5 (40%) (Reynders et al., 2009). However, L773R patient cells have usual levels of other subunits and 66% less Cog4 than wild-type cells (Ng et al., 2011). Computationally generated protein models were used to investigate how the stability of these Cog4 mutants could be affecting COG subunit levels.

First, the AlphaFold predicted Cog4 structure was assessed (Figure 3). AlphaFold predicts structures based on protein databank (PDB) models, primary amino acid sequences, and structures of homologous sequences. Previously, the structure of the Cog4 C-terminus from residue 536 to 785 was solved by X-ray crystallography (Richardson et al., 2009). The predicted Cog4 aligns well with the solved structure of the Cog4 C-terminus, which accounts for 31% of the entire Cog4 structure (Figure 3A). The confidence scoring system used by AlphaFold suggests that the majority of the Cog4 prediction was confident or highly confident, including the previously solved structure of C-terminal section where the CDG mutations were found (Figure 3B). AlphaFold also provides a predicted aligned error (PAE) plot that can be used to assess inter-domain accuracy. The PAE plot indicates the expected positional error for scored residues (x-axis) when the AlphaFold generated structure and known structures were aligned at the aligned residue (y-axis). If residues x and y are found in two different domains, when the PAE values (x, y) are low, it suggests a defined orientation, whereas if the value is high, it suggests that the orientation is unreliable (Varadi et al., 2022). The scale provided shows that the lighter the shade of green, the higher the PAE value. In the Cog4 structure, scored residues from 200 to 785 have high PAE value with aligned residues 1 to 200 (Figure 3C). This suggests that the orientation of the N-terminal helical region (residues 1-165) with respect to the rest of the structure is unreliable. The N-terminal region was found to be involved in protein-protein interactions

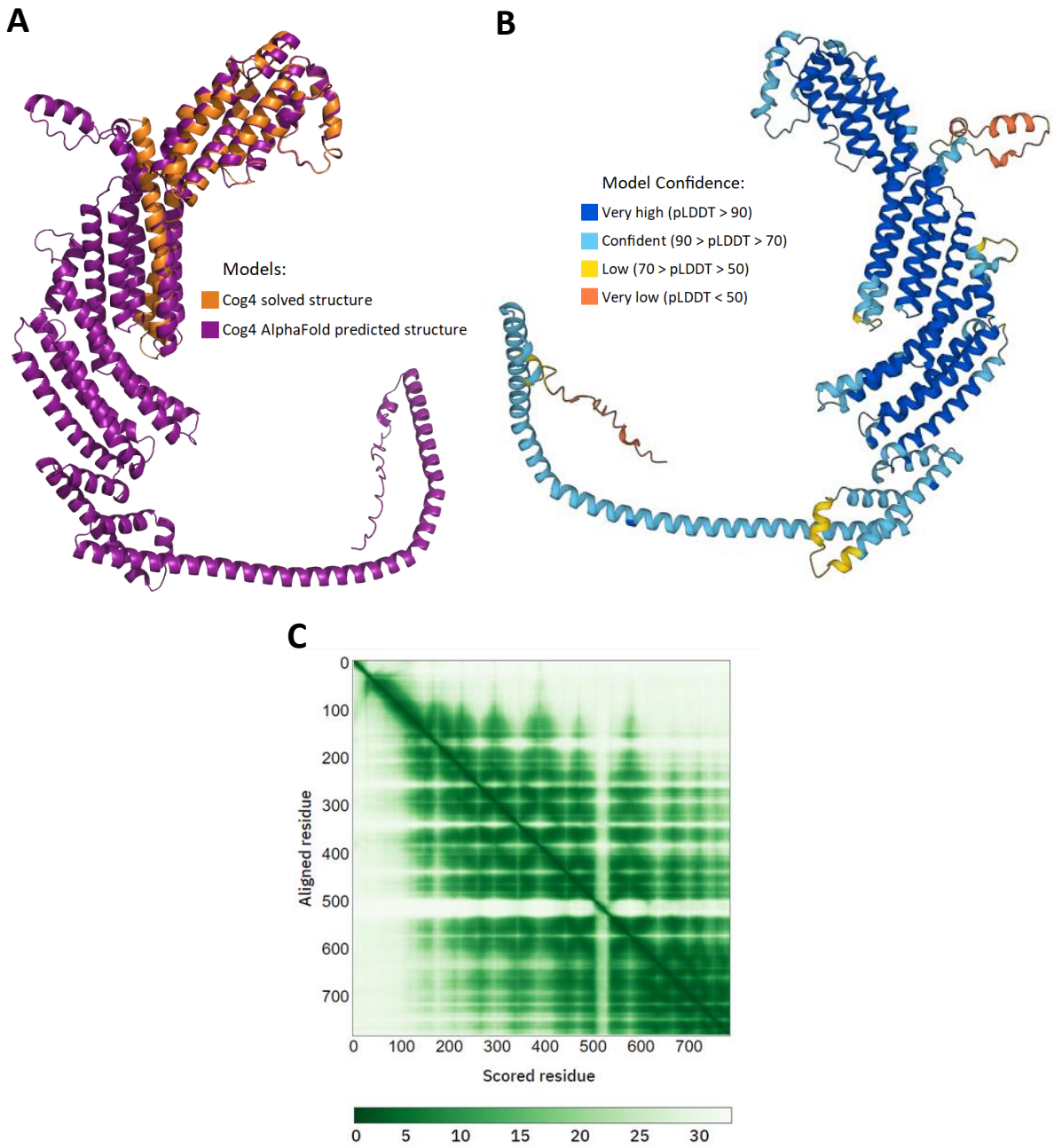

Figure 3 - The AlphaFold predicted Cog4 structure. A) AlphaFold prediction of Cog4 (purple) was aligned to solved structure of Cog4 C-terminus (orange). B) Cog4 predicted by AlphaFold was coloured by confidence of Alpha fold prediction (pLDDT) with dark blue and light blue representing very high and high confidence respectively, and yellow and orange representing low and very low respectively. C) PAE plot for Cog4 AlphaFold predicted structure showing expected error in aligning structural domains with green representing a low error value and white representing high error value.

(Laufman et al., 2009). Therefore, in this chapter the unreliability of the N-terminal region positions in relation to the rest of the AlphaFold predicted structure was considered in analysis of the N-terminal region as a potential Rab binding site.

The CDG mutation positions R729 and L773 were identified on the Cog4 structure, L773 was found at position L769 on the Cog4 structure (Figure 4A). When looking at the surface of Cog4, R729 appears to be surface facing; however, L773 is buried (Figure 4B). The conservation of the CDG mutation positions was examined with a multiple sequence alignment (MSA) of Cog4 sequences from mostly model organism species. Conservation is scored on a scale of zero to 11, the higher the score the more conserved the amino acids in the column, a score of 10 indicates any mutations in the column have conserved properties.

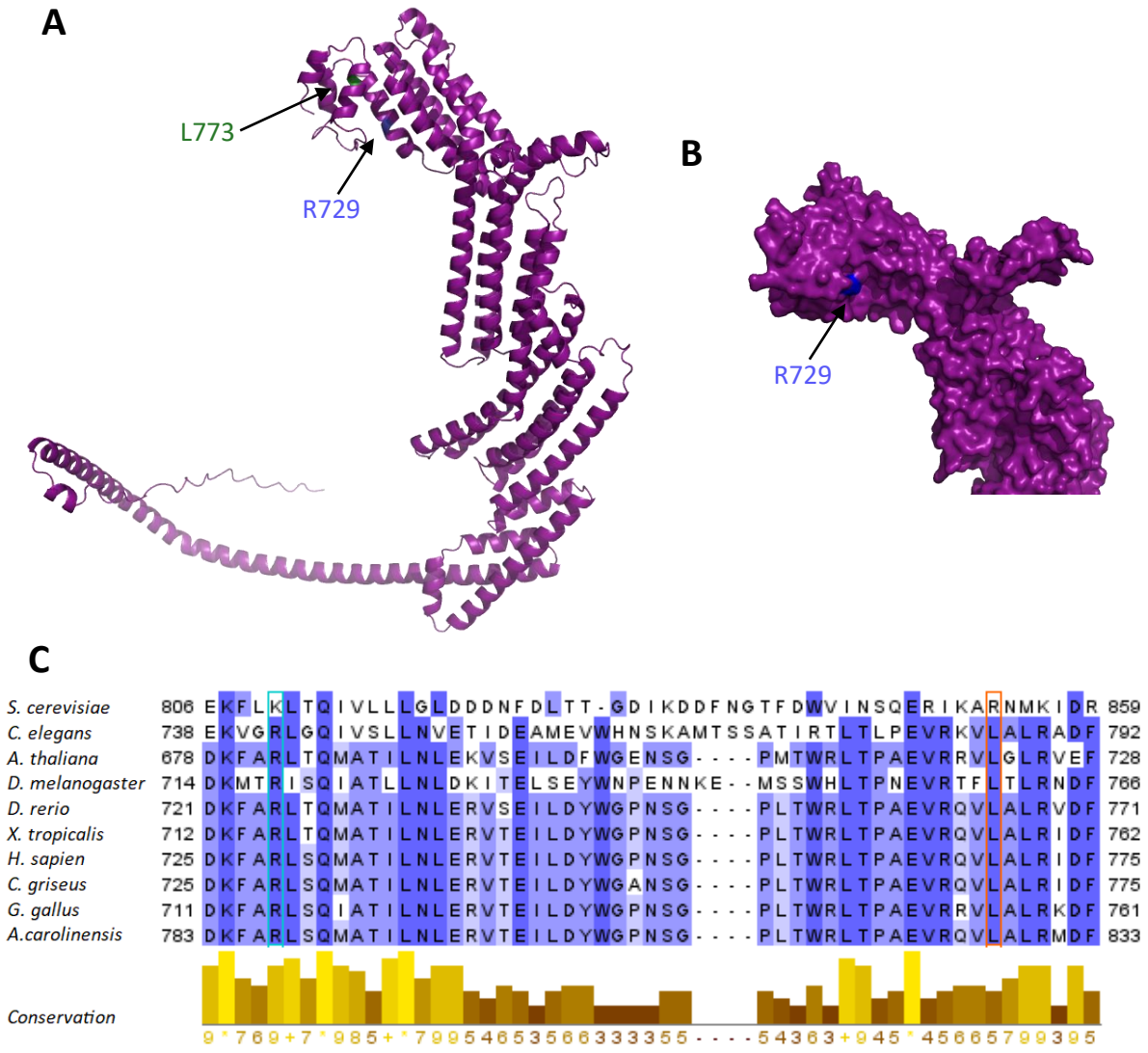

Figure 4 – CDG mutation positions on Cog4 structure and conservation. A) Cog4 structure labelled with CDG mutation positions R729 and L773, blue and green, respectively. L773 was found at position L769 on the Cog4 structure. B) The surface of the Cog4 structure with R729 labelled in blue. C) ClustalW multiple sequence alignment (MSA) of Cog4 from model organism species and CHO cell line species: *H. sapiens, G. gallus, X. tropicalis, D. rerio, D. melanogaster, A. thaliana, C. elegans, C. griseus, S. cerevisiae,* and *A. carolinensis.* MSA coloured by percentage match to consensus sequence. R729 highlighted in blue box and L773 (at position L769) in an orange box. Conservation score annotation is based on the number of conserved physico-chemical properties.

Both CDG mutation positions were well conserved with R729W scoring 10 and L773 scoring five. The region around the mutation positions was also well conserved suggesting it is integral to Cog4 structure and function (Figure 4C). Taken together this suggests the R729 position is integral to the structure of Cog4 and its ability to interact with other subunits. Whereas the L773 position may be able to be occupied by similar amino acids and maintain the structural integrity of Cog4.

To assess the stability of Cog4 with CDG mutations the Cog4 structure and mutation positions were input into DDMut. For each mutation Cog4 models displaying amino acids and their interactions within the structure were produced, alongside a ΔΔG value and based on that an assessment of whether the mutation is stabilising (ΔΔG ≥ 0 kcal/mol) or destabilising (ΔΔG < 0 kcal/mol) the structure. The ΔΔG value represents a change in thermodynamic stability of the protein upon mutation. L773R and R729W had ΔΔG values of -2.12 kcal/mol and -2.1 kcal/mol respectively and were destabilising.

Cog4 L773 was found at the end of an alpha helix and forms both hydrogen bonds and hydrophobic interactions with surrounding residues (Figure 5A). Leucine prefers to be buried in protein hydrophobic cores. When leucine at position 773 was mutated to arginine the number of hydrophobic interactions decreased (Figure 5B). As hydrophobic interactions are known to stabilise tertiary structure, a reduction in their number could explain a reduction in ΔΔG. Taking this into consideration with the lower levels of Cog4 in CDG patients with L773R mutations, the instability caused by L773R could cause a reduction in Cog4 from an instable tertiary structure. L773 is an internal residue (Figure 4B), therefore instability of the tertiary structure here could result in degradation of Cog4. Cog4 R729 is located on an alpha helix in the C-terminus and is known to form a salt bridge network with E688 and E764. This network is thought to stabilise the tertiary structure of Cog4 and could therefore affect Cog4 functions if mutated. As R729 is at the centre of the salt bridge network replacing with the bulkier and rigid tryptophan is thought to disrupt intramolecular interactions. Contacts between Q732 and R729 are also thought to stabilise the tertiary structure of Cog4 (Richardson et al., 2009). In wild-type Cog4 R729 forms hydrogen bonds and hydrophobic interactions with R759 and L760 on another helix (Figure 5C). Both

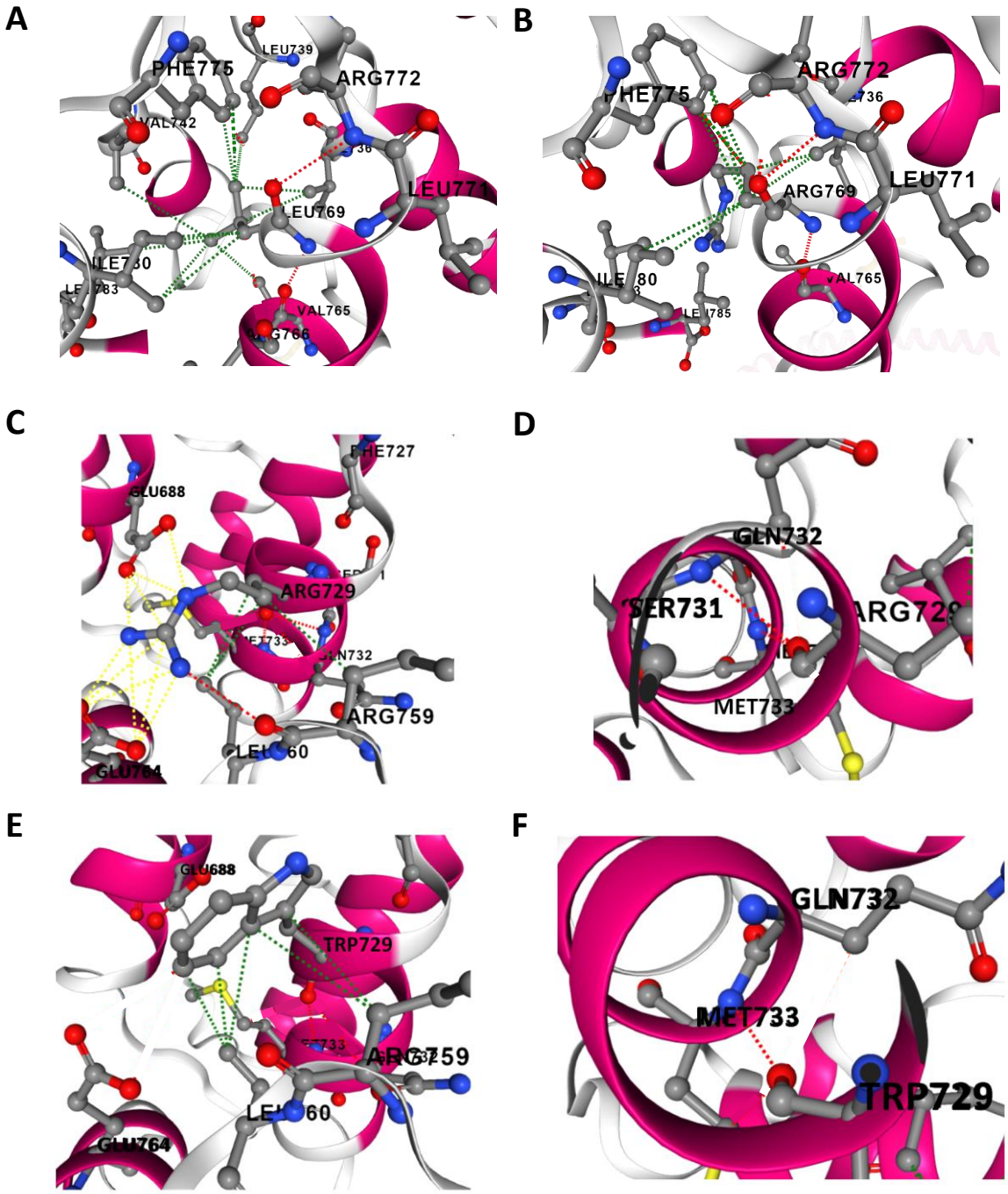

Figure 5 – Visualisations of bonds found in WT-Cog4 and when each CDG mutation was introduced. Each structure was coloured by secondary structural features, pink indicates helix. Interactions are denoted by dotted lines, green represents hydrophobic interactions, red denotes hydrogen bonds and yellow represent ionic bonds. A) WT-Cog4 showing L773 (L769) interactions. B) Mutant Cog4 showing L773R (L769R) interactions. C) WT-Cog4 showing R729 interactions. D) WT-Cog4 showing R729 hydrogen bond interactions within the alpha helix. E) Mutant Cog4 showing R729W interactions. F) Mutant Cog4 showing R729W hydrogen bond interactions within the alpha helix.

hydrogen bonds and hydrophobic interactions are known to stabilise tertiary structure.

Hydrogen bonds form between Q732, M733 and R729 which were found on the same alpha

helix (Figure 5D). Hydrogen bonds are known to maintain an alpha helical shape, therefore stabilising secondary structure. When the arginine at position 729 was mutated to tryptophan all connections with E688 and E764 were lost. The number of hydrophobic interactions W729 had with residues on other helices increased when compared to R729 (Figure 5E). W729 formed a hydrogen bond with M733 however, there was no hydrogen bond between W729 and Q732 (Figure 5F). The position of R729 on the surface of Cog4 may explain the difference between the effects of L773R and R729W on levels of other COG subunits (Figure 4B). Taken together this suggests the R729W mutation could be altering the internal stability of Cog4 thus changing the ability of Cog4 to interact with other COG complex subunits resulting in low levels of Cog4 and other subunits.

Rab docking to Cog4 revealed preferential binding to the N-terminal helix.

Cog4 interacts with other COG complex members and is known to interact with various Golgi proteins involved in membrane trafficking. To assess the current knowledge of Cog4 interactions a STRING diagram was created, this shows the experimentally determined interactions of Cog4 with other Golgi resident proteins. These Golgi proteins were divided into the following groups: GARP (Golgi-associated retrograde protein) complex members, SNAREs, Rabs, and SM proteins. Many of these Golgi proteins interact with each other, however Rab30 is only known to interact with Cog4 (Figure 6A). Investigation of Cog4 interaction with Rab30 has revealed the GTP bound form of Rab30 is preferred by Cog4 (Miller et al., 2013). All the SNAREs known to interact with Cog4 also interact with most other proteins in the diagram, as well as with at least one of the Rabs, Rab43 or Rab33 (Figure 6A). To investigate the interaction between Cog4 and the SNARE STX5 truncated Cog4 was used to determine potential binding sites. It was found that residues 1 to 222 of Cog4 alone could interact with STX5 (Shestakova et al., 2007). This interaction site for STX5 was later refined to residues 84 to 153 of Cog4 (Laufman et al., 2009).

As evidence suggests the Cog4 N-terminus participates in protein interactions amino acids commonly involved in protein interactions were identified on Cog4 to determine if these occur more often at the N-terminus. For N-terminal analysis the N-terminus was defined as residues 1 to 165 to cover the N-terminal alpha helix protruding from the rest of the

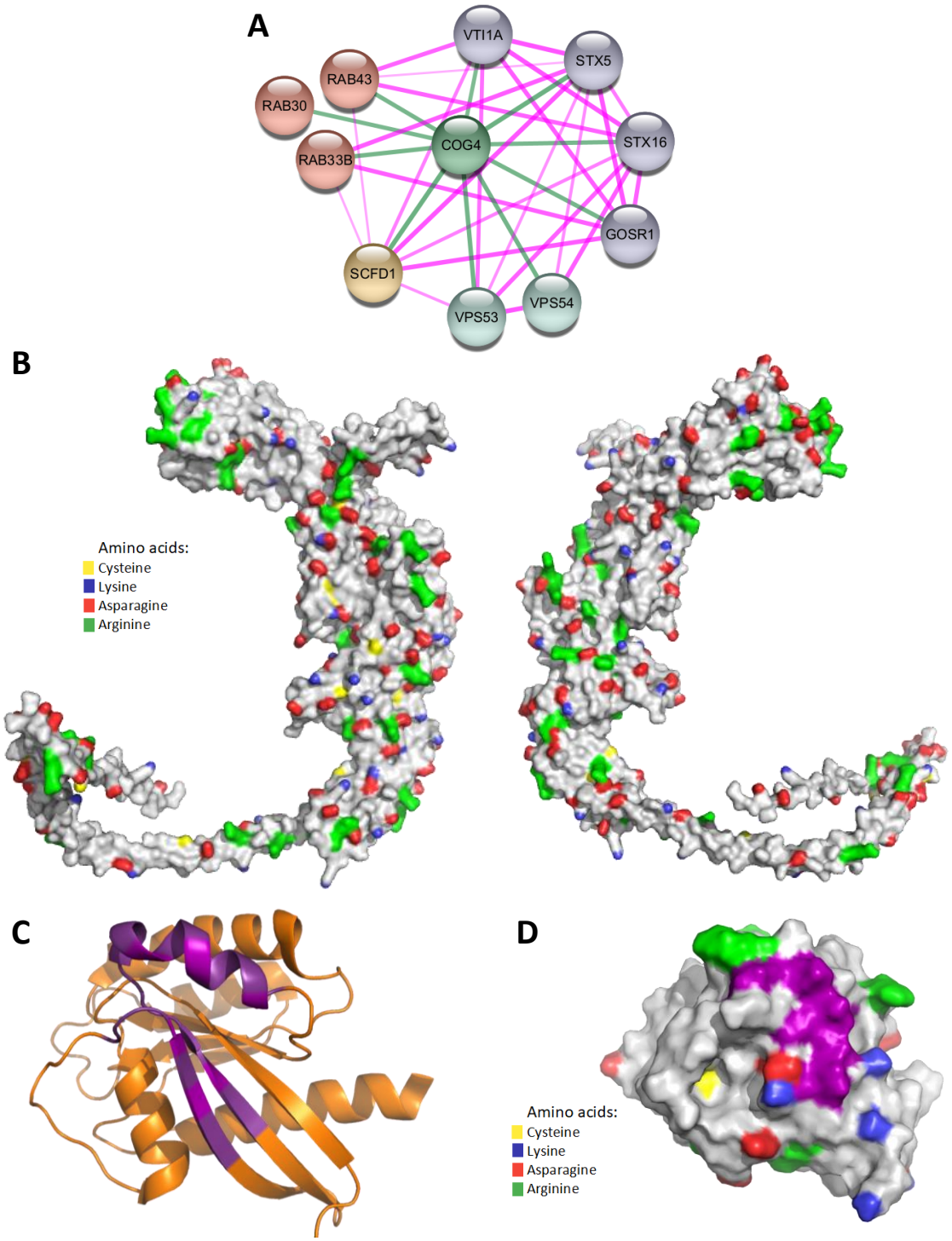

Figure 6 – Cog4 interacts with Golgi proteins. A) STRING diagram created with Cog4 interactions identified through experimental evidence with a confidence score of 0.45, line thickness indicates interaction confidence. Nodes were coloured based on protein type: GARP complex members, SNAREs, Rabs, and an SM protein. B) Cog4 structure surface with amino acids commonly involved in protein-protein interactions highlighted. Yellow represents cysteine, blue represents lysine, red represents asparagine, and green represents arginine. C) Rab8a structure with interaction site coloured in purple. D) Rab8a structure surface with interaction site coloured in purple and amino acids commonly involved in protein-protein interactions highlighted. Yellow represents cysteine, blue represents lysine, red represents asparagine, and green represents arginine.

structure and the region known to interact with STX5. Cysteine is known to form hydrogen bonds and disulfide bridges, within the N-terminus (21%) there were 25% of Cog4 cystines. Asparagine has a high propensity to hydrogen bond, 24% of the total asparagine in Cog4 were found in the N-terminus. Arginine is frequently involved in protein active or binding sites, within the N-terminus 23% of Cog4 arginines were found. Cysteine, Asparagine and Arginine were all more prominent in the N-terminus than the rest of Cog4. However, for lysine which promotes protein stability through several interactions, there were only 17% found at the N-terminus (Figure 6B). To determine the significance of the proportion of amino acids commonly involved in protein interactions found on Cog4 the amino acids were also identified on Rab8. The universal interaction site of Rabs has been identified on Rab8, which consists of a hydrophobic binding pocket and a triad of aromatic amino acids (Rai et al., 2016). For proportion calculations the interaction site was defined as the site around the hydrophobic binding pocket and a triad of aromatic amino acids, this site covered 7% of the total Rab8 protein (Figure 6C). All but one of the amino acids involved in protein interactions made up more than 7% of the Rab8 interaction site, Asparagine (11%), cysteine (0%), arginine (18%), lysine (11%) (Figure 6D). Taken together these suggest the Cog4 N-terminus has more capability to take part in binding events than the rest of Cog4. There appears to be a collection of Asparagine and Arginine around the start of the long helix (Figure 6B). This suggests this may be an active site for protein binding and could potentially be involved in binding multiple proteins.

Previously, Cog4 mutants found to have altered interactions with Rabs were identified in the Ungar lab. Cog4-L36P has a reduced interaction with GTP-Rab30 (Spencer, 2022). A Cog4 mutant was found to have an enhanced interaction with Rab41, there are three mutations that may be responsible, either alone or in combination. The three mutations are R34C, Q67R and F137S (unpublished) (Figure 7A). These mutations are found on the N-terminus of Cog4, this suggests the N-terminus may be the site of these Rab interactions. To investigate the significance of these mutations the conservation of each position was assessed. Overall residue conservation in the N-terminus appears to increase as residue number increases. Most of the Rab interaction mutation positions were conserved, L36 scored 10, suggesting this is an important position for Cog4 stability or function. Whereas conservation of the other positions was lower, Q67 and F137 scored six and five respectively however, R34 has a

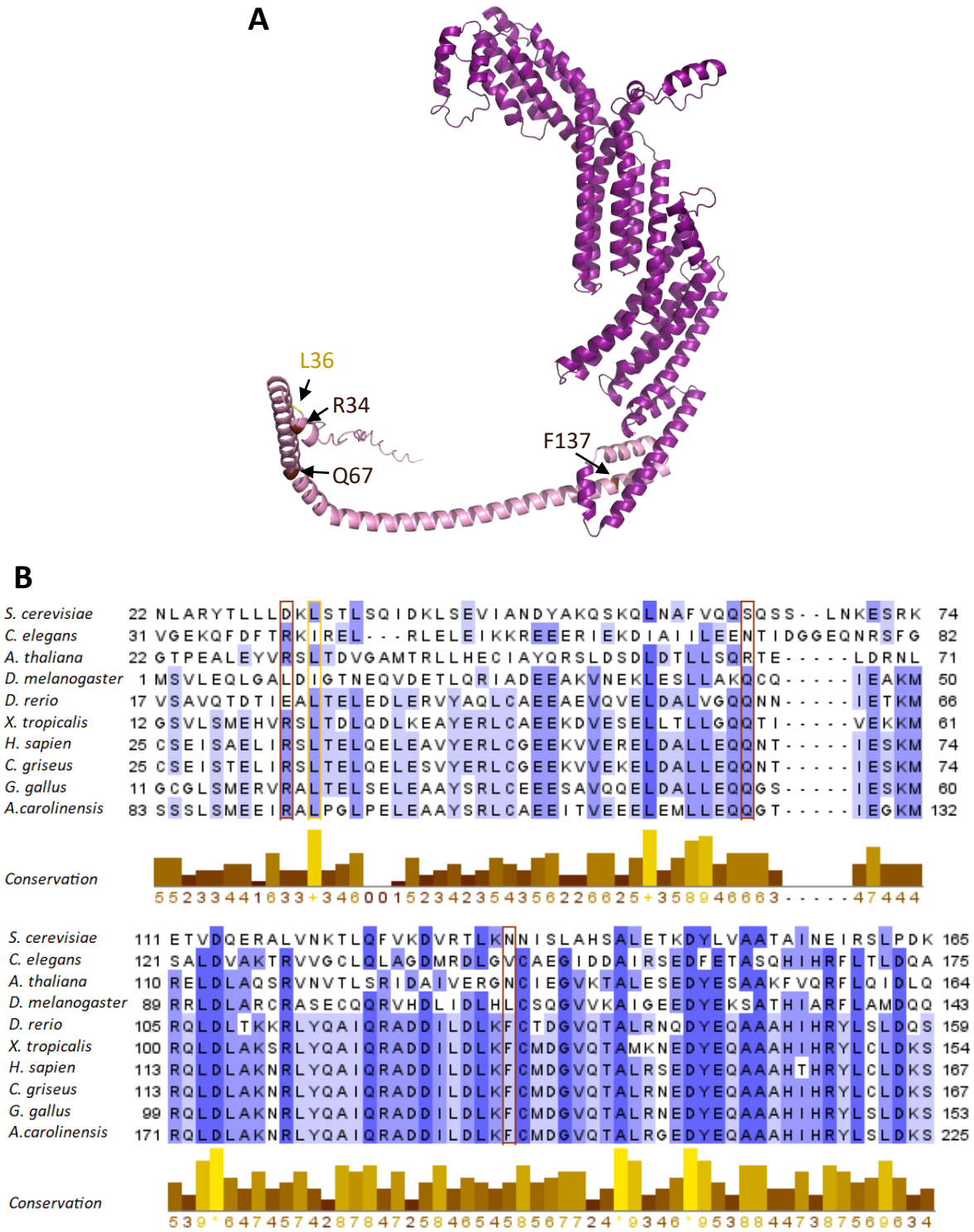

Figure 7 – Positions of Cog4 mutations linked to Rab binding and conservation. A) Cog4 structure labelled with position of amino acids that when mutated are linked to Rab30 binding (L36) and Rab 41 binding (R34, Q67, F137). B) ClustalW multiple sequence alignment (MSA) of Cog4 from model organism species and CHO cell line species: *H. sapiens, G. gallus, X. tropicalis, D. rerio, D. melanogaster, A. thaliana, C. elegans, C. griseus, S. cerevisiae and A. carolinensis.* MSA coloured by percentage match to consensus sequence. R34, Q67 and F137 were highlighted in brown boxes and L36 in a yellow box. Conservation score annotation is based on the number of conserved physico-chemical properties.

low conservation score of 3 (Figure 7B). These scores suggest the R34 could be mutated without altering the stability of Cog4 therefore may be involved in Rab41 binding directly. Alternatively, Q67 and F137 could be conserved because they participate in the Rab interaction and their mutation leads to a stronger interaction. Both R34C and Q67R mutations produce amino acids that are more likely to be involved in protein-protein interactions. This suggests these mutations could be causing the interaction between Cog4 and Rab41 by increasing the propensity of the N-terminus to interact.

Electrostatic interactions play an important role in protein-protein interactions, charged and polar residues have been suggested to enhance complex stability (Tsai et al., 1997). APBS was used to calculate the electrostatic surface potential of Cog4, GTP-Rab30 and GTP-Rab41 to assess if the N-terminus is more likely to interact with these Rabs. The universal interaction site of Rabs consists of a hydrophobic binding pocket and a triad of aromatic amino acids (Itzen and Goody, 2011). These sites were identified by aligning Rab8, which has known residues of these sites, with the Rab30 and Rab41 models (Figure 8A and 8B; respectively) (Rai et al., 2016). Rab30 has 15% positively charged residues and 12% negatively charged, with the interaction surface appearing mostly positively charged (Figure 8C). Rab41 has 30% of positive and negative residues, with more positively charged regions appearing at the interaction surface (Figure 8D). Cog4 has a majority negative surface potential, with some positive areas along the N-terminus (Figure 8E). Taken together these suggest the negatively charged Cog4 N-terminus could attract the positively charged interaction surfaces of Rab30 and Rab41. All the mutation positions linked to Rab binding were found in negatively charged regions of Cog4. As R34C results in the loss of the positive charge this would not alter the overall properties of the region. Q67R gains a positive charge which would increase the overall positivity of the N-terminus. Taken together these suggest the R34C mutation may be more favourable than Q67R for enhancing electrostatic interactions with Rab41 as R34C increases the overall negativity of the N-terminus therefore attracting Rabs.

To investigate the potential mechanism behind Cog4 mutants binding to Rabs, the Cog4 structure was docked to Rab30 and Rab41. Docking was completed using Hdock, where the interaction site of Rab30 and Rab41 previously described was specified (Figure 8A and 8B; respectively). The resulting docked models suggested many possible interaction positions on

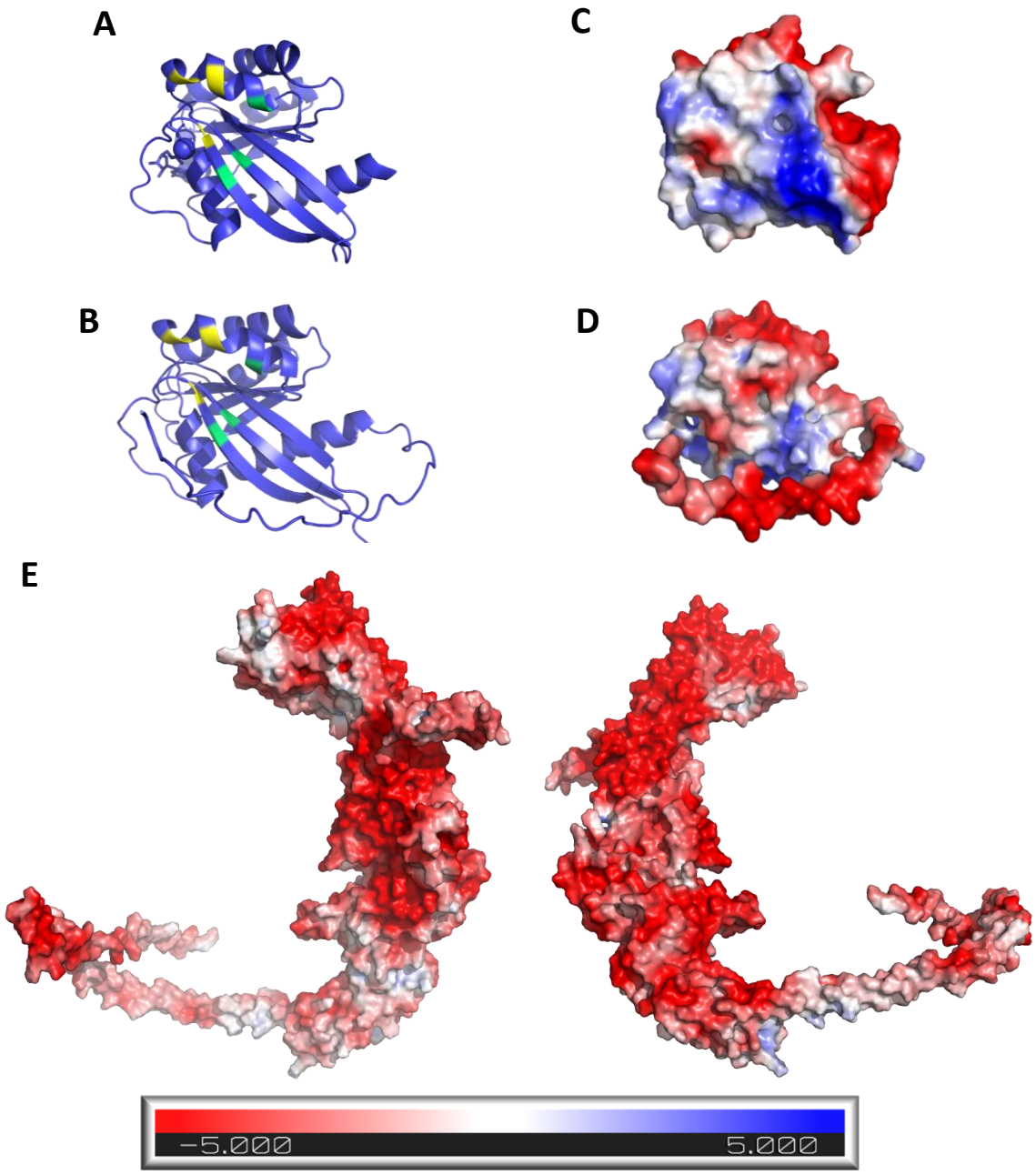

Figure 8 – Electrostatic potential of Rab30, Rab41 and Cog4 structures. A) Rab30 structure highlighted with residues involved in the hydrophobic pocket in yellow and triad of aromatic amino acids in green. C) Rab41 structure highlighted with residues involved in hydrophobic pocket in yellow and triad of aromatic amino acids in green. D) Electrostatic potential ±5 kT/e of Rab30. E) Electrostatic potential ±5 kT/e of Rab41. E) Electrostatic potential ±5 kT/e of Cog4.

Cog4 for Rab30 and Rab41 (Figure 9A and 9B; respectively). For the analysis of the docked Rabs the N-terminal helical region was defined as residues 1 to 165. The percentage of surface area taken up by the alpha helical region was 21% of the total surface area of Cog4. For the Rab30 models 28% bound to the alpha helical region, suggesting the binding of Rabs to the alpha helical region is not by chance. However, no Rab30 models bound around the

L36 residue (Figure 9A). Out of the models generated with the Rab41 structure 34% bound to the alpha helical region, out of these none around R34, 4% docked around Q67 and 12% around F137 (Figure 9B). This suggests Rab41 is more likely than Rab30 to bind the furthest end of the N-terminal helical region from the main body of Cog4 and that Rab41 may be directly interacting with the mutation(s) but L36 may not be the site of Rab30 interaction.

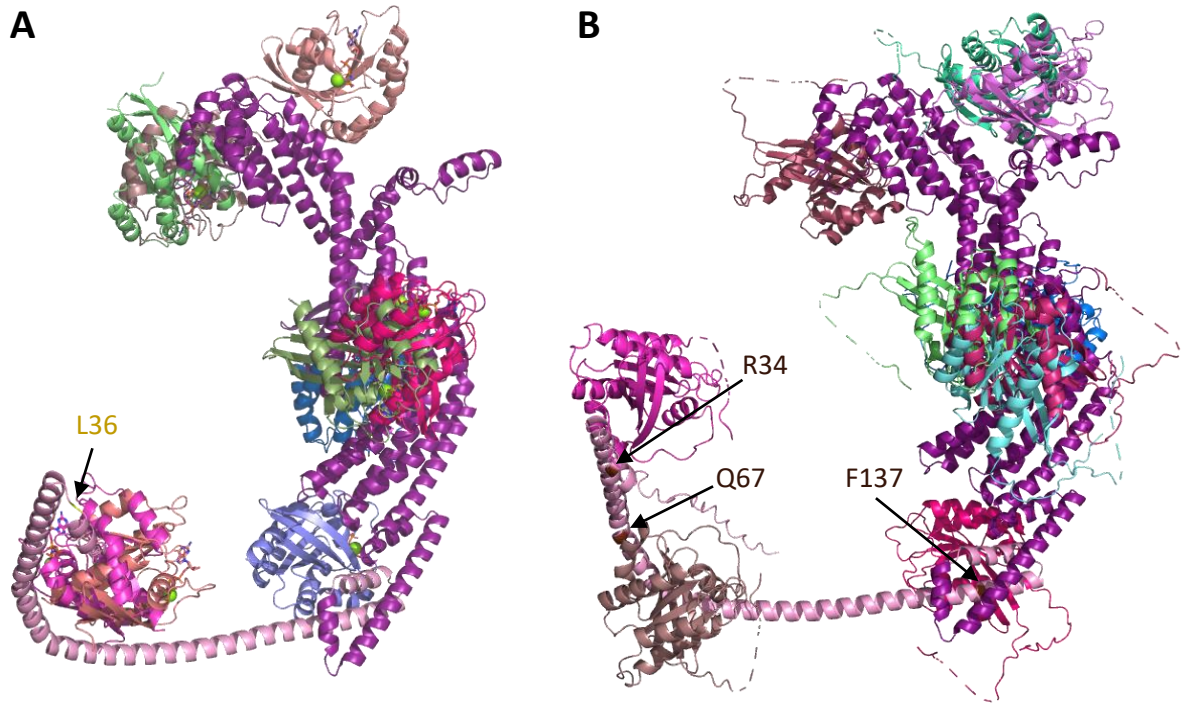

Figure 9 – Possible docking sites of Rab30 and Rab41 to Cog4. A) Rab30 docking at representative sites along the Cog4 structure. B) Cog4 with Rab30 docked at representative sites along the structure.

Stability of Cog4 mutants may indicate their effects on Rab interactions.

The stability of Cog4 mutants known to alter Rab interactions, L36P, R34C, Q67R, F137S, was investigated using DDMut. For each mutation Cog4 models displaying amino acids and their interactions within the structure were produced, alongside a ΔΔG value and based on that an assessment of whether the mutation is stabilising or destabilising the structure. The ΔΔG values indicate whether the overall stability is changing which might affect the ability of Cog4 to bind to Rabs. L36P had a ΔΔG of -2.05 kcal/mol and was destabilising. R34C had a ΔΔG of 0.07 kcal/mol and was stabilising, whereas Q67R and F137S had ΔΔG values of -0.02 kcal/mol and -0.29 kcal/mol respectively thus were destabilising.

In wild-type Cog4, L36 forms hydrogen bonds and hydrophobic interactions with residues on the coiled coil helices in this region (Figure 10A). When the leucine at position 36 was mutated to proline clashes occur with residues on the helices. Proline is known to break helices, suggesting the L36P mutation causes a loss of the coiled coil conformation (Betts and Russell, 2003). The L36P mutation also results in the loss of most hydrophobic interactions (Figure 10B). As hydrogen bonds and hydrophobic interactions are known to stabilise tertiary structure, a reduction in their number could explain a reduction in ΔΔG. This combination of changes to amino acid interactions due to L36P could suggest an increase in flexibility between the helices. An increased flexibility at the N-terminus could restrict Rab30 binding, suggesting a direct involvement of L36 in Rab30 binding or consequential change to the Rab30 binding site.

R34 is found in a smaller helix adjacent to the long helical region extending from the main body of Cog4 (Figure 10C). When arginine at position 34 was mutated to cysteine there was an increase in hydrogen bond when compared to wild type (Figure 10D). Hydrogen bonds are known to support alpha helices therefore stabilising secondary structure. The number of hydrogen bonds increases with the R34C mutation, suggesting it is more stable than wildtype. Q67 and F137 form hydrogen bonds with surrounding residues (Figure 10E and 10G respectively). When glutamine at position 67 was mutated to arginine there are fewer hydrogen bonds than wild type alongside the formation of an ionic bond (Figure 10F). Arginines are quite frequent in protein active or binding sites, suggesting this exposed arginine could be involved in Rab41 binding. When phenylalanine at position 137 was mutated to serine there was a loss of hydrophobic interactions (Figure 10H). As hydrophobic interactions are known to stabilise tertiary structure, a reduction in their number would explain a reduction in ΔΔG. The phenylalanine side chain is non-reactive and rarely involved in protein function, however it can play a role in substrate recognition. The hydroxyl group of serine is known to be reactive and can form hydrogen bonds with a variety of polar substrates, and serine is known to be involved in protein active sites (Betts and Russell, 2003). Taken together this could suggest F137S could alter an existing Rab binding site to allow for Rab41 binding. The Rab41 mutations had negligible effect on the overall stability of Cog4, this suggests they are directly involved in binding Rab41.

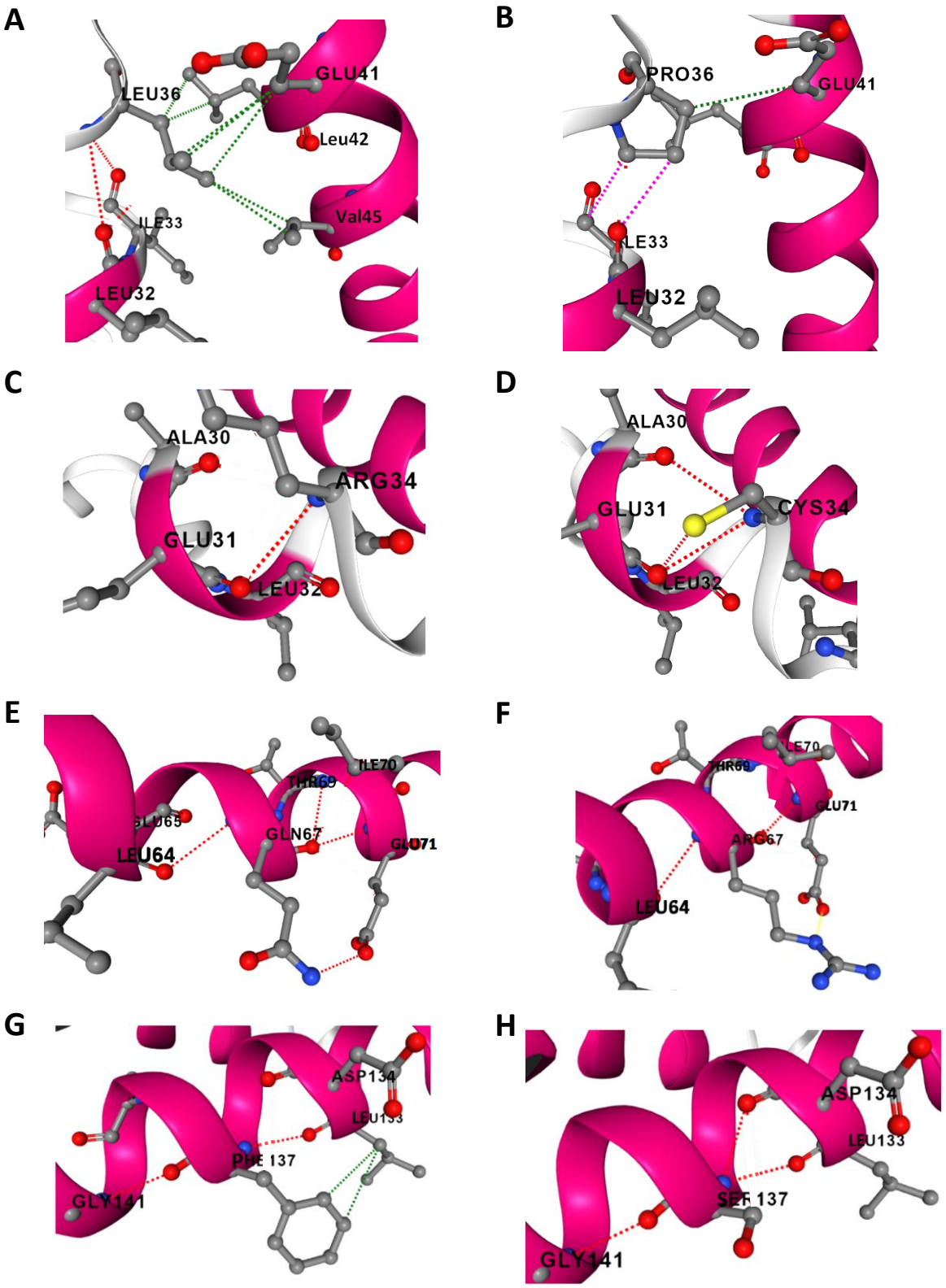

Figure 10 - Visualisations of bonds found in WT-Cog4 and when each Rab binding mutation was introduced. Cog4 structure coloured by secondary structural features; pink indicates helix. Interactions are denoted by dotted lines, pink denotes a clash, green represents hydrophobic interactions, red denotes hydrogen bonds, and yellow represents ionic bonds. A) WT-Cog4 showing L36 interactions. B) Mutant Cog4 showing P36 interactions. C) WT-Cog4 showing R34 interactions. D) Mutant Cog4 showing R34C interactions. E) WT-Cog4 showing Q67 interactions. F) Mutant Cog4 showing Q67R interactions. G) WT-Cog4 showing F137 interactions. H) Mutant Cog4 showing F137S interactions.

Summary

In this chapter a computationally generated Cog4 structure was used to investigate Cog4 mutations that alter Cog4 levels or Rab binding ability. The stability of Cog4 mutants was analysed using DDMut. Mutations that have been linked to CDGs, R729W and L773R, had similar destabilising effects on stability. Lower levels of Cog4 and other subunits in R729W patient cells may be due to the position of R729 on the exterior of the Cog4 structure possibly being involved in interactions with other subunits. STRING analysis showed Cog4 interacts with various Golgi resident proteins, mainly Rabs and SNAREs. The N-terminus of Cog4 is known to interact with STX5 and Sly1 (Laufman et al., 2009). Docking was conducted using Hdock to identify possible binding sites of Rabs on the Cog4 structure. This suggested the Rab structures preferred binding to the N-terminus of Cog4. The percentage of surface area taken up by the alpha helical region was 21% of the total surface area of Cog4. For the Rab30 models 28% bound to the alpha helical region however, no Rab30 models bound around the L36 residue. Out of the models generated with the Rab41 structure 34% bound to the alpha helical region, out of these 16% docked to a mutation site. This suggests Rab41 is directly interacting with the mutation positions but L36 may not be directly involved in the Rab30 interaction. Mutations altering Rab binding were also assessed for their effect on stability. Cog4-L36P was found to destabilise the structure, suggesting the L36 position is not directly involved in Rab30 binding. Whereas Rab41 interaction associated Cog4 mutations have low ΔΔG values suggesting negligible effect on the overall stability of Cog4. This suggests these mutation positions may be directly involved in binding Rab41.

Chapter 2: Analysing effects of Cog4 mutations on Herceptin glycosylation

Introduction

Mutations of Cog4 identified in patients with congenital disorders of glycosylation (CDGs) can cause defects in glycosylation. L773R and R729W Cog4 mutations are known to reduce levels of Cog4 which can alter COG complex function in aiding membrane tethering in the Golgi (Shestakova et al., 2006; Richardson et al., 2009; Ng et al., 2011). *Medial*-Golgi resident enzymes are incorrectly localised in COG subunit knockdowns (Shestakova et al., 2006). R729W is known to alter sialylation and galactosylation in patient cells (Reynders et al., 2009). The L36P Cog4 mutation was identified in the Ungar lab and found to have altered whole cell glycosylation. Previous transfections of the Cog4-L36P stable CHO cell line resulted in increased production of Herceptin when compared to a wild type cell line (Spencer, 2022). Therefore, this chapter aims to investigate how Cog4 mutations L36P, R729W and L773R influence glycosylation of mAbs. CHO cell lines containing Cog4-WT (wild type), Cog4-KO (Knock out) and mutant Cog4 were transiently transfected with Herceptin. The Herceptin produced in these cell lines was glycan profiled, allowing the effects of Cog4 mutations to be compared to Cog4-WT and Cog4-KO.

Molecular Biology methods

E. coli transformation

Competent DH5α *E. coli* cells (New England Biolabs C2987) were transformed with plasmids expressing GFP (Green fluorescent protein) (pcDNA3-eGFP originally purchased from Clontech (pEGFP-C3)) and Herceptin (Cobra Biologics) for amplification. The Herceptin plasmid was from Cobra Biologics. pcDNA3-eGFP plasmid was originally purchased from Clontech (pEGFP-C3). Transformation tubes were pre-cooled on ice, whilst *E. coli* was defrosted on ice for 10 minutes. 50 µL of *E. coli* and ~100 ng of plasmid DNA were added to pre-cooled tubes and incubated on ice for 30 minutes. Bacteria were then heat shocked for 90 seconds at 42°C and returned to ice for 5 minutes. 800 µl to 1 ml of Lysogeny Broth (LB) media was added to each tube before shaking incubation at 37°C for 1 hour. Following this

the *E. coli* were plated on LB plates containing ampicillin (2% Agar, 2% Lysogeny Broth, 100 mg/ml Ampicillin) and spread with glass spreader. Plates were incubated at 37°C overnight.

DNA purification

For mammalian cell transfection, plasmid DNA was extracted from *E. coli* using a Qiagen midi-prep plasmid purification kit (Qiagen 12123). 5 mL of selective LB was inoculated with *E. coli* for overnight culture. The following day overnight cultures were set up with 100 mL of LB media inoculated with 500 µL E. coli from the initial 5 mL overnight culture. DNA was then extracted following the Qiagen midi-prep plasmid purification kit protocol. Concentration and purity of extracted DNA was measured using a nanodrop.

Herceptin production methods

Cell lines

CHO-K1 cell lines containing Cog4 variants Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R were used for Herceptin production. Cog4-KO-WT and Cog4-KO-L36P (Cog4-WT and Cog4-L36P respectively) cells were engineered and adapted to suspension (Spencer, 2022). The Cog4-KO cell line was engineered by Estere Seinkmane, and the Cog4-R729W and Cog4-L773R cell lines were engineered by Will Townley. The Cog4-KO, Cog4-R729W and Cog4-L773R cell lines were adapted to suspension by Ben West. The Cog4 protein levels were the same in Cog4-WT cells as Cog4-L36P, Cog4-R729W and Cog4-L773R cells.

Suspension cell culture

Suspension cells were cultured in FortiCHO medium (Gibco A1148301), supplemented with 8 mM GlutaMAX (ThermoFisher 35050061) and 1% v/v pen/strep. Frozen stocks of Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cell lines were thawed in a 37°C water bath, then 1/1.5 mL thawed cell stock was added to T25 with 7 mL media. The following day cultures were transferred to a 15 mL falcon tube and centrifuged at 1,000 xg for 5 minutes, the supernatant containing DMSO (Dimethylsulfoxide) was discarded. Cell pellets were then resuspended in fresh media.

Cell culture flasks were incubated at 37 °C in a humidified incubator with 5% CO_2 on an orbital shaker at 80 rpm. Culture cell density and viability were regularly monitored using a

Vi-Cell XR automatic cell counter (Beckman Coulter), 300 µL cells were incubated with 300 µL TrypLE (Gibco 12604021) for 5 mins at 37°C before counting. Cells were cultured in T25 flasks and were passaged when cells grew up to $3x10^6$ cell/mL then re-seeded at a density of $2.5x10^5$ cells/mL. When cultures had reached 90% viability cells were cultured in Erlenmeyer 125 mL flasks with 30 mL media for at least two passages prior to transfection.

Transient transfections with Herceptin plasmid

The day prior to transfection, Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cells were seeded to a density of 1 x 10^6 cells/mL in 15 mL OptiPRO serum free media (ThermoFisher 12309019) containing 8 mM GlutaMAX (ThermoFisher 35050061). On the following day cells were transfected using 25 µL FreeStyleMAX (ThermoFisher 16447100) transfection reagent diluted in 600 µL OptiPRO containing 8 mM GlutaMAX, the same volume of media was used to dilute 18 µg Herceptin and 18 µg GFP DNA. Immediately following dilution, the DNA and FreeStyleMAX solutions were mixed with a 1:1 (v:v) ratio, after which the mixtures were incubated at room temperature for 15 minutes. The transfection mixtures were then added to the cells prepared the previous day. For the following 5 days samples were collected from the cultures to monitor viability, amount of Herceptin production and transfection efficacy using GFP transfected cells. On the last day, the remaining media was collected and frozen.

Enzyme-linked immunosorbent assay (ELISA) for Herceptin

Samples were collected from transiently transfected cultures and diluted in PBS to be analysed by ELISA. Serial dilutions of human IgG standards (Sigma Aldrich I3506), ranging from 20 ng to 100 ng in PBS were also prepared. Pierce™ protein A-coated white 96-well plates (ThermoFisher 15130) which were pre blocked with SuperBlock blocking buffer (ThermoFisher 37515) were washed three times with PBST. Samples and standards were loaded to the coated plates in triplicate. Herceptin was recognised by HRP-conjugated (Horseradish peroxidase) chicken polyclonal antibody for human IgG (Abcam 112454), diluted 1:1000 in SuperBlock blocking buffer (containing 0.05% Tween-20) and detected with Supersignal ELISA pico luminol detection agent (ThermoFisher 1859677). Incubation steps were completed at room temperature on an orbital shaker for 1 hour. Plates were read on a CLARIOstar (BMG LABTECH) plate reader at 425 nm. Measurements from samples

were calibrated against a standard curve from human IgG standard measurements. Cell counting was conducted using Vi-Cell XR – Cell Counter (Beckman Coulter) and used to calculate Herceptin production per cell. Amount of Herceptin in samples taken for purification was also calculated. A graph was generated and the Kruskal Wallis statistical test was conducted using R Statistical Software (v4.3.1; R Core Team, 2023) (Wickham, 2011; Wickham, 2016; Wickham et al., 2019; Shuangbin Xu et al., 2021; Pedersen, 2022; Giraudoux, 2023; Ushey et al., 2023; Wickham et al., 2023; Wickham, Vaughan and Girlich, 2023). The Kruskal-Wallis test was chosen as the group sizes can vary.

Purification of Herceptin

Supernatant containing Herceptin was collected on day five of the transient transfection. Herceptin was purified from supernatant samples using PreDictor™ Robocolumn™ MabSelect Sure™ (GE Healthcare 28-9861-07) protein A columns. Samples were loaded to the column, followed by three wash steps, and then eluted in 50 mM Sodium Acetate (pH 3.5). Herceptin concentration after purification was measured using Lunatic UV/Vis Absorbance Spectrometer (Unchained Labs) with BSA (Bovine serum albumin) standard. Purified Herceptin was then concentrated using Amicon Ultra - 0.5 mL 10 kDa Centrifugal Filters (Merck UFC501024) at 14000 rcf for 12 minutes total followed by 2 minutes with the filter inverted to elute sample. Amount of Herceptin following purification was calculated and compared to amount of Herceptin prior to purification, a graph representing this comparison were generated. The Kruskal Wallis statistical test was conducted, and graph was made using R Statistical Software (v4.3.1; R Core Team, 2023) (Wickham, 2011; Wickham, 2016; Wickham et al., 2019; Shuangbin Xu et al., 2021; Pedersen, 2022; Giraudoux, 2023; Ushey et al., 2023; Wickham et al., 2023; Wickham, Vaughan and Girlich, 2023). The Kruskal-Wallis test was chosen as the group sizes can vary.

Glycan profiling of Herceptin

Glycans were released from concentrated purified Herceptin and labelled with RapiFluor using RapiFluor-MS *N*-Glycan Kit (Glycoworks 176003606). Labelled glycan samples were run on a BioAccord liquid chromatography coupled with mass spectrometry (LC-MS) System (Waters). Sample integration was performed using UNIFI software (Waters) calibrated using Dextran standards, glycan peaks were assigned using mass confirmation and theoretical GU

value. Profiles were collected for each cell line from the three transfection experiments and the abundance of each glycan identified was calculated. Glycan abundances were then normalised to 100% and the average and error from the three transfection experiments were calculated. Graphs were generated using R Statistical Software (v4.3.1; R Core Team, 2023) (Wickham, 2011; Wickham, 2016; Wickham et al., 2019; Shuangbin Xu et al., 2021; Pedersen, 2022; Firke, 2023; Giraudoux, 2023; Signorell, 2023; Ushey et al., 2023; Wickham et al., 2023; Wickham, Vaughan and Girlich, 2023).

Results

Herceptin production was not significantly different across Cog4 cell lines.

The cell lines used to produce Herceptin in transient transfection experiments were Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R. Transfection experiments were repeated three times and samples were collected on the final day of transfection to be used to calculate Herceptin production per cell from ELISA measurements (Figure 11). For two samples, from COG4-WT and COG4-L36P cell lines, ELISA measurements were too low to calculate the amount of Herceptin present in the samples accurately. There was no significant difference in Herceptin production between cell lines (Kruskal: 2.29; df 4; p-value 0.68). Mean Herceptin production of COG4-R729W and COG4-L773R cell lines appears a little higher than WT, suggesting these Cog4 mutations could increase Herceptin production (Figure 11). There was a large variation in the mean Herceptin produced in COG4-L36P cells, therefore this data cannot confirm whether COG4-L36P cells have increased production of Herceptin compared to wild type cells.

To examine Herceptin glycan profiles supernatant from the five cell lines was collected on the final day of the transfection. Herceptin was purified from supernatant with protein A, concentrated, deglycosylated and labelled, and profiled by LC-MS. To determine the efficiency of the Herceptin purification, the amount of Herceptin was calculated before and after purification and recovery rate was calculated. The amount of Herceptin before purification was calculated from ELISA measurements. After protein A purification and concentration of Herceptin by centrifugation the amount of Herceptin was calculated from UV/Vis Absorbance Spectrometer measurements. From the Herceptin amounts the recovery

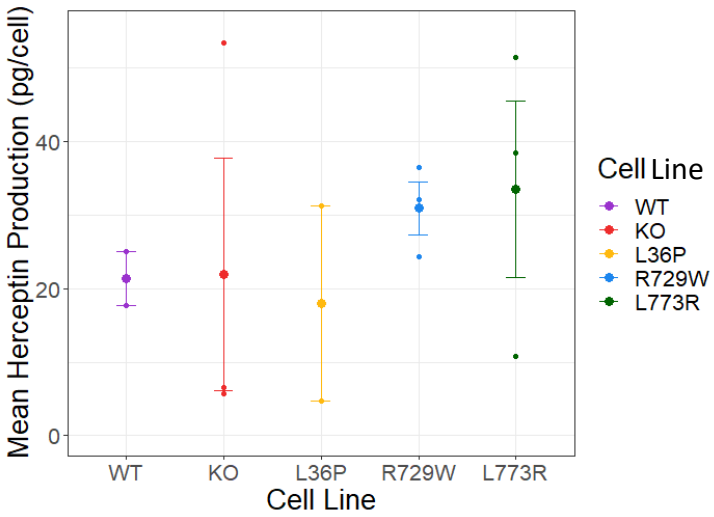

Figure 11 - Mean Herceptin production. Quantification of Herceptin production from transient transfection experiments using Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cell lines. Production of Herceptin was calculated per cell (pg/cell) using ELISA measurements of culture supernatant on the final day of transfection. Three biological repeats are represented for all cell lines except WT and L3P which have two biological repeats. Standard error of the mean was used for error bars. No significant difference was found between cell lines (Kruskal: 2.29; df 4; p-value 0.68).

rate of the purification and concentration procedure was calculated (Figure 12A). Recovery rates could not be calculated for the two samples, from COG4-WT and COG4-L36P cell lines, where ELISA measurements were too low to calculate the amount of Herceptin present in the samples accurately. Recovery rate for all COG4-WT and COG4-L36P samples was lower than other cell lines, however no one cell line had significantly better recovery than others (Kruskal: 6.17; df 4; p-value 0.18). Alternatively, recovery rate could be linked to the amount of Herceptin at the start of the purification process. Purified Herceptin was plotted against unpurified Herceptin, this shows more unpurified Herceptin results in more purified Herceptin (Figure 12B). The minimum amount of Herceptin needed for glycan profiling using LC-MS was 0.015mg, 11 of the samples passed that criterion.

Cog4 mutant cell lines had varying effects on Herceptin glycosylation.

Glycan profiles were generated from Herceptin from the 11 samples with enough Herceptin for profiling produced in cell lines Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R. Glycan profiles were produced by LC-MS analysis of purified and concentrated Herceptin. The percentage abundance of glycans found on Herceptin from each cell line

A

Cell line	Amount of Herceptin		Percentage recovery rate of purification
	Unpurified (mg)	Purified (mg)	
WT 1	0.431	0.039	9%
WT 2	0.051	0.002	4%
WT 3		0.007	
KO 1	0.791	0.116	15%
KO 2	0.143	0.015	10%
KO 3	0.545	0.080	15%
L773R 1	0.406	0.072	18%
L773R 2	0.259	0.030	12%
L773R 3	0.439	0.056	13%
R729W 1	0.866	0.135	16%
R729W 2	0.411	0.028	7%
R729W 3	0.710	0.059	8%
L36P 1	0.623	0.043	7%
L36P 2		0.002	
L36P 3	0.073	0.006	8%

B

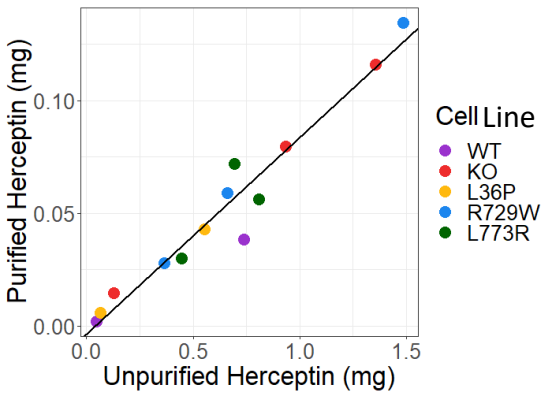

Figure 12 - Amount of Herceptin before and after purification. Quantifications for Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cell lines from three transient transfection experiments. Amount of Herceptin before purification was calculated from ELISA measurements of culture supernatant on the final day of transfection. Purified Herceptin amount was calculated from UV/Vis Absorbance Spectrometer measurements following protein A purification and concentration by centrifugal filter. Three biological repeats are represented for all cell lines except WT and L3P which have two biological repeats. A) Recovery rate was calculated from the amount of Herceptin before and after purification. B) Purified Herceptin was plotted against unpurified Herceptin, and a line of best fit was plotted at y = 0.087217x + -0.003305. No significant difference in recovery was found between cell lines (Kruskal: 6.17; df 4; p-value 0.18).

were averaged from the two or three transfection experiments. To determine the types of glycans found on Herceptin from each cell line glycans were categorised based on monosaccharide composition and number of antennae and the amount of glycans in each category was calculated (Figure 13). For example, an oligomannose glycan will also be either biantennary or triantennary and the amount of this glycan would contribute to the total of each category. Partially processed glycans are $GlcNAc_3Man_3$ structures that have been processed without having the second antennae GlcNAc. Even though there were predominantly biantennary glycans on Herceptin from each cell line, mutant Cog4 cell lines produced Herceptin with fewer biantennary glycans than Cog4-WT with Cog4-KO Herceptin having the lowest abundance of biantennary glycans. Most of the glycans found on Herceptin also have a core fucose modification, again mutant Cog4 Herceptin glycans have less core fucose than Cog4-WT with Cog4-KO Herceptin having the lowest. Herceptin

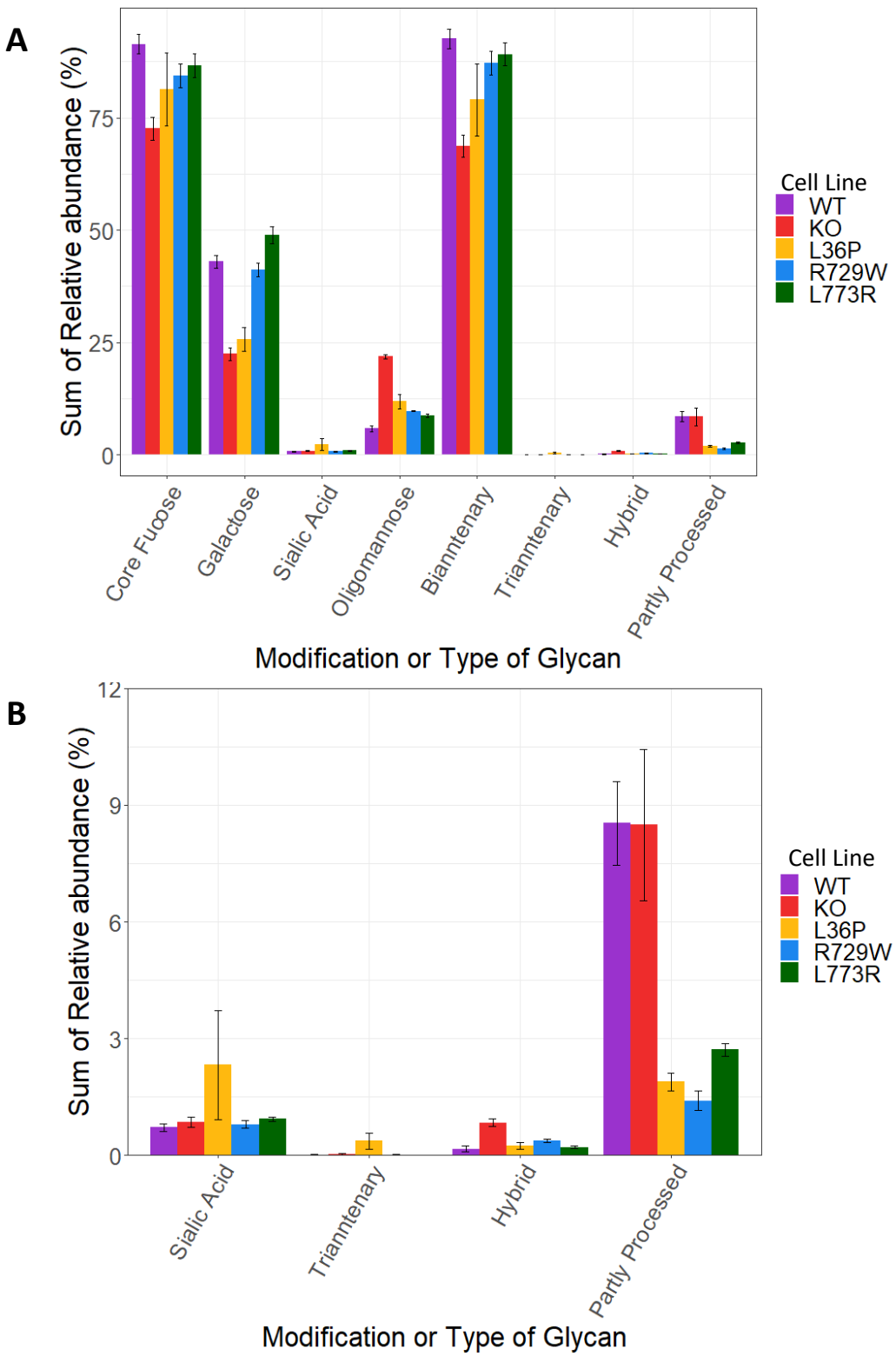

Figure 13 - Types of glycans found on Herceptin produced in different Cog4 cell lines. Sum of relative abundances of the different types and modifications of Glycans were calculated from glycan profiles of Herceptin produced in three separate transient transfection experiments using Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cell lines. Modifications include core fucose, galactose, and sialic acid. Types of glycans include Oligomannose, Biantennary, Triantennary, Hybrid and partly processed. Standard error of the mean was used for error bars.

produced in Cog4-KO and Cog4-L36P cell lines appear to have fewer glycans with galactose modifications than Herceptin from Cog4-WT cells. Cog4-KO cells also produce Herceptin with higher abundance of oligomannose glycans than Cog4-WT cells (Figure 13A). Herceptin produced in Cog4-L36P cells has more sialic acid than Cog4-WT cells. More triantennary glycans were found on Herceptin produced in Cog4-L36P cells and more Hybrid glycans were found on Herceptin produced in Cog4-KO cells than Cog4-WT cells (Figure 13B). Glycans from Herceptin produced in Cog4-KO cells are more likely to be oligomannose or hybrid and less likely to be biantennary, this suggests Cog4-KO cells have less mannose cleavage by mannosidase enzymes which are localised to the *cis*-Golgi and *medial*-Golgi. Cog4-KO cell lines produce fewer galactosylated glycans on Herceptin than Cog4-WT, but levels of sialylation were within error of the Cog4-WT amount. Galactosylation precedes sialylation and these occur in the *trans*-Golgi. Therefore, a decrease in galactosylation but not sialylation could be the result of fewer glycans being available to be modified by galactosylation enzymes suggesting the *medial*-Golgi localised enzymes may not be recycled as effectively in Cog4-KO cells. Cog4-L36P cells produce Herceptin glycans with more sialic acid than Cog4-WT cells, but galactose levels are the same. Cog4-L36P cells also produce fewer biantennary glycans than Cog4-WT cells. Taken together this could suggest the L36P mutation causes reduced recycling of glycosylation enzymes from the *trans*-Golgi to the *medial*-Golgi. The types of glycans found on Herceptin produced in Cog4-L773R and Cog4-R729W cells closely resemble Cog4-WT. However, Herceptin produced in Cog4-L773R cells were more likely to have galactosylated and sialylated glycans than Cog4-WT cells. This suggests the L773R mutation causes increased recycling of *trans*-Golgi enzymes.

Cog4 mutant cell lines cause glycosylation defects.

The differences of individual glycans produced by the Cog4 cell lines can indicate changes to enzyme sorting because of Cog4 mutation or knockout. To investigate potential enzyme sorting changes Herceptin profiles from each cell line were examined. Cog4 mutations result in altered glycan profiles of Herceptin, some mutations also caused large differences in the most abundant glycans (Figure 14). Glycan processing by glycosylation enzymes is dependent on the abundance of glycans to be processed. Therefore, it is unnecessary to determine if there is any statistical difference in glycans produced in different mutant cell lines as difference may be from the production or use of a glycan. Instead glycans should

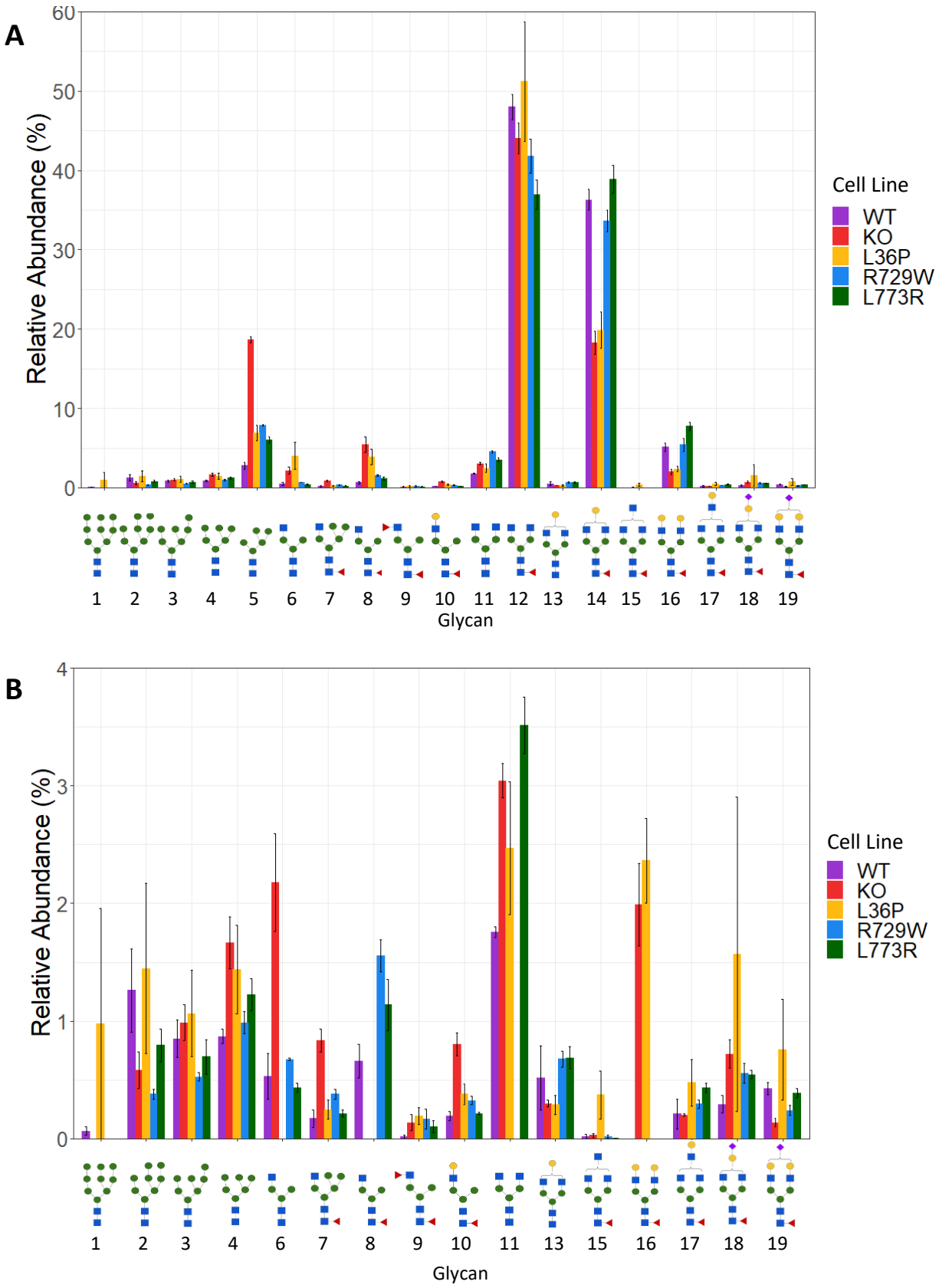

Figure 14 - Glycan profiles of Herceptin produced in different Cog4 cell lines. Herceptin was glycan profiled from three separate transient transfection experiments using Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R cell lines. Relative abundances were determined for the three experiments normalised to 100% and averaged for each cell line. Standard error of the mean was used for error bars.

be compared on a basis of how they were produced or used and the associated glycans should be assessed to determine how the Cog4 mutations are altering abundances.

Cog4-KO produced a greater abundance of $GlcNAc_2Man_5$ and smaller glycans, such as glycans 6, 7 and 8, than Cog4-WT (Figure 14A). This does not suggest lower mannosidase activity instead, but it suggests recycling of glycosylation enzymes beyond the *cis*-Golgi may be compromised. Galactosylated glycans 13, 14 and 16 have lower levels in Cog4-KO than Cog4-WT (Figure 14). Levels of sialylated glycan 18 were higher in Cog4-KO cell line than WT, whereas levels of sialylated glycan 19 were lower (Figure 14B). Taken together this suggests the *medial*-Golgi and *trans*-Golgi enzymes are less effectively recycled. Cog4-L36P cells produce more of glycans $GlcNAc_2Man_5$, $GlcNAc_2Man_6$, and $GlcNAc_3Man_3Fuc$ (Glycans 5, 6, and 8 in Figure 14), which suggests reduced recycling of glycosylation enzymes in later Golgi cisternae. Galactosylated glycans 13, 14 and 16 were at lower levels from Cog4-L36P cell lines than Cog4-WT (Figure 14). Abundances of glycan 18 and 19 from Cog4-L36P cells were higher but within error of Cog4-WT cells (Figure 14B). Taken together this suggests the L36P mutation causes decreased recycling of *medial*-Golgi glycosylation enzymes.

Cog4-R729W and Cog4-L773R cell lines produce very similar glycan profiles. However, Glycans 6, 8, 10 and 11 were more abundant from Cog4-R729W cells whereas glycans 14, 16, 17 and 19 were more abundant from Cog4-L773R cells. Glycans 14, 16, and 17 containing galactose modifications were also more abundant in Cog4-L773R cell lines than Cog4-WT (Figure 14). Preceding these glycans in processing are the galactosylated glycans 10 and 13, which have similar abundances in Cog4-WT and Cog4-L773R cell lines (Figure 14B). Taken together this suggests an increase in recycling of galactosylation enzymes to the *trans*-Golgi. Biantennary glycans can be formed from glycans 6 and 8. These glycans were more prevalent from Cog4-R729W cells than Cog4-L773R cells. From glycan 6 the biantennary glycan 11 is formed, this was more prevalent from Cog4-R729W cells than Cog4-L773R cells, and both had more than Cog4-WT cells (Figure 14A). This suggests these mutant cells process $GlcNAc_4Man_3$ less than Cog4-WT cells, however as Cog4-L773R cells have more abundance of glycan $GlcNAc_4Man_3GalFuc$, $GlcNAc_4Man_3Gal_2Fuc$ and $GlcNAc_4Man_3GalSiaFuc$ (Glycans 14, 16, and 18 in Figure 14), this could suggest increased galactosylation enzyme recycling. From glycan 8 the biantennary glycan 12 is formed, this was more prevalent from Cog4-WT cells than Cog4-R729W cells than Cog4-L773R cells

(Figure 14). This suggests that $GlcNAc_3Man_3Fuc$ (Glycan 8 in Figure 14) processing is not as efficient particularly in Cog4-R729W cells. Taken together these suggest the R729W mutation causes decreased recycling of *medial*-Golgi glycosylation enzymes.

Summary

In this chapter the effect of Cog4 knock out and mutations L36P, R729W and L773R on glycosylation of mAbs was investigated. CHO cell lines Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R were transiently transfected with Herceptin. Herceptin production was found to be not significantly different across the cell lines, however most of the cell lines showed variation in production. Two cell lines had higher production than the other cell lines, with stable Herceptin expression these cell lines may prove to be significantly better producers. The variation in the quantification of production per cell in the Cog4-L36P cell lines was particularly large, and only two repeat experiments could be used to calculate the production. Therefore, this broad range of production rate could be the result of issues with the transfection of Herceptin in one of the experiments resulting in an outlying data point. Herceptin was purified by protein A and concentrated with centrifugal filters. The more Herceptin used for the purification process the better the recovery rate. Herceptin from each cell line was glycan profiled. Glycans GlcNAc4Man3Fuc and GlcNAc4Man3FucGal are most abundantly found on Herceptin made in all the cell lines (Figure 14). These glycans are typically found in high abundances on Herceptin (Yu, Ahn and Gilar, 2010). Therefore, the changes to glycosylation caused by Cog4 mutant cell lines may not alter the efficacy of the Herceptin product. However, the Cog4 mutations do appear to have various consequences on glycosylation enzyme sorting. Examining the different types and modifications of glycans from each cell line revealed potential decreases in *trans*-Golgi to *medial*-Golgi recycling of glycosylation enzymes in Cog4-KO and Cog4-L36P cell lines. This was supported by investigating the different glycans identified on Herceptin from these cell lines. Through glycan profiling the glycans produced in Cog4-WT and Cog4-L36P cell lines were found to be significantly different. Another difference identified was an increase in galactosylation in the Cog4-L773R cells from Cog4-WT cells, which suggests an increase in recycling of galactosylation enzymes to the *trans*-Golgi.

Discussion

Summary

Improvement of the production and efficacy of monoclonal antibodies (mAbs) is important. mAbs are glycosylated and glycans dictate mAb stability, functionality, and efficacy through influencing interactions with mAb biological targets and the immune system (Nose and Wigzell, 1983; Mimura et al., 2000; Wu et al., 2010). mAbs have heterogeneous glycan profiles and glycosylation can vary between mAb batches. Glycan heterogeneity is the result of glycosylation having no template and being dictated by glycosylation enzyme organisation in the Golgi. Glycosylation enzyme localisation within the Golgi cisternae is controlled by intra Golgi vesicle transport. The COG complex is a critical component of the glycosylation machinery, due to its involvement in the retrograde trafficking of glycosylation enzymes within the Golgi. The COG complex is known to interact with Rabs to coordinate Golgi trafficking (Miller et al., 2013). Understanding the factors that govern glycosylation and its machinery, is important for improving mAb production and efficacy. The work presented in this thesis has examined the role of mutations in COG complex subunit Cog4 in the glycosylation of mAbs. A computationally generated Cog4 structure was used to examine Cog4 interactions with Rabs and Cog4 mutant stability. CHO Cog4 mutant cell lines were transfected with Herceptin to determine how the Cog4 mutants affected mAb glycosylation.

In chapter one the functions of Cog4 mutations were investigated using a computationally generated Cog4 structure. Stability analysis using DDMut of Cog4 mutations linked to CDGs, R729W and L773R, suggested the lower levels of Cog4 and other subunits found in CDG patient cells is related to the destabilising effect of the mutations (Table 1). STRING analysis was conducted to examine what kind of interactions Cog4 is involved in. The N-terminus of the Cog4 structure is known to interact with STX5 and Sly1, both of which participate in Golgi vesicle trafficking (Shestakova et al., 2007; Laufman et al., 2009). The N-terminus of Cog4 was analysed for features involved in protein-protein interactions. Rab30 and Rab41 structures were docked to the Cog4 structure to determine where these were likely to bind and if the mutation positions could be directly involved in these interactions. Cog4 protein stability following mutations involved in Rab interactions was examined, this suggested L36P

Cog4 mutation	Cog4 level (as % of WT)	Other COG subunits effected	Known Glycosylation defects	Source	Effect on Stability of Cog4	ΔΔG (kcal/mol)	Glycsoylation defects seen in this study
L36P	100%	Not tested	No difference	(Spencer, 2022)	Destabalising	-0.23	N-acetylglucosamine, Galactosylation
R729W	20%	1, 2, 3, 5	Sialylation	(Reynders et al., 2009)	Stablising	0.34	N-acetylglucosamine
L773R	44%	-	Sialylation, Galactosylation	(Ng et al., 2011; Miura et al., 2005)	Destabalising	-1.41	Galactosylation

Table 1 - Summary of Cog4 mutations investigated in this study. For each mutation the levels of COG subunits in patient fibroblasts and whether the mutation effects levels of other COG subunits have been included. Glycosylation defects identified in patients with each of mutations are also included. Sources provided for Cog4 levels, subunits effected and known glycosylation defects columns. From this study stability of Cog4 and glycosylation defects have been included.

and F137S destabilised Cog4 structure, potentially effecting Rab binding through structural changes rather than being directly involved in binding (Table 2).

In chapter two the effect of Cog4 knock out (Cog4-KO) and mutations L36P, R729W and L773R on glycosylation of mAbs was investigated. CHO cell lines Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R were transiently transfected with Herceptin. Herceptin production was found not to be significantly different across the cell lines. Herceptin was purified by protein A and concentrated with centrifugal filters. The more Herceptin used for the purification process the better the recovery rate. Herceptin from each cell line was glycan profiled. Examining the different types and modifications of glycans from each cell line revealed potential decreases in *trans*-Golgi to *medial*-Golgi recycling of glycosylation enzymes in Cog4-KO and Cog4-L36P cell lines. Galactosylation was found to be increased in the Cog4-L773R cells compared to Cog4-WT cells, which suggests an increase in recycling of galactosylation enzymes to the *trans*-Golgi (Table 1).

Cog4 mutation	Effect on Rab binding	Effect on Stability of Cog4	ΔΔG (kcal/mol)
R34C	Increased Rab41	Stablising	0.08
L36P	Decreased Rab30	Destabalising	-0.23
Q67R	Increased Rab41	Stablising	0.12
F137S	Increased Rab41	Destabalising	-0.19

Table 2 - Summary of Cog4 mutations involved in Rab interactions investigated in this study. For each mutation effects on Rab binding and stability of Cog4.

Discussion of key results

Investigating Cog4 mutants using computationally generated structures.

Since the Cog4 structure used in this thesis is computationally generated the conclusions made about the areas of the structure not solved by X-ray crystallography would need further study to confirm. Although AlphaFold is confident in its prediction of the Cog4 structure, the interpretation of the PAE plot for Cog4 suggests the orientation of the N-terminal helix with respect to the rest of the protein is unreliable (Figure 3). As this N-terminal region is predicted as a long extension from the rest of the protein it could be that this is relatively flexible in its positioning to allow for proper function, for example in binding to STX5 and Sly1 (Shestakova et al., 2007; Laufman et al., 2009). The N-terminal domain of COG complex subunits has been aligned to subunits of other membrane tethering complexes, exocyst and GARP, revealing sequence similarities that suggest similar modes of action (Whyte and Munro, 2002). The yeast Rab sec4 recruits the exocyst tethering complex (Guo et al., 1999). Together these suggest the Rab binding nature of the Cog4 N-terminus could be shared with other membrane tethering complexes, potentially supporting a similar mode of action. In this thesis the Cog4 N-terminus was explored as a site of Rab binding, examining other tethering complex Rab binding could help identify a mode of action for Cog4.

R729W patient cells had 80% less Cog4 than wild-type cells and lower levels of Cog1 (25%), Cog2 (40%), Cog3 (25%), and Cog5 (40%) (Reynders et al., 2009). However, L773R patient cells have usual levels of other subunits and 66% less Cog4 than wild-type cells (Ng et al., 2011). When the CDG mutation L773R was identified it was noted that the deficiency in Cog4 caused by this mutation was not destabilising the COG complex, it was suggested this could be due to the mutation lacking a direct interaction with the other subunits (Ng et al., 2011). This is supported by the observation of the L773 position being on the interior of Cog4 (Figure 4B). Protein stability of Cog4 mutants was examined using DDMut which provides computational predictions of stability. Therefore, to confirm the interpretations made in this thesis the stability would need to be experimentally measured. R729W had a ΔΔG of -2.1 kcal/mol and a destabilising effect on the Cog4 structure (Figure 5). Lower levels

of Cog4 and other subunits in R729W patient cells may be due to the position of R729 on the exterior of the Cog4 structure possibly being involved in interactions with other subunits. Alternatively, the N-terminus participates in subunit interactions. A study elucidating the architecture of the yeast COG complex suggests the Cog4 N-terminus is involved in interacting with Cog1-3 (Ha et al., 2016). Therefore, the R729W mutation could be causing loss of function of Cog4 after lobe A has formed resulting in lobe A being targeted for degradation.

Previous studies of the effect of R729W on the structure of Cog4 suggest the mutation disrupts a salt bridge network which stabilises the structure in that region. It was also suggested that the replacement with tryptophan may cause additional stress on the region due to its inflexible bulky nature. This resulted in the hypothesis that the region around R729 is important for proper functioning of the COG complex and R729W affects the stability and/or positioning of the region (Richardson et al., 2009). Taken together with the observation of lower Cog4 and other subunit levels in R729W patient cells this suggests the R729W mutation could be changing the ability of Cog4 to interact with other COG complex subunits.

The observations that cysteine, asparagine, and arginine were more prominent in the N-terminus than the rest of Cog4 does not demonstrate the N-terminus is the site of interaction with other tethering factors (Figure 6B). However, it does support the already available evidence that the N-terminus is the site of protein-protein interactions. Residues 1 to 222 of Cog4 alone were found to interact with STX5 (Shestakova et al., 2007). This interaction site for STX5 was later refined to residues 84 to 153 of Cog4, and the SM protein Sly1 was found to interact with the first 83 residues of Cog4 (Laufman et al., 2009). Cog4-L36P has a reduced interaction with GTP-Rab30 (Spencer, 2022). Three N-terminal mutations, R34C, Q67R and F137S, were also found to have an enhanced interaction with Rab41 (unpublished). This suggests that Rab interactions may occur at the N-terminus like the STX5 and Sly1 interactions. The Cysteines, Asparagines and Arginines identified on the N-terminus could be incorporated into further studies of Rab and other tethering component interactions.

Most of the Rab interaction mutation positions were conserved, L36 scored 10, suggesting this is an important position for Cog4 stability or function. Whereas conservation of the

other positions was lower, Q67 and F137 scored six and five respectively however, R34 has a low conservation score of 3 (Figure 7B). These scores suggest the R34C is not directly involved in Rab41 binding and could be mutated without altering the stability of Cog4 therefore may not be involved in Rab41 binding. The species without conserved Cog4 at these positions are *D. rerio, D. melanogaster, A. thaliana* and *C. elegans* (Figure 7B). These are not vertebrates like the rest of the species in the sequence alignment suggesting the function of these amino acids may not be shared by species that are invertebrates. Rab41 is found in *D. rerio* but not in *D. melanogaster, A. thaliana* and *C. elegans.* Therefore, these positions may be conserved in vertebrates to prevent Rab41 binding, which is not necessary for invertebrates without Rab41. The positions of Q67 and F137 are conserved for *D. rerio* Cog4 which also has Rab41, this could suggest that Q67R and F137S are involved in Rab41 binding.

The percentage of surface area taken up by the alpha helical region was 21% of the total surface area of Cog4. For the Rab30 models 28% bound to the alpha helical region, suggesting the binding of Rabs to the alpha helical region is not by chance. However, no Rab30 models bound around the L36 residue (Figure 9A). Out of the models generated with the Rab41 structure 34% bound to the alpha helical region, out of these none around R34, 4% docked around Q67 and 12% around F137 (Figure 9B). This suggests Rab41 is directly interacting with the mutation positions but L36 may not be directly involved in the Rab30 interaction. With more robust findings on how Cog4 interacts with binding partners and the rest of the COG complex this docking could be further refined. However, modelling these interactions using Hdock is limited by rigid body docking which cannot account for molecules experiencing large conformational changes upon binding. To overcome this residue restraints during binding can be input into Hdock. Cog4 interactions with other COG subunits may also cause conformational changes, therefore further experimental studies could provide a greater understanding of Cog4 protein-protein interactions that could be applied to computational modelling. Post-translational modifications (PTMs) are not included on computational protein structures, and these can influence protein interactions. Therefore, any known PTMs should be accounted for with residue restraints in further modelling studies.

This would be followed with sections of the Cog4 being removed experimentally to confirm which Cog4 region Rabs were binding. The interaction site for STX5 is found at residues 84 to 153 of Cog4, and the SM protein Sly1 was found to interact with the first 83 residues of Cog4 (Laufman et al., 2009). Docking STX5 and Sly1 to Cog4 with the experimentally determined sites specified could allow the method of Cog4 interaction with partners to be refined and applied to Rab30 and Rab41 docking. Cog4 mutant structures could also be computationally generated to compare docking with Rab30 and Rab41 between mutant and wild-type Cog4 models. This would determine how the Cog4 mutations could be altering the possible Rab binding sites.

Examination of the stability of the Cog4 structure with mutations associated with Rab interactions may not help determine the sites of Rab binding however could indicate a change in the stability of Cog4 that affects Rab binding. L36P had the highest ΔΔG of -2.05 kcal/mol amongst the Rab interaction mutations. This value was determined as destabilising, suggesting lower levels of Cog4 would be found in cells with this mutation. In previous experiments with Cog4-L36P CHO cell lines these were found to have lower levels of Cog4 than Cog4-WT cells, but this was not a significant difference. (Spencer, 2022). To further refine the role of the Cog4 N-terminus the amounts of other COG subunits in the Cog4-L36P cells could be measured. R34C and Q67R had ΔΔG values of 0.07 kcal/mol and -0.02 kcal/mol respectively, whereas F137S had a ΔΔG of -0.29 kcal/mol. Rab41 interaction associated Cog4 mutations have low ΔΔG values suggesting negligible effect on the overall stability of Cog4. This suggests these mutation positions may be directly involved in binding Rab41. Taken together with the percentage of Rab41 docked at each mutation site this suggests Q67R or F137S as the interaction sites for Rab41.

Analysing effects of Cog4 mutations on Herceptin glycosylation

Herceptin was produced in CHO cell lines Cog4-WT, Cog4-KO, Cog4-L36P, Cog4-R729W and Cog4-L773R. Herceptin production was not significantly different across the cell lines, however most of the cell lines showed variation in the quantification of production per cell. A previous transfection using the Cog4-WT, Cog4-KO, and Cog4-L36P cell lines suggested the Cog4-L36P cell line had increased production of Herceptin compared to Cog4-WT (Spencer,

2022). This increased Herceptin production from the Cog4-L36P cell line was not observed in the transfection data presented in this thesis (Figure 11). The variation in the quantification of production per cell in the Cog4-L36P cell lines was particularly large, and only two repeat experiments could be used to calculate the production. Therefore, this broad range of production rate could be the result of issues with the transfection of Herceptin in one of the experiments resulting in an outlying data point. Repeating these transfection experiments would determine if either of the data points were outliers. Cell lines producing a higher yield of mAbs would be advantageous, however the efficacy of these mAbs would need to be assessed. One study compared high producer and low producer CHO cell line proteomes to identify the differences in growth, metabolism and secretion that may contribute to increased mAb production. The efficacy of mAbs produced in these cell lines was then compared. Despite the differences in these cell lines the mAbs they produced showed comparable biological activity. The proteins identified as differentially expressed in this study included *trans*-Golgi proteins such as a galactose transferase which was downregulated and Rab32 which was upregulated (Pérez-Rodriguez et al., 2021). The difference in these Golgi proteins suggests altering cells to increase mAb production could influence the availability of glycosylation enzymes, however in this study the altered glycosylation did not impact the efficacy of the mAb product. The glycosylated heavy chain of Herceptin produced from previous transfections of ΔCog4WT, ΔCog4L36P and Cog4-L36P cell lines was analysed by mass spectrometry. There was no difference in glycosylation found between ΔCog4WT and ΔCog4L36P, and ΔCog4L36P and Cog4-L36P (Spencer, 2022). Therefore, the changes to glycosylation caused by Cog4 mutation may not alter the efficacy of Herceptin. In this thesis Herceptin produced in Cog4-WT and Cog4-L36P cell lines was glycan profiled and a few differences were found between these profiles, mostly in low abundance glycans (Figure 14). The low abundance glycans would not have been detected with the less sensitive ESI-MS mass spectrometric approach used previously to compare the ΔCog4WT, ΔCog4L36P and Cog4-L36P cell lines (Spencer, 2022).

Purification efficiency and recovery rate were determined using the amount of Herceptin before and after purification. The amount of Herceptin before purification was calculated from ELISA measurements. Some of this data was varied within cell lines, for example the Cog4-KO and Cog4-L36P (Figure 12). Therefore, conducting a western blot with these

samples would validate the values determined by ELISA. After protein A purification and concentration of Herceptin by centrifugation the amount of Herceptin was calculated from UV/Vis Absorbance Spectrometer measurements. Herceptin was purified from supernatant therefore, UV/Vis Absorbance Spectrometer could not be used to determine unpurified Herceptin concentration as other proteins are present in the supernatant. Recovery rates of Herceptin from these experiments were low (Figure 12). Hence, having an accurate value for the unpurified Herceptin will aid in calculating if after purification there is the minimum amount of purified Herceptin needed for profiling. Therefore, for further glycan profiling experiments implementing the validation of the amount of Herceptin before purification would improve the accuracy of this value.

Purified and concentrated Herceptin from 11 samples had enough Herceptin to be glycan profiled using LC-MS. Glycans $GlcNAc_4Man_3Fuc$ and $GlcNAc_4Man_3FucGal$ are most abundantly found on Herceptin made in all the cell lines (Figure 14). These glycans are typically found in high abundances on Herceptin (Yu, Ahn and Gilar, 2010). Therefore, the changes to glycosylation caused by Cog4 mutant cell lines may not alter the efficacy of the Herceptin product. The Cog4 mutations appear to have various consequences on glycosylation enzyme sorting. A study of COG depletion in HeLa cells found mis localisation of *medial*-Golgi and *trans*-Golgi enzymes including MAN2, MGAT1, galactosyltransferase and sialyltransferase suggesting the COG complex is specifically regulating recycling of glycosylation enzymes (Pokrovskaya et al., 2011). Another study used computational modelling to predict changes in glycosylation enzyme sorting between cell lines. The modelling identified MAN2, MGAT1 and galactosyltransferase rates as decreasing in a Cog4-KO HEK cell line compared to wild type, this prediction was then validated by western blotting (Fisher et al., 2019). Taken together with the results presented in this thesis this suggests Cog4-KO cells have less effective recycling of *medial*-Golgi and *trans*-Golgi enzymes.

Glycan profiling of whole cell glycosylation has been conducted for Cog4-WT, Cog4-KO and Cog4-L36P cell lines (Spencer, 2022). As the profiling of these cell lines was for the whole cell it is difficult to directly compare to the Herceptin glycan profiles. For example, there will be less partially processed glycans on Herceptin which is secreted from the cell therefore most glycans should be fully processed. However, the differences between Cog4-WT, Cog4-

KO and Cog4-L36P whole cell glycans can be compared to the differences between Cog4-WT, Cog4-KO, and Cog4-L36P Herceptin glycans. The whole cell glycan profiling of Cog4-WT, Cog4-KO and Cog4-L36P showed a significantly higher proportion of oligomannose glycans from Cog4-WT and Cog4-L36P cell lines than Cog4-KO. However, $GlcNAc_2Man_5$ was more abundant from Cog4-KO and Cog4-L36P cell lines than the Cog4-WT cell line. The Cog4-KO cell line also produced significantly more core fucosylated glycans than Cog4-WT and Cog4-L36P cell lines (Spencer, 2022). This was not seen in the glycan profiling data presented in this thesis, Cog4-KO produced more oligomannose than Cog4-L36P which produced more than Cog4-WT (Figure 13), most of these were $GlcNAc_2Man_5$ (Figure 14). This suggests in Cog4-KO cells processing by Man1 is occurring but there are issues with MGAT1 and possibly Man2 processing. Core fucosylation was the lowest for the Cog4-KO cell line (Figure 13).

The identification of the Cog4-R729W mutation prompted glycan profiling of CDG patient serum to compare glycosylation defects caused by Cog4-R729W to mutations in other COG complex subunits. A decrease in sialylation was identified in all the COG mutant cells, with Cog4-R729W and Cog1 patients being the least effected. Trafficking defects were also examined Cog4-R729W and Cog8 patient cells had less severe retrograde transport delays (Reynders et al., 2009). In this thesis the glycosylation of Cog4-R729W cells was compared to other Cog4 mutations. Cog4-R729W cells had no difference in sialylation compared to Cog4-WT cells but did have an increase in galactosylation. This suggests the R729W mutation may causes decreased recycling of *trans*-Golgi glycosylation enzymes.

Cog4-L773R cells produced glycans with more terminal galactose than Cog4-WT cells (Figure 13). L773R patient cells were glycan profiled and compared wild-type cells glycan profile, this revealed an accumulation of $GlcNAc_2Man_5$ which was interpreted as impairment of an early processing step. This profiling also showed an increase in core-fucosylated complex chains suggesting disruption to galactosylation and sialylation (Miura et al., 2005). The glycan profiles presented in this thesis suggest Cog4-L773R cell lines may have an increase in recycling of galactosylation enzymes to the *trans*-Golgi.

Herceptin usually has oligomannose glycans however these are mostly below 5% abundance. There are partially processed glycans found on Herceptin, $GlcNAc_3Man_3$, $GlcNAc_3Man_3Fuc$ and $GlcNAc_3Man_3FucGal$, these have below 5% abundance (Yu, Ahn and Gilar, 2010). Whereas the glycan profiles presented in this thesis had abundances of below

5% for most oligomannose species, however $GlcNAc_2Man_5$ was found more frequently in every cell line besides Cog4-WT. There are also more types of partially processed glycans, these were mostly below 5% abundance. Having a high abundance of oligomannose glycans in the glycan profile of mAbs negatively impacts the complement-dependent cytotoxicity (CDC) activity of the mAb due to a lower binding affinity for the C1 complex. Whereas mAbs with high abundances of oligomannose glycans have been linked to increased antibody-dependent cellular cytotoxicity (ADCC) activity (Boune et al., 2020). However, oligomannose glycans cannot be fucosylated, which has been linked to improved FcγR binding which enhances ADCC (Chiu et al., 2019). Therefore, the lack of fucosylation, not the presence of oligomannose glycans may cause the enhanced ADCC activity. Kanada et al. have studied the effect of knocking out the FUT8 enzyme on the activity of antibodies with oligomannose glycans. The study found the oligomannose glycans reduced the antibodies affinity to FcγR and the C1 complex, causing reduced ADCC and CDC activity (Kanda et al., 2007). Terminal galactose can improve CDC and C1 complex interactions (Boune et al., 2020). Cog4-L773R cells produced glycans with more terminal galactose than Cog4-WT cells (Figure 13). Therefore, the Cog4-L773R mutation could be beneficial to incorporate into CHO cell lines used in the manufacturing of Herceptin.

Future work

Further research is needed to comprehensively understand the mechanisms by which Cog4 contributes to glycosylation and its implications for cellular function and pathology. Recommendations for experiments to further validate observation made from the data presented in this thesis have been made throughout this discussion. Other experimentation to further the understanding of the mechanisms by which Cog4 contributes to glycosylation are proposed here.

Firstly, the Cog4 mutations responsible for the enhanced Rab41 interaction will need to be elucidated. A yeast two hybrid approach examines the interactions of mutant Cog4 and Rab41 in isolation. A co-immunoprecipitation approach would allow the interaction to be studied in the presence of the other COG subunits. A Cog4 mutant cell line could then be developed to investigate the effects of the enhanced interaction with Rab41 on glycosylation of mAbs produced in CHO cell lines. Herceptin or other mAbs produced in the

various Cog4 mutant cell lines could also be tested for CDC and ADCC to determine if their efficacy is affected by the alterations in glycosylation. Fc receptor binding can be analysed using flow cytometry and surface plasmon resonance (SPR).

Previous studies of Cog4 interactions with STX5 and Sly1 observed Cog4 independent of the rest of the complex. Thermal shift assays would allow the strength of the interactions along with the stability of mutant Cog4 to be studied in the context of the rest of the COG complex. CDG mutant Cog4 could be examined in a thermal shift assay to compare the stability of Cog4-R729W and Cog4-L773R experimentally. To further investigate Cog4 interaction sites, those identified through Rab30 and Rab41 docking could be experimentally mutated and thermal shift assays used to determine the stability of the mutants and the stability of the interactions with Rab30 and Rab41. Cysteine, arginine, and asparagine known to be involved in interactions could be computationally or experimentally mutated to refine the possible binding sites for Rab30 and Rab41 interactions. Cog4 could also be mutated to alter the electrostatic potential of possible binding sites to determine if electrostatics play a role in Cog4 and Rab interactions. Understanding the mechanics of Cog4 protein interactions and how mutations effect these may identify further targets for study that allow mAb glycosylation to be manipulated. Manipulating mAb glycosylation will allow greater control over which glycan types appear on mAbs which influence mAb efficacy.

Other COG subunits are known to alter glycosylation, for example the Cog8 mutation Y537X has been found to alter sialylation and galactosylation (Foulquier et al., 2007). Terminal galactose has been found to improve CDC and C1 complex interactions (Boune et al., 2020). The mutations found in other COG subunits could be investigated as potential targets for mAb glycosylation improvements. If COG subunit mutations could be applied in tandem to a cell line without effecting COG complex integrity a combination of glycosylation effects may be observed. Manipulating glycosylation using multiple COG subunits could further improve the mAb efficacy of applying just one mutation.

List of Abbreviations

ADCC	Antibody-dependent cellular cytotoxicity
CDC	Complement-dependent cytotoxicity
CDG	Congenital disorder of glycosylation
CHO	Chinese hamster ovarian
COG	Conserved oligomeric Golgi complex
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
Fab	Antigen binding fragment
Fc	Fragment crystallisable
FcγR	Fcγ receptor
FutT	Fucosyltransferase
Fut8	Fucosyltransferase 8
GalT	Galactosyltransferase
GDP	Guanidine diphosphate
GlcNAc	*N*-acetylglucosamine
GTP	Guanidine triphosphate
HEK	Human embryonic kidney
IgG	Immunoglobulin G
KO	Knock out
LB	Lysogeny Broth
LC-MS	Liquid chromatography coupled with mass spectrometry
mAb	Monoclonal antibodies
Man	Mannose
Man1	Mannosidase 1
Man2	Mannosidase 2
MGAT1	Alpha-1,3-mannosyl-glycoprotein 2-beta-Nacetylglucosaminyltransferase
MGAT2	Alpha-1,6-mannosyl-glycoprotein 2-beta-Nacetylglucosaminyltransferase
MSA	Multiple sequence alignment
PAE	Predicted aligned error
PDB	Protein databank
PNGaseF	Peptide N-glycosidase F
SiaT	Sialyltransferase
SM	Sec1/munc19
SNARE	Soluble NSF attachment protein receptor
STX	Syntaxin
WT	Wild type

References

Andersen, D. C. et al. (2000). Multiple Cell Culture Factors Can Affect the Glycosylation of Asn-184 in CHO-Produced Tissue-Type Plasminogen Activator. *Biotechnol Bioeng*, 70, pp.25–31.

Bateman, A. et al. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51 (D1), pp.D523–D531.

Betts, M. J. and Russell, R. B. (2003). *Bioinformatics for Geneticists: Chapter 14 - Amino Acid Properties and Consequences of Substitutions*. Barnes, M. and Gray, I. (Eds). John Wiley & Sons, Ltd.

Blackburn, J. B. et al. (2016). COG complex complexities: Detailed characterization of a complete set of HEK293T cells lacking individual COG subunits. *Frontiers in Cell and Developmental Biology*, 4 (MAR), p.188025.

Borys, M. C., Linzer, D. I. H. and Papoutsakis, E. T. (1993). Culture pH affects expression rates and glycosylation of recombinant mouse placental lactogen proteins by Chinese hamster ovary (CHO) cells. *Bio/technology (Nature Publishing Company)*, 11 (6), pp.720–724.

Borys, M. C., Linzer, D. I. H. and Papoutsakis, E. T. (1994). Ammonia affects the glycosylation patterns of recombinant mouse placental lactogen-I by chinese hamster ovary cells in a pH-dependent manner. *Biotechnol Bioeng*, 43 (6), pp.505–514.

Boune, S. et al. (2020). Principles of *N*-Linked Glycosylation Variations of IgG-Based Therapeutics: Pharmacokinetic and Functional Considerations. *Antibodies 2020, Vol. 9, Page 22*, 9 (2), p.22.

Carchon, H. A. et al. (2004). Diagnosis of Congenital Disorders of Glycosylation by Capillary Zone Electrophoresis of Serum Transferrin. *Clinical Chemistry*, 50 (1), pp.101–111.

Casetta, B. et al. (2020). A new strategy implementing mass spectrometry in the diagnosis of congenital disorders of N-glycosylation (CDG). *Clinical Chemistry and Laboratory Medicine*, 59 (1), pp.165–171.

Chen, S. et al. (2011). Trs65p, a subunit of the Ypt1p GEF TRAPPII, interacts with the Arf1p exchange factor Gea2p to facilitate COPI-mediated vesicle traffic. *Molecular Biology of the Cell*, 22 (19), pp.3634–3644.

Chiu, M. L. et al. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies 2019, Vol. 8, Page 55*, 8 (4), p.55.

Curling, E. M. A. et al. (1990). Recombinant human interferon-gamma. Differences in glycosylation and proteolytic processing lead to heterogeneity in batch culture. *Biochemical Journal*, 272 (2), p.333.

D'Souza, Z., Taher, F. S. and Lupashin, V. V. (2020). Golgi inCOGnito: From vesicle tethering to human disease. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1864 (11), p.129694.

Federici, M. et al. (2013). Analytical lessons learned from selected therapeutic protein drug comparability studies. *Biologicals: journal of the International Association of Biological Standardization*, 41 (3), pp.131–147.

Fisher, P. et al. (2019). Modeling Glycan Processing Reveals Golgi-Enzyme Homeostasis upon Trafficking Defects and Cellular Differentiation. *Cell reports*, 27 (4), pp.1231-1243.e6.

Fotso, P. et al. (2005). Cog1p plays a central role in the organization of the yeast conserved oligomeric Golgi complex. *Journal of Biological Chemistry*, 280 (30), pp.27613–27623.

Foulquier, F. et al. (2006). Conserved oligomeric Golgi complex subunit 1 deficiency reveals a previously uncharacterized congenital disorder of glycosylation type II. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (10), pp.3764–3769.

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. *Human Molecular Genetics*, 16 (7), pp.717–730.

Freeze, H. H. (2007). Congenital Disorders of Glycosylation: CDG-I, CDG-II, and Beyond. *Current Molecular Medicine*, 7 (4), pp.389–396.

Giraudoux, P. (2023). *pgirmess: Spatial Analysis and Data Mining for Field Ecologists*.

Glick, B. S., Elston, T. and Oster, G. (1997). A cisternal maturation mechanism can explain the asymmetry of the Golgi stack. *FEBS Letters*, 414 (2), pp.177–181.

Goochee, C. F. and Monica, T. (1990). Environmental effects on protein glycosylation. *Bio/technology (Nature Publishing Company)*, 8 (5), pp.421–427.

Goud, B., Liu, S. and Storrie, B. (2018). Rab proteins as major determinants of the Golgi complex structure. *Small GTPases*, 9 (1–2), p.66.

Guo, W. et al. (1999). The exocyst is an effector for Sec4p, targeting secretory vesicles to sites of exocytosis. *The EMBO Journal*, 18 (4), pp.1071–1080.

Guo, W., Tamanoi, F. and Novick, P. (2001). Spatial regulation of the exocyst complex by Rho1 GTPase. *Nature Cell Biology 2001 3:4*, 3 (4), pp.353–360.

Ha, J. Y. et al. (2016). Molecular architecture of the complete COG tethering complex. *Nature Structural & Molecular Biology 2016 23:8*, 23 (8), pp.758–760.

Hanson, Q. M. and Barb, A. W. (2015). A perspective on the structure and receptor binding properties of immunoglobulin G Fc. *Biochemistry*, 54 (19), pp.2931–2942.

Hayes, G. L. et al. (2009). Multiple rab GTPase binding sites in GCC185 suggest a model for vesicle tethering at the *trans*-Golgi. *Molecular Biology of the Cell*, 20 (1), pp.209–217.

Hayter, P. M. et al. (1992). Glucose-limited chemostat culture of Chinese hamster ovary cells producing recombinant human interferon-gamma. *Biotechnol Bioeng,* 39 (3), pp.327–335.

Hebert, D. N., Garman, S. C. and Molinari, M. (2005). The glycan code of the endoplasmic reticulum: Asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends in Cell Biology*, 15 (7), pp.364–370.

Higel, F. et al. (2016). N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *European Journal of Pharmaceutics and Biopharmaceutics*, 100, pp.94–100.

Holst, M. J. and Saied, F. (1995). Numerical solution of the nonlinear Poisson–Boltzmann equation: Developing more robust and efficient methods. *Journal of Computational Chemistry*, 16 (3), pp.337–364.

Holst, M. and Saied, F. (1993). Multigrid solution of the Poisson—Boltzmann equation. *Journal of Computational Chemistry*, 14 (1), pp.105–113.

Hughes-Jones, N. C. and Gardner, B. (1979). Reaction between the isolated globular sub-units of the complement component C1q and IgG-complexes. *Molecular immunology*, 16 (9), pp.697–701.

Itzen, A. and Goody, R. S. (2011). GTPases involved in vesicular trafficking: Structures and mechanisms. *Seminars in Cell & Developmental Biology*, 22 (1), pp.48–56.

Jenkins, N. and Curling, E. M. A. (1994). Glycosylation of recombinant proteins: problems and prospects. *Enzyme and microbial technology*, 16 (5), pp.354–364.

Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, 596 (7873), pp.583–589.

Jurrus, E. et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein science: a publication of the Protein Society*, 27 (1), pp.112–128.

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), pp.104–118.

Kayser, V. et al. (2011). Glycosylation influences on the aggregation propensity of therapeutic monoclonal antibodies. *Biotechnology Journal*, 6 (1), pp.38–44.

Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review Biochemistry*, VOL. 54, pp.631–664.

Laufman, O. et al. (2009). Direct interaction between the COG complex and the SM protein, Sly1, is required for Golgi SNARE pairing. *The EMBO Journal*, 28 (14), pp.2006–2017.

Laufman, O., Hong, W. J. and Lev, S. (2013). The COG complex interacts with multiple Golgi snares and enhances fusogenic assembly of SNARE complexes. *Journal of Cell Science*, 126 (6), pp.1506–1516.

Lipiński, P. and Tylki-Szymańska, A. (2021). Congenital Disorders of Glycosylation: What Clinicians Need to Know? *Frontiers in Pediatrics*, 9, p.715151.

Liu, S. et al. (2016). Identification of Rab41/6d effectors provides an explanation for the differential effects of Rab41/6d and Rab6a/a' on Golgi organization. *Frontiers in Cell and Developmental Biology*, 4, p.175512.

Livingstone, C. D. and Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer applications in the biosciences: CABIOS*, 9 (6), pp.745–756.

Lu, R. M. et al. (2020). Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science 2020 27:1*, 27 (1), pp.1–30.

Madden, T. L., Tatusov, R. L. and Zhang, J. (1996). [9] Applications of network BLAST server. *Methods in Enzymology*, 266, pp.131–141.

Mariño, K. et al. (2010). A systematic approach to protein glycosylation analysis: a path through the maze. *Nature Chemical Biology 2010 6:10*, 6 (10), pp.713–723.

Matsuura-Tokita, K. et al. (2006). Live imaging of yeast Golgi cisternal maturation. *Nature 2006 441:7096*, 441 (7096), pp.1007–1010.

McBride, H. M. et al. (1999). Oligomeric complexes link Rab5 effectors with NSF and drive membrane fusion via interactions between EEA1 and syntaxin 13. *Cell*, 98 (3), pp.377–386.

Miller, V. J. 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). *Journal of Biological Chemistry*, 288 (6), pp.4229–4240.

Mimura, Y. et al. (2000). The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Molecular Immunology*, 37 (12–13), pp.697–706.

Miura, Y. et al. (2005). Clinical and biochemical characterization of a patient with congenital disorder of glycosylation (CDG) IIx. *Journal of Pediatrics*, 147 (6), pp.851–853.

Monica, T. J., Goochee, C. F. and Maiorella, B. L. (1993). Comparative biochemical characterization of a human IgM produced in both ascites and in vitro cell culture. *Bio/technology (Nature Publishing Company)*, 11 (4), pp.512–515.

Morrison, H. A. et al. (2008). RegUlation of Early Endosomal Entry by the Drosophila tumor suppressors rabenosyn and Vps45. *Molecular Biology of the Cell*, 19 (10), pp.4167–4176.

Ng, B. G. et al. (2011). Identification of the first COG–CDG patient of Indian origin. *Molecular Genetics and Metabolism*, 102 (3), pp.364–367.

Nose, M. and Wigzell, H. (1983). Biological significance of carbohydrate chains on monoclonal antibodies. *Proceedings of the National Academy of Sciences*, 80 (21), pp.6632–6636.

Nyberg, G. B. et al. (2000). Metabolic effects on recombinant interferon-gamma glycosylation in continuous culture of Chinese hamster ovary cells. *Biotechnol Bioeng*, 44 (5), pp.770–784.

Ohtsubo, K. and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell*, 126 (5), pp.855–867.

Ondruskova, N. et al. (2021). Congenital disorders of glycosylation: Still “hot” in 2020. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1865 (1), p.129751.

Pedersen, T. L. (2022). *patchwork: The Composer of Plots*.

Pérez-Rodriguez, S. et al. (2021). Compartmentalized Proteomic Profiling Outlines the Crucial Role of the Classical Secretory Pathway during Recombinant Protein Production in Chinese Hamster Ovary Cells. *ACS Omega*, 6 (19), pp.12439–12458.

Piedade, A. et al. (2022). Epidemiology of congenital disorders of glycosylation (CDG)—overview and perspectives. *Journal of Rare Diseases 2022 1:1*, 1 (1), pp.1–19.

Planinc, A. et al. (2016). Glycan characterization of biopharmaceuticals: Updates and perspectives. *Analytica Chimica Acta*, 921, pp.13–27.

Podos, S. D. et al. (1994). LDLC encodes a brefeldin A-sensitive, peripheral Golgi protein required for normal Golgi function. *Journal of Cell Biology*, 127 (3), pp.679–691.

Pokrovskaya, I. D. et al. (2011). Conserved oligomeric Golgi complex specifically regulates the maintenance of Golgi glycosylation machinery. *Glycobiology*, 21 (12), p.1554.

Quintana, E. et al. (2009). Screening for congenital disorders of glycosylation (CDG): Transferrin HPLC versus isoelectric focusing (IEF). *Clinical Biochemistry*, 42 (4–5), pp.408–415.

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Rabouille, C. and Klumperman, J. (2005). The maturing role of COPI vesicles in intra-Golgi transport. *Nature Reviews Molecular Cell Biology 2005 6:10*, 6 (10), pp.812–817.

Rai, A. et al. (2016). Bmerb domains are bivalent rab8 family effectors evolved by gene duplication. *eLife*, 5

Reichert, J. M. et al. (2023). *Therapeutic monoclonal antibodies approved or in review in the EU or US.*

Reynders, E. et al. (2009). Golgi function and dysfunction in the first COG4-deficient CDG type II patient. *Human Molecular Genetics*, 18 (17), pp.3244–3256.

Richardson, B. C. et al. (2009). Structural basis for a human glycosylation disorder caused by mutation of the COG4 gene. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (32), pp.13329–13334.

Rizzo, R. et al. (2013). The dynamics of engineered resident proteins in the mammalian Golgi complex relies on cisternal maturation. *Journal of Cell Biology*, 201 (7), pp.1027–1036.

Sarmay, G. et al. (1992). Mapping and comparison of the interaction sites on the Fc region of IgG responsible for triggering antibody dependent cellular cytotoxicity (ADCC) through different types of human Fc gamma receptor. *Molecular immunology*, 29 (5), pp.633–639.

Schrödinger, L. (2015). *The PyMOL Molecular Graphics System*.

Shannon, P. et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13 (11), pp.2498–2504.

Shestakova, A. et al. (2007). Interaction of the conserved oligomeric Golgi complex with t-SNARE Syntaxin5a/Sed5 enhances intra-Golgi SNARE complex stability. *Journal of Cell Biology*, 179 (6), pp.1179–1192

Shestakova, A., Zolov, S. and Lupashin, V. (2006). COG Complex-Mediated Recycling of Golgi Glycosyltransferases is Essential for Normal Protein Glycosylation. *Traffic*, 7 (2), pp.191–204.

Shuangbin Xu et al. (2021). Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers. *Frontiers in Genetics*, 12, p.774846.

Simonsen, A. et al. (1999). The Rab5 effector EEA1 interacts directly with syntaxin-6. *Journal of Biological Chemistry*, 274 (41), pp.28857–28860.

Sinka, R. et al. (2008). Golgi coiled-coil proteins contain multiple binding sites for Rab family G proteins. *Journal of Cell Biology*, 183 (4), pp.607–615.

Spencer, H. (2022). *Characterisation of Cell Lines with Engineered Golgi Organisation*. University of York.

Stavenhagen, K. et al. (2013). Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry*, 48 (6), pp.627–639.

Suzuki, E. et al. (2007). A Nonfucosylated Anti-HER2 Antibody Augments Antibody-Dependent Cellular Cytotoxicity in Breast Cancer Patients. *Clinical Cancer Research*, 13 (6), pp.1875–1882.

Tsai, C. J. et al. (1997). Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect. *Protein Science*, 6 (1), pp.53–64.

Ungar, D. et al. (2005). Subunit architecture of the conserved oligomeric Golgi complex. *Journal of Biological Chemistry*, 280 (38), pp.32729–32735.

Ushey, K. et al. (2023). *rstudioapi: Safely Access the RStudio API*.

Varadi, M. et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50 (D1), p.D439.

Voynov, V. et al. (2009). Dynamic Fluctuations of Protein-Carbohydrate Interactions Promote Protein Aggregation. *PLOS ONE*, 4 (12), p.e8425.

Wang, J. et al. (2005). *Crystal Structure of Rab30 in complex with a GTP analogue: RCSB PDB - 2EW1*.

Waterhouse, A. M. et al. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25 (9), pp.1189–1191.

Whyte, J. R. C. and Munro, S. (2002). Vesicle tethering complexes in membrane traffic. *Journal of Cell Science*, 115 (13), pp.2627–2637.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40 (1), pp.1–29.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, H. et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4 (43), p.1686.

Wickham, H. et al. (2023). *dplyr: A Grammar of Data Manipulation*.

Wickham, H., Vaughan, D. and Girlich, M. (2023). *tidyr: Tidy Messy Data*.

Willett, R., Ungar, D. and Lupashin, V. (2013). The Golgi puppet master: COG complex at center stage of membrane trafficking interactions. *Histochemistry and Cell Biology*, 140 (3), pp.271–283.

Wu, S. J. et al. (2010). Structure-based engineering of a monoclonal antibody for improved solubility. *Protein Engineering, Design and Selection*, 23 (8), pp.643–651.

Wu, X. et al. (2004). Mutation of the COG complex subunit gene COG7 causes a lethal congenital disorder. *Nature Medicine 2004 10:5*, 10 (5), pp.518–523.

Yan, Y. et al. (2020). The HDOCK server for integrated protein–protein docking. *Nature Protocols 2020 15:5*, 15 (5), pp.1829–1852.

Yu, Y. Q., Ahn, J. and Gilar, M. (2010). Waters application note: Trastuzumab Glycan Batch-to-Batch Profiling using a UPLC/FLR/Mass Spectrometry Platform. *Waters Application note*.

Zeevaert, R. et al. (2008). Deficiencies in subunits of the Conserved Oligomeric Golgi (COG) complex define a novel group of Congenital Disorders of Glycosylation. *Molecular Genetics and Metabolism*, 93 (1), pp.15–21.

Zhang, L., Luo, S. and Zhang, B. (2016). Glycan analysis of therapeutic glycoproteins. *mAbs*, 8 (2), pp.205–215.

Zhou, Y. et al. (2023). DDMut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Research*, 51 (W1), pp.W122–W128.