Mechanism of action of enzymes involved in rare glycogen storage diseases

Laura Marr

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Biological Sciences
School of Molecular and Cellular Biology

May 2022

Intellectual Property and Publication Statement

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Work in chapters 3, 4 and 5 contain data from one publication and one book chapter:

Marr, L., Biswas, D., Daly, L. A., Browning, C., Vial, S., Maskell, D. P., Hudson, C., Bertrand, J., Pollard, J., Ranson, N. A., Khatter, H., Eyers, C. E., Sakamoto, K., & Zeqiraj, E. Mechanism of glycogen synthase inactivation and interaction with glycogenin. *In review.*

The candidate performed molecular biology, protein production, electron microscopy, differential scanning fluorimetry and mass photometry experiments. Additionally, the candidate designed experiments and interpreted data independently and with guidance from the supervisor(s). The candidate wrote the first draft of the manuscript with guidance from E.Z. and additional input from all authors.

Marr, L., Biswas, D., Sakamoto, K., & Zeqiraj, E. Large Scale Protein Production and Activity Assay Protocols for human Glycogen Synthase-Glycogenin Complex. *In review.*

The candidate and D.B. co-wrote the first draft of the manuscript with input from E.Z. and K.S.

Acknowledgements

First of all, I would like to thank my supervisor Dr Elton Zeqiraj. His patience, guidance and knowledge have made this work possible. I have learnt so much during this journey and I am very grateful for his support. I would also like to thank my co-supervisor Prof. Neil Ranson who's knowledge and advice helped greatly along the way. I would like to thank the MRC DiMeN DTP and Vertex Pharmaceuticals Ltd for the funding and opportunity to work on this project.

A big thank you goes to all members of the Zeqiraj lab, past and present: Miriam Walden, Upasana Sykora, Safi Kani Masandi, Martina Foglizzo, Daniel Maskell, Mohd Syed Ahanger, Emma Cowan, Lisa Campbell, Laura Musgrove, Francesca Chandler and Linda Makhlouf. Each and everyone one them have made this experience enjoyable and I couldn't have asked for better colleagues. Thanks also to all members of Miall and Astbury level 6 for creating a great working environment.

Thank you to my collaborators at the Novo Nordisk Foundation Center, Dipsikha Biswas and Kei Sakamoto, and the University of Liverpool, Leonard Daly and Claire Eyers for contributing to this work. It has been great to work with you all.

A lot of this work would not be possible without the help from ABSL team over the years: Rebecca Thompson, Emma Hesketh, Louie Aspinall, Yehuda Halfon, Joshua White, Charlotte Scarff, Tom O'Sullivan and Martin Fuller. Thank you to the amazing EM community at Leeds, who's advice and knowledge have been invaluable. I would also like to acknowledge the PIXC facility for their help and advice.

A special thank you to Fran, Ben, Molly, Veronica, Amy, Alex, Maria, Josh, Aisling and Sam. From lunchtime chats and afternoon coffees to weekend hiking and nights out. Without their friendship I would have lived, laughed and loved a lot less over the past 4 years.

Finally, I would like to thank my family, who have not only supported me and shown great interest in this journey, but have encouraged me to pursue everything I wish for. Last but certainly not least, a special thank you goes to Rob for being there for me no matter what. I wouldn't have got through this without you.

Abstract

Glycogen is the major glucose reserve in eukaryotes, acting as an energy source and maintaining glucose homeostasis. Dysregulation of glycogen metabolism leads to diseases, such as glycogen storage diseases, including neurodegeneration and lysosomal dysfunction. Glycogen synthesis is carried out by glycogenin (GN), glycogen synthase (GS) and glycogen branching enzyme (GBE). GN initiates glycogen synthesis through autoglucosylation to produce a primer glucose chain of 8-12 units. GS interacts with GN to elongate the primer chain and GBE introduces branches, producing globular glycogen particles. GS is activated by glucose-6-phosphate (G6P) binding and inhibited by phosphorylation.

The work described in this thesis aims to understand how GS and GN coordinate to synthesise glycogen and further investigate the complex regulation of GS. Currently, there is no structural information of this complex and no structure of human GS. In addition, the mechanism of inactivation by phosphorylation remains elusive.

Structural and biochemical techniques were used to investigate the GS-GN complex. Optimisation of the expression and purification of the full-length human GS-GN complex allowed the first structural analysis of this complex by electron microscopy. Both low- and high-resolution 3D reconstructions attained reveals the stoichiometry of the complex and elucidates the mechanism of GS inactivation by phosphorylation. Overall, the work in this thesis provides new insights into glycogen synthesis regulation and facilitates studies of glycogen-related diseases.

Table of Contents

Abbreviations

ADP	Adenosine diphosphate
AGL	Glycogen debranching enzyme
AMPK	Adenosine monophosphate activated protein kinase
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AU	Arbitrary units
BH	Binding helix
BSA	Bovine serum albumin
C-terminal	Carboxy-terminal
CAMKII	Ca^{2+}/calmodulin-dependent protein kinase II
cAMP	3'-5'-cyclic adenosine monophosphate
cAZY	Carbohydrate-Active enZymes
CK1	Casein kinase 1
CK2	Casein kinase 2
Cryo-EM	Cryogenic electron microscopy
crYOLO	Cr-You Only Look Once
cryoSPARC	cryo-EM single-particle ab initio reconstruction and classification
CTF	Contrast transfer function
CV	Column volume
Da	Dalton
DNA	Deoxyribonucleic acid
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
DYRK	Dual specificity tyrosine-phosphorylation-regulated kinase
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EER	Electron-event representation
EGTA	Ethyleneglycol- *bis*(β-aminoethyl)-N,N,N',N'-tetraacetic Acid
EM	Electron microscopy
EMDB	Electron Microscopy Data Bank
FSC	Fourier shell correlation
FT	Fourier transform
g	Gram (or gravity)
G1P	Glucose-1-phosphate
G6P	Glucose-6-phosphate

GAA	Acid-alpha-glucosidase
GBE	Glycogen branching enzyme
GFP	Green fluorescent protein
GLUT	Sodium independent transporter
GN	Glycogenin
GP	Glycogen phosphorylase
GPCR	G-protein-coupled-receptor
GS	Glycogen synthase
GSD	Glycogen storage disease
GSK3	Glycogen synthase kinase-3
GST	Glutathione-S-transferase
GT	Glycosyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HK	Hexokinase
HPLC	High performance liquid chromatography
HRP	Horseradish peroxidase
IPTG	Isopropyl β- d-1-thiogalactopyranoside
Lambda PP	Lambda protein phosphatase
LB	Luria Broth
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
M	Molar
MCS	Multi cloning site
min	minute
MOPS	3-(N-Morpholino)propanesulfonic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MST	Microscale thermophoresis
N-terminal	Amino-terminal
NsEM	negative stain electron microscopy
ºC	Degrees Celsius
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PAS	Periodic-acid Schiff
PASK	Per-Arnt-Sim domain containing serine/threonine kinase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction

PDB	Protein Data Bank
PGM	Phosphoglucomutase
Phenix	Python-based Hierarchical Environment for Integrated Xtallography
PhK	Phosphorylase kinase
PKA	Protein kinase A
PKC	Protein kinase C
PMSF	Phenylmethylsulfonyl fluoride
PP1	Protein phosphatase 1
PTG	Protein targeting glycogen subunit
RELION	Regularised likelihood optimisation
RMSD	Root mean square deviation
rpm	Revolutions per minute
s	second
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
TB	Terrific broth
TBE	Tris-boric acid-EDTA
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
Tm	Melting temperature
TSBT	Tris-buffered saline with Tween
UDP	Uridine diphosphate
UDP-G	Uridine diphosphate glucose
UGP	UDP-glucose pyrophosphorylase
UTP	Uridine triphosphate
UV	Ultraviolet
V	Volts
v/v	Volume to volume
w/v	Weight to volume
WT	Wild type
YFP	Yellow fluorescent protein

List of Figures

List of Tables

1 Chapter 1: Introduction

1.1 Glycogen metabolism

1.1.1 Overview

Glycogen is a branched polymer of glucose which functions as a primary energy store in eukaryotes[1]. Glycogen was discovered in 1857 by Claude Bernard, through the isolation of the molecule from liver[2]. A year later, Kekulé established glycogen as a polymer of glucose $(C_6H_{10}O_5)_n$ [2]. In humans, glycogen is primarily stored in the muscle and liver tissues, although it is also found in the brain, kidney, adipose tissue and heart[1]. This accessible energy reserve is formed when glucose levels are high, and can be released for utilisation within cells when required, or can be released systemically when glucose levels are low. Glycogen is created through the action of three enzymes: glycogenin (GN), glycogen synthase (GS) and glycogen branching enzyme (GBE)[1] (Figure 1-1). Glycogen is degraded by glycogen phosphorylase (GP) and glycogen debranching enzyme (AGL). Dysregulation of glycogen metabolism can lead to various diseases, including glycogen storage diseases, cancer and neurodegeneration[3,4].

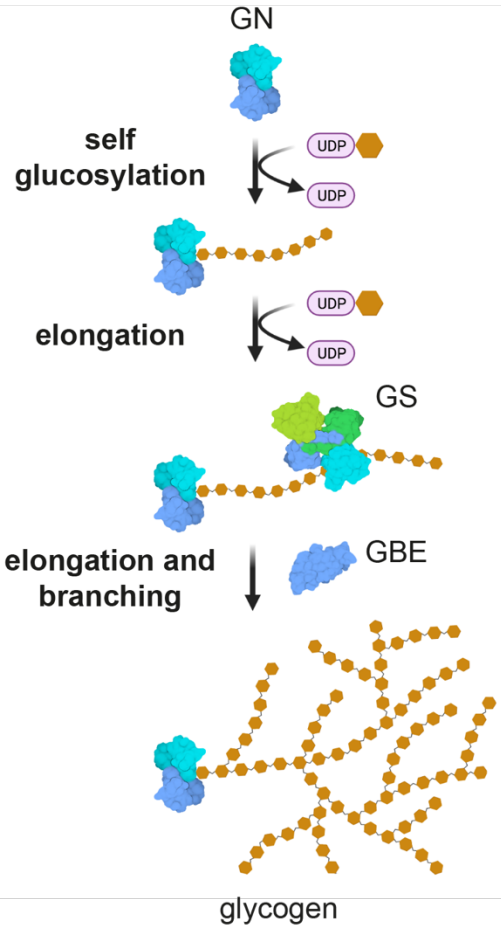

Figure 1-1: Basic schematic of glycogen synthesis

Glycogenin (GN) initiates glycogen synthesis through self glucosylation, using UDP-glucose as a glucose donor to produce a primer glucose chain. Glycogen synthase (GS) can elongate this chain, and glycogen branching enzyme (GBE) introduces branch points to create the final globular glycogen particle. Not to scale. For simplicity, a single glucose chain is shown from one GN dimer.

1.1.2 Glycogen structure

The glucose residues which form glycogen are linked by both α-1,4 and α-1,6-glycosidic linkages. The primary polymerisation is α-1,4-linkages to create linear chains, and branch points are introduced by α-1,6-linkages[5] (Figure 1-2a). A well accepted model for glycogen is a tiered model, containing GN in the centre. GN is surrounded by inner B-chains which are branched, and outer A-chains which are unbranched (Figure 1-2b). The model was based on the calculated spacing of the branch points, the average length of the unit chains and the 'Whelan model'[6]. The theoretical maximum number of tiers is 12, as the addition of a 13th tier would add an impossible density of glucose residues, as well as preventing the addition of more glucose residues due to steric hindrance[7,8]. Using this model, a full size glycogen particle can contain up to ~55,000 glucose molecules, organised in 12 tiers, with a diameter of ~44 nm and a theoretical molecular weight of ~10^7 kDa[5]. Transmission electron microscopy (EM) visualising glycogen in skeletal muscle revealed that the average size of a glycogen particle is 25 nm, which is 7 tiers[9]. This study also found that the largest size of glycogen had a diameter of 44 nm, which agrees with the maximum predicted size based on mathematical modelling[7-9]. Each glycogen granule containing polymerised glucose and various associated regulatory proteins are thought to form an organelle-like structure, termed the "glycosome"[10].

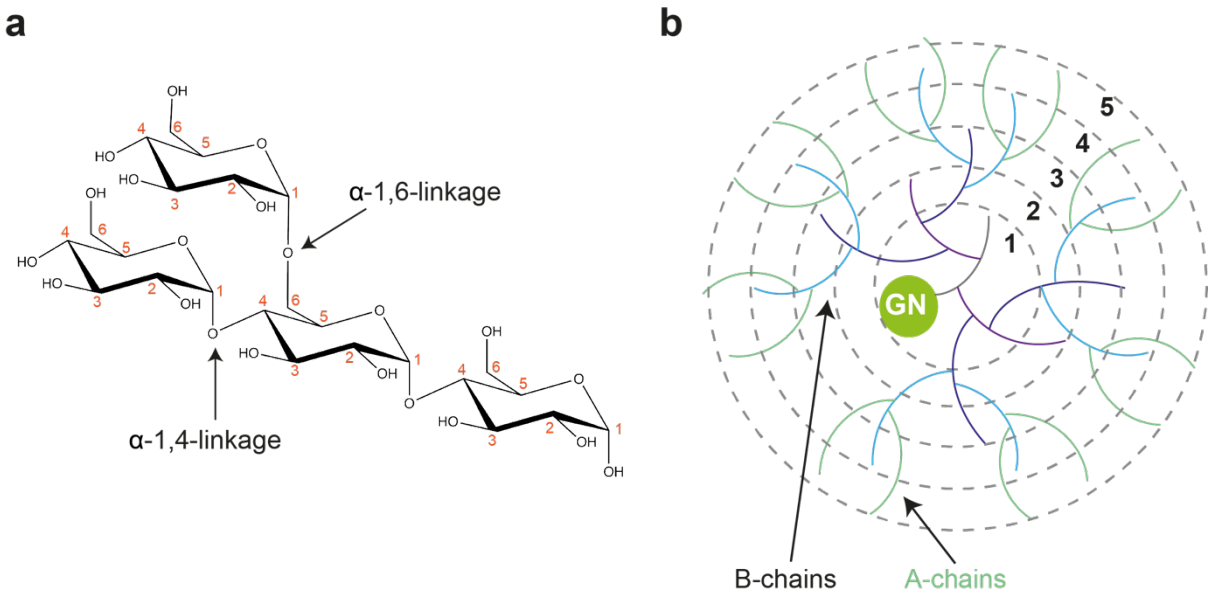

Figure 1-2: The structure of glycogen

a Glycogen is a glucose polymer, where the glucose residues are connected by α-1,4-linkages, to create linear chains, and also α-1,6-linkages to introduce branch points. The carbon numbering is shown in red. **b** The current accepted 'Whelan model' of glycogen is organised in tiers. Glycogenin (GN) is in the centre, surrounded by inner B-chains which are branched (blue, purple and grey), and outer A-chains which are unbranched (green). Only 5 out of the possible 12 tiers are shown for simplicity. Not to scale.

Glycogen granules are composed of protein and glycogen, and three types of granules have been identified by EM: α-, β- and γ- particles[11] (Figure 1-3). γ-particles are the smallest, and several γ-particles can bind to form β-particles. The β-particles are found in the muscle and the liver[12,13]. In the liver, these β-particles aggregate together to form larger α-particles, which can be up to 200 nm in diameter[14] (Figure 1-3). The mechanism of α-particle formation from β-particles is not well understood[13]. It has been postulated that proteins are responsible for the linking of β-particles, as glycogen particles are coupled with several proteins[15-17]. This was suggested by a study that showed changes in glycogen particle size and shape in the presence of proteases that remove proteins from glycogen[18]. In addition, another study observed changes in glycogen size via acid hydrolysis[19]. The protein responsible for joining β-particles was suggested to be GN, as proteome analysis found GN on the surface of glycogen particles[15,20]. However, this idea is contested as a gene-knockout experiment revealed that GN was not essential for glycogen accumulation in eukaryotes[21], as well as the detection of α-particles in a rodent model with GN deficiency[22]. In summary, the formation of α-particles is still uncertain and requires further investigation[23].

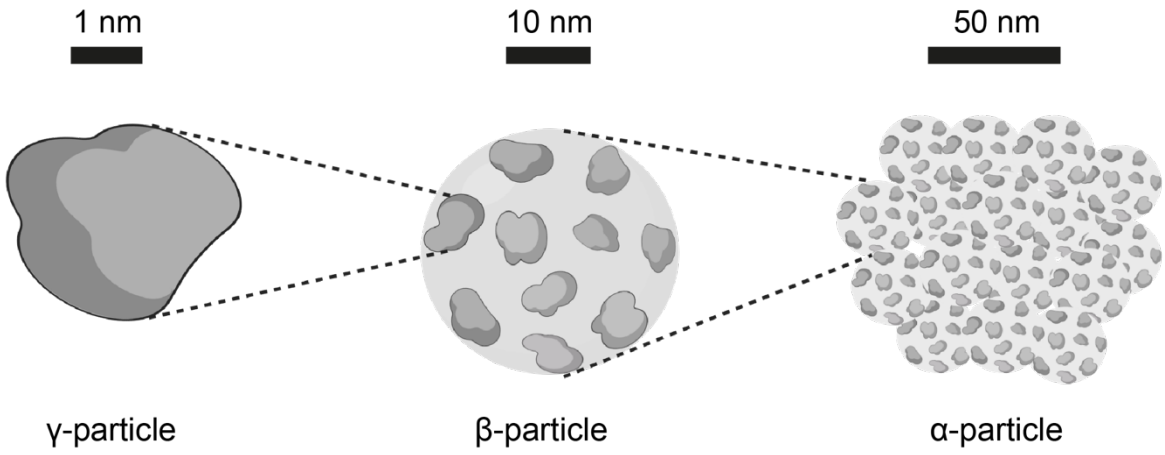

Figure 1-3: The three types of glycogen particles

The smallest structure of glycogen is the γ-particle. Several γ-particles bind to form a β-particle. Many β-particles bind to form larger α-particles. Adapted from Prats, Gómez-Cabello & Hansen, 2011[24].

1.1.3 Glycogen function and localisation

Carbohydrates are one of the main sources of energy for many forms of life. The polymerisation of glucose is a universal mechanism of energy storage, creating starch in plants and glycogen in eukaryotes[1,25]. The structure of glycogen (discussed in section 1.1.2) allows this macromolecule to function efficiently as an energy source. The tiered, branched structure (Figure 1-2) reveals many non-reducing ends on the surface of the molecule, leaving them accessible to glycogen degrading enzymes, and thus allowing rapid mobilisation of glucose. This also allows the quick incorporation of glucose into glycogen when required. In addition, the polymerisation of glucose and the branching structure allows the formation of compact, soluble granules, thus minimising the elevation of osmotic pressure in cells. Overall, the organised structure gives glycogen a great metabolic advantage for its function as an energy store.

In humans, the main stores of glycogen are located in the skeletal muscle and the liver. The physiological role of glycogen differs in these various locations. In the muscle, glycogen is fuel that is required during exercise and muscle contraction. Whereas in the liver, glycogen aids in whole-body glucose homeostasis. However, inside these cells, glycogen particles are not homogenously distributed. In muscle fibres, glycogen is found in three main pools and preferential use of specific pools for different cellular functions has been observed[9,24,26-29]. It is not clear what determines this compartmentalisation of glycogen and the local regulatory mechanisms associated with this.

1.1.4 Glycogen synthesis

Glycogen is created through the actions of three enzymes: GN, GS and GBE[1]. GN initiates the process through autoglucosylation to produce a primer glucose chain. This glucose chain is extended by GS and branch points are introduced by GBE[5] (Figure 1-1).

Glycogen synthesis can occur in either a 'direct' or 'indirect' pathway. The 'direct' pathway requires the transport of glucose into cells, by a large family of transport proteins named glucose transporters[30,31]. Whereas the 'indirect' pathway, also known as the glycolytic/gluconeogenic pathway, utilises gluconeogenic precursors[32].

Two main categories of glucose transporters have been identified: sodium-glucose linked transporters (SGLTs) and facilitated diffusion glucose transporters (GLUTs)[31]. There are several different GLUT isoforms, and each may play a specific role in glucose metabolism in different cell types[30]. GLUT1 is a ubiquitously expressed glucose transporter and is regulated by cellular stress[30]. GLUT2 is expressed in liver, intestine, pancreatic β-cells and kidney[33]. GLUT3 is present in both dendrites and axons and is the main neuronal glucose transporter[34]. GLUT4 is the insulin regulated glucose transporter, where in muscle and fat cells insulin causes GLUT4 to move to the cell surface in order to stimulate uptake of glucose[35].

The fate of intracellular glucose depends on nutrient availability. When nutrients are low, glucose can be used for glycolysis or oxidative phosphorylation. However, at times of nutritional abundance, intracellular glucose is converted into glucose-6-phosphate (G6P) by either hexokinase (HK) in the muscle, or glucokinase in the liver (Figure 1-4). G6P can be converted to glucose-1-phopshate (G1P) by phosphoglucomutase (PGM). Subsequently, UDP-glucose pyrophosphorylase (UGP) can catalyse the formation of UDP-glucose (UDP-G) from G1P and uridine triphosphate (UTP) (Figure 1-4). UDP-G is the substrate of GS, thus leading to glycogen synthesis through the concerted action of GS and GBE (Figure 1-4).

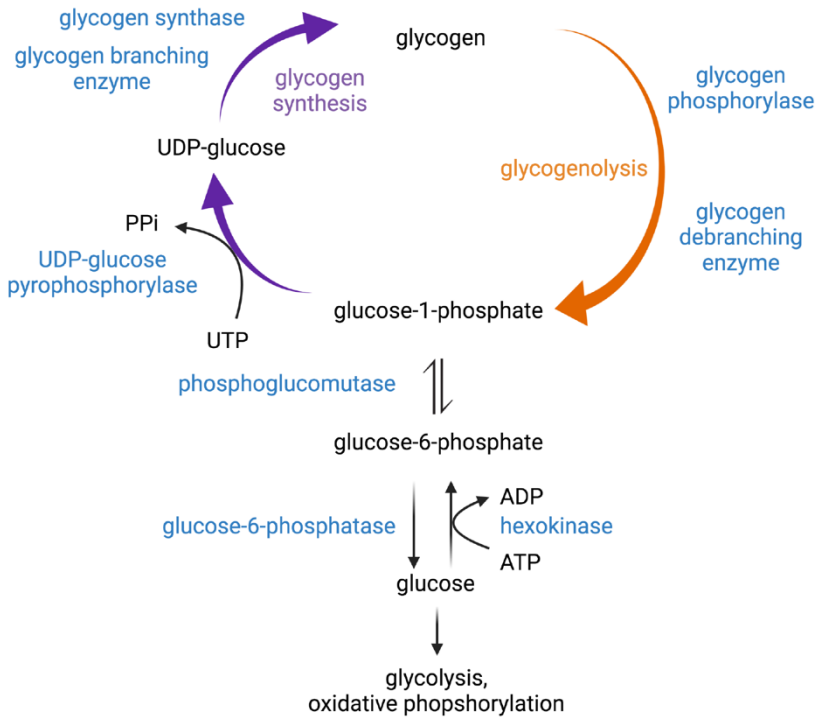

Figure 1-4: Schematic overview of glycogen synthesis and breakdown

Glycogen is broken down by glycogen phosphorylase (GP) and glycogen debranching enzyme (AGL), producing glucose-1-phosphate (G1P). G1P can then be converted into glucose-6-phosphate (G6P) by phosphoglucomutase (PGM). G6P can then be converted into glucose by glucose-6-phoshatase to be used for glycolysis and oxidative phosphorylation. Alternatively, glucose can be converted into G6P by hexokinase (HK) and then G1P can be created which can then be converted into UDP-G by UDP-glucose pyrophosphorylase (UGP). This UDP-G is used as a sugar donor by glycogen synthase (GS) to produce glycogen with the aid of glycogen branching enzyme (GBE).

1.1.5 Glycogen breakdown

Glycogen is broken down into glucose to provide energy and also to maintain blood glucose levels. Glycogen breakdown, or glycogenolysis, is carried out by GP and AGL (Figure 1-5).

GP breaks α-1,4-glycosidic bonds and phosphorylates glucose, releasing G1P in the first step of glycogenolysis[36] (Figure 1-4 and Figure 1-5). GP exists as three isoforms: glycogen phosphorylase liver (PYGL), glycogen phosphorylase brain (PYGB) and glycogen phosphorylase muscle (PYGM). These isoforms differ in their regulatory properties and physiological role depending on their location[37]. The G1P that is released by this process can be converted into G6P by PGM. In this way, G6P can then be utilised for glycolysis

and oxidative phosphorylation (Figure 1-4). G1P can also be converted into UDP-G by UGP, as described previously, to feed forward into glycogen synthesis (Figure 1-4).

The breakdown of α-1,4-linked glucose residues by GP halts four residues before a branch point (an α-1,6-linked residue). Therefore, AGL is needed to continue glucose release. AGL has α-1,4 glucanotransferase and α-1,6 glucosidase activities and continues glycogen breakdown in two steps[38,39] (Figure 1-5). Its glucanotransferase activity transfers three glucose residues from the branch to a neighbouring non-reducing end, forming α-1,4-linkages. The remaining α-1,6-linked glucose residue on the original branch can be removed by its glucosidase activity (Figure 1-5).

Glycogen can also be broken down by acid α-glucosidase (GAA) in the lysosome. GAA can break both α-1,4 and α-1,6 linkages in glycogen, and deficiency of this enzyme leads to accumulation of lysosomal glycogen[40]. This disease is Glycogen storage disease type II, also named Pompe disease, and is discussed further in section 1.5.1.1.

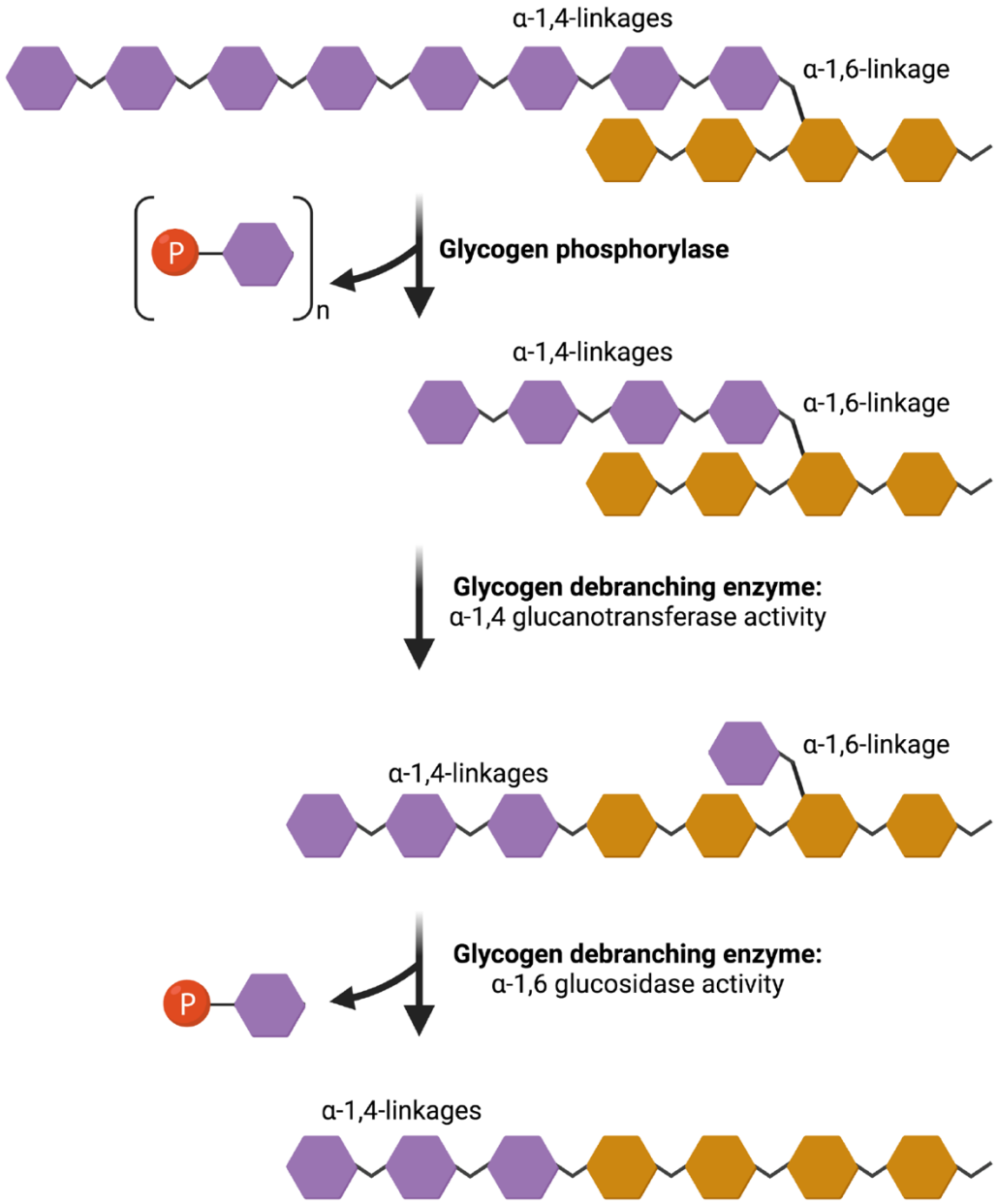

Figure 1-5: Schematic of glycogen breakdown

Glycogen breakdown begins with glycogen phosphorylase (GP) which breaks α-1,4-linked glucose residues and results in the creation of phosphorylated glucose, G1P. GP halts 4 glucose residues before a branch point (an α-1,6-glycosidic bond). Glycogen debranching enzyme (AGL) has α-1,4 glucanotransferase activity, which transfers three α-1,4-linked glucose residues from the branch onto a non-reducing end. The α-1,6 glucosidase activity then cleaves the α-1,6-glycosidic bond, resulting in G1P and full removal of the branch. Not to scale.

1.2 Glycosyltransferases

1.2.1 Families

Glycosyltransferases (GTs) are responsible for the transfer of a sugar moiety from sugar donors to specific acceptor molecules, and thus they catalyse glycosidic bond formation. GTs are classified into families based on their amino acid sequence[41,42]. The families of GTs are continually updated. The Carbohydrate-Active enZymes (CAZy)[43] database reports novel families when they are identified. Currently, there are 114 GT families (database last updated 2022-03-14, accessed 2022-03-29). The two main enzymes discussed and investigated in this thesis are GN and GS. GN belongs to the GT8 family whereas eukaryotic GS belongs to the GT3 family. At the time of writing this thesis, GS is the only member of the GT3 family, however GN is one of eight members of the GT8 family[43].

1.2.2 General structure

The two predominant folds that GTs contain are the GT-A or the GT-B fold, both of which consist primarily of α/β/α sandwiches[44]. These α/β/α sandwiches, also known as Rossmann fold domains, are associated with binding nucleotides[45]. There is also a third, less common fold, the GT-C fold[46].

The GT-A fold comprises of a single Rossmann fold domain (Figure 1-6a). Another feature of a GT-A fold is the presence of a Asp-x-Asp (DxD) motif and the requirement for a divalent cation for activity[44,47]. The main function of the DxD motif is to interact with the phosphate groups of the sugar donor, through coordination of a divalent cation, which is typically Mn^{2+}. The GT-B fold comprises of two separate Rossmann fold domains, with an interdomain cleft containing the active site (Figure 1-6b). Unlike the GT-A fold, they do not contain a DxD motif and they do not require divalent cations for activity.

More recently, a deep-learning framework has been developed to identify, classify and predict GT folds based solely on secondary structure annotations[48], in contrast to the traditional identification based on primary sequence[42]. This allows a greater understanding of the evolutionary relationships of GTs, which has thus far been hindered due to wide sequence variation between GTs[48]. In addition, it allowed the identification of clusters of families within the same fold type, which could not be previously identified through primary sequence analysis[48]. Moreover, this deep-learning framework provides new insights into GT structure and function.

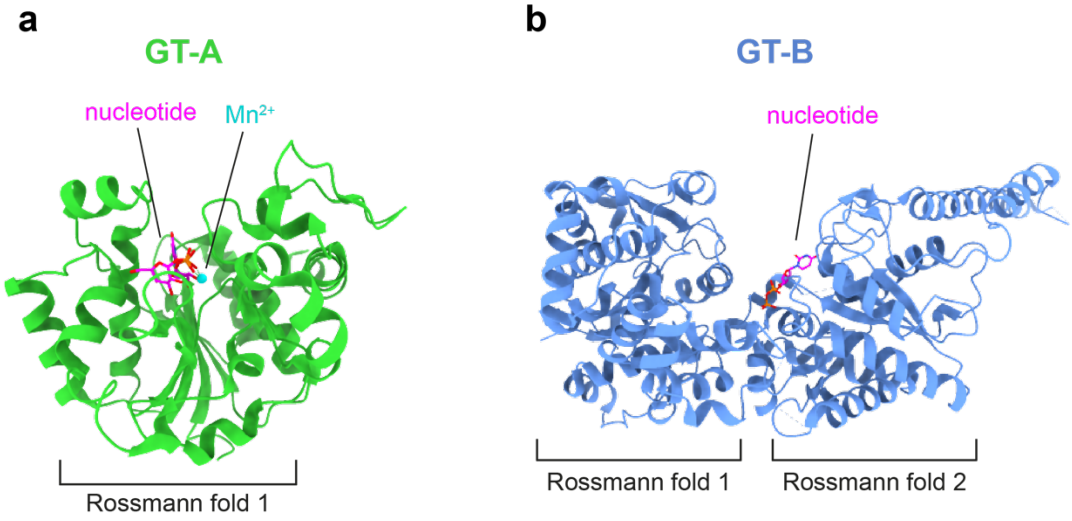

Figure 1-6: The two main folds of glycosyltransferases are the GT-A and GT-B fold

a A ribbon diagram of human glycogenin-1 (PDB ID 3T7O) that contains a GT-A fold. The nucleotide is shown in magenta, the Mn^{2+} is shown in cyan and the approximate location of the Rossmann fold domain is shown. **b** A ribbon diagram of yeast glycogen synthase (PDB ID 3O3C) that contains a GT-B fold. The nucleotide is shown in magenta and the approximate location of the two Rossmann fold domains are shown.

1.2.3 Mechanism

There are two main stereochemical results for the action of GTs, either inversion or retention of the anomeric configuration with respect to the donor sugar[49] (Figure 1-7). However, the catalytic mechanism of a particular GT is not dictated by its fold, as both inverting and retaining GTs have been identified within both GT-A and GT-B fold classes[49].

Although the molecular mechanisms of GN and GS catalysis has been under investigation for the past two decades, a consensus in labelling such reactions as S_Ni- or S_N2-based has not been reached.

Briefly, the S_Ni mechanism for GTs is partly based on the interaction between the hydroxyl group of the acceptor and the departing phosphate group of the sugar donor[50,51]. The leaving group undergoes decomposition, which subsequently becomes a nucleophile (Figure 1-8a). This results in retention of the stereochemistry (Figure 1-7), which is thought to be because of the high rate of decomposition of the leaving group and the subsequent nucleophilic attack (Figure 1-8a).

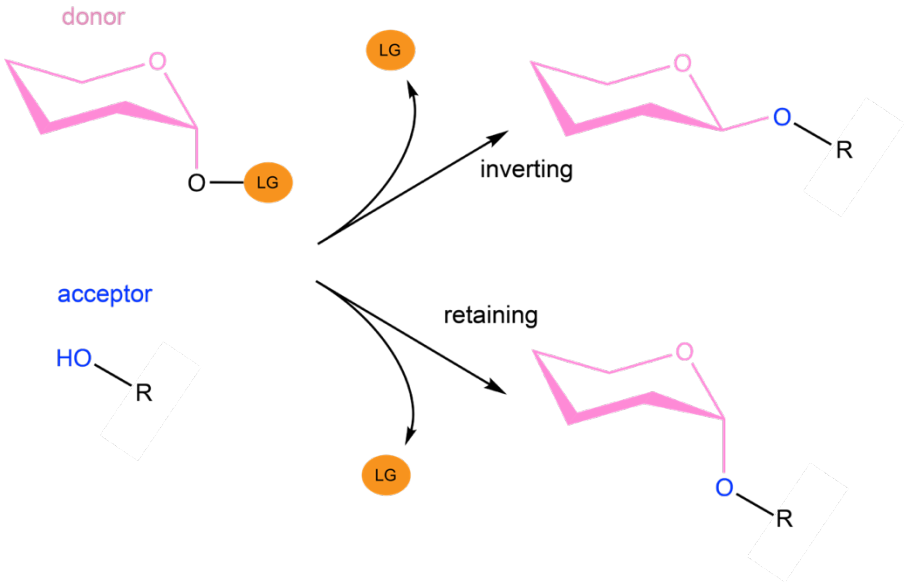

Figure 1-7: Mechanism of glycosyltransferases

Glycosyltransferases (GTs) catalyse group transfer with either inversion or retention of the donor group stereochemistry. GTs use an activated donor that contains a leaving group (LG), as shown by the orange circle named LG. The inverting mechanism (top) inverts the stereochemistry of the donor group. The retaining mechanism (bottom) keeps the stereochemistry of the donor group.

The S_N2 mechanism involves two nucleophilic displacement reactions and is a retaining mechanism (Figure 1-7). The mechanism needs a catalytic nucleophile in the enzyme that initiates the reaction, and an appropriately positioned general acid/base to assist the second nucleophilic attack (Figure 1-8b). An E-X_7-E motif that is conserved across eukaryotic GS enzymes was identified and mutagenesis studies led to the proposal that the conserved glutamate residues function as the nucleophile and general acid/base catalyst in GS[52]. Mutagenesis revealed the first Glu residue in this motif to be a fundamental nucleophile, whereas the second Glu residue provides important but not essential catalytic assistance. However, further investigation is required to identify such reactions as S_Ni or S_N2 mechanisms, as well as identifying the exact role of each Glu residue in the catalytic reaction.

GS is a processive glycosyltransferase, meaning it remains associated with its polymeric product during multiple rounds of catalysis. This processive mechanism is shared with other glycosyltransferases, including cellulose synthase, the enzyme that generates cellulose, which is a linear β-1,4-linked glucose polymer[53].

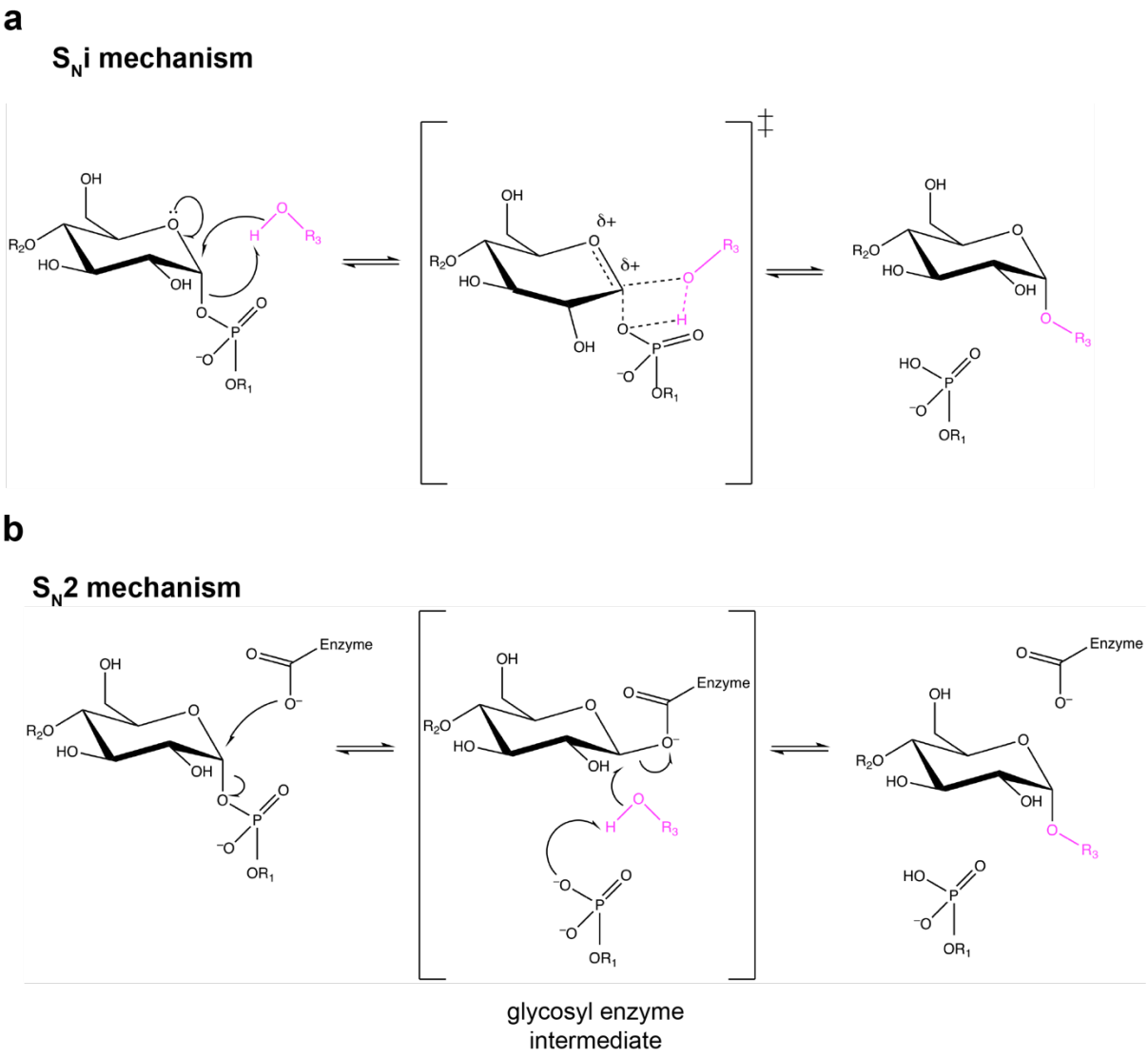

Figure 1-8: The S_Ni and S_N2 mechanisms of glycosyltransferases

a The S_Ni mechanism involves one catalytic nucleophile from the donor molecule. **b** The S_N2 mechanism involves double nucleophilic enzyme substitution and the formation of a glycosyl enzyme intermediate.

1.3 Enzymes involved in glycogen synthesis

1.3.1 Glycogenin

The concept that a primer is required for the synthesis of glycogen arose in the 1930s following the discovery of glycogen phosphorylase[54]. The search continued, and in 1957 GS was discovered[55]. Later in the 1970s, several observations led to the proposal that glycogen synthesis started on a protein backbone, and that the first step of this process was the transfer of glucose from UDP-G onto multiple sites of the protein acceptor[56]. The second step proposed was that the protein acceptor could add additional glucose residues

until the chain was long enough for extension to continue by GS[56] (Figure 1-9 and Figure 1-1). Following this proposed mechanism, it was shown that this protein primer remains attached to the glycogen particle[57,58]. Further investigation revealed that this protein, named GN, was linked to glucose via a tyrosine residue[59,60], which was later identified as Tyr194 in rabbit GN[61] (Tyr195 in humans) (Figure 1-9). It was found that GN could be isolated either bound to glycogen, or co-purified with GS[60,62,63]. GN was found to be an auto-glucosylating transferase which is Mn-dependent, and the product of autoglucosylation is the substrate for further glucosylation by GS[64]. There are two isoforms of GN, GN1 and GN2, encoded by the *GYG1* and *GYG2* genes respectively[65]. While *GYG1* is widely expressed, *GYG2* is restricted to the liver, heart and pancreas[65,66]. However, non-primates only have a single GN gene which is expressed ubiquitously[67].

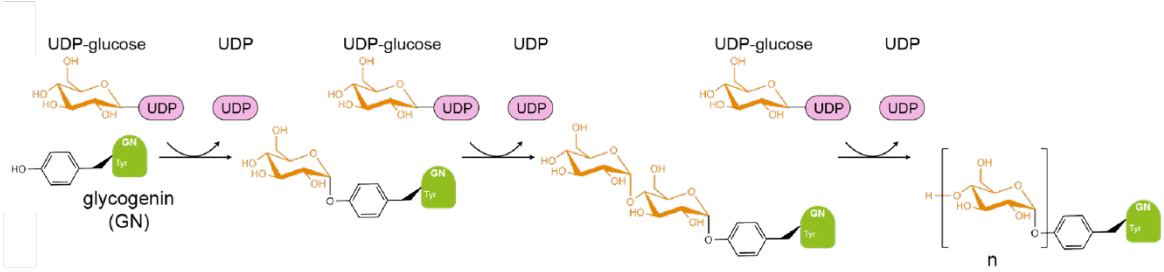

Figure 1-9: Auto-glucosylation of glycogenin

Catalytic mechanism of GN auto-glucosylation. UDP-G is used as a glucose donor to glucosylate GN on a conserved Tyr (residue 195 in humans). Additional glucose residues can be added via α-1,4-glycosidic linkages, to create a primer glucose chain of 8-12 residues.

As mentioned previously (see section 1.2), GN belongs to the GT8 family of GTs. GN has a retaining mechanism and contains a GT-A fold which is characterised by a Rossmann fold domain at the N-terminus. Structural studies revealed that GN forms a dimer[68], where each monomer comprises a globular domain, containing the Rossmann fold, a variable length linker and a conserved C-terminus (Figure 1-10a and b). The C-terminus comprises a highly conserved region that is the minimal targeting region for binding GS[69,70]. The variable length linker region is not conserved in amino acid composition or length across various species and isoforms (Figure 1-10).

It has been shown that rabbit GN can bind to actin, through a common actin binding motif 'DNIKKKL'[71]. This is consistent with the observation that glycogen granules and cytosolic filaments are associated[10]. This actin binding motif is highly homologous in human GN1 ('DNIKRKL'), indicating an importance of this fragment[71]. Moreover, it is believed that the

birth of new glycogen granules is associated with, and perhaps regulated by, the actin cytoskeleton. However, further investigation is required to fully understand this process[72].

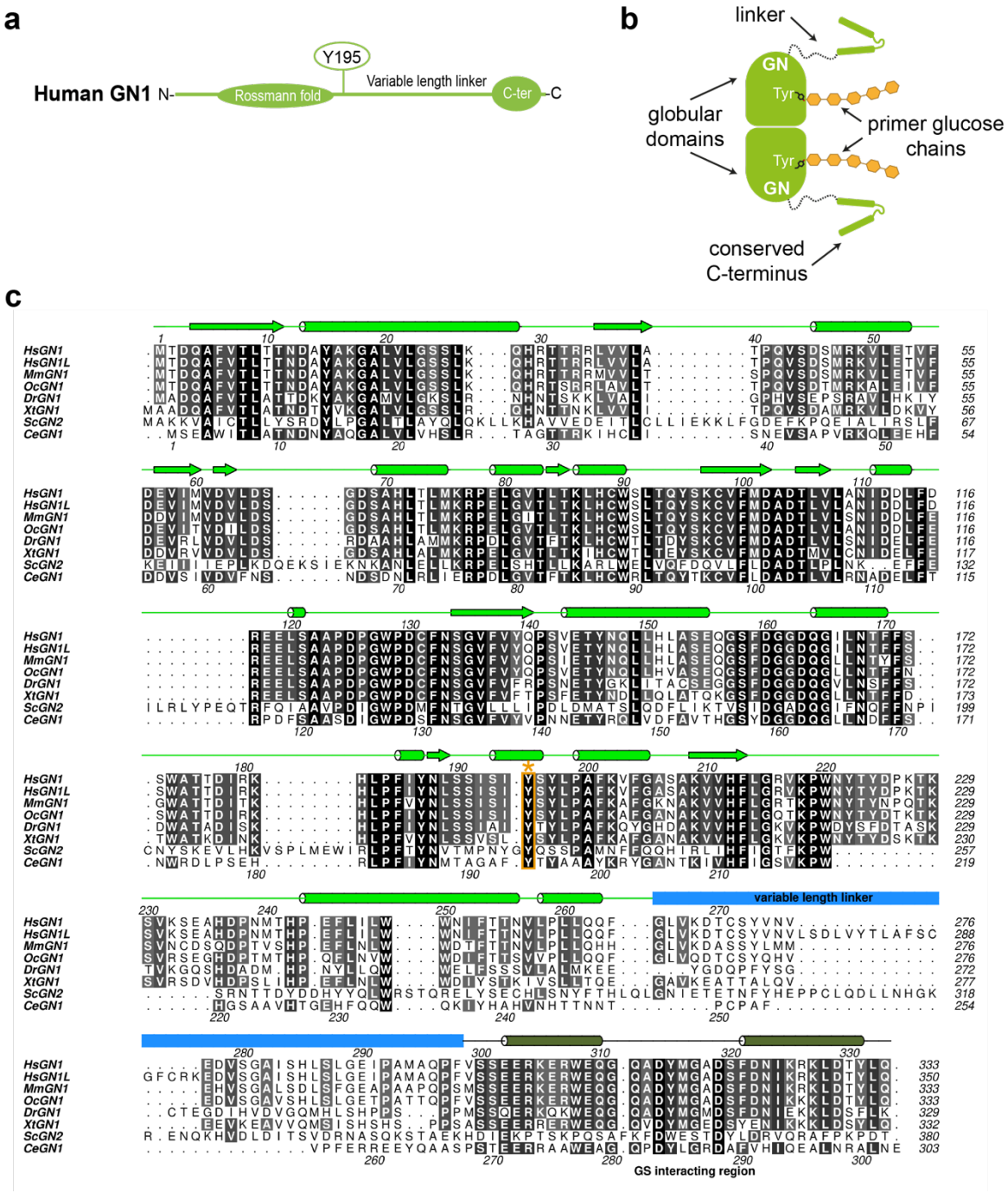

Figure 1-10: Domain architecture of glycogenin

a Schematic of the domain architecture of glycogenin (GN), containing an N-terminal Rossmann fold domain, a variable length linker region and a conserved C-terminus. The autoglucosylating Tyr residue (195 in humans) is shown. **b** Cartoon representation of a GN dimer, showing the globular

domains, linker region and primer glucose chains as a result of autoglucosylation. **c** Multiple sequence alignment of GN from the indicated species (black = conserved, white = not conserved). Sequence alignment was performed with MUSCLE and edited and displayed using ALINE. The predicted secondary structure for HsGN is shown at the top of the alignment and coloured in green (analysed by PSIPRED). The human and *C. elegans* GN sequences are numbered at the top and bottom of the alignment respectively. The variable length linker is coloured in blue, the C-terminal GS interaction region is coloured in dark green and the tyrosine that is auto-glucosylated (Y195 in human) is boxed in orange and labelled with an asterix. *Hs, H. sapiens; Mm, M. musculus; Oc, O. cuniculus; Dr, D. rerio. Xt, X. tropicalis; Sc, S. cerevisiae; Ce C. elegans.*

Variable lengths of the linker region in GN can be created through alternative splicing during gene expression[73]. It was shown *in vitro* that the length of the linker region correlates with the size of the glycogen particle produced[70]. Shorter linkers generated smaller glycogen particles, whereas longer linkers resulted in larger particles with wider size distributions[70]. However, whether this is true in a biological system is unknown and thus the role of the linker region requires further investigation.

It could be hypothesised that GN governs the maximal amount of glucose present, seen as though it is the initiating particle. However, this is not the case as there was found to be no correlation between increased glycogen levels and GN protein in rat skeletal muscle[74]. This is consistent with the observation that overexpression of GN in COS-1 cells and rats had no effect on total glycogen levels[75,76].

The widely accepted concept is that GN is essential for *de novo* glycogen synthesis. However, in glycogenin-1 deficient patients, glycogen is present in skeletal muscle[77], thus challenging the widely accepted theory. Controversially, one study found GN1 to be dispensable for glycogen synthesis in humans, as GN1 deficiency was not compensated for by the upregulation of GN2, yet glycogen could still be synthesised[78]. However, the glycogen formed in the absence of GN1 was abnormal in structure, and this is the cause of a glycogen storage disease. This is supported by the finding that GN knockout mice are still able to synthesise glycogen[22], suggesting that GS can synthesise glycogen from free glucose, although this glycogen is abnormal. This suggests a role of GN in regulating glycogen synthesis in muscle, but this still is unclear and requires further investigation[78].

The catalytic mechanism of GN, which unusually catalyses its own glucosylation, has been studied in various different ways. It was initially shown that both intra-subunit (same GN

monomer) and inter-subunit (different GN monomer) mechanisms were required for full catalysis, but initial glucosylation could not be completed through intra-mechanisms[79] (Figure 1-11). Crystal structures of catalytic snapshots of human GN revealed that the first four glucose residues are added in an intra-subunit reaction, whereas elongation of longer chains requires catalytic action in an inter-subunit reaction[80] (Figure 1-11). However, the precise mechanics of GN autoglucosylation have been hindered by the inability to study homogenous glycoforms of GN[81]. Recently, a new chemical tool was developed to mimic the intermediate forms of the autoglucosylation mechanism of GN, to provide greater understanding of the initiation of glycogen synthesis[81]. This study added a reactive 'tag' onto to GN, by replacing the hydroxyl group of Tyr195 with an iodide moiety, which is a substrate for palladium mediated cross-coupling reactions to add sugar-mimic templates to generate various GN glycoforms[81]. The findings of this study are consistent with previous reports[80], and suggested a potential multiphase initiation, with a step-by-step 'inter, intra, inter' mechanism of action for GN self glucosylation. Where the initial auto-glucosylation occurs via inter-subunit action, then up to five sugar residues can be added via intra-monomer reactions and the final (up to 11) glucose residues are added in an inter-subunit manner[81].

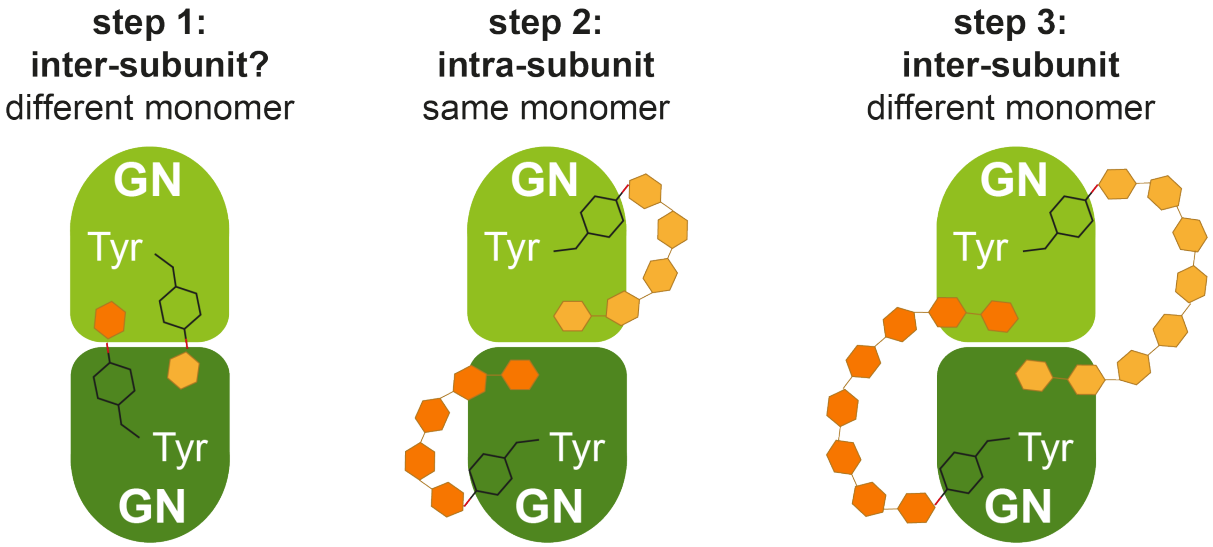

Figure 1-11: Intra- and inter- subunit autoglucosylation of glycogenin

Glycogenin (GN) autoglucosylation has been proposed to either be intra-subunit (same monomer) or inter-subunit (different monomer). Each GN monomer is coloured different shades of green. Glucose is represented as hexagons, with different shades corresponding to glucose chains from different GN monomers. The autoglucosylating tyrosine (Tyr195 in humans) is shown. Step 1 is

predicted to occur in an inter-subunit reaction. Step 2 has been shown to occur in an intra-subunit reaction and step 3 occurs in an inter-subunit reaction. Not to scale.

1.3.2 Glycogen synthase

GS was discovered in 1957 by Leloir and Cardini[55]. As mentioned previously, eukaryotic GS is a member of the GT3 family of GTs and contains a GT-B fold (see section 1.2). This fold is characterised by two Rossman fold domains, one at the N- and one at the C-terminus, with an interdomain cleft that contains the active site[46,82]. Human GS is found as two isoforms, GS1 and GS2, encoded by the *GYS1* and *GYS2* genes respectively. These are differentially expressed, with *GYS1* being expressed predominantly in the skeletal muscle and most other cell types where glycogen is present, while *GYS2* is expressed exclusively in the liver[83-85]. Similar to GN, GS uses UDP-G to extend glucose chains through the addition of α-1,4-linked glucose residues[1], and in concert with GBE this produces glycogen (Figure 1-12). GS is tightly regulated and is the rate limiting enzyme in glycogen biosynthesis. It is activated by G6P binding and inhibited by phosphorylation, which are discussed in more detail in sections 1.4.1.1 and 1.4.1.2 respectively.

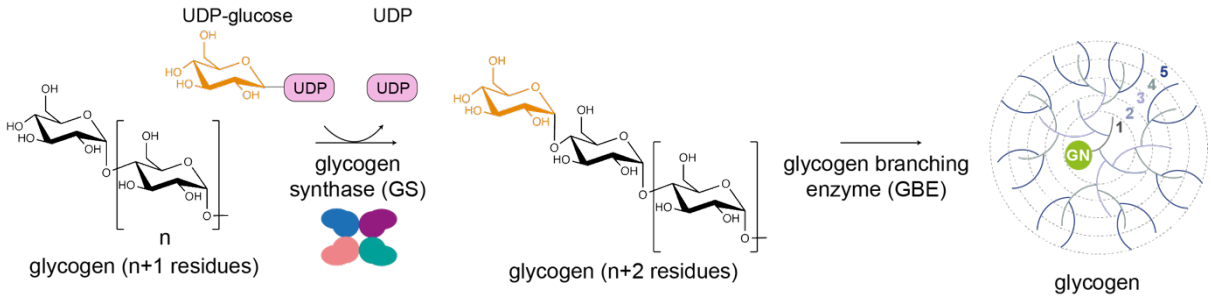

Figure 1-12: The catalytic mechanism of glycogen synthase (GS) and glycogen branching enzyme (GBE)

GS uses UDP-G as a glucose donor to elongate the primer glucose chain attached to GN, through the addition of α-1,4-linked glucose residues. GBE introduces branch points to form the "tiered" glycogen particle.

The first structure of GS, solved in 2004, was a crystal structure from the bacterium *Agrobacterium tumefaciens* (*A. tumefaciens*) (AtGS), in the presence and absence of adenosine diphosphate (ADP)[82] (Figure 1-13). This confirmed that GS contains two Rossmann fold domains with an interdomain cleft, and it was found that GS had

crystallised in an 'open' state revealing a deep and wide cleft containing the active site[82]. GS was found to be dimeric and ADP binds to a pocket adjacent to the interdomain cleft, which induced small changes in the overall structure[82]. This structure also revealed that the overall fold and active site of GS is very similar to GP, highlighting that the opposite reactions of glycogen synthesis and breakdown are catalysed by homologous enzymes[82].

Following this, two more GS crystal structures were solved, from *Escherichia coli* (*E. coli*) [86] and *Pyrococcus abyssi*[87], revealing GS as monomeric and trimeric respectively. All these GS structures are from bacteria, which belong to the GT5 family of GTs. The GT5 family is not regulated by phosphorylation and allosteric activation, unlike eukaryotic GS which belongs to the GT3 family. It was not until 2010 when a eukaryotic GS structure was solved, a crystal structure of Gsy2 from *Saccharomyces cerevisiae (S. cerevisiae)* (referred to as yeast GS in this thesis) in the presence and absence of the allosteric activator G6P[88] (Figure 1-13). Unlike previous structures, this revealed that GS forms a tetramer through two helices from each protomer mediating the tetramerisation interface. G6P binding induces large conformational changes, which are discussed in more detail in section 1.4.1.1. Other crystal structures of yeast GS in complex with various ligands including UDP-G[89] and maltooctaose[90], as well as an enzyme that mimics the phosphorylated state[91] have been solved. In addition, another eukaryotic GS structure, GS1 from *Canorhabditis elegans*[92] (*C. elegans*), in complex with the C-terminus of GN was solved and is further discussed in section 1.3.3.

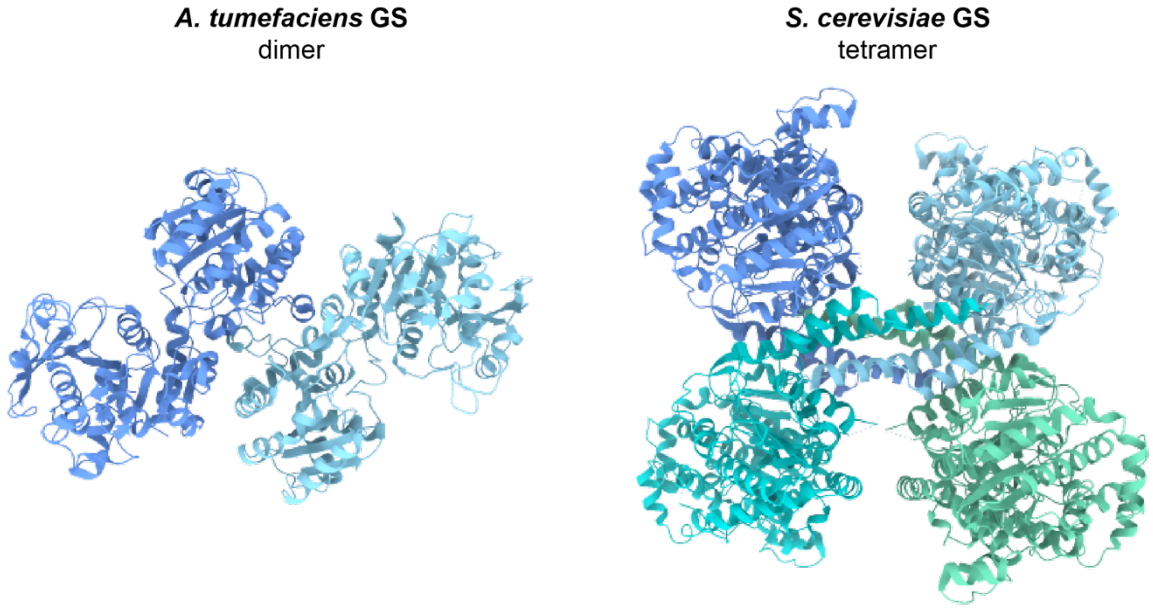

Figure 1-13: Oligomeric states of different glycogen synthase (GS) crystal structures
The first GS structure to be solved from *A. tumefaciens* is dimeric (PDB ID 1RZV) (left), whereas GS from *S. cerevisiae* is tetrameric (PDB ID 3NAZ) (right). Each protomer is coloured differently.

Substrate binding was elucidated by a crystal structure of yeast GS in complex with G6P and UDP-G, which showed UDP-G binding in the interdomain cleft containing the active site which is different to the allosteric site where G6P binds[89] (Figure 1-14). Although this was a catalytically inactive enzyme (E169Q mutant) UDP-G was hydrolysed, suggesting non-productive hydrolysis occurred during crystallisation[89] (Figure 1-14).

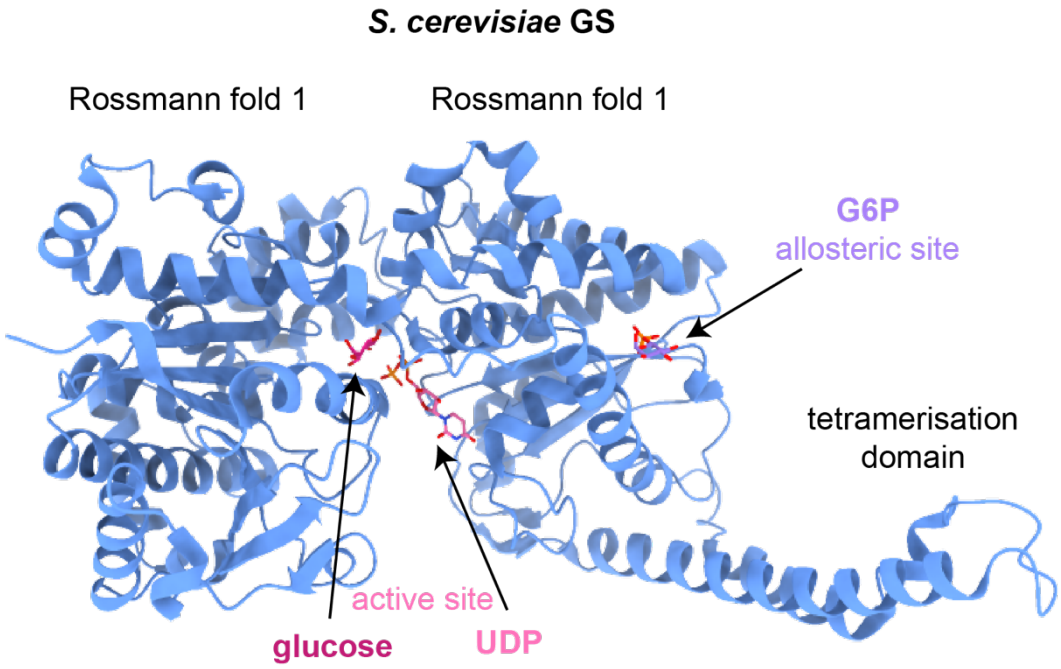

Figure 1-14: The active site and allosteric site of yeast GS

For simplicity, one monomer of the tetrameric yeast GS is shown (PDB ID 4KQM). UDP and glucose bind in the active site, which is located in a cleft between the two Rossmann fold domains. The glucose-6-phoshate (G6P) allosteric site is in a different location. The two helices which are involved in oligomerisation are labelled as 'tetramerisation domain'.

While the structures thus far are very informative in understanding GS oligomerisation and allosteric activation by G6P, the mechanism of inactivation by phosphorylation is still unknown. Thus, due to the lack of sequence conservation at the N- and C- terminal tails which contain the phosphorylation sites (Figure 1-15), how human GS is regulated still remains elusive. At the start of this project (October 2018), the structure of the human enzyme was not solved.

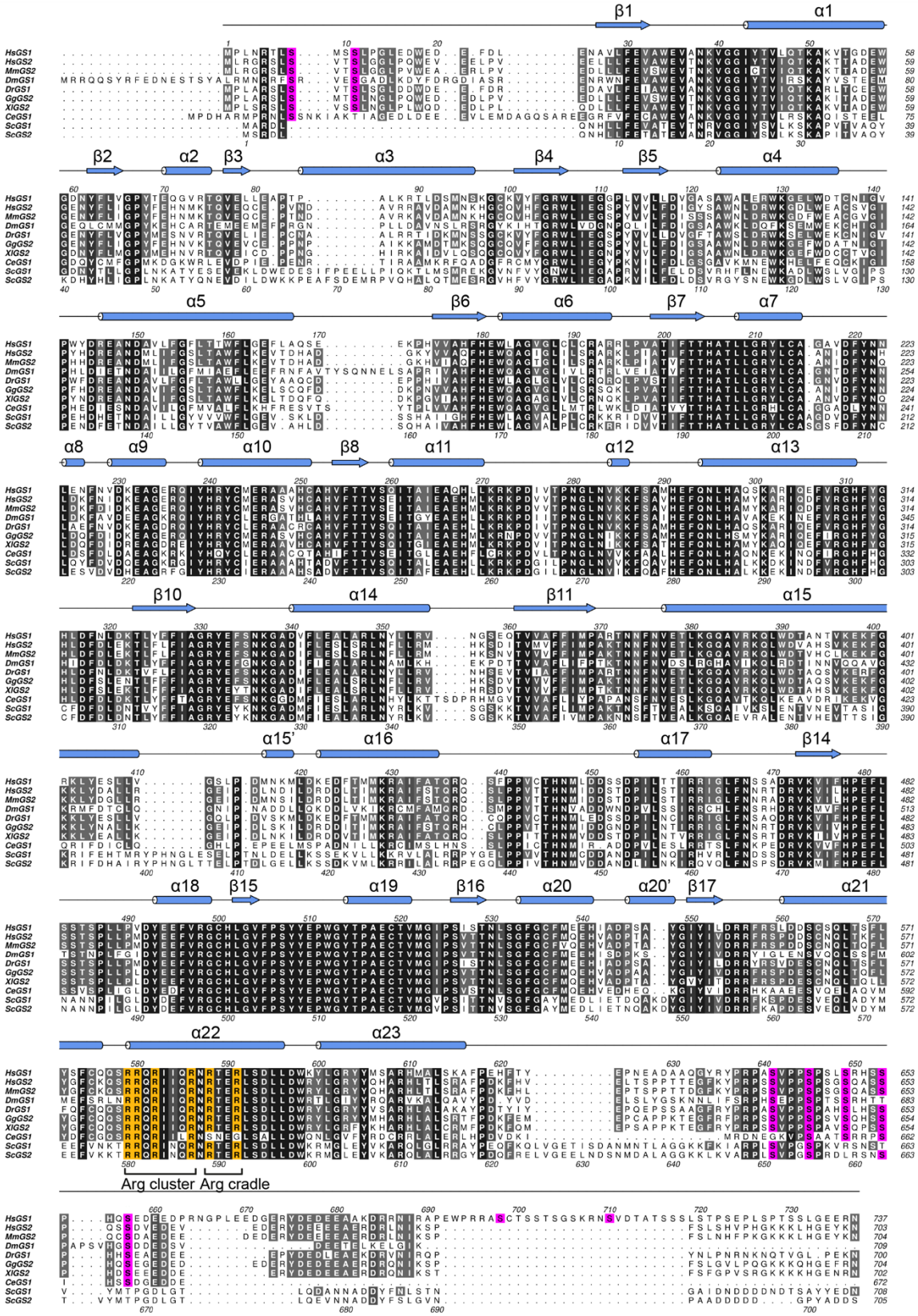

Figure 1-15: Sequence conservation of glycogen synthase

Multiple sequence alignment of glycogen synthase (GS) from the indicated species (black = conserved, white = not conserved). Sequence alignment was performed with MUSCLE and edited and displayed using ALINE. The secondary structure for human GS (HsGS) is shown at the top of

the alignment and coloured in blue. The human and yeast GS sequences are numbered at the top and bottom of the alignment respectively. Known *in vivo* phosphorylation sites are shown in magenta. Residues that form the arginine cluster and arginine cradle are coloured in orange and labelled. *Hs, H. sapiens; Sc, S. cerevisiae; Ce, C. elegans; Dm, D. melanogaster; Mm, M. musculus; Xl, X. laevis; Gg, G. gallus; Dr, D. rerio.*

Thus far, it is not known exactly how GS elongates glucose chains and also how GS remains tightly associated to glycogen. Previous studies using rabbit muscle GS proposed a two site model, composed of a catalytic site and a polysaccharide-binding site, and high catalytic activity is only achieved upon occupancy of both sites[93]. Following this, oligosaccharide-binding sites were identified in the N-terminal domain of *E. coli*[94] and *Pyrococcus abyssi*[95] GS. Structural analysis of GS from yeast identified multiple glycogen binding sites on the surface of GS (Figure 1-16). Two were located at the C-terminus, one at the N-terminus and one adjacent to the active site[90] (Figure 1-16). Each of these sites are predicted to contribute different functions to GS activity. With some sites allowing tight association between GS and glycogen, and one site involved in positioning of the non-reducing end of the glucose chain during catalysis[90]. These studies shed some light on how GS elongates the glucose chain, however, exactly how GS elongates the primer glucose chain from GN is still unclear.

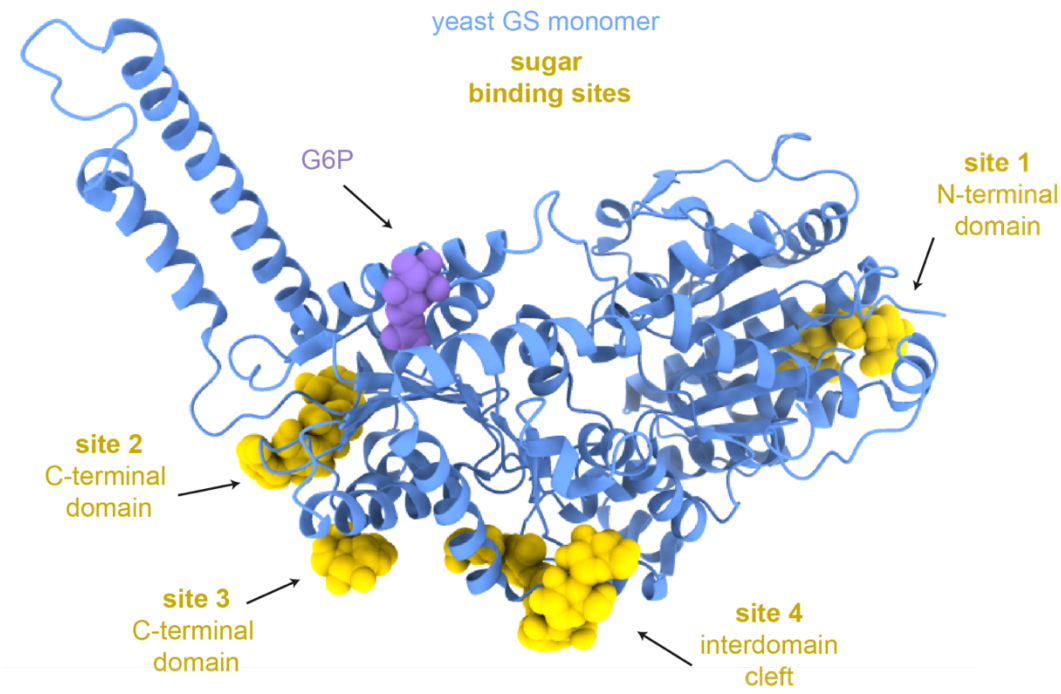

Figure 1-16: Sugar binding sites in yeast GS

A yeast GS monomer in the activated state, with glucose-6-phosphate (G6P) coloured in purple and shown in space-filling atoms. Four sugar binding sites have been identified, and these are coloured in yellow and shown in space-filling atoms (PDB ID 3RSZ and 3RT1).

1.3.3 Glycogen synthase – glycogenin interaction

When GS from rabbit skeletal muscle was first purified in 1976, it co-purified with an additional protein[96], which was later identified as GN[63]. It was not until 2006 when the interaction interface between these two enzymes was further investigated. Biochemical characterisation using deletion constructs and affinity chromatography (a.k.a. pulldown analyses) revealed that the extreme C-terminus of GN is sufficient for strong binding to GS, although other interaction interfaces may exist[69]. It also showed that the glucose polymer attached to GN is not responsible for the interaction with GS[69]. This interaction interface was structurally characterised by a crystal structure of GS from *C. elegans* in complex with a GN C-terminal peptide (GN[34]), revealing the minimal binding region to GS of GN[70] (Figure 1-17). This interaction occurs in a location different to the pockets described for UDP-G or G6P. Four GN[34] peptides bind to the GS tetramer, one for each protomer, and they each form a helix-turn-helix consisting of binding helix 1 (BH1) and BH2[92] (Figure 1-17). However, it was also shown that full-length GN had a higher affinity to GS than just the C-terminal peptide, perhaps due to an avidity effect, but additional interaction interfaces could not be ruled out[70]. In addition, mutation of GN[34] in mouse GN

reduced its interaction with GS and glycogen content in cells, showing that the interaction is essential for glycogen synthesis[70]. The observation that four GN C-terminal peptides can bind to a GS tetramer suggests that when GN is in its dimeric form, two GN dimers can bind to one GS tetramer[97]. However, unlike its yeast and *C. elegans* counterparts, human GS cannot be expressed in bacteria and also requires co-expression of GN[98,99], which potentially suggests a different interaction for the human proteins. At the start of this project (October 2018), the interaction between human GS and GN was not established.

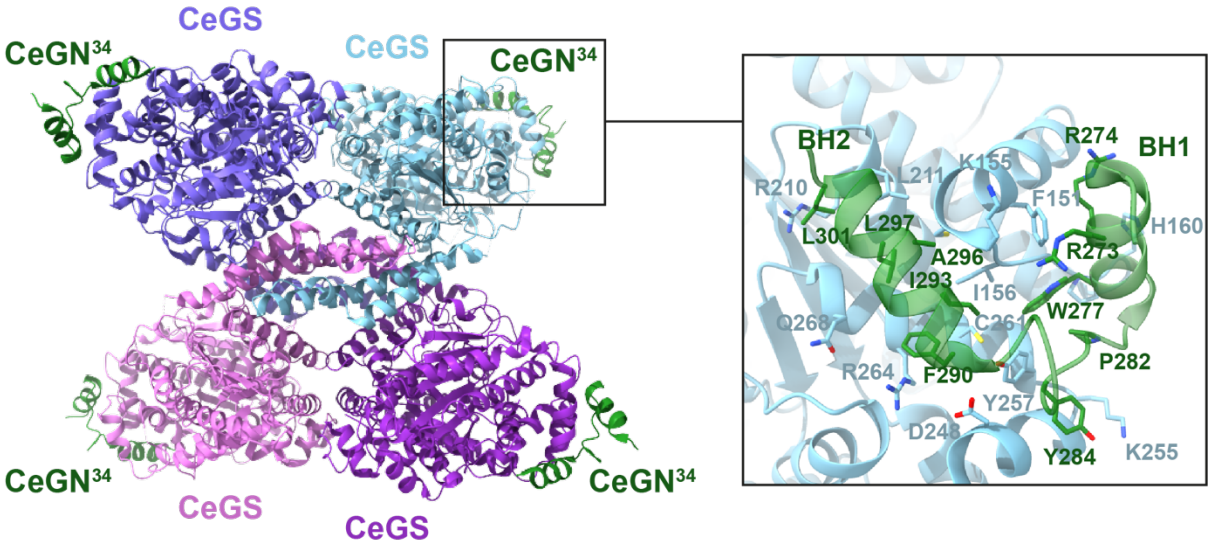

Figure 1-17: *C. elegans* GS in complex with the C-terminus of GN

The crystal structure of *C. elegans* GS tetramer, with each protomer coloured differently (CeGS), bound to four C-terminal GN peptides (CeGN[34]) (left). The interaction interface between CeGS and CeGN[34] (right). CeGN[34] forms a helix-turn-helix, consisting of binding helix 1 (BH1) and BH2. Atoms involved in the interaction are shown as sticks and labelled.

1.3.4 Glycogen branching enzyme

GBE introduces branch points in the growing glucose chain in order to create the desired tiered, branched and soluble structure of glycogen. GBE does this by cleaving an α-1,4-glycosidic residue, approximately every 8-14 glucose residues, on chains more than six glucose residues from the non-reducing end. GBE can then transfer this cleaved oligosaccharide, onto the same or a neighbouring chain, via an α-1,6-linkage (Figure 1-18)[100].

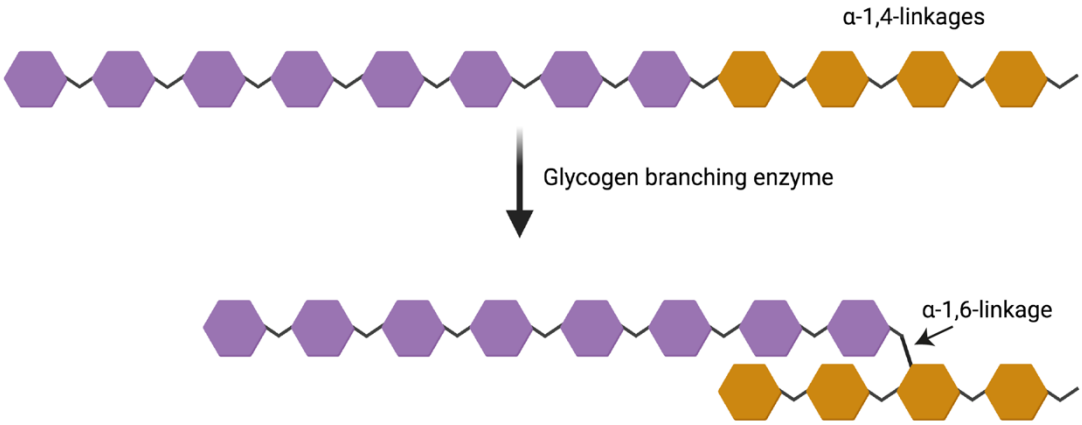

Figure 1-18: Schematic representation of the mechanism of glycogen branching enzyme
Glycogen branching enzyme (GBE) cleaves an α-1,4-glycosidic residue and transfers this onto another glucose polymer chain, but via an α-1,6-linkage, thus creating branch points. Hexagons represent glucose molecules, and orange shows the original chain whereas purple shows the residues which are relocated. Not to scale.

The degree of branching in a polysaccharide dictates its solubility, for example the highly branched glycogen is soluble in aqueous solution, whereas starch, which has less frequent branch points, is less soluble in aqueous solution. Thus, the branching of glycogen is an essential feature in the formation of functional glycogen. And indeed, glycogen storage disease type IV is caused by defects in GBE which leads to poorly branched, insoluble glycogen[101].

1.4 Regulation of glycogen metabolism

1.4.1 Glycogen synthase regulation

GS is the rate limiting enzyme in glycogen synthesis. GS can undergo allosteric activation by G6P binding and also covalent inhibition by phosphorylation, which are discussed in more detail in sections 1.4.1.1 and 1.4.1.2 respectively.

1.4.1.1 Allosteric activation

The binding of G6P has been structurally characterised by crystal structures of GS from yeast, revealing one G6P molecule can bind to one GS protomer[88]. The pocket is formed of six residues, including His286, Lys290, His500, Arg580, Arg583 and Arg587, which are conserved in eukaryotes and interact with the phosphate group of G6P through hydrogen bonding interactions[88] (Figure 1-19a). Gln283, Asn284 and His280 form hydrogen bonds

to the glucose moiety, and unlike the phosphate binding site, this site is disordered prior to G6P binding[90,91] (Figure 1-19a). G6P binding causes large conformational changes in comparison to the basal state, to result in a high activity state including opening between the two Rossmann fold domains, which leads to a more accessible active site[88] (Figure 1-19). G6P pushes the regulatory helices apart, measuring 16 Å apart in comparison to the 12 Å distance in the basal state structure[88] (Figure 1-20).

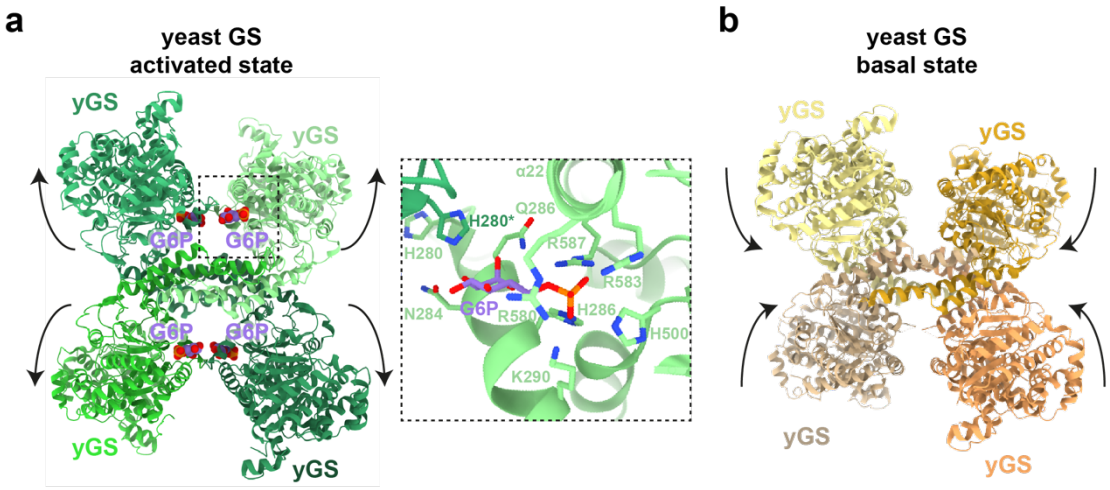

Figure 1-19: Crystal structures of yeast GS in the basal and activated state

a Crystal structure of yeast GS in the activated state (PDB 5SUK). The glucose-6-phoshate (G6P) binding site is shown by the dashed box and the interaction interface is shown with residues involved shown as sticks (right). * denotes a residue from the neighbouring subunit. The arrows represent the conformational change caused by G6P binding in comparison to the basal state shown in **b**. **b** Crystal structure of yeast GS in the basal state (PDB ID 3NAZ). Arrows show direction of conformational changes in comparison to the activated state shown in **a**.

A proposed arginine cluster, containing 6 conserved arginines (Arg579 - Arg591 in humans), on the helix α22 (Figure 1-15) is widely known as the regulatory helix, and it has an important role in the regulation of GS. Mutagenesis and structural studies of yeast GS have shown that G6P binds beneath the regulatory helices (α22), and specifically interacts with the middle two arginines (Arg582 and 586) in the arginine cluster[88,102] (Figure 1-19a).

The role of the regulatory helices has been of interest in order to understand GS regulation. Mutagenesis proposed a role for the first two arginine residues (Arg579 and R580) in the response to inhibitory phosphorylation at the yeast residue T668[102]. In addition, this study also suggested that the final two arginine residues (Arg588 and Arg591) play a role in

stabilising the basal/dephosphorylated state of the enzyme[102]. The movement of these helices in the G6P activated state in comparison to the basal state, further highlights a role for α22 in regulation of G6P, potentially through altering accessibility to the active site. Interestingly, G6P can activate GS regardless of phosphorylation state[103,104].

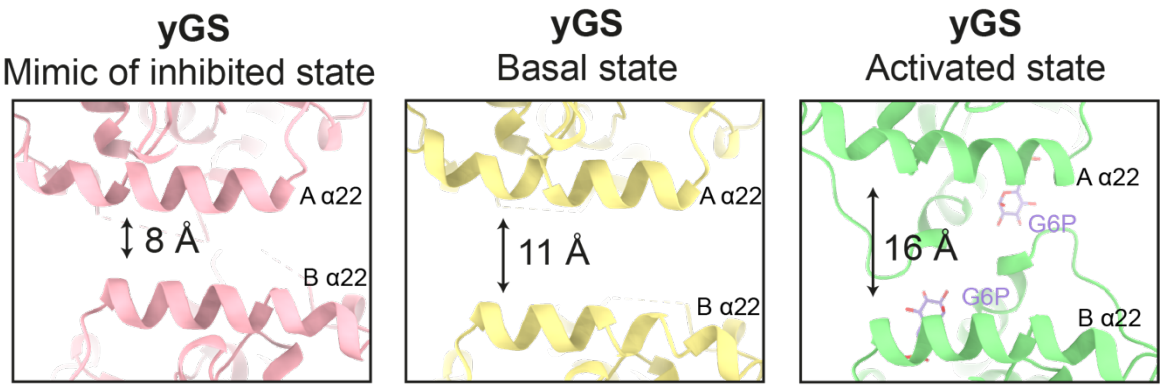

Figure 1-20: The regulatory helices of yeast GS

Comparison of distances between regulatory helices of adjacent monomers of the low activity inhibited mimic (PDB ID 5SUL), basal state (PDB ID 3NAZ) and glucose-6-phosphate (G6P) activated (PDB ID 5SUK) yeast GS (yGS) crystal structures. Quoted distances were measured from Cα of Arg592 (chain A) and -Cα of Arg581 (chain B) yeast residues.

1.4.1.2 Inhibitory phosphorylation

GS is a model example of 'multisite phosphorylation', which was first described in 1971[105], and since then many efforts have been made to determine the extent of this phosphorylation, the kinases involved and how multiple phosphorylation occurs. Nine phosphorylation sites have been identified on human muscle GS, which lie on the N- and C- terminal tails (Figure 1-21). However, human liver GS only has seven phosphorylation sites as it does not contain the two final C-terminal sites, 1a and 1b (Figure 1-21).

Kinases that can phosphorylate these sites *in vitro* have also been identified (Figure 1-21 and Table 1-1). These kinases include: AMP activated protein kinase (AMPK)[106]; PhK, phosphorylase kinase[107]; CAMKII, Ca^{2+}/calmodulin-dependent protein kinase II[108]; PKA, protein kinase A[109]; PKC, protein kinase C[110]; CK1, casein kinase 1[111,112]; GSK3, glycogen synthase kinase 3 [113,114]; PASK, Per-Arnt-Sim domain containing serine/threonine kinase[115]; DYRK, dual specificity tyrosine-regulated kinase[116] and CK2, casein kinase 2 [117].

It is now known that phosphorylation of GS occurs in a hierarchical manner, whereby the phosphorylation of one site becomes the recognition motif for subsequent phosphorylation[118]. The discovery of this phenomenon arose when it was found that sites 3a, 3b and 3c could not be phosphorylated *in vitro* by GSK3 unless site 5 was phosphorylated[113] (Figure 1-21). Later studies also showed this[119], as well as the finding that CK1 preferred phosphorylated sites, although phosphorylated substrate is not an absolute requirement[111,119]. Mutagenesis experiments identified sites 2, 2a, 3a and 3b to be most important in the inhibition of rabbit muscle GS[75,120], in addition to further emphasis for a dominant role of site 3a[121].

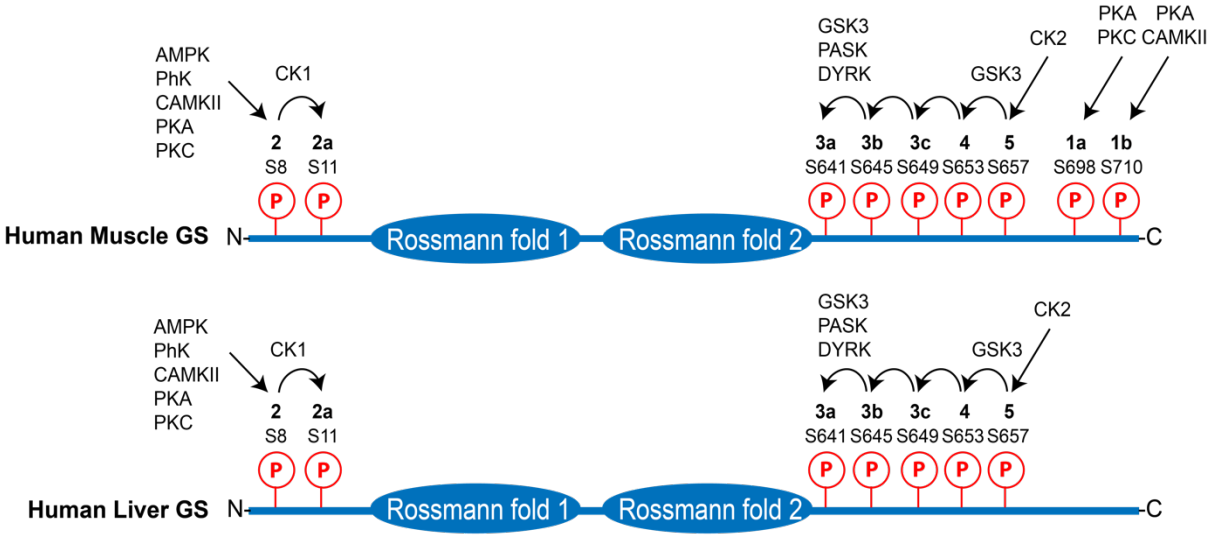

Figure 1-21: Human GS phosphorylation sites

Domain architecture of human muscle GS (top) and human liver GS (bottom). Known *in vivo* phosphorylation sites of GS are shown in red and are labelled with residue number and classical nomenclature (in bold). Known *in vitro* kinases are shown and the hierarchical fashion is indicated by the arrows. Not to scale. AMPK, AMP activated protein kinase; PhK, phosphorylase kinase; CAMKII, Ca2+/calmodulin-dependent protein kinase II; PKA, protein kinase A; PKC, protein kinase C; CK1, casein kinase 1; GSK3, glycogen synthase kinase 3; PASK, Per-Arnt-Sim domain containing serine/threonine kinase; DYRK, dual specificity tyrosine-regulated kinase; CK2, casein kinase 2.

Table 1-1: Human GS phosphorylation sites and corresponding *in vitro* kinase(s)

*Sites 1a and 1b are only found in muscle glycogen synthase

Phosphorylation site	Residue (human)	*In vitro* kinase(s)
2	Ser8	AMPK, CAMKII, PhK, PKA, PKC
2a	Ser11	CK1
3a	Ser641	GSK3, PASK, DYRK
3b	Ser645	GSK3
3c	Ser649	GSK3
4	Ser653	GSK3
5	Ser657	CK2
1a*	Ser698	PKA, PKC
1b*	Ser710	PKA, CAMKII

At the start of this project (October 2018), no structure of the phosphorylated, inhibited form of GS had been solved. However, in an attempt to understand the inhibited state, a crystal structure of yeast GS enzyme that mimicked the phosphorylated state was solved[91]. This double mutant (R589A + R592A) had an activity level similar to that of phosphorylated yeast GS yet could still be activated by G6P[91]. The structure was found to be in a tighter conformation in comparison to the basal state enzyme, with an N-terminal domain closure restricting access to the catalytic cleft[91].

However, due to the lack of similarity in the GS phosphoregulatory region across species, it is difficult to translate postulations from yeast and *C. elegans* to human. In fact, yeast Gsy2 only contains three of the seven human C-terminal phosphorylation sites and completely lacks N-terminal phosphorylation (Figure 1-15). The more closely related *C. elegans* GS contains four of the seven human C-terminal phosphorylation sites as well as both of the N-terminal sites (Figure 1-15). Insights into the GS N-terminus from *C. elegans* GS revealed close proximity to the regulatory helix, and a mechanism of reorganisation of the N-terminal tail induced by phosphorylation was predicted to occur to allow interaction between the phosphoregulatory region and the Arg cluster[70] (Figure 1-22).

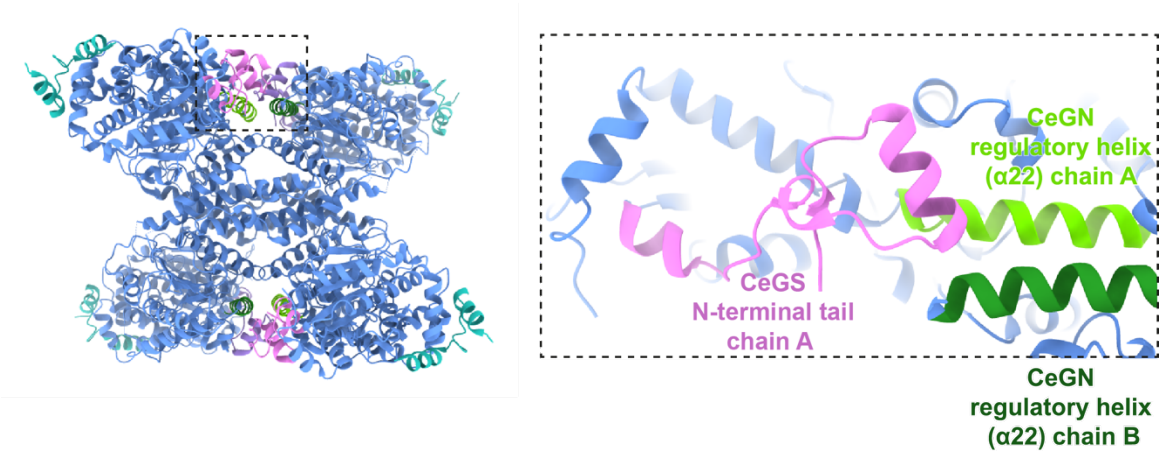

Figure 1-22: N-terminal tail of *C. elegans* GS

The *C. elegans* GS tetramer (blue) in complex with then C-terminus of GN (turquoise) (PDB ID 4QLB). The N-terminus is coloured in pink (chains A/C) or purple (chains B/D) and the regulatory helices are shown in green or dark green (left). The area boxed is shown on the right. The N-terminal tail is situated towards the regulatory helices from the same and the adjacent GS protomer.

A three-state model has been proposed for the regulation of GS, in light of enzyme kinetic and structural data[88,91,102]. The three states being an inhibited phosphorylated state, a basal, dephosphorylated state and a G6P activated state, potentially with substrate accessibility being responsible for activity. In a mimic of the phosphorylated state, the regulatory helices (α22) are closest together at 8 Å and the active site is closed[91]. In the basal state, the regulatory helices are further apart at 12 Å [88]. When G6P is bound and GS is in a high activity state, the regulatory helices are 16 Å apart and the active site is open[88,91] (Figure 1-20).

1.4.1.3 Phosphatase regulation

GS is also regulated by dephosphorylation by protein phosphatase 1 (PP1) and lambda protein phosphatase (lambda PP). It has previously been shown that lambda PP, which releases phosphates from phosphorylated serine, threonine and tyrosine residues, can dephosphorylate GS *in vitro*, leading to increased GS activity[99,122].

PP1 consists of a catalytic subunit (PP1c) and a regulatory subunit, with the latter subunit having a role in recruiting the phosphatase to its targets. Common motifs of the seven regulatory subunits identified are the 'RVSF' motif for PP1c binding, a glycogen-binding motif and a GS binding motif[123-125].

PP1 can interact with major enzymes involved in glycogen metabolism, including GS, PhK and GP, through its protein targeting to glycogen (PTG) subunit[126]. Specifically, PTG can bind to glycogen thus bringing the catalytic subunit of PP1 (PP1c) into close proximity to GS to cause its dephosphorylation[126]. Overexpressing the PTG gene increases glycogen accumulation by recruiting PP1 to glycogen where it can increase GS activity through dephosphorylation[127].

Another glycogen targeting subunit of PP1, R_{GL}/G_M, the first of which to be identified[128], is exclusively expressed in striated muscle and can recruit the substrate, GS[123]. It has involvement in the activation of GS in response to exercise[129]. A prevalent variant in the gene that encodes R_{GL} impairs glycogen synthesis and decreases human skeletal muscle glycogen levels, and thus there is evidence to support a critical role for these regulatory subunits in governing glycogen accumulation in humans[130,131].

1.4.2 Regulation by insulin

GS was the first example of an intracellular target for insulin[132]. In the fed state, insulin binding to the insulin receptor stimulates GLUT4 translocation to the membrane, leading to glucose uptake. HK converts this glucose into G6P, which can either allosterically activate GS or be converted into UDP-G to be used by GS as a substrate, thus GS is able to synthesise glycogen (Figure 1-23). Insulin binding also leads to the phosphorylation of GSK3, through pathways involving PI3K and Akt, which causes GSK3 inactivation and thus prevents GS inactivation, allowing glycogen synthesis to occur (Figure 1-23). Insulin can also reduce the activity of other kinases, including AMPK and PKA[133], again leading to increased GS activity (Figure 1-23). The mechanism of insulin mediated PP1 regulation is unknown. It has been suggested that insulin causes PP1 activation by recruiting PP1 to its substrates[134]. Through PTG recruitment of not only GS, but also GP and PK, PP1 can stimulate glycogen synthesis by GS activation and also dephosphorylates and thus inhibits GP and PK[135] (Figure 1-23).

The multiple pathways of insulin mediated regulation of glycogen metabolism indicates a level of redundancy, for example it has been shown that glycogen content is not disrupted in GSK3 knockout mice[136]. In diabetes, abnormal insulin function is associated with defective glycogen metabolism[1].

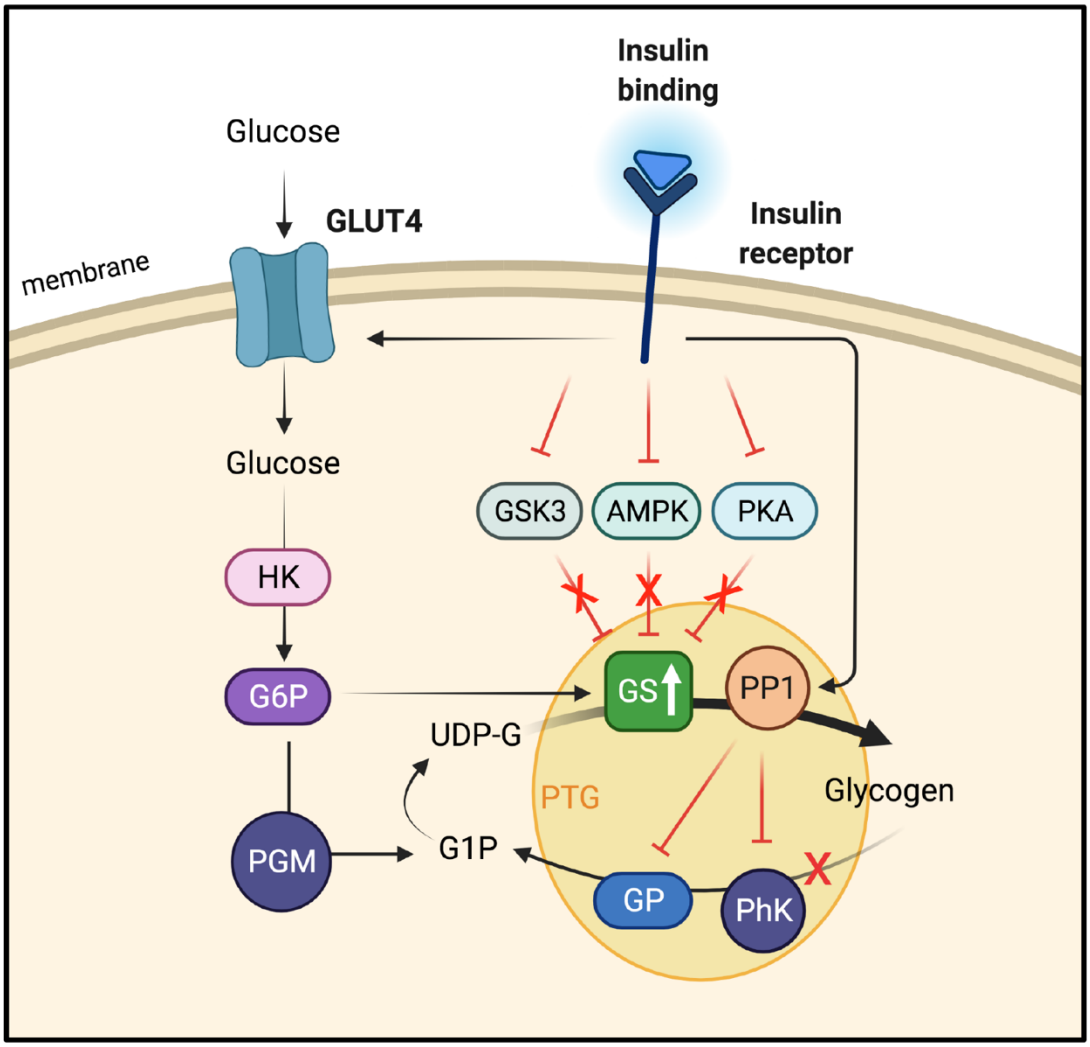

Figure 1-23: Insulin mediated regulation of glycogen metabolism

Schematic representation of the proposed regulation of glycogen synthesis by insulin. Insulin binding to the insulin receptor inhibits glycogen synthase kinase 3 (GSK3), AMP-kinase (AMPK) and protein kinase a (PKA), which in turn prevents in the inhibition of glycogen synthase (GS) and thus stimulates glycogen synthesis. Insulin stimulates glucose uptake through GLUT4, which is then converted to glucose-6-phoshate (G6P) by hexokinase (HK). G6P can allosterically activate GS, and can also be converted to G1P by phosphoglucomutase (PGM). G1P can be converted to UDP-glucose (UDP-G) which is the substrate for GS. PP1 is recruited to various substrates via protein targeting glycogen subunit (PTG), including to GS which in turn is activated, and also glycogen phosphorylase (GP) and phosphorylase kinase (PhK), which in turn are inhibited. Inhibition of GP and PhK prevent glycogen breakdown, leading to the accumulation of glycogen. Adapted from Saltiel, 2021[137].

1.4.3 Regulation by glucagon and adrenaline

Other hormones that are involved in the regulation of glycogen metabolism are glucagon and adrenaline/epinephrine, which in contrast to insulin, promote glycogenolysis. Glucagon functions in the liver, whereas adrenaline mainly acts in the muscle. During fasting, blood glucose and insulin levels fall which results in secretion of glucagon from pancreatic α-cells and binding of glucagon to the glucagon receptor, which is a member of the G-protein-coupled-receptor (GPCR) family. This activates liver adenylate cyclase, which converts ATP to cAMP, thus increasing cAMP levels and stimulating PKA. PKA in turn phosphorylates and inactivates GS, thus halting glycogen synthesis (Figure 1-24). PKA can also activate PhK, leading to GP activation and thus an increase in glycogen degradation[138] (Figure 1-24).

Adrenaline also leads to the increase in cAMP levels and subsequent activation of PKA, through binding to the β-adrenergic receptor, another class of GPCRs[139]. It has been proposed that PKA phosphorylates R_{GL}, decreasing the activity of PP1[140]. In addition, PKA can phosphorylate and activate the PP1 inhibitory protein in response to adrenaline stimulation[141], leading to the inactivation of PP1. Moreover, GS remains phosphorylated and inactive and glycogen synthesis cannot occur.

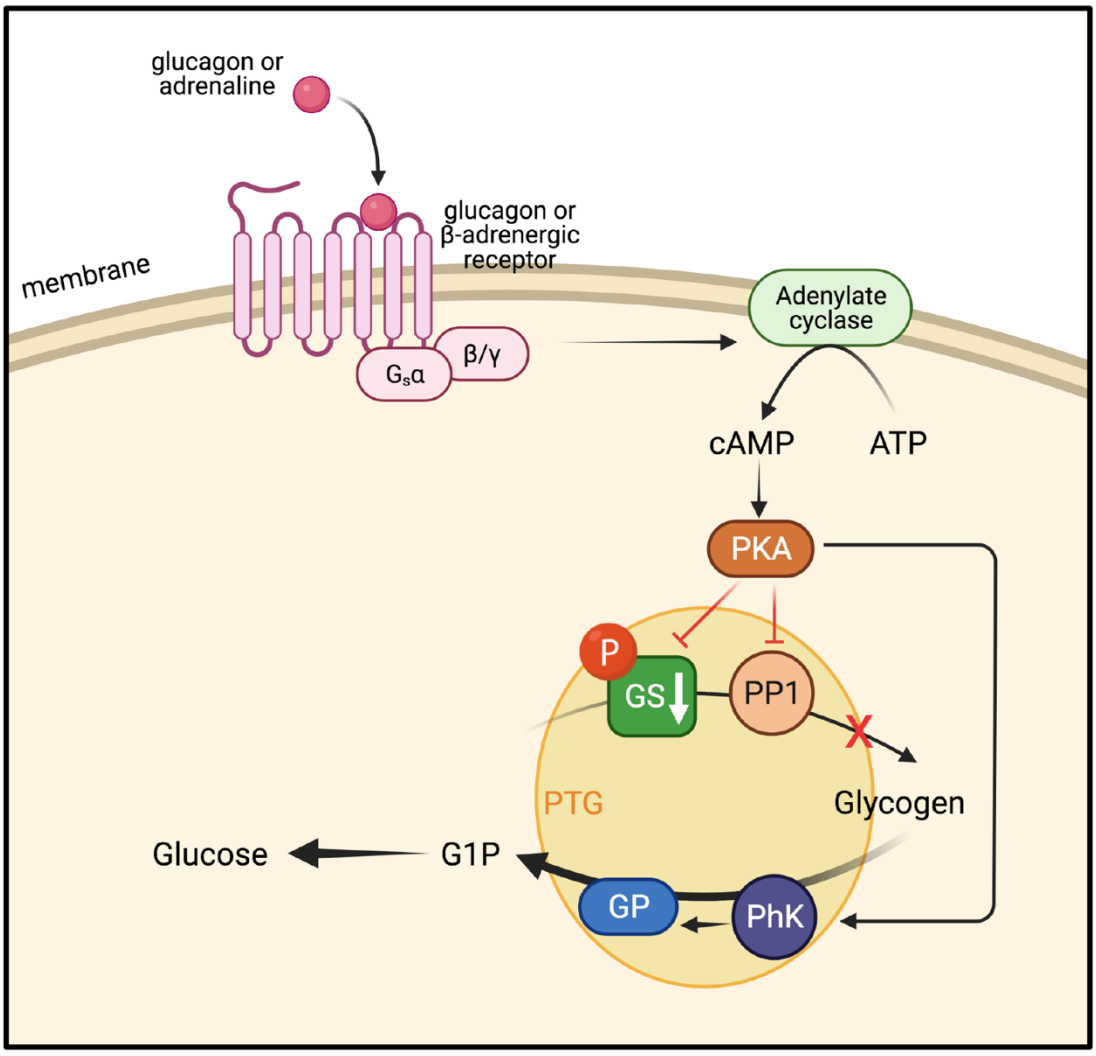

Figure 1-24: Glucagon and adrenaline regulation of glycogen metabolism

Schematic representation of the proposed regulation of glycogen synthesis by glucagon and adrenaline. Binding of glucagon and adrenaline to the glucagon or β-adrenergic receptor leads to adenylate cyclase converting ATP to cAMP. cAMP causes the activation of protein kinase A (PKA) which inactivates both glycogen synthase (GS) and protein phosphatase 1 (PP1), which are brought into close proximity via protein targeting glycogen (PTG), thus halting the synthesis of glycogen. PKA also activates phosphorylase kinase (PhK), which activates glycogen phosphorylase (GP), thus promoting the degradation of glycogen to glucose-1-phosphate (G1P) and glucose.

1.4.4 Regulation by exercise

One of the major roles of glycogen is to provide energy during exercise, and thus following glycogen depletion during exercise these glycogen stores need to be replenished. However, both activation and inactivation of GS has been observed in response to exercise[29,142].

Similar to insulin, exercise stimulates the uptake of glucose via GLUT4, leading to activation of GS via the pathway shown in Figure 1-23. How exercise mediates GS phosphorylation and inhibition is not well understood. One proposed action is the dephosphorylation and thus activation of GS via R_{GL} recruitment of PP1c. However, an opposing regulatory pathway which leads to GS inactivation by kinases including AMPK, PKA and PhK has also been suggested[143]. AMPK activation occurs when muscular contraction causes breakdown of ATP which increases cAMP levels, which as well as activating AMPK also activates GP. Thus, this mechanism causes inactivation of GS and stimulates breakdown of glycogen. However, more investigation is required to understand this complex mechanism.

It is thought that the biggest contributing factor of glycogen metabolism during exercise is the activation of GP to stimulate glycogen breakdown, and during exercise recovery it is a combination of both GS activation and GP inhibition to stimulate glycogen synthesis[144].

In addition, it has been shown that exercise causes increased GN gene expression in human skeletal muscle, suggesting a potential role for GN in the regulation of glycogen metabolism in response to exercise[145]. However more investigation is required to ascertain the exact role of GN in this process.

1.5 Dysregulation of glycogen metabolism

1.5.1 Glycogen storage diseases

To date, there are 16 described glycogen storage diseases (GSDs), characterised by the enzyme deficiency that causes the diseases, where abnormal glycogen is a main characteristic. Usually defects in enzymes directly involved in glycogen synthesis or breakdown are the major causes of the disease. But they can also be caused by mutations/defects in enzymes that affect glycogen degradation or glucose utilisation. In addition, defects/mutations in regulatory proteins of glycogen metabolism can also lead to some types of GSDs, for example Lafora disease, which is discussed in more detail in section 1.5.1.2. Although skeletal muscle and liver are the two main locations for glycogen storage, disease manifestation is present widely. Symptoms of GSDs can include muscle weakness, fatigue, cardiomyopathy and heart failure, and unfortunately the fatal outcome of death at early ages is not uncommon in GSD patients. The different GSDs are summarised in Table 1-2, and some are discussed further in sections 1.5.1.1 and 1.5.1.2.

Table 1-2: Glycogen storage diseases

* Fanconi-Bickel syndrome is no longer classed as a GSD

Disease name	GSD type	Enzyme (gene) deficiency	Enzyme function	Symptoms
-	type 0	Glycogen synthase 2 (GYS2)	Glycogen synthesis, catalyses transfer of glucose from UDP-glucose	Growth failure, muscle cramping
Von Gierke's disease	type I	Glucose-6-phosphatase (G6PC)	Catalyses the hydrolysis of glucose-6-phosphate to glucose	Growth failure
Pompe disease	type II	Acid alpha glucosidase (GAA)	Breakdown of glycogen in the lysosome	Muscle weakness, heart failure, early death
Cori's disease	type III	Glycogen debranching enzyme (AGL)	Breakdown of glycogen	Myopathy
Andersen disease	type IV	Glycogen branching enzyme (GBE1)	Glycogen synthesis, adds branch points to the growing glycogen chain	Muscle weakness, myopathy, early death
McArdle disease	type V	Muscle glycogen phosphorylase (PYGM)	Muscle glycogen breakdown, release of G1P from cleaving -1-4-glycosidic bonds	Exercise induced cramps
Hers' disease	type VI	Liver glycogen phosphorylase (PYGL)	Liver glycogen breakdown, release of G1P from cleaving alpha-1,4-glycosidic bonds	Growth retardation
Tarui's disease	type VII	Muscle phosphofructokinase (PFKM)	Glycolysis, converts fructose-6-phosphate to fructose 1,6-bisphosphate	Muscle weakness, exercise intolerance
-	type VIII	Hepatic glycogen phosphorylase kinase (PYKL)	Phosphorylates and activates liver glycogen phosphorylase, which subsequently breaks down glycogen	Growth retardation
-	type IX	Phosphorylase kinase (PHKA2)	Phosphorylates and activates muscle glycogen phosphorylase, which subsequently breaks down glycogen	Delayed motor development, exercise intolerance
-	type X	Phosphoglycerate mutase (PGAM2)	Glycolysis, converts 3-phosphoglycerate to 2-phosphoglycerate	Muscle weakness
-	type XI	Muscle lactate dehydrogenase (LDHA)	Glycolysis, catalyses the interconversion of pyruvate and L-lactate	Fatigue, muscle pain and cramps during exercise
Fanconi-Bickel syndrome*	type XI *	Glucose transporter (GLUT2)	Glucose transporter in β-cells of pancreatic islets and hepatocytes	Fragile bones, growth retardation
Red cell aldolase deficiency	type XII	Aldolase A (ALDOA)	Catalyses the conversion of fructose 1,6-bisphosphate to glyceraldehyde 3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP)	Exercise intolerance
-	type XIII	β-enolase (ENO3)	Catalyses the conversion of 2-phosphoglycerate to phosphoenolpyruvate	Muscle pain, exercise intolerance
-	type XV	Glycogenin-1 (GYG1)	Glycogen synthesis, initiates the process by autoglucosylation	Progressive muscle weakness
Lafora disease	-	Laforin (EPM2A) or malin (EMP2B)	Glycogen metabolism regulatory proteins	Neurodegeneration, seizures, adolescent death

1.5.1.1 Pompe disease (GSD type II)

Pompe disease is an autosomal recessive disease caused by mutations in the GAA gene, causing a deficiency in acid-α-glucosidase protein, which is the enzyme responsible for breaking down glycogen in the lysosomes[146] (Table 1-2). Moreover, glycogen accumulation occurs in the lysosomes leading to lysosomal dysfunction and thus patients suffer from muscle weakness, heart failure, and in the infantile disorder, an early death by ~2 years of age[147]. GAA mutations that cause Pompe disease are compound heterozygous, and the majority are missense mutations, although a small number of deletion mutations have also been identified[148].

Pompe disease is currently treated by enzyme replacement therapy (ERT), with human recombinant GAA[149]. This has so far been successful by significantly slowing the rate of disease progression and life-saving in the infant-onset patients. However, it seems that the beneficial effects of ERT plateau within a few years of treatment, thus there remains a need for more effective treatments for Pompe disease patients[150,151].

Moreover, some pharmaceutical companies are targeting GS as potential substrate reduction therapy (SRT) to treat Pompe disease, in the aim of reducing glycogen synthesis, thus alleviating the accumulation of glycogen in the lysosomes (Figure 1-25).

A *GAA/GYS1* double knockout mouse demonstrated a significant reduction in glycogen levels in the skeletal muscle and liver, as well as complete correction of cardiomegaly, thus suggesting that abolition of muscle glycogen synthesis may lead to improvement of the defects seen in Pompe disease[152]. This was further investigated through suppression of muscle GS1 in Pompe mice, which showed significant decrease of aberrant glycogen accumulation in the heart, diaphragm and quadriceps, without effecting liver glycogen[153]. It has also been shown that reduction in muscle glycogen is well tolerated, and moreover supports the safety of using SRT as a treatment for Pompe disease[130].

However, it is important to note that glycogen distribution widely differs across animals. In humans, ~85% of glycogen is stored in the muscle, with only ~15% located in the liver[154]. In dogs, the distribution between muscle and liver glycogen is more similar, with ~48% and ~51% respectively[155]. Whereas in rodents, glycogen is predominantly stored in the liver, with rats storing ~88%[156] and mice storing ~71%[157] of glycogen in the liver. Moreover, it is important to take these findings into consideration when testing and analysing potential therapeutics regarding glycogen metabolism.

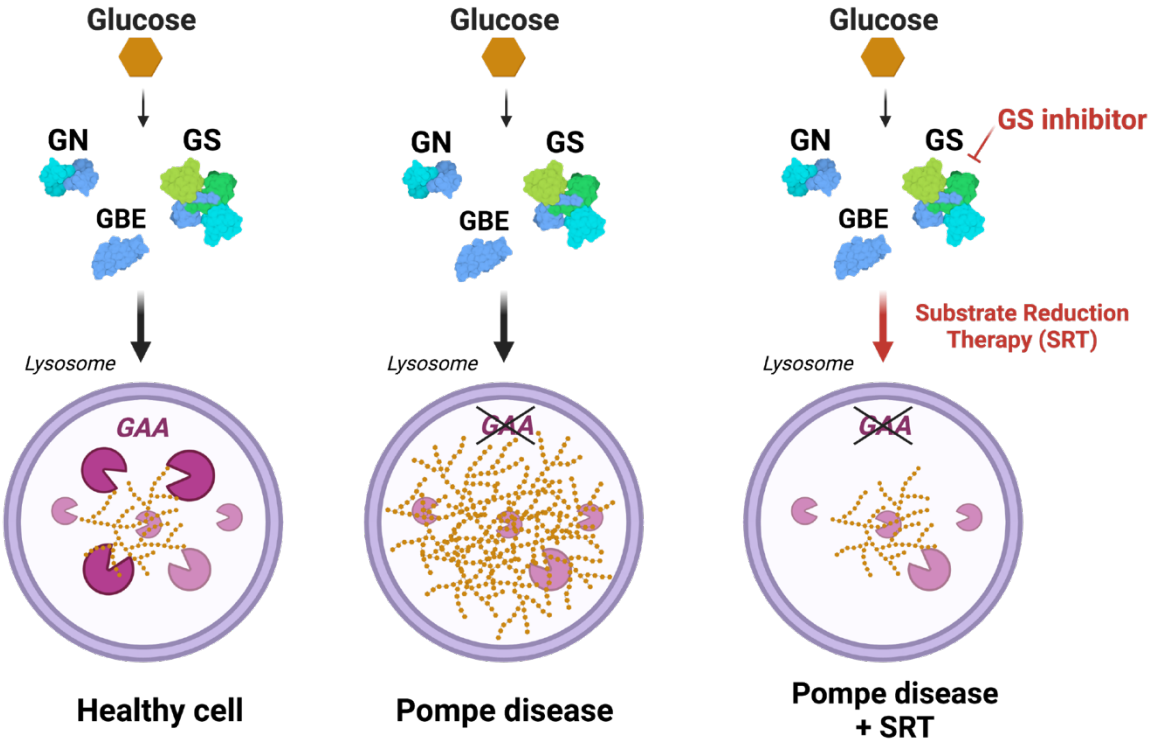

Figure 1-25: Schematic representation of Pompe disease and substrate reduction therapy (SRT)

Glucose (displayed as an orange hexagon) is converted into glycogen (displayed as chains of orange hexagons) through the action of glycogenin (GN), glycogen synthase (GS) and glycogen branching enzyme (GBE). Some glycogen is stored in lysosomes. In a healthy cell, acid alpha glucosidase (GAA) breaks down glycogen in the lysosomes (left). In Pompe disease, GAA is deficient which leads to lysosomal glycogen accumulation (middle). Substrate reduction therapy (SRT) by targeting GS to reduce its activity is a potential therapeutic by reducing glycogen accumulation in the lysosomes (right).

1.5.1.2 Lafora disease

Lafora disease is a neurodegenerative disorder, typically with onset in adolescent years followed by death within ten years[158]. A characteristic of Lafora disease is the presence of inclusion bodies that comprise of poorly branched, insoluble glycogen particles named Lafora bodies[159]. These glycogen accumulations are found in organs with high glucose metabolism, including the heart, liver and brain. In the brain, Lafora bodies are present in the majority of neurons, specifically in their dendrites and cell bodies[159,160]. 90% of Lafora disease cases are accounted for by mutations in either Epilepsy progressive myoclonus type 2A (*EPM2A*) and Epilepsy progressive myoclonus type 2B (*EPM2B*)[161,162]. *EPM2A* encodes a dual specificity protein phosphatase named laforin, that contains a

carbohydrate binding domain that allows tight binding to glycogen [163-165]. *EPM2B* encodes an E3 ubiquitin ligase named malin, which can interact with laforin and catalyses its polyubiquitination[166,167]. These two proteins function as a complex to regulate GS activity[168] as well as translocation of GLUTs[169] to indirectly affect glycogen accumulation.

It is still not clear exactly how mutations in *EMP2A* and *EMP2B* cause Lafora disease. One suggested role for laforin is that of a repair protein to remove the phosphate that is mistakenly incorporated into glycogen by GS, and thus deficiency of laforin leads to hyperphosphorylated Lafora bodies[170-172] (Figure 1-26). A potential role for malin in Lafora disease is the ubiquitination of enzymes involved in glycogen metabolism[173-175], including PTG[176] (Figure 1-26). Despite an increased understanding in the molecular basis of Lafora disease in recent years, many questions remain unresolved.

Recently it has been found that targeting *GYS1* mRNA with antisense therapy could prevent and halt the progression of Lafora disease[177]. Additional studies also showed that knocking out GS1 later in disease progression halted LB formation and decreased neuroinflammation, suggesting this as a potential therapeutic but also highlights a requirement for early disease diagnosis as late suppression of GS did not halt the accumulation of LBs[178,179].

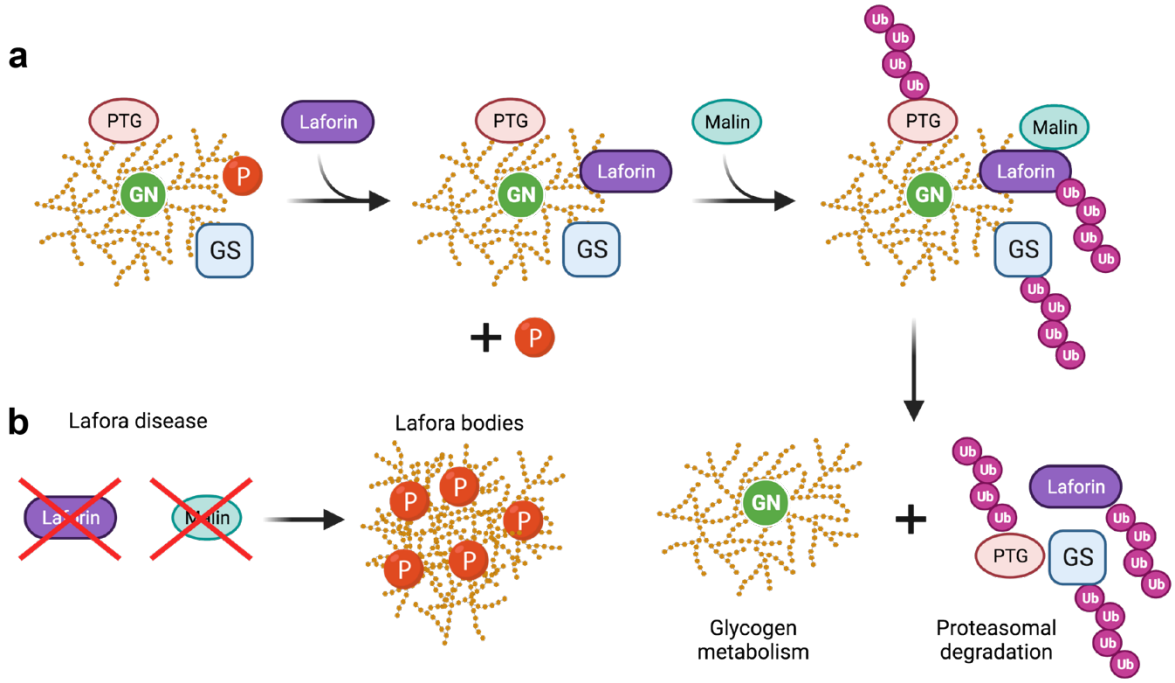

Figure 1-26: A potential role for laforin and malin in glycogen metabolism

a A schematic representation of the proposed role of laforin and malin. Laforin can be recruited to glycogen, where it removes phosphates. Malin binds to laforin and can subsequently ubiquitinate laforin, glycogen synthase (GS) and the PTG subunit of PP1, causing their proteasomal degradation

thus allowing glycogen metabolism to proceed normally. Adapted from Romá-Mateo, Sanz & Gentry, 2012[173]. **b** In Lafora disease, defects in laforin and malin cause hyper phosphorylated, poorly branched glycogen particles named Lafora bodies.

1.5.2 Ageing and senescence

Ageing is associated with glycogen accumulation in various locations, including the brain neurons, liver and muscle[180-184]. In addition, glycogen biogenesis is often implicated in age-related diseases and cancer. However, the dysregulation of glycogen metabolism in age-related diseases is unclear, as decrease of glycogen synthesis activity has been observed[185,186], as well as the conflicting observation of accumulation of glycogen[180,187]. GSK3, which is a multifunctional kinase, of which GS is a substrate (see section 1.4.1.2), has been identified as defective in multiple age-related diseases[184,188,189]. In senescent cells (cells which do not divide but do not die) GSK3 becomes inactivated, which leads to GS activation and thus resulting in glycogen accumulation[184]. In addition, in ageing macrophages and microglia, glycogen synthesis is promoted through the activation of GS[190].

Until recently, it was unclear how abnormal, unbranched glycogen was distinguished from normal glycogen. However, it was revealed that the E3 ubiquitin ligase HOIL-1 can recognise the unbranched glycogen and cause its monoubiquitylation *in vitro*[191]. Polyubiquitylation may then occur by other E3 ubiquitin ligases, and targeting the abnormal glycogen for degradation[191].

1.6 Chapter summary

Glycogen is a polymer of glucose, forming a macromolecular particle with multiple associated proteins bound covalently and non-covalently. This energy store can be created when glucose levels are high, ready for glucose to be released and used for glucose homeostasis and also during exercise. GN initiates glycogen synthesis and this primer glucose chain can be extended by GS and GBE creates branch points. GN and GS form a complex and GS activity is tightly regulated, and the molecular basis for GS-GN association and GS activity regulation are the main investigation in this thesis. Glycogen metabolism is implicated in many diseases, including 16 rare diseases named GSDs.

1.7 Project aims

At the start of this project (October 2018), little was known about how GS and GN coordinate to synthesise glycogen. The mechanism of GS inactivation, and how the opposing pressures of allosteric activation and covalent inhibition are coupled, were also unknown. In addition, the interest of targeting GS as a therapeutic for GSDs is severely hindered by the lack of structural insights into human GS. Thus, the main objective of this work was to structurally and biochemically investigate the full-length human GS-GN complex. My results shown in chapters 3, 4 and 5 collectively aim to answer the following key questions:

- How do GS and GN coordinate during glycogen synthesis?
- What is the structure of human GS?
- What is the mechanism of human GS inhibition?
- How are the contrasting pressures of GS activation and inhibition coupled?

2 Chapter 2: Materials and methods

2.1 Preparation of plasmids

2.1.1 Vectors

The GS-GN complex was produced in both pFL[192] and pFastBac vectors. Co-expression or a bicistronic vector is required to successfully express both GS and GN, as GS alone does not have a high level of expression. In the pFL vector, GS was cloned into multi cloning site 1 (MSC1) and GN was cloned into MCS2, which was performed by Dr Elton Zeqiraj. GN was expressed with an N-terminal 6 x His tag. The GN(Y195F) mutation was generated to produce a non-glucosylated species, as Y195 is the site of human GN autoglucosylation. The mutation was engineered by site directed mutagenesis by Dr Elton Zeqiraj.

For co-expression of GS and mutants the genes encoding human GS1 and human GN1 were cloned in pFastBac HTB vectors, both with a 6 x His purification tag followed by a His-tagged tobacco etch virus (TEV) site at the N-terminus.

Lambda PP and protein phosphatase 1 alpha (PPP1CA) were both purchased from MRC PPU Reagents and Services in the pGEX6P1 plasmid. Lambda PP contains an N-terminal Glutathione-S-transferase (GST) tag and a C-terminal 6 x His tag. PPP1CA has an N-terminal GST tag.

2.1.2 Preparation of competent *E. coli* cells

This protocol was used to prepare *E. coli* strains DH5α and BL21(DE3). Glycerol stock of the cell were streaked onto a Luria Broth (LB)-Agar plate and incubated at 37 ºC overnight. One colony from this plate was then used to grow a 5 mL culture in LB at 37 ºC overnight. 1 mL of this overnight culture was then used to inoculate 100 mL of LB, which was subsequently grown at 37 ºC until it reached an optical density measured at a wavelength of 600 nm (OD_{600}) of 0.2-0.6 arbitrary units (AU). The culture was incubated on ice for 10 min, and then pelleted by centrifugation at 2700*g* for 10 min at 4 ºC. The cells were then resuspended in 33 mL of buffer containing 100 mM rubidium chloride, 50 mM manganese (II) chloride tetrahydrate, 30 mM potassium acetate, 10 mM calcium chloride dihydrate and 15% (w/v) glycerol. The cells were incubated on ice for 15 min, followed by centrifugation at 580*g* for 15 min at 4 ºC. Cells were then resuspended in 4 mL buffer containing 10 mM 4-Morpholinepropanesulfonic acid, 3-(N-Morpholino)propanesulfonic acid (MOPS), 10 mM

rubidium chloride, 75 mM calcium chloride dihydrate and 15% (w/v) glycerol. Cells were incubated on ice for 15 min before creating 50 µL aliquots and storing at -80 ºC.

2.1.3 DNA transformation into *E. coli* cells

This protocol was used for plasmid DNA transformation into competent *E. coli* strains DH5α, BL21(DE3) and One Shot™ Stbl3™ cells. Competent cells were thawed on ice and ~50-100 ng of plasmid DNA was added and incubated on ice for 30 min. Cells were then heat-shocked in a water bath at 42 ºC for 90 s, then incubated on ice for 15 min. 1 mL of sterile LB media was added and the transformation mixture was left shaking at 37 ºC for 1 hour. The transformation mixture was then spun at 17,000*g* for 2 min, 700 µL of the supernatant was removed. The remaining mixture was resuspended and plated onto LB agar plates containing the appropriate antibiotic and incubated at 37 ºC overnight.

The following protocol was used for plasmid DNA transformation into competent DH10MultiBac Turbo cells. Competent cells were thawed on ice and 500-1000 ng of plasmid DNA was added and incubated on ice for 30 min. Cells were then transferred to a BioRad GenePulser cuvette and electroporated (1.8 kV, 1 pulse). The cells were then transferred to a sterile 1.5 mL tube and incubated on ice for 10 min. 450 µL of LB was added to the mixture and incubated at 37ºC for 4-6 hours. Neat and serial dilutions of the transformation mixture (1:10, 1:100, 1:1000) were plated onto LB-agar plates containing 10 µg/mL gentamycin, 50 µg/mL kanamycin, 10 µg/mL tetracycline, 4 µg/mL Isopropyl β-d-1-thiogalactopyranoside (IPTG), 83 µg/mL X-gal and the appropriate plasmid antibiotic. Plates were protected from light and incubated at 37 ºC for 48 hours.

2.1.4 Agarose gel electrophoresis

Tris-boric acid-ethylenediaminetetraacetic acid (EDTA) (TBE) was made as a 10X stock solution, which was diluted to 1X solution which contains 0.089 M Tris, 0.089 M Borate and 0.002 M EDTA. 1% (w/v) agarose gels were prepared by dissolving agarose in TBE with SYBR™ safe DNA gel stain (Invitrogen). Loading dye (Invitrogen) was added to DNA samples before loading into the gel which was subsequently ran at 100 V. Gels were visualised by UV light.

2.1.5 Purification of plasmid DNA

Plasmids were purified from 5 mL overnight growth of *E. coli* strains DH5α using QIAGEN Mini-Prep Plasmid kit according to the manufacturers protocol. Concentration was determined using a Nanodrop 8000 Spectrophotometer (ThermoScientific).

2.1.6 Purification of bacmid DNA

Following the DH10Bac™ transformation described in section 2.1.3, 5 white colonies per construct were streaked onto a LB-Agar plate (containing 10 μg/mL gentamycin, 50 μg/mL kanamycin, 10 μg/mL tetracycline, 4 μg/mL IPTG, 83 μg/mL X-gal and the appropriate plasmid antibiotic) and also used to grow a 5 mL overnight culture at 37 ºC. A blue colony was also streaked onto the LB-Agar plate as a negative control, and the plate was subsequently incubated at 37 ºC protected from light. After confirmation of a white phenotype on the re-streaked plate, the corresponding overnight growth was pelleted by centrifugation at 2,900*g* for 10 min. Buffers from the QIAGEN miniprep kit were used to purify the bacmid DNA. The supernatant was discarded and the pellet resuspended in 300 μL P1 buffer, followed by addition of 300 μL lysis (P2) buffer and incubation for no longer than 5 min. Followed by addition of 300 μL neutralisation (N3) buffer and inversion until a homogenous mix is achieved. The mixture was spun at 17,000*g* for 12 min at room temperature, and the supernatant was transferred to a fresh 1.5 mL tube, which was spun at 17,000*g* for 8 min at room temperature. The supernatant was transferred to a fresh 1.5 mL tube and 700 μL isopropanol was added and the tube was inverted slowly, followed by centrifugation at 17,000*g* for 12 min at room temperature. The isopropanol was removed without disrupting the pellet and 200 μL 70% ethanol was added to the pellet followed by centrifugation at 17,000*g* for 8 min at room temperature. The 70% ethanol was removed and 50 μL 70% ethanol was added. In a tissue culture hood, the 70% ethanol was removed and the pellet left to dry for 10 min. 35 μL filtered H_2O was added and the pellet was left to dissolve for 10 min. After mixing, the concentration was measured and the presence of the DNA insert was confirmed by performing a PCR reaction with insert primers and run the PCR product on an agarose gel. Successful bacmids were subsequently used for transfection into *Spodoptera frugiperda* (*Sf*9) cells.

2.1.7 Cloning of GN(Y195F)

170 μg of GS-GN(Y195F) in pFL vector and an empty pFastBac HTB vector were digested with SfoI and NotI (New England Biolabs) at 37 ºC for 2 hours. The digested DNA was run on a 1% agarose gel (as described in 2.1.4). The appropriate bands were excised from the gel and purified using the QIAGEN gel extraction kit according to the manufacturers protocol. To calculate how much insert to use in the ligation reaction, this equation was used: ng insert = 3(ng of vector)(kb insert)/kb vector, where ng of vector = 30. Using a 3:1 ratio of insert to vector, 20 ng of insert and 6 ng of vector was used in a ligation reaction with T4 DNA ligase (Roche) for 1 hour at room temperature. The reaction mix was then

transformed into One Shot™ Stbl3™ competent cells, using the protocol described in 2.1.3.

2.1.8 Site directed mutagenesis

A PCR based site directed mutagenesis approach was used to create point mutations. For site directed mutagenesis experiments, the WT plasmid (10 ng) was used as a template and kod hot start polymerase (Novagen) was used. The PCR reaction mix was initially denatured at 95 ºC for 2 min, followed by 20 cycles of denaturation (95 ºC for 20 s), annealing (highest primer melting temperature, 10 s) and elongation (70 ºC, 3 min). And subsequently a final elongation step at 70 ºC for 4 min. To remove the parent plasmid, *Dpn1* (New England Biolabs) was added to the PCR reaction mix and the digestion was carried out at 37 ºC for 4 hours. The *Dpn1* digested mixture was PCR cleaned up with the QIAGEN kit (as per manufacturers protocol), and then the subsequent DNA was transformed into *E. coli* DH5α cells (See section 2.1.3). Typically, 5 mL cultures from colonies were grown overnight at 37 ºC with shaking at 200 rpm. Plasmids were purified using the miniprep QIAGEN kit (as per manufacturers protocol). All of the alterations were confirmed by DNA sequencing.

All primers used for site-directed mutagenesis are listed in Appendix 1.

2.2 Protein expression and purification

2.2.1 *E. coli*

E. coli cells were cultured in either LB (Fisher Scientific) or Terrific broth (TB) (Fisher Scientific) for large scale growths. LB contains 10 g tryptone, 5 g yeast extract and 10 g sodium chloride per 1 L of media. TB contains 12 g tryptone, 24 g yeast extract, 9.4 g di-potassium phosphate, 2.2 g mono-potassium phosphate and 4 mL glycerol per 1 L of media. Media was sterilised by autoclaving at 121ºC and 15 psi for 20 min. Appropriate antibiotic was added when necessary to sterile media. Cultures were grown at 37 ºC with shaking at 200 rpm. Generally, cultures were grown by inoculating media with an overnight growth to OD_{600} of 0.6-0.8 AU. Cultures were then induced by addition of 0.5 mM IPTG and incubated overnight at 18ºC at 180 rpm. Cultures were then harvested by centrifugation at 6,000*g* for 20 min and stored at -80 ºC.

For solid media, 1.5% (w/v) agar was added to LB prior to sterilisation. The appropriate antibiotic was added to liquid LB-Agar media, which was subsequently mixed before plates were poured.

2.2.2 Insect cell culture

*Sf*9 cells were cultured in Sf-900™ II SFM media (Gibco), supplemented with 2 mM L-glutamine (Gibco) and 1X antibiotic-antimycotic (Gibco). *Trichoplusia ni* (*Tni*) cells were cultured in ESF-921 media (Expression Systems), supplemented with 1X antibiotic-antimycotic (Gibco). Maintenance cultures were grown in suspension in 125 mL flasks at 27 ºC and their growth was checked every 48-72 hours. When confluent, cells were diluted to ~1.0 x 10^6 cells/mL. shaking at 120 rpm.

Recombinant progeny (P1) was created by transfecting recombinant bacmid into *Sf9* cells. 5 mL of cells at a density of 1.0 x 10^6 cells/mL (in Sf-900™ II SFM media without antibiotic-antimycotic) were seeded into T25 flasks and incubated for 30 min at 27 ºC in a static incubator. 30 μL of DNA at 250 ng/μL was added to 300 μL Sf-900™ II SFM media containing no antibiotic-antimycotic and separately 10 μL of transfection reagent (Roche) was added to 100 μL of the same media. The DNA mix was then added to the transfection reagent mix and left to incubate at room temperature for 30 min. After the subsequent incubation times, the 440 μL DNA/transfection reagent mix was added to the attached cells, and incubated for 24 hours in a static incubator at 27 ºC. After 24 hours, the media was removed and replace with 6 mL Sf-900™ II SFM media supplemented with antibiotic-antimycotic. 7 days after the initial transfection, cells were visualised. If DH10MulitBac Turbo-YFP cells were used, YFP levels can be visualised (in this case using a GFP filter). High levels of YFP indicated a successful transfection (Figure 2-1). The P1 virus was then harvested by centrifugation at 1000*g* for 5 min at 4 ºC. The supernatant was stored at 4 ºC protected from light.

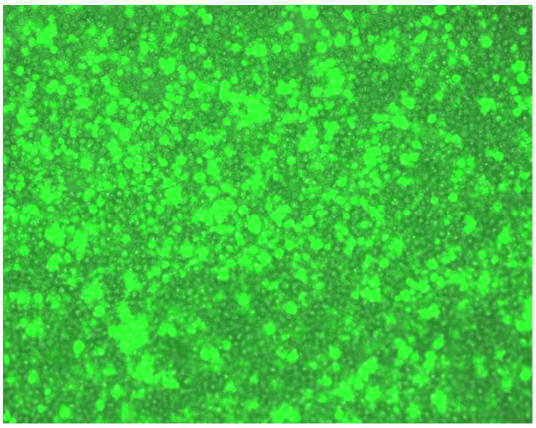

Figure 2-1: *Sf9* cells after transfecting bacmid DNA containing YFP to create P1 virus
YFP levels can be visualised, in this case using a GFP filter, indicating success of transfection.

P2 virus was produced by adding P1 virus (165 μL per 50 mL cells) to *Sf9* cells at a density of 1.5 x 10^6 cells/mL. The supernatant was harvested 72 hours post infection by centrifugation at 1000*g* for 5 min. P2 virus was used to inoculate either *Sf9* or *Tni* cells at a density of 2.0 x 10^6 cells/mL (1:10 ratio). For example, if using the bicistronic single vector pFL, add 60 mL P2 virus to 540 mL cells (Figure 3-1a). For co-infection, a dilution of nine-part GS virus to one-part GN virus was employed, for example 108 mL GS P2 virus and 12 mL GN P2 virus (total of 120 mL virus) to 480 mL cells (Figure 3-1b). Cells were harvested 48 hours post infection by centrifugation at 500*g* for 15 min at 4 ºC with slow acceleration and deacceleration and *Sf9* cell pellets were subsequently washed in 1X PBS. A schematic of the process of insect cells expression is shown in Figure 3-1.

2.2.3 SDS-PAGE analysis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels for electrophoresis were made in-house. Either 10% or 12% gels were made, using the recipe shown in Table 2-1. NuPAGE 3-8% Tris acetate gels (Invitrogen) were also used.

Table 2-1: SDS-PAGE gel recipe.
SDS, sodium dodecyl sulfate; APS, ammonium persulfate and TEMED, Tetramethylethylenediamine.

	Resolving		**Stacking**
	10%	**12%**	
dH_2O (mL)	11.75	9.77	5.4
1.5 M Tris-HCl pH 8.8 (mL)	7.34	7.34	/
1.5 M Tris-HCl pH 6.8 (mL)	/	/	2.25
Acrylamide (30%) (mL)	9.77	11.7	1.35
SDS (10%) (μL)	292.5	292.5	90
APS (10%) (μL)	146.25	146.25	45
TEMED (μL)	14.62	14.62	4.5
Total (mL)	30.76	30.70	9.14

2.2.4 Purification of GS-GN, GS-GN(Y195F) and GS mutants

Cell pellets containing GS-GN were resuspended in lysis buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 0.075% β-mercaptoethanol, 1 mM benzamidine, 0.8 mM phenylmethyl sulfonyl fluoride (PMSF), 0.3 mg/mL lysozyme). Cells were lysed by sonication (1 s on, 3 s off for a total of 5 min) on ice and the lysate was

cleared by centrifugation at 35,000*g* for 30 min at 4 °C. The clarified lysate was sonicated again (1 s on, 3 s off for a total of 1 min), followed by filtering with a 0.45 µM filter (MerckMillipore). Filtered lysate was loaded onto a pre-equilibrated 1 mL or 5 mL HisTrap HP column (GE Healthcare) charged with Ni^{2+}. The loaded column was washed with four column volumes (CV) of low salt buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 0.075% β-mercaptoethanol, 1 mM benzamidine), followed by four CV washes of high salt buffer (50 mM Tris-HCl pH 7.6, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 0.075% β-mercaptoethanol, 1 mM benzamidine) and finally 4 CV washes in low salt buffer. The column was then attached to the AKTA system (GE Healthcare) and washed with low salt buffer. The protein was then eluted by applying an imidazole gradient with elution buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 300 mM imidazole, 10% glycerol, 0.075% β-mercaptoethanol, 1 mM benzamidine). The fractions containing protein were analysed by SDS-PAGE and then pooled and dialysed overnight (10,000 MWCO SnakeSkin dialysis tubing (Thermo Scientific)) at 4 °C in dialysis buffer with TEV protease added (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 20 mM imidazole, 10% glycerol, 0.075% β-mercaptoethanol, 1 mM benzamidine). The dialysed protein was re-loaded onto the HisTrap column equilibrated with low salt buffer, for a Ni subtraction step. TEV cleaved protein was eluted in the flow through and low salt washes. The flow through and first low salt wash were pooled and concentrated using a VIVASPIN20 30,000 MWCO (Sartorius, Generon), followed by centrifugation at 17,000*g* for 15 min at 4 °C. Protein was then injected onto a 16/600 or 10/300 Superdex200 column (GE Healthcare) equilibrated with gel filtration buffer (25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 150 mM NaCl, 1 mM Tris(2-carboxyethyl)phosphine (TCEP), 10% glycerol). Fractions containing protein were analysed by SDS-PAGE, and fractions containing GN were pooled, concentrated and stored at -80 °C. Some fractions containing GS-GN complex were stored separately at -80 °C and the remaining protein was pooled and concentrated before being stored at -80 °C.

2.2.5 Purification of lambda protein phosphatase

Cell pellets containing lambda PP were purified by nickel affinity chromatography as previously described (see section 2.2.4). No TEV was added during dialysis to keep the His-tag attached to the protein, and protein was injected onto a S200 16/600 column (GE Healthcare) equilibrated with gel filtration buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). Fractions containing protein were analysed by SDS-PAGE, and separate peaks were pooled, concentrated and stored at -80 °C.

2.2.6 Periodic-acid Schiff staining

Pierce™ Glycoprotein staining kit (ThermoScientific) was used to detect glucosylated species using the periodic-acid Schiff (PAS) staining method. SDS-PAGE gels were made and run as stated in 2.2.3. Staining was carried out according to the manufacturer's protocol. Specifically, the SDS-PAGE gel was soaked in 50% methanol for 30 min. Followed by two washes in 3% acetic acid for 10 min each with gentle agitation. The gel was then gently agitated in oxidising solution for 15 min, followed by three washes in 3% acetic acid for 5 min each with gentle agitation. The gel was then washed in Glycoprotein stain for 15 min, followed by a wash in reducing solution for 5 min. The gel was then washed extensively with 3% acetic acid and ultrapure water, and subsequently stored in 3% acetic acid. The gel was imaged using a scanner.

2.2.7 Determination of protein concentration

Absorbance at 280 nm was measured using a Nanodrop 8000 Spectrophotometer (ThermoScientific). Concentration was determined on the basis of the Beer-Lambert Law, ($A=\varepsilon cl$), where A is the absorbance of the protein sample, ε is the extinction coefficient ($M^{-1}cm^{-1}$) as predicted by the ExPASy-ProtParam[193] server based on the amino acid sequence, and l is the path length (cm). This gives the protein concentration in M, which can be converted into a mg/mL concentration by multiplying by the molecular weight.

2.3 Electron microscopy

The general pipeline for cryo-electron microscopy (cryo-EM) analysis begins with the sample and then subsequent negative stain EM (nsEM). Followed by cryo-EM grid preparation, data collection and image processing. A general schematic of these steps is shown in Figure 2-2. Each step will be discussed in more detail in the following sections: 2.4, 2.5 and 2.6.

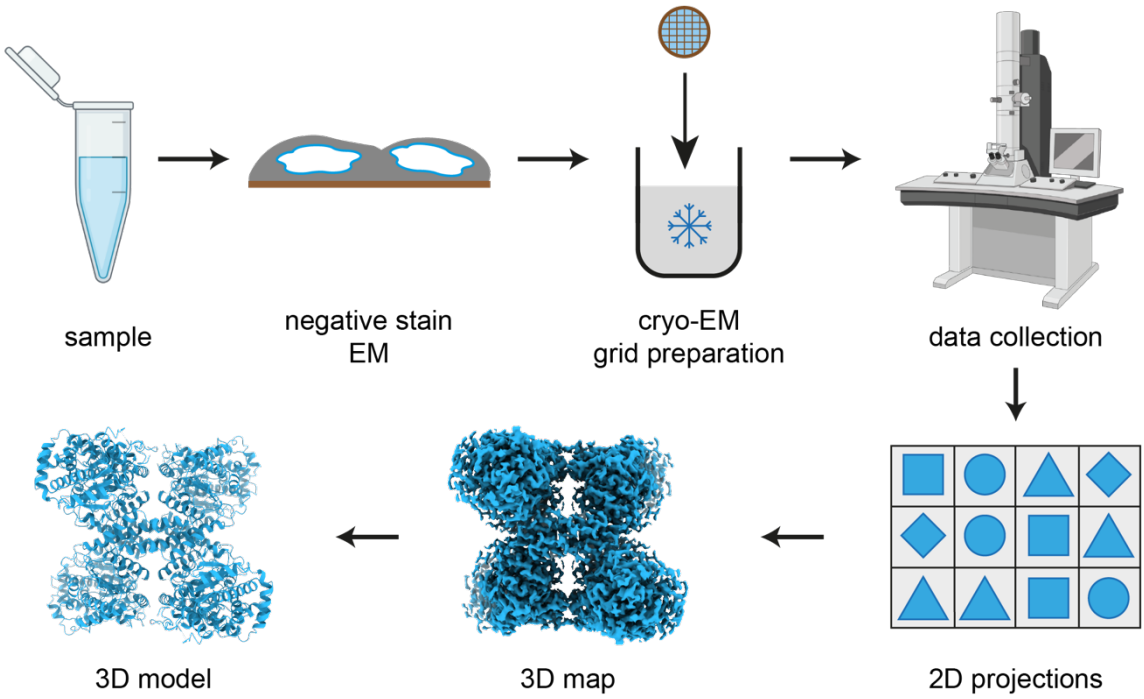

Figure 2-2: General electron microscopy pipeline, from sample to structure

The sample is first analysed by negative stain EM. Followed by preparation and optimisation of cryo-EM grids. Data can then be collected in an electron microscope. Data processing can then be performed to generate 2D projections. These 2D projections are used to generate a 3D map. A 3D model can then be built into the electron density in the 3D map. Adapted from Doerr, 2016 [194].

2.4 Negative stain electron microscopy

When embarking on an electron microscopy project, it is often useful to perform nsEM analyses first. NsEM allows visualisation of your protein of interest under the electron microscope by embedding the protein molecules in a coating of heavy metal solution (Figure 2-3). These heavy metals are electron dense, and thus can absorb electrons more readily than the biological specimen, which provides high contrast in micrographs[195]. Moreover, when visualising negative stain grids in an electron microscope, the protein appears lighter in comparison to the dark background (Figure 2-3). This technique is limited in resolution, obtaining maps in the range of ~20-40 Å[195]. Nonetheless, it is useful for assessing samples for cryo-EM analysis, giving an indication of sample purity and heterogeneity. It can also be useful in gaining insight into biological specimens through low-resolution 3D reconstructions[196].

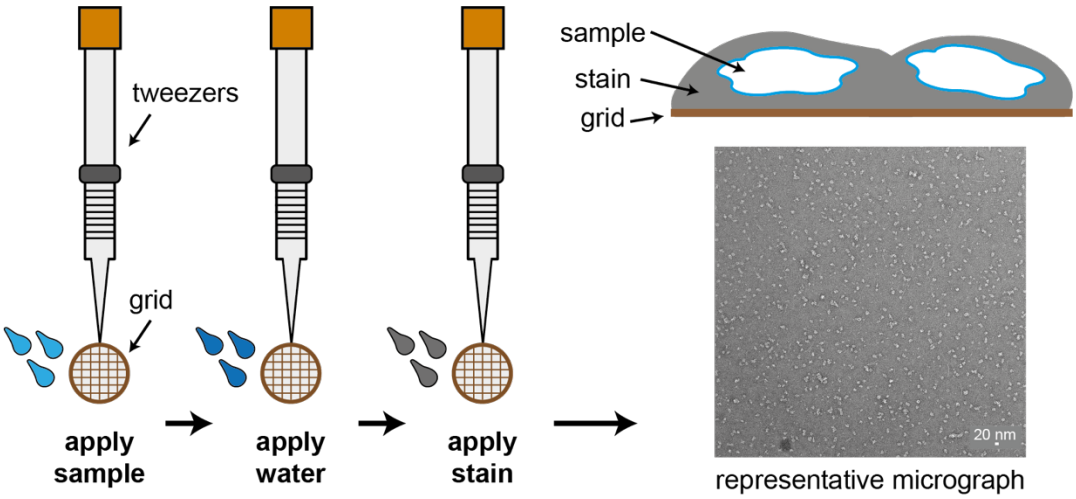

Figure 2-3: Overview of negative stain EM sample application and resulting micrograph. To a glow discharged grid, sample, water and heavy metal stain is applied. This coats the protein molecules in the stain, meaning when the grid is visualised under the electron microscope the protein molecules appear lighter against a dark background.

2.4.1 Sample preparation

In order to visualise your protein of interest in an electron microscope, it needs to be applied to a support structure, which is a grid[197]. Standard grids are 3 mm in size, and contain a metal mesh (copper or gold) with a support substrate, which is most commonly a thin layer of amorphous carbon[196,197]. These properties cause the surface of the grid to be hydrophobic. Before sample application, the grid needs to be glow discharged to render the surface hydrophilic. This glow discharge process removes any remaining hydrocarbons, resulting in a hydrophilic surface[198]. Thus, preventing the sample from being repelled from the surface of the grid. To the glow discharged grid, sample can be applied and incubated for 1 min. The grid can then be washed with H_2O before being stained with the heavy metal solution[195]. A schematic of this process is shown in Figure 2-3.

Specifically, the procedure used in this thesis is that the protein sample was diluted in buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, 10% glycerol) to concentrations between 0.01 and 0.02 mg/mL immediately before grid preparation. Carbon-coated copper grids (Formvar/Carbon 300 mesh Cu, Agar Scientific) were glow-discharged for 30 s with 10 mA and 0.39 mBar pressure (PELCO easiGlow, Ted Pella). Grids were incubated for 1 min with 6 µL sample, washed with H_2O three times and stained twice with 2% (w/v) uranyl acetate for 20 s. Excess liquid was removed by blotting with filter paper.

2.4.2 Data collection

Negative stain data was collected on a FEI Technai F20 electron microscope operated at 120 keV, equipped with a FEI Ceta (CMOS CCD) camera. Data was collected at a nominal magnification of 29,000 x and a pixel size of 3.51 Å.

2.4.3 Image processing

RELION 3.0 was used for processing of negative stain-EM data. Contrast transfer function (CTF) parameters were determined using gCTF. Approximately 2,000 particles were manually picked, extracted with a box size of 104 $Å^2$, then subjected to reference-free 2D classification to produce initial references to be used for auto-picking. The parameters for auto-picking were optimised and particles were extracted. The extracted particles were used for iterative rounds of reference-free 2D classification. Based on visual inspection, the best quality 2D class averages were selected to generate a *de novo* 3D initial model, which was used as a reference in unsupervised 3D classification. These classes were then subjected to 3D refinement to generate a final EM density map.

2.5 Cryo-electron microscopy

2.5.1 Sample preparation

For cryo-EM analysis, the sample must first be prepared for imaging, and grid preparation is a major focus in the cryo-EM field. To preserve the sample in a near-native state, the sample must be suspended in a layer of vitreous ice[199]. Grids are commonly prepared using the basic method developed by Jacques Dubochet in the 1980s, but using devices such as the Vitrobot[200,201]. However, more recently other approaches have been developed, such as spraying. This aims to solve problems often seen with traditional blotting methods, such as protein denaturation and dissociation at the air-water interface. In this thesis, both standard and spraying approaches are used, and these methods are described in sections 2.5.1.1 and 2.5.1.2 respectively.

2.5.1.1 Vitrobot

Quantifoil R2/1 Cu300 or Quantifoil R1.2/1.3 Cu300 (Quantifoil Micro Tools) grids were glow-discharged using the GloQube (Quorum) at 40 mA for 30 s. A FEI Vitrobot IV was equilibrated at 4 °C at 100% relative humidity. GS-GN complex was diluted in buffer (25 mM HEPES pH7.5, 150 mM NaCl and 1 mM TCEP) containing 1.5-8% glycerol immediately before 3 µL was applied to the grid. This was followed by blotting and plunge-freezing into liquid ethane cooled by liquid nitrogen. The blotting force and blotting time can be controlled, which modifies the ice thickness. A schematic of this process is shown in Figure 2-4.

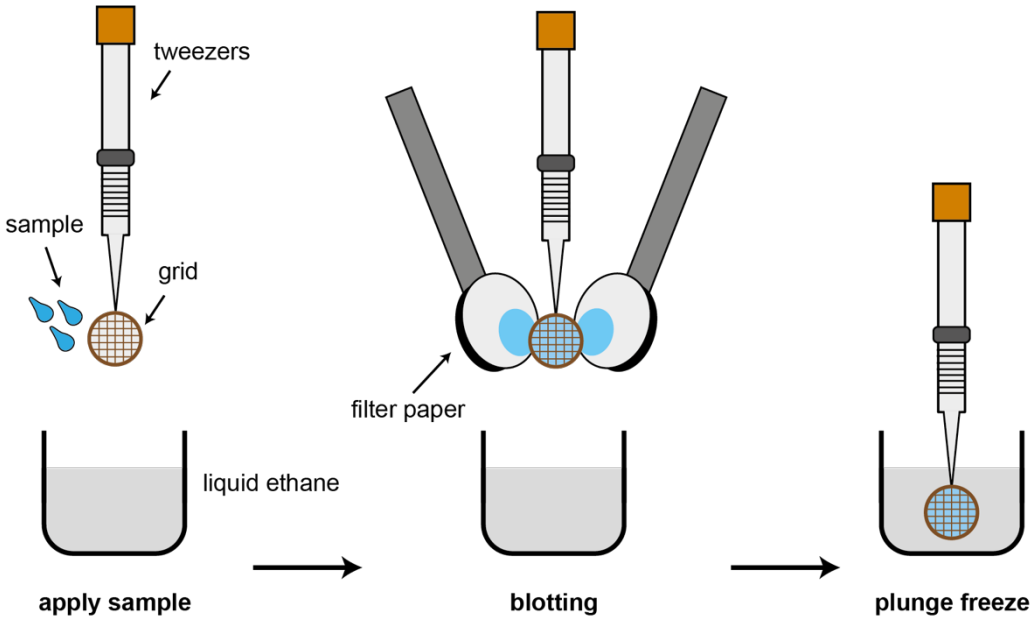

Figure 2-4: Schematic of cryo-EM grid preparation using the vitrobot

A glow-discharged grid is held in a humidity-controlled compartment by a pair of tweezers. Sample is then applied to the grid, followed by blotting using filter paper to create a thin layer of the sample. The sample is then vitrified by plunge freezing into liquid ethane and subsequently stored in liquid nitrogen.

2.5.1.2 Chameleon

The Chameleon (SPT Labtech) was used to produce grids in an alternative approach to the standard method. The Chameleon uses a dispensing tip to apply a stream of droplets onto a 'self-wicking' nanowire grid. This system aims to decrease protein denaturation at the air-water interface and improve overall grid quality and reproducibility through automated grid preparation[202]. See Chapter 4.5.1 for further details.

The nanowire grids were glow discharged with 12 mA for a total of 80 s. 3 µL of sample at 0.68 mg/mL in buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 10% glycerol was aspirated onto the grid. Wicking speed was 260 ms before the grid was plunged into liquid ethane.

2.5.2 Electron microscope

Following cryo-EM grid preparation, the grids are inserted into the electron microscope for visualisation of the protein of interest. For nsEM analyses a FEI Tecnai F20 was used for imaging, and for cryo-EM analyses a Titan Krios was used. The microscope contains an electron source, which emits electrons which are subsequently directed through a series of electromagnetic lenses and apertures. This allows for magnification of the sample

before imaging at the detector. The electron path is kept under high vacuum to prevent unwanted interactions with air and the electrons. A basic schematic of an electron microscope is shown in Figure 2-5. More recently, the detector in question is often a direct electron detector, and the development of such detectors was one of the major reasons behind the 'resolution revolution'. This is due to the ability to record individual electron events which improves the signal-to-noise ratio. In addition, the high frame rate of these detectors allows for the collection of images as a series of frames or a 'movie'. In this thesis, various detectors were used to collect cryo-EM data. This includes FEI Falcon III and 4 direct electron detectors and K2 (Gatan) direct electron detector. Electron dose and length of collection various across datasets. The data collection parameters used in this study are shown in Table 2-2.

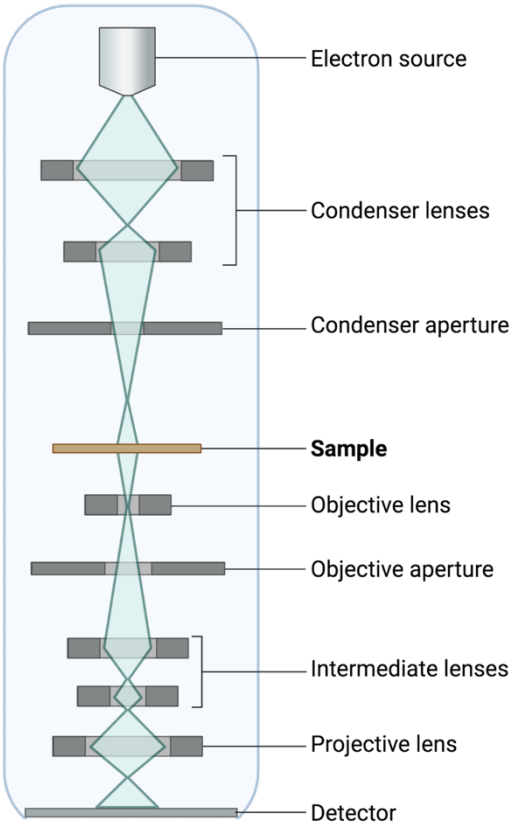

Figure 2-5: Schematic representation of an electron microscope

The electron beam (green) is directed through a series of lenses and apertures to create a magnified image of the sample. Adapted from Orlova & Saibil, 2011[203].

Table 2-2: Cryo-EM data collection parameters

	GS-GN(Y195F)	**GS-GN**	**GS-GN(Y195F)**	**GS-GN(Y195F) crosslinked**	**GS-GN(Y195F)**	**GS-GN(Y195F) + 25-fold G6P**	**GS-GN(Y195F) + 100-fold G6P**	**GS-GN(Y195F)**	**GS(R588A+R591A) - GN(Y195F)**
Chapter	4.2	4.4	4.5	4.5	4.6	4.7	4.7	5.2	5.4
Grid preparation	Vitrobot	Vitrobot	Chameleon	Vitrobot	Vitrobot	Vitrobot	Vitrobot	Vitrobot	Vitrobot
Detector	Falcon III	Falcon III	K2	Falcon 4	K2	K2	Falcon III	Falcon 4	Falcon III
Energy Filter (eV)	/	/	/	10	/	/	/	10	/
Voltage (kV)	300	300	300	300	300	300	300	300	300
Magnification (x)	75,000	75,000	130,000	165,000	165,000	130,000	75,000	165,000	75,000
Pixel size (Å)	1.065	1.065	1.07	0.71	0.85	1.07	1.065	0.71	1.065
Dose per physical pixel per second	53.9	60	7.3	4.7	5	7.5	49.2	4.7	56
Total dose ($e/Å^2$)	85	85	63.7	35	72	66	87	34.8	35.9
Initial particle images (no.)	780,290	250,250	82,357	186,533	157,656	57,708	231,003	1,883,188	158,126
Final particle images (no.)	160,687	36,972	11,351	/	61,998	21,497	36,570	739,232	/
Resolution (Å)	4.0	6.0	10.2	/	4.2	4.1	7.1	2.6	/

2.5.3 Image processing overview

Following the recording of images by the electron microscope, the dataset must be processed in order to obtain a structure. In the work described in this thesis, this was performed by a process termed single particle analysis. The principle behind this process is that the protein is frozen in its native state in multiple orientations. Movies collected must be motion corrected and the Contrast Transfer Function (CTF) estimated. Particles in the micrographs are then picked and a particle stack is generated. These particles are classified, based on their several orientations, to create 2D class averages. These 2D projections are combined to create several 3D reconstructions of the protein, in a process named 3D classification. These reconstructions are refined and sharpened, as well as the use of other tools to increase the resolution such as Bayesian polishing. The final high-resolution map can then be used to build an atomic model of the protein. An overview of this process is shown in Figure 2-6. Each of these steps is discussed in more detail in the following sections. The first steps of this process were carried out 'on the fly', which allows data processing to begin in parallel with data collection[204]. This allowed the monitoring of the quality of data being collected, as well as improving the speed at which data can be processed.

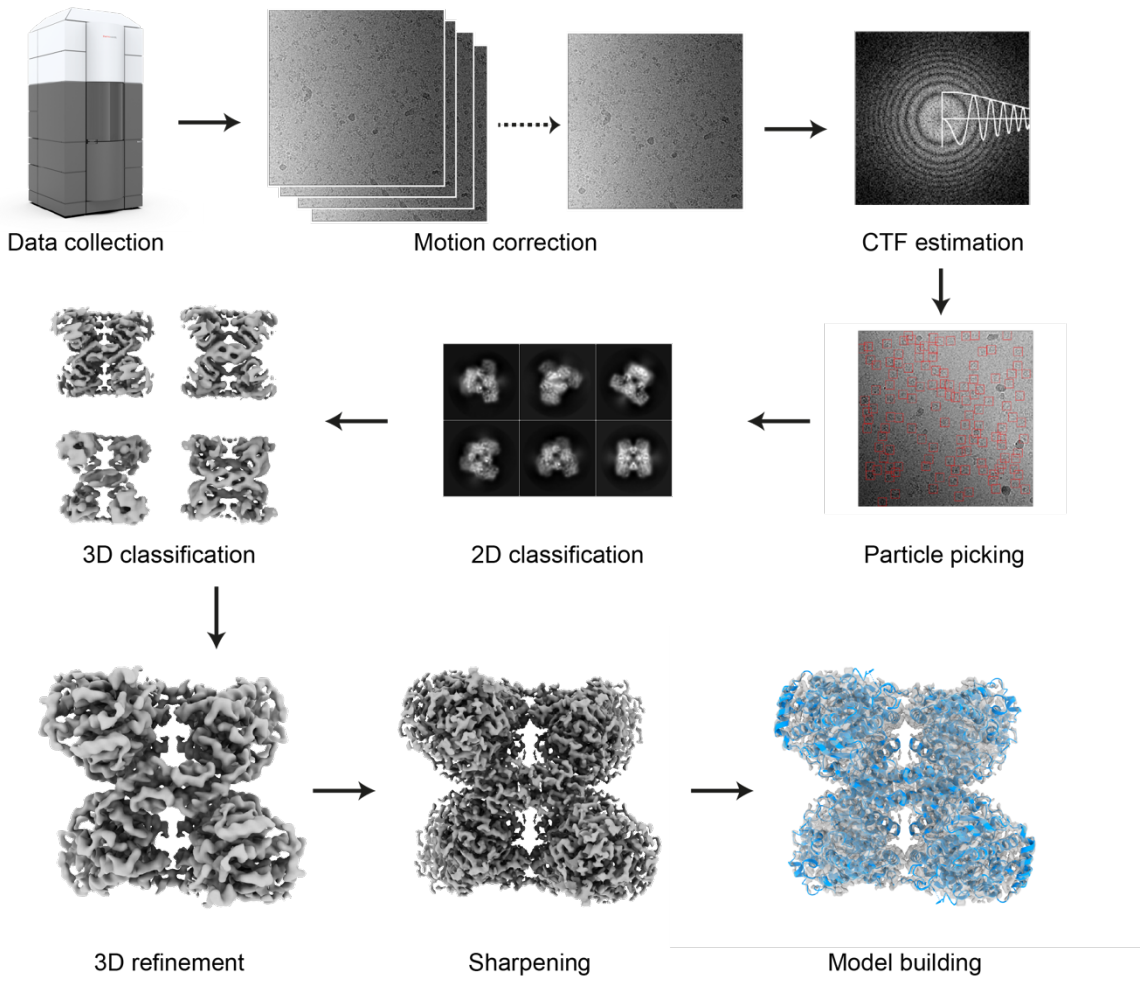

Figure 2-6: Overview of the cryo-EM image processing pipeline

Data is collected in the form of movies which are motion corrected to micrographs. CTF estimation is performed and then particles can be picked. Particles are classified based on their orientation, and a range of good quality classes are taken forward for 3D classification. 3D refinement and sharpening, along with other tools such as particle polishing, can be performed to generate a final map that allows model building.

2.5.3.1 Motion correction

As mentioned previously, electron detectors collect a series of frames, also called movies. The first step of data processing is to align these individual movie frames and correct for the beam-induced movement in the sample by a process termed motion correction, which can be completed by software such as MotionCor2[204,205]. The alignment of the frames creates an image, or a micrograph, that is the overall sum of all the frames[206]. This corrects for large motions across the image, but does not consider the movement of individual particles. For datasets collected on the Falcon 4 detector, RELION's own implementation

of motion correction was used, due to the different file format of electron-event representation (EER) in comparison to MRC files[207].

2.5.3.2 Contrast Transfer Function estimation

The vitrified biological specimen imaged produces minimal amplitude contrast and weak phase shift, thus resulting in little phase contrast[203]. Applying defocus during data collection can increase the contrast in the images, however, this affects the CTF of the images. In order to generate a reliable 3D map, the original information of the images must be corrected using accurate CTF parameters. Moreover, the CTF must be estimated, and the more accurate this is the higher the potential resolution of the reconstruction. CTF estimation is performed after motion correction, and there are several programs available to perform CTF estimation, including Gctf[208] and CTFFIND[209].

The CTF is a frequency-dependent oscillating function which measures the amplitude of an electron wave at a certain resolution (spatial frequency). The power spectrum of an image shows a pattern named Thon rings, which can be explained as the effect of the CTF. The CTF is represented as a sine wave which can be plotted through the Thon rings, which oscillates between positive and negative maxima, crossing zero (which indicates frequencies that contain zero information) (Figure 2-7a, b and c). At higher defocus, the oscillation of the CTF is more severe which results in more information being lost due to the wave passing through zero more often (Figure 2-7b and d). To mitigate the loss of information, data is collected at a range of defocus values, meaning the zeros of each particular CTF will be filled by data from a different defocus (Figure 2-7e). Phases of the sine waves can be flipped to ensure all amplitudes are positive, meaning a sum of the waves can be calculated (Figure 2-7f).

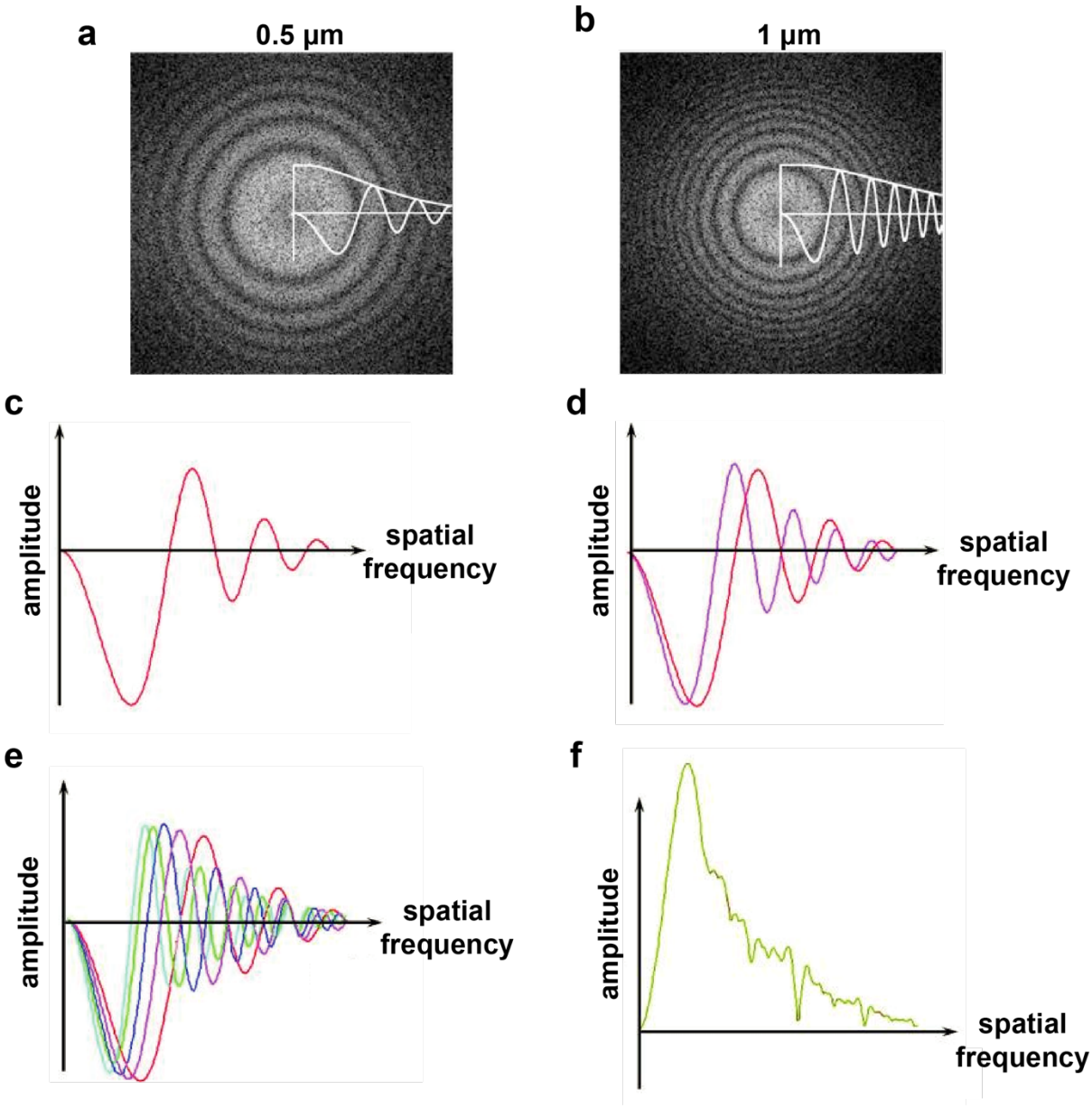

Figure 2-7: CTF estimation

a Thon rings can be observed on the 2D power spectrum at 0.5 µm, and the sine wave can be fitted through this pattern. **b** At a higher defocus of 1 µm, oscillation of the CTF is more severe. **c** A CTF curve showing spatial frequency against amplitude, an oscillating curve can be observed. **d** Two CTF curves shown at two different defocus values, where the pink curve is at a higher defocus. **e** CTF curves at a range of defocus values, as would occur in a data collection to ensure no resolution have zero information when the data is averaged. **f** The sum of all the curves shown in **e** after phases have been flipped. Adapted from Orlova & Saibil, 2011[203].

2.5.3.3 Particle picking

The position of each particle within each micrograph has to be defined. The coordinates for each particle can be defined manually, however this is a slow and arduous process,

therefore automatic particle picking tools are available, including RELION auto-picking[210], cryoSPARC template picking[211] and crYOLO[212].

crYOLO is an automated particle picking software, that can be trained for a specific dataset, or a general model can be used[212]. This allows for fast and efficient particle picking. In this thesis, the general model was used as it effectively and successful picked particles on all micrographs. The template picker in cryoSPARC can auto-pick particles using a template[211]. In this thesis, this tool was utilised to pick particles that resembled the full GS-GN complex after crosslinking (see section 4.5.2.1). 2D projections of the final nsEM map of full complex were created, which were used as templates to perform automated picking. This successfully picked particles on all micrographs. After particle picking, the particles are extracted with a specified box size, and subsequently used for 2D classification.

2.5.3.4 2D classification

2D classification aligns and averages particles with similar orientations, improving the low signal-to-noise ratio that individual particles have. In addition, 2D classification can separate the protein of interest from unwanted 'junk particles', including ice contamination and protein aggregates that may be selected during automated picking. The approach to 2D classification varies between programs. The main program used in this thesis, RELION, uses a reference-free maximum likelihood approach[213]. Initially, the references used are random and all particles are aligned to these references. Each particle is assigned a probability weighting, based on the likelihood they belong to each class based on the degree of matching to a particular reference. This allows separation of 'junk particles', as these unwanted particles often align poorly with several classes, resulting in 'fuzzy' 2D class averages.

2.5.3.5 3D classification

The maximum likelihood classification described above can also be performed in three dimensions, also known as 3D classification in RELION. The user provides an initial model of the protein of interest and specifies the number of classes to be generated. This process is useful in separating different conformational states of the protein within the dataset. Similar to 2D classification, the probabilities of the particle contributing to the different 3D classes is calculated. The classification ends when the number of particles which change class stabilises, resulting in a number of classes with varying numbers of particles[214]. 3D

classification can also be performed on a particular area or domain by focused 3D classification using a mask encompassing the area of interest.

2.5.3.6 3D refinement

Following generation of a homogenous subset of particles, a final 3D reconstruction step is performed. The “projection-slice theorem” can explain how 2D projections can be combined into a 3D reconstruction. This theorem states that the Fourier transform of a 2D projection corresponds to a central slice of an object in 3D space[215] (Figure 2-8). Thus, the multiple orientations of the 2D projections can be used to generate a 3D structure by performing a reverse Fourier transform (Figure 2-8). A typical approach to this is the “projection-matching” approach, where the individual particle images are compared to an initial model that is used to generate reprojections at multiple orientations. The angular assignment for the back projections can then be used to calculate the 3D map, by placing the 2D Fourier transforms into the 3D transform and subsequently calculating the inverse transform[203].

In RELION, data are split into two independent subsets, and each subset is refined separately. The Fourier Shell Correlation (FSC) measures the correlation between the two half maps in Fourier space, and gives information on the quality of the 3D map. Therefore, the resolution of a map can be determined by how well the two half maps correlate to one another at a given resolution, as displayed by the FSC curve. FSC values of 0.143, also known as the FSC gold-standard[216,217], as well as 0.5[218] are commonly used as cut-off levels. Similar to 3D classification, focused 3D refinement can be performed using a mask containing an area of interest.

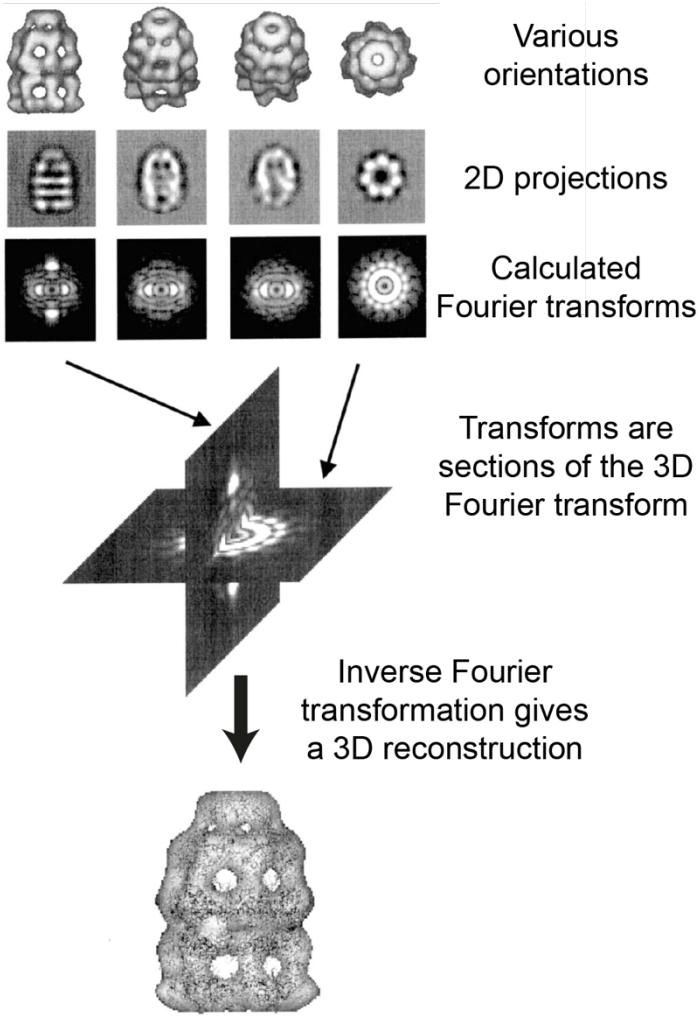

Figure 2-8: The projection-slice theorem

An overview of the principle of 3D reconstruction from 2D projections. 2D projections of various orientations are shown. The calculated Fourier transforms of these projections correspond to central slices through the 3D Fourier transform. A 3D reconstruction can be generated by performing an inverse Fourier transformation. Figure adapted from Saibil, 2000[219].

2.5.3.7 PostProcessing

The map obtained from 3D refinement often does not resemble the reported resolution. This is due to the loss of high-resolution features caused by radiation damage or imperfections in the imaging system[220]. These high-resolution features can be recovered by B-factor sharpening, in which the high-resolution signal is enhanced[216]. This was performed in a PostProcessing step in RELION. A mask is created which compasses the electron density of the 3D refinement map based on visual inspection, and is subsequently used in PostProcessing.

2.5.3.8 CTF Refinement and Bayesian polishing

Other methods to improve the quality of the 3D reconstruction is CTF refinement and Bayesian polishing. CTF refinement can correct anisotropic magnification, higher-order aberrations and per-particle defocus. Similar to motion correction Bayesian polishing corrects for movement, but for individual particles[210]. This can estimate and correct the motion tracks for each particle and subsequently create adequately weighted averages of the movie frames, which can improve the signal to noise ratio and thus improve the resolution of the 3D map.

2.5.3.9 cryoSPARC 3D variability analysis

To analyse and visualise any flexibility in the cryo-EM maps, the cryoSPARC[211] tool cryoSPARC 3D variability analysis[221] was used (see section 4.3.2). This tool can identify discrete or continuous flexibility within datasets which enables visualisation of flexibility[221]. This analysis aims to compute the eigenvectors of the 3D covariance of a set of particles. Eigenvectors of the 3D covariance are trajectories in 3D space in which there is significant variability within the dataset[221]. Moving along these trajectories gives all the different conformations present in a particular eigenvector. Solving multiple eigenvectors revealed 3D reconstructions that define a family of conformations in the dataset.

2.6 Model building

An atomic model was built after a high-quality final 3D map was obtained. There are various starting points for model building depending on the resolution of 3D reconstruction. At very high resolution, a model can be built *de novo*. With lower resolution maps, existing structures can be rigid body fitted to allow their approximate location to do determined. Typical pathways involve iterative rounds of manual modelling building in Coot[222] and automated refinement in programs such as Phenix[223] or REFMAC[224]. In this thesis, Coot v0.9.2-pre[222] and real space refinement in Phenix v1.19[223] were used to build the human GS structure (Table 2-3).

A preliminary human GS model was generated by AlphaFold[225] (accessed October 2021). This preliminary model was rigid body fitted into the cryo-EM density in UCSF Chimera[226] and subsequently used for model building into the human GS electron density.

A model of the last 34 residues of human GN was created by Phyre2[227]. This was placed in the approximate location using the *C. elegans* GS-GN[34] crystal structure (PDB ID 4QLB) in Chimera[226] and then fitted into the map by rigid body fitting. This model was

subsequently used for manual model building and real space refinement into the electron density.

Table 2-3: Model refinement statistics

	GS-GN(Y195F) PDB 7ZBN EMDB-14587
Refinement	
Model resolution (Å)	2.7
FSC threshold	0.5
Model composition	
Non-hydrogen atoms	41588
Protein residues	2608
Ligands	0
B factors (Å^2)	
Protein (min/max/mean)	5.81/46.06/19.42
Map-model CC	
CC (mask)	0.81
CC (volume)	0.76
CC (peaks)	0.69
CC (box)	0.71
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angle (°)	0.551
Validation	
Molprobity score	1.71
Clashscore	10.45
Rotamers outliers (%)	0
Ramachandran plot	
Favored (%)	97.01
Allowed (%)	2.99
Outliers (%)	0

2.7 Structural visualisation and figure construction

UCSF Chimera[226] v1.12 and ChimeraX[228] v1.3 were used for structural visualisation and figure construction. Sequence alignments were performed used MUSCLE[229] and edited and displayed using ALINE v1.0.025[230].

2.8 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF), also known as thermal stability assays, measures protein unfolding by determining changes in fluorescence at increasing temperatures. Using a fluorescent dye that binds to hydrophobic residues, a folded protein shows no fluorescence. Upon increasing the temperature, the protein begins to unfold which exposes hydrophobic residues, thus, an increase in fluorescence is observed. The melting temperature (T_m) of the protein can be calculated using the derivative or Boltzmann methods.

DSF assays were performed using an Applied Biosystems QuantStudio 3 Real-Time PCR system. SYPRO™ Orange (Invitrogen) was used as a fluorescence probe. Proteins were diluted in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP to a final concentration of 1 µM. Varied concentrations of G6P were added and the reaction was incubated at room temperature for 30 min. SYPRO™ Orange was diluted in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP to a final concentration of 2.5 X in a total reaction volume of 20 µL. The temperature was raised in 0.018 °C intervals from 20 °C to 95 °C. Data were analysed using Protein Thermal Shift v1.4.

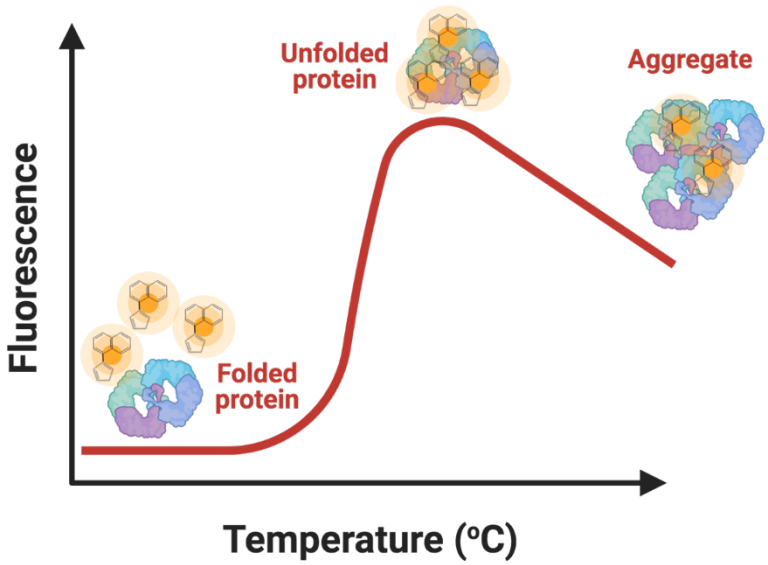

Figure 2-9: Schematic representation of differential scanning fluorimetry analysis

At low temperature (x-axis), the protein is folded and so the fluorescent dye (SYPRO™ orange) cannot access hydrophobic residues within the interior of a folded protein and thus the fluorescence

reading (y-axis) is low. Upon increasing the temperature, the protein begins to unfold, exposing hydrophobic residues which SYPRO™ orange can bind, leading to a gradual increase in fluorescence. At high temperatures, once all the protein has unfolded aggregates form, which causes a reduction in fluorescence.

2.9 Tandem MS/MS, phosphorylation mapping (completed by Dr Leonard Daly)

Tandem mass spectrometry/mass spectrometry (MS/MS) was used to identify phosphorylated residues in the GS-GN complex. A schematic overview of this process is shown in Figure 2-10. Concentrated purified protein complexes (6.75 µg) were diluted 30-fold in 25 mM ammonium bicarbonate pH 8.0 before being subject to reduction with dithiothreitol and alkylation with iodoacetamide, as previously described[231]. The eluent was equally divided into three for digestion with either: 33:1 (w/w) trypsin gold (Promega), 25:1 (w/w) chymotrypsin (Promega), or 10:1 (w/w) elastase (Promega), using the manufacturer's recommended temperatures for 18 hours with 600 rpm shaking. Digests were then subject to in-house packed, strong cation exchange stage tip clean-up, as previously described by[232]. Dried peptides were solubilised in 20 µl of 3% (v/v) acetonitrile and 0.1% (v/v) TFA in water, sonicated for 10 min, and centrifuged at 13,000*g* for 15 min at 4 °C being separated using an Ultimate 3000 nano system (Dionex) by reversed-phase HPLC, over a 60-minute gradient, as described in[231]. All data acquisition was performed using a Thermo Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific), with higher-energy C-trap dissociation fragmentation set at 32% normalised collision energy for 2+ to 5+ charge states. MS1 spectra were acquired in the Orbitrap (60K resolution at 200 *m/z*) over a range of 350 to 1400 *m/z*, AGC target = standard, maximum injection time = auto, with an intensity threshold for fragmentation of $2e^4$. MS2 spectra were acquired in the Orbitrap (30K resolution at 200 *m/z*), AGC target = standard, maximum injection time = dynamic. A dynamic exclusion window of 20 s was applied at a 10 ppm mass tolerance. Data was analysed by Proteome Discoverer 1.4 using the UniProt Human reviewed database (updated April 2020) with fixed modification = carbamidomethylation (C), variable modifications = oxidation (M) and phospho (S/T/Y), instrument type = electrospray ionization–Fourier-transform ion cyclotron resonance (ESI-FTICR), MS1 mass tolerance = 10 ppm, MS2 mass tolerance = 0.01 Da, and the *ptm*RS node on; set to a score > 99.0.

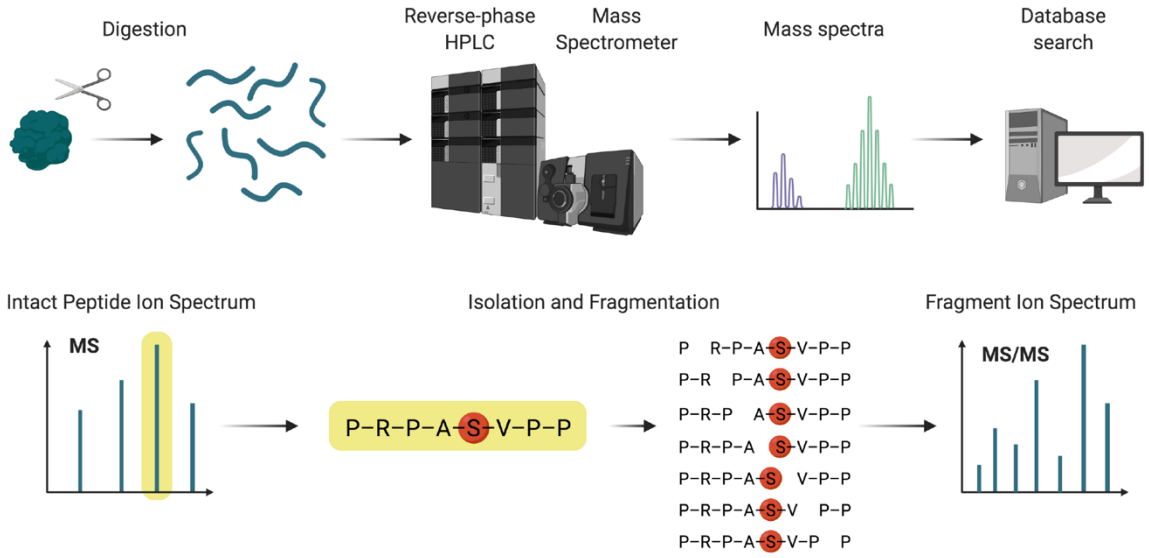

Figure 2-10: Tandem mass spectrometry analysis of phosphorylation

The sample is digested by proteolytic enzyme, and the subsequent peptides are separated by reverse-phase High Performance Liquid Chromatography (HPLC). The peptides enter a mass spectrometer as they elute from HPLC. Mass spectra is generated and peptide matching is completed using databases. From the intact peptide ion spectrum generated peptides can be isolated and fragmented. The fragment ions are detected in a MS/MS scan, in which the position of the phosphorylated residue (circled in red) can be identified. Adapted from Dephoure *et al*, 2013 [233].

2.10 Protein identification Mass Spectrometry (completed by Dr Leonard Daly)

10 µg of purified protein was separated by SDS-PAGE (10% resolving, 4% stacking) before colloidal Coomassie staining overnight and thorough washing in milliQ water[234]. A scalpel was used to excise the major band at ~85 kDa, and incremental bands spanning 43-55, 55-72, 95-130 and 130+ kDa. Bands were washed in 500 µL HPLC H_2O for 10 min shaking at 1500 rpm, room temperature. Bands were then washed in 500 µL of 100 mM ammonium bicarbonate with water bath sonication, as before, for 10 min, before an equal volume of HPLC acetonitrile was added and sonication repeated. Previous two wash steps were repeated until the gel pieces were clear. 100 µL of reduction solution (4 mM dithiothreitol in 50 mM ammonium bicarbonate) was added to each gel slice and heated to 60 °C for 10 min with 600 rpm shaking. A final concentration of 16.4 mM iodoacetamide was added and incubated in darkness at room temperature for 30 min, before quenching by addition of 100 mM dithiothreitol to make a final concentration of 7 mM. All liquid was removed before dehydrating the gel slice by addition of 100 µL HPLC acetonitrile and

shaking 1500 rpm at room temperature for 15 min. Dehydration was repeated until gel slices were opaque white and left open lid to dry completely (~15 min). 0.5 µg of trypsin in 40 mM ammonium bicarbonate was added to the dehydrated gel slices and incubated room temperature for 15 min. Residual liquid was removed and 100 µL of incubation solution (40 mM ammonium bicarbonate, 5% acetonitrile) added and incubated overnight at 37 ^{o}C with 600 rpm shaking. An equal volume of acetonitrile was added and left to shake for an additional 30 min. Gel slices were briefly centrifuged and supernatant collected. Supernatant was dried to completion, resuspended and analysed by LC-MS/MS as described before (see section 2.9).

2.11 Mass photometry

Mass photometry is an in-solution mass measurement method. The light scattered by a molecule in contact with the interface (a cover slide) interferes with the light reflected at that interface. This interference contrast can be recorded, and the contrast scales linearly with mass, i.e. a darker contrast is indicative of a larger molecular weight. Thus, with the use of molecular weight standards, the molecular weight of your protein of interest can be estimated. A schematic of the principle of mass photometry is shown in Figure 2-11.

Mass photometry experiments were performed using a Refyn One^{MP} mass photometer. Immediately prior to mass photometry measurements, proteins were diluted in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP for a final concentration of 50 nM. For each measurement, (16 µL) buffer was added to a well and the focus point was found and adjusted when necessary. Protein (4 µL) was then added to the buffer droplet, the sample was mixed and movies of 60 s were recorded using AcquireMP. Data were analysed using DiscoverMP.

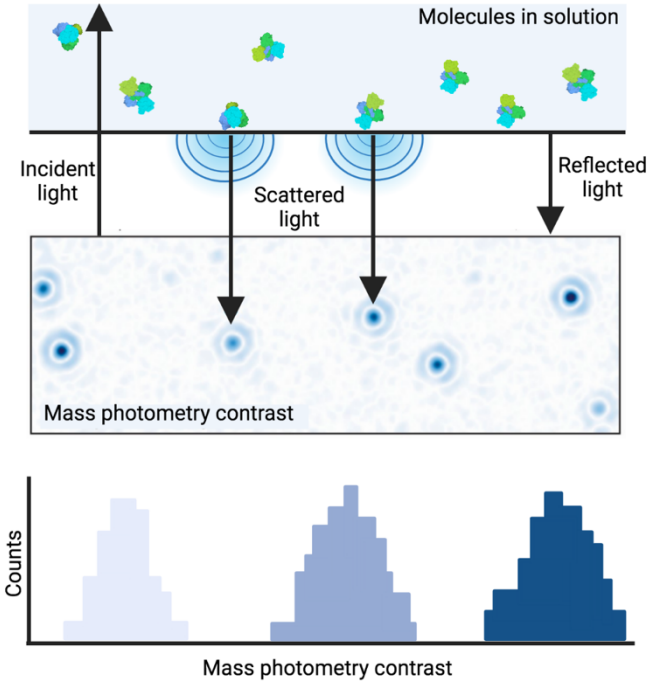

Figure 2-11: Schematic representation of the principle of mass photometry
Molecules in solution scatter light at the measurement interface by interfering with the light reflected at the interface. This produces mass photometry contrast, which scales linearly with mass. Adapted from Young, G. *et al.*, 2018 [235].

2.12 *In vitro* dephosphorylation

GS-GN complex was dephosphorylated in reactions containing equal amounts of PP1 and lambda PP in 25 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, 1 mM $MnCl_2$ and 10% glycerol for 30 min at 30 °C. For subsequent DSF experiments, the phosphatases were removed by incubating the reactions with GST beads for 1 hour at 4 °C. Reactions were then passed through an equilibrated 0.45 µm Spin-X column (Costar, 0.45 µm cellulose acetate) and eluted by centrifugation at 16,000*g* for 2 min.

2.13 GS activity assay (completed by Dr Dipsikha Biswas)

1 µg of purified protein was diluted in ice cold lysis buffer (270 mM sucrose, 50 mM Tris HCl (pH 7.5), 1 mM EDTA, 1 mM ethylene glycol tetraacetic acid (EGTA), 1% (v/v) Triton X-100, 20 mM glycerol-2-phosphate, 50 mM NaF, 5 mM $Na_4P_2O_7$, 1 mM DTT, 0.1 mM PMSF, 1 mM benzamidine, 1 mg/mL microcystin-LR, 2 mg/mL leupeptin, and 2 mg/mL pepstatin A) to a total volume of 100 µL. 20 µL of the protein was added to 80 µL of the

assay buffer (25 mM Tris-HCl (pH 7.8), 50 mM NaF, 5 mM EDTA, 10 mg/mL glycogen, 1.5 mM UDP-glucose, 0.125% (v/v) 2-mercaptoethanol and 0.15 mCi/mmol D-[^{14}C]-UDP-glucose (American Radiolabelled Chemicals, Inc., ARC 0154) with 0 or 12.5 mM G6P. Reactions were incubated for 20 min at 30 °C with mild agitation at 300 rpm. The reactions were stopped by spotting 90 µL of the reaction mix onto 2.5 cm x 2.5 cm squares of filter paper (Whatman 3MM) which were immediately immersed in ice cold 66% ethanol and left to incubate with mild agitation for 20 min. The filter papers were washed thrice more in 66% ethanol for 20 min each and rinsed in acetone. The dried filters were subjected to scintillation counting.

For the GS activity assays, data are reported as mean ± standard error of the mean (SEM) and statistical analysis was performed using GraphPad Prism software v9.3.1. As indicated in the respective figure legends, one-way or two-way analysis of variance was performed with Tukey's post hoc test. Statistical significance was set at $p<0.05$.

2.14 Immunoblotting (completed by Dr Dipsikha Biswas)

Total GN antibody (S197C, third bleed) was obtained from MRC-PPU Reagents and Services. Total GS (#3893) and pGS S641 (#47043) antibodies were from Cell Signaling Technologies. pGS S641/S645 (#07-817) is from MerckMillipore. Affinity-purified pGS S8 antibody (YZ5716) was custom-generated by (YenZym Antibodies Brisbane, CA, USA) by immunisation with a combination of phosphorylated peptides of the mouse GS1 (residues 2–14: PLSRSL-*S-VSSLPG-Ahx-C-amide, in which the prefix * denotes the phosphorylated residue) and human GS1 (residues 2–14: PLNRTL-*S-MSSLPG-Ahx-C-amide). Ahx and Cysteine (C) were added at the C terminal of the antigen peptides as linker/spacer and for conjugation to carrier protein, respectively. Secondary antibodies (#711-035-152 and #713-035-147) were obtained from BioRad. Glucose-6-phosphate (G6P) (#10127647001) is from Roche. All other chemicals if not noted otherwise are from Sigma Aldrich.

Purified proteins were denatured in Laemmli buffer at 100 °C for 5 min. 100 ng of the protein was separated by SDS-PAGE on 4-10% gel and transferred onto nitrocellulose membranes (#926-31090, LiCOR). Membranes were blocked for 45 min at room temperature in 5% skim milk (Sigma, 1153630500) followed by incubation in TBST (10 mM Tris (pH 7.6), 137 mM NaCl, and 0.1% (v/v) Tween-20) containing 5% (w/v) BSA and the primary antibody overnight at 4 ºC. The membranes were incubated for 45 min in HRP conjugated secondary antibodies diluted 1:10,000 in 3% skim milk, after extensive washing

in TBST. The membranes were imaged using enhanced chemiluminescence (ECL) reagent (GE Healthcare). For total protein staining of blots, Revert™ 700 Total Protein Stain (LiCOR) was used.

2.15 Glutaraldehyde Crosslinking

The GS-GN(Y195F) complex was crosslinked by adding 1 µL protein to 8 µL buffer containing 25 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM TCEP, 10% glycerol and various concentrations of glutaraldehyde. This reaction was incubated either at 25 ºC or on ice for varying times, before quenching with 500 mM Tris-HCl pH 7.5. The reaction was buffer exchanged into 25 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM TCEP using an equilibrated 0.5 mL Zeba desalting spin column (GE Heathcare). Samples were visualised on NuPAGE 3-8% Tris-Acetate gels (Invitrogen) and Hi Mark protein standard (Invitrogen) was used.

2.16 Microscale Thermophoresis

The Monolith NT.115 instrument was used to analyse G6P binding to the GS-GN complex. GS-GN containing a His-tag on GN was labelled with Monolith His-tag Labeling Kit RED-tris-NTA 2nd generation (NanoTemper Technologies). Proteins were assayed at 50 nM, with the addition of various concentrations G6P, reaching up to 500 mM, in buffer containing 25 mM HEPES pH 7.5, 150 mM NaCl and 0.01% Tween-20. GraphPad Prism v7 was used for data analysis to calculate a K_d value.

2.17 *In vitro* deglycosylation

GS-GN was incubated with α-amylase from human saliva (Sigma) to deglycosylate the complex. Reactions contained 4 µM GS-GN with either 500 mU or 1 U α-amylase in buffer containing 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 5 mM $CaCl_2$. Reactions were incubated for 30 min, 1 hour or 2 hours at 37 ºC and were terminated by the addition of SDS-PAGE loading dye.

3 Chapter 3: Protein purification and structural analysis of the full-length GS-GN complex

3.1 Introduction

The molecular details of how GS interacts with GN through the C-terminus of GN were described in 2014 using the *C. elegans* enzymes as a model[69,70]. However, at the start of this project (October 2018), it was not known how the full-length human proteins interact and thus the mechanism in which they coordinate to produce glycogen remained elusive. In order to biochemically and structurally investigate the human GS-GN complex, sufficient amounts of active protein need to be produced. However, unlike its yeast and *C. elegans* counterparts, human GS cannot be expressed in *E. coli*. In addition, GS appears to require stabilising interactions with GN for large scale recombinant protein production, and it is possible to produce human GS by co-expression with full length GN or the minimal GS-interacting region (GN^{34})[70,98,99,236]. In this chapter, the expression and purification of the full-length GS-GN complex and the non-glucosylated GS-GN(Y195F) complex is described. Along with bioanalytical and structural analyses to determine the stoichiometry of the complex, which gained insight into how these two proteins interact and cooperate to produce glycogen.

3.2 Expression and purification of the GS-GN complex

3.2.1 Expression of the GS-GN complex

To successfully express both GS and GN, a bicistronic vector or co-expression is required, as GS alone does not have a high level of expression. In addition, the human complex cannot be expressed in *E. coli*, and thus an insect cell expression system was used. In this thesis, both pFL and pFastBac vectors were used (Figure 3-1). Although the pFL vector produces a stoichiometric complex, as GN is more easily expressed than GS, the size-exclusion chromatography step of the purification shows excess GN as well as the GS-GN complex (Figure 3-2). Moreover, a co-expression method using GS and GN in separate pFastBac vectors was optimised, which allows for balancing differences in expression (Figure 3-1). Utilisation of a 9-part GN to 1-part GS virus during infection for protein production results in reduced excess of GN during the purification (Figure 3-1 and Figure 3-2). This also allows for easier manipulation by site directed mutagenesis.

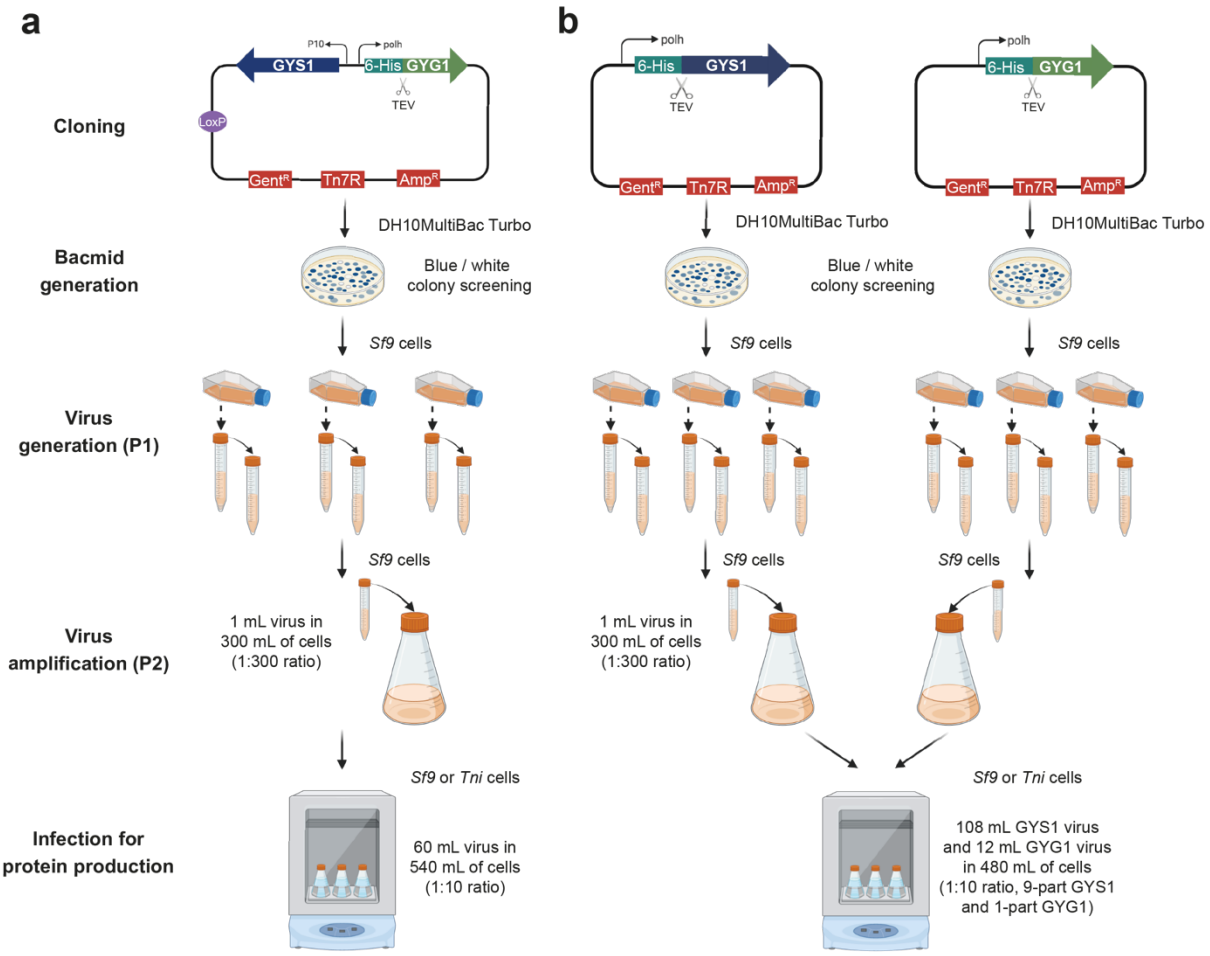

Figure 3-1: Overview of expressing GS and GN in insect cells

a pFL plasmid can be used to co-express GS and GN, by the *GYS1* and *GYG1* genes respectively, in a single vector. **b** pFastBac vectors can be used to generate separate viruses for GS and GN, and co-express these viruses to produce the complex.

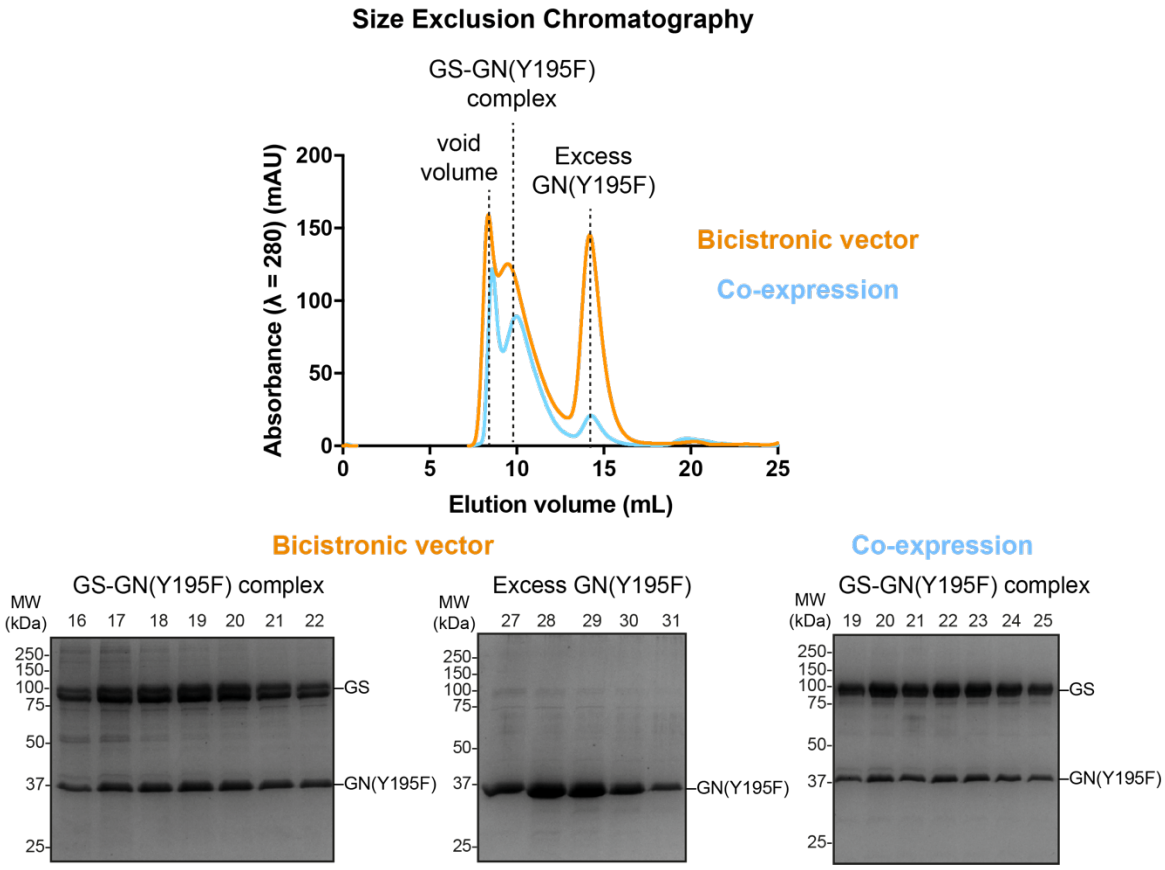

Figure 3-2: Optimisation efforts for the expression of the GS-GN complex

Size exclusion chromatogram traces from purifications using either a bicistronic vector or co-expression. Corresponding SDS-PAGE gels showing the fractions around the named peaks. Using a co-expression method reduces the amount of excess GN that is produced using a bicistronic vector.

3.2.2 Purification of the GS-GN(WT) complex

The full-length human GS-GN complex was expressed and purified from insect cells. Following cell lysis, sonication and centrifugation, the lysate was loaded onto an HisTrap HP Ni column and subsequently eluted by imidazole. The resulting protein was dialysed overnight with TEV protease to cleave the His-tag, and TEV protease was removed by running the dialysed protein through the same HisTrap HP column. These steps were performed by Vertex Pharmaceuticals Ltd. At this stage, the protein was snap frozen and stored at -80°C. The sample was then further purified by size-exclusion chromatography. The protein eluted in two peaks on a Superdex 200 10/300 column, peak 1 (fraction 18) and peak 2 (fraction 20) (Figure 3-3). Curiously, a sharp band migrating at the expected molecular weight for GS was observed, however, GN appeared to migrate as a smear and I hypothesised that this was glucosylated GN.

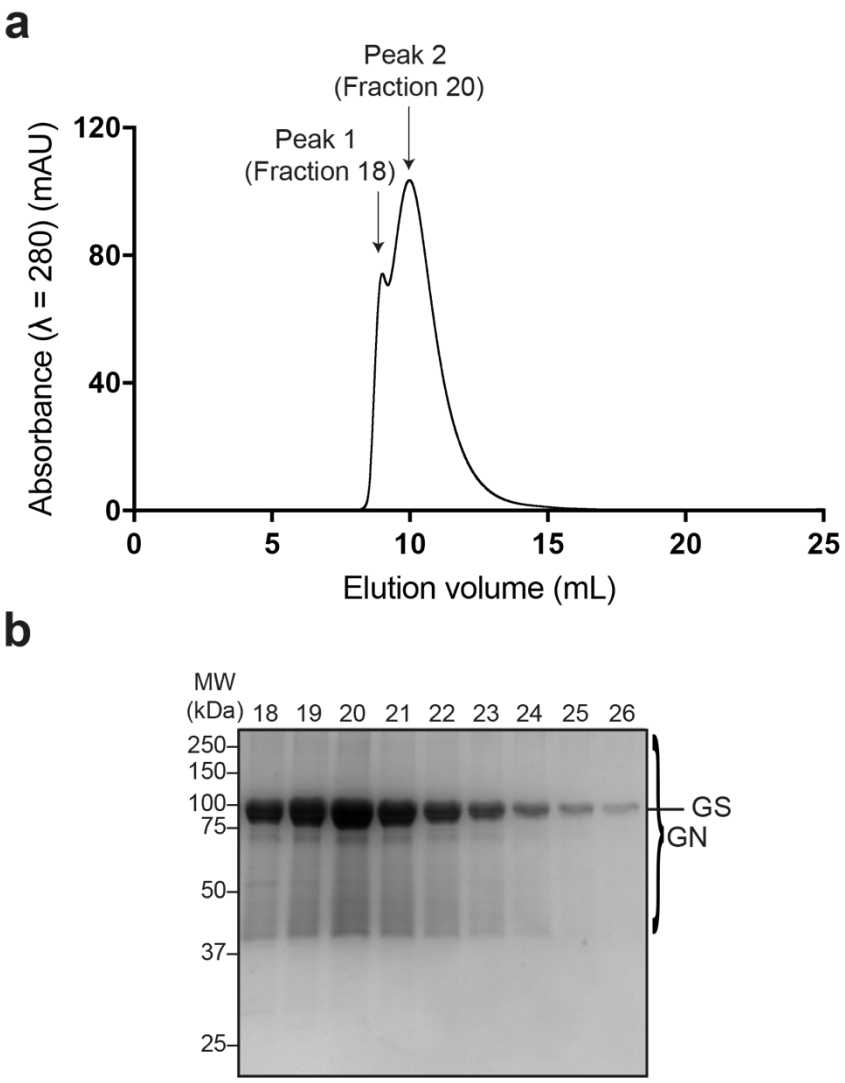

Figure 3-3: Size exclusion chromatography of GS-GN(WT) complex

a Size exclusion chromatogram from a Superdex 200 10/300 column. The GS-GN protein complex elutes in two peaks, as indicated. **b** SDS-PAGE analysis of the indicated fractions. A band ~100 kDa is labelled as GS and the background 'smear' is labelled as GN.

3.2.3 Purification of the GS-GN(Y195F) complex

In order to remove the glucosylation associated with the WT complex, in the aim of producing a more homogenous complex, the GN auto-glucosylating tyrosine 195[60,237] was mutated to a phenylalanine (Y195F), which resulted in a non-glucosylated GN species, as shown by a single band for GN migrating at the expected size (~37.5 kDa) on an SDS-PAGE gel (Figure 3-4b). Similar to the WT complex, the Y195F was purified by nickel affinity chromatography, the His tag was cleaved by TEV protease and both TEV and un-cleaved proteins were subsequently removed by nickel subtraction (Figure 3-4). A final size exclusion chromatography step was introduced for further purification and as a buffer

exchange step (Figure 3-5a and b). Similar to GS-GN(WT), the GS-GN(Y195F) complex also elutes in two peaks in size exclusion chromatography. However, unlike the WT complex, peak 1 is in the void volume. This protein is mainly aggregated which can be seen by nsEM analysis, although a few single particles were visible (Figure 3-5c). Peak 2 was the expected GS-GN(Y195F) complex, and nsEM revealed the vast majority of these particles are monodisperse particles (Figure 3-5c).

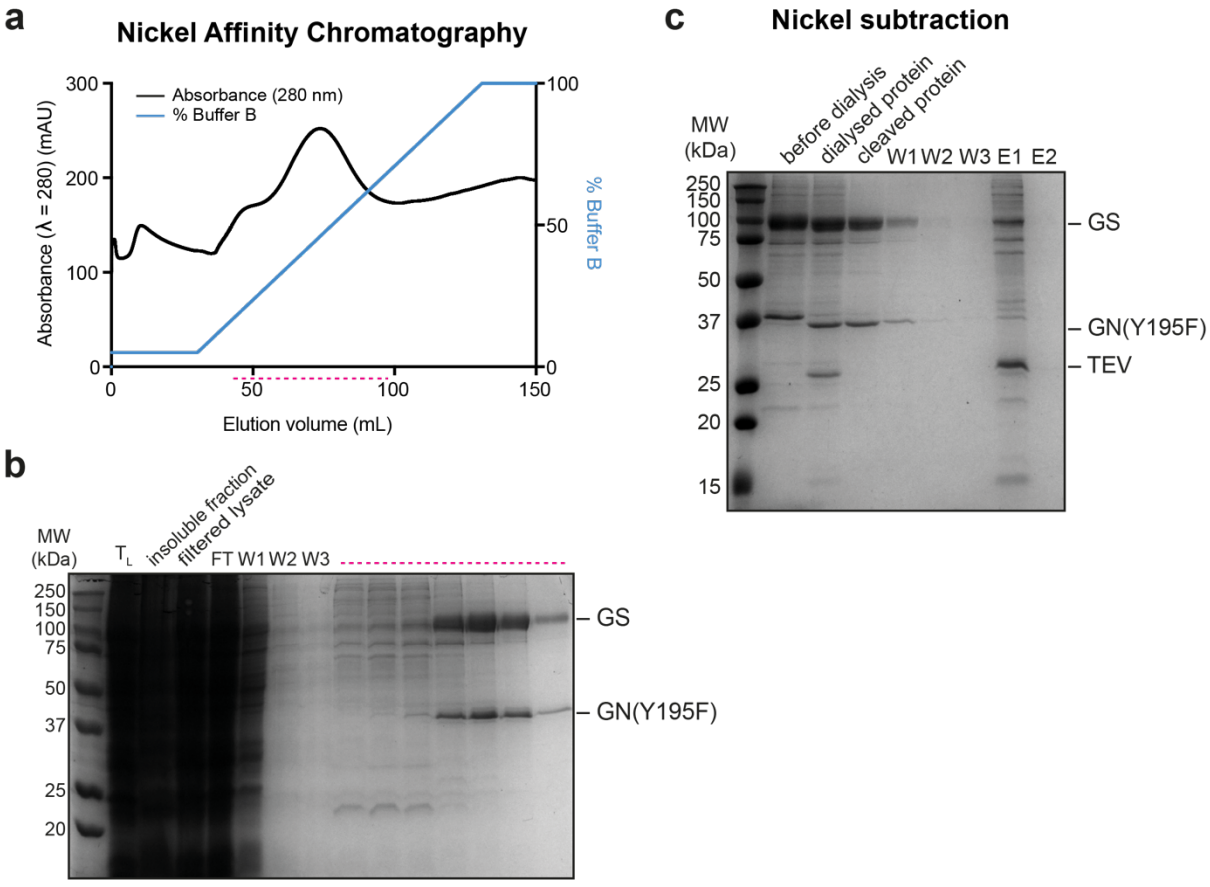

Figure 3-4: Nickel affinity and subtraction steps in the purification of GS-GN(Y195F)

a Chromatogram of the Nickel affinity purification. The pink dotted line shows which area was subsequently ran on an SDS-PAGE gel. **b** SDS-PAGE gel of the steps from the Nickel affinity purification. Samples from the total lysate (T_L), insoluble fraction, filtered lysate, flow through (FT) and subsequent washes (W1, W2, and W3) are shown. The pink dotted line indicates the fractions from the chromatogram. **c** SDS-PAGE gel showing cleavage of the His-tag on GS-GN(Y195F) by His tagged-Tobacco Etch Virus (TEV) protease, and subsequent removal of the His-tagged TEV protease.

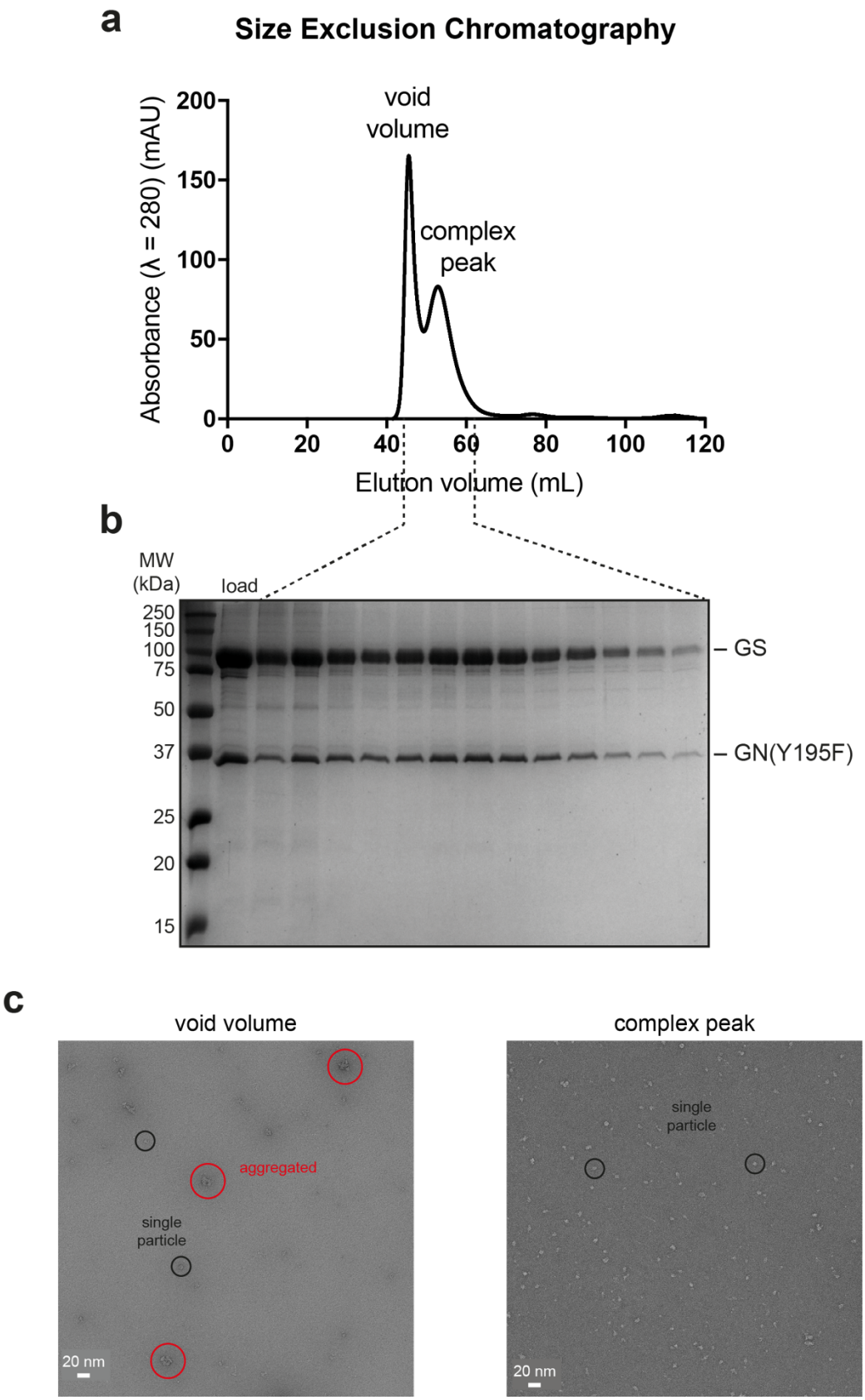
a
Size Exclusion Chromatography
Absorbance (λ = 280) (mAU)
200
150
100
50
0
void volume
complex peak
0
20
40
60
80
100
120
Elution volume (mL)
b
MW (kDa)
load
250
150
100
75
50
37
25
20
15
– GS
– GN(Y195F)
c
void volume
aggregated
single particle
20 nm
complex peak
single particle
20 nm

Figure 3-5: Size exclusion chromatography of the GS-GN(Y195F) complex

a Size exclusion chromatogram from running a Superdex S200 16/600 column. **b** SDS-PAGE analysis of the size exclusion chromatography step. The dotted line shows which area of the chromatogram was run on the gel. **c** Negative stain EM of the void volume and complex peak, to show mainly aggregated particles in the void in comparison to single particles in the complex peak.

3.2.4 GN is glucosylated

There is a stark difference in the preparations of GS-GN and GS-GN(Y195F) as seen by SDS-PAGE. With the WT complex, Coomassie staining shows a background 'smearing effect' as opposed to a sharp GN band at ~37.5 kDa in the Y195F complex (Figure 3-6a). Moreover, it is essential to analyse this smearing effect to investigate the purity of this sample. Periodic acid-Schiff (PAS) staining and immunoblotting was used to analyse the GS-GN WT and Y195F complexes, and reveals that the smear corresponds to glucosylated GN (GN^{Glc}). This reveals that the use of an insect cell expression system resulted in purification of a highly glucosylated complex (Figure 3-6).

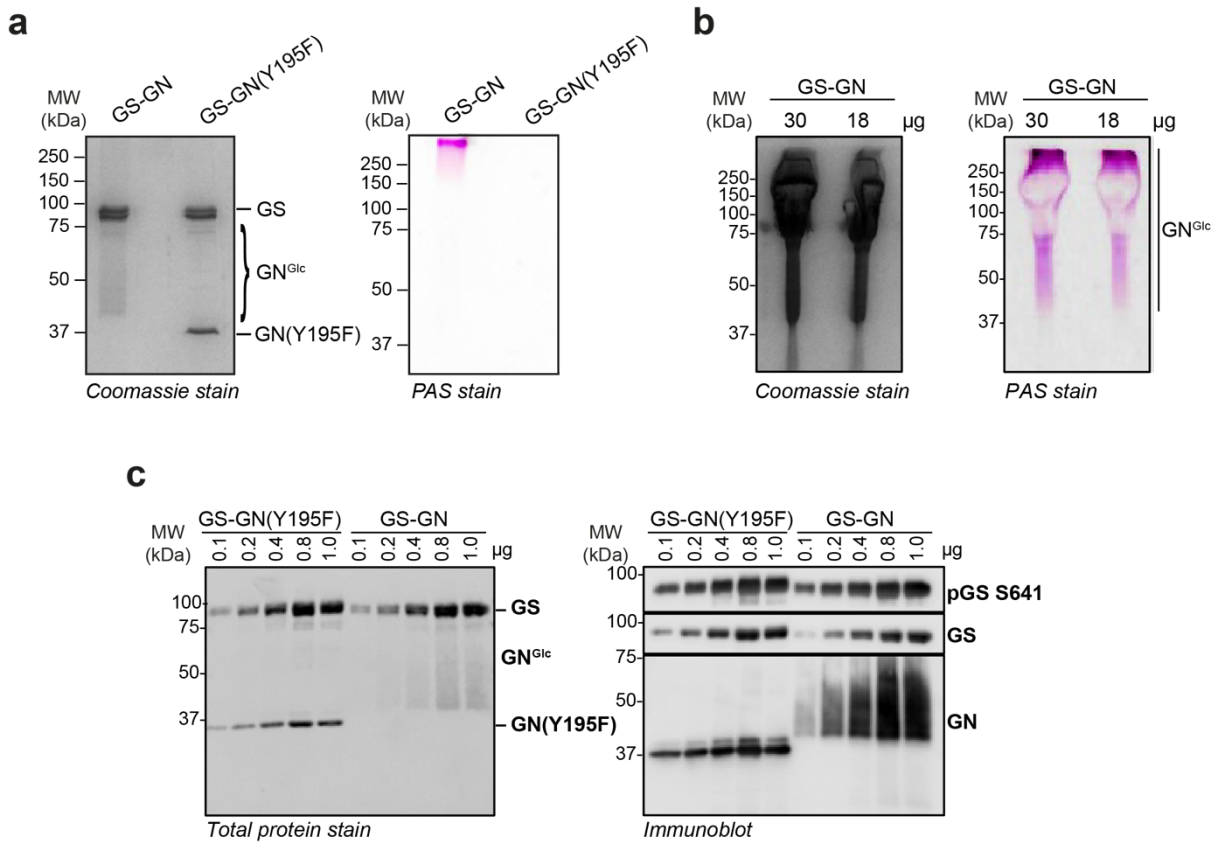

Figure 3-6:The GS-GN complex expressed in insect cells is glucosylated

a SDS-PAGE analysis of GS-GN and GS-GN(Y195F) complexes, where 1.2 µg is loaded (left) and periodic acid-Schiff (PAS) staining of both complexes (right). **b** SDS-PAGE analysis (left) and PAS

staining (right) of the GS-GN complex at higher concentrations. **c** Total protein stain (left) and immunoblot (right) for GS-GN and GN-GN(Y195F) complexes at a range of concentrations.

In-gel protease digestion of the GS-GN complex at different MW ranges (encompassing mass ranges 43-55 kDa, 55-72 kDa, 95-130 kDa and greater than 130 kDa) in combination with tandem mass spectrometry confirmed that the background 'smearing' effect is due to GN, as GN was identified as one of the major components at each interval (Figure 3-7). Consistent with this, treatment of GS-GN preparations with α-amylase (endo-α-1,4-d-glucan hydrolase, which can cleave α-1,4-glycosidic bonds) results in the disappearance of the smeared bands and the appearance of a single, sharp band migrating at the expected molecular weight for GN1 (~37.5 kDa), as well as the absence of glucosylated species after PAS staining (Figure 3-8).

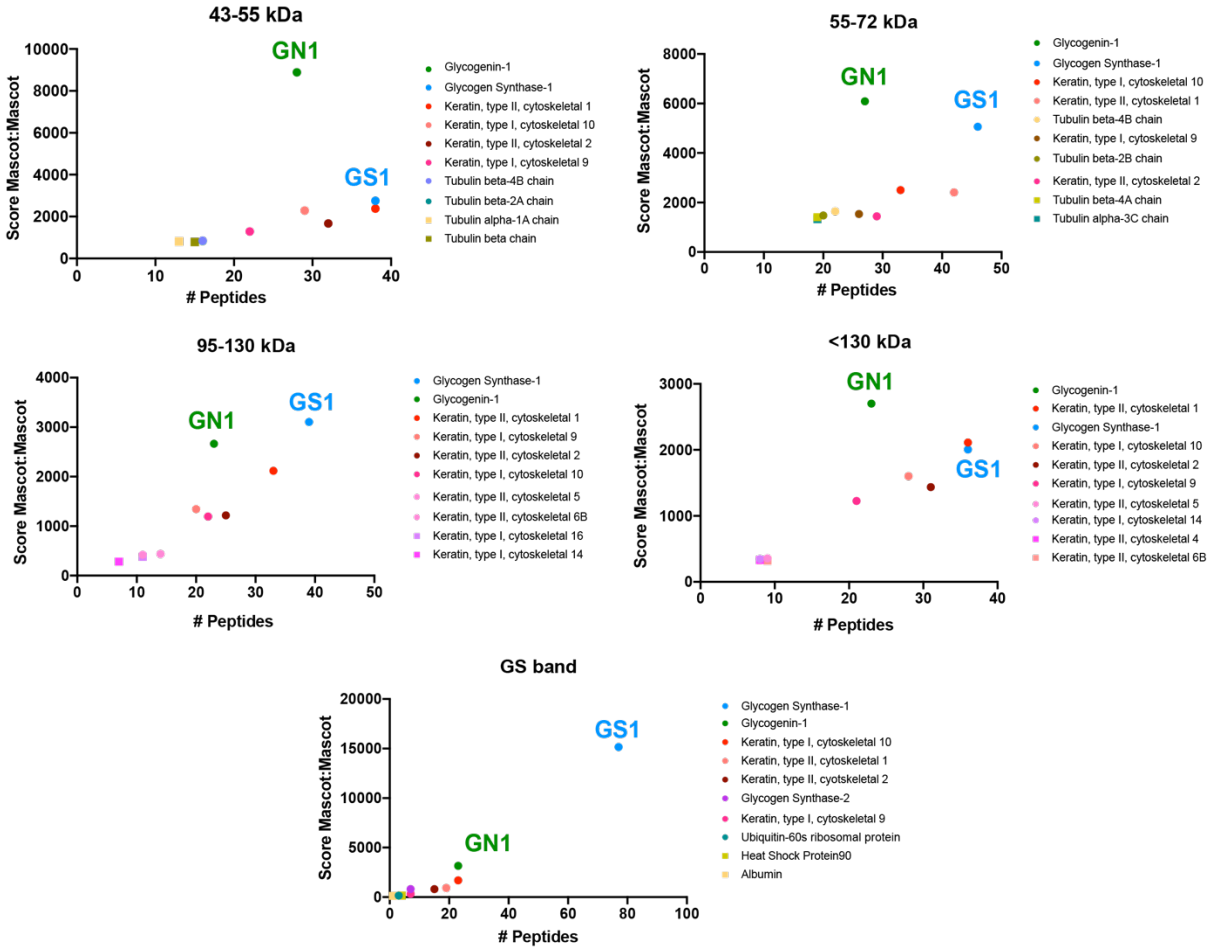

Figure 3-7: Protein Identification by Mass Spectrometry of GS-GN

Peptide mass fingerprinting analysis by mass spectrometry of trypsin-digested gel pieces of the indicated molecular weight regions. The following ranges were analysed: 43-55 kDa (top left), 55-

72 kDa (top right), 95-130 kDa (middle left), greater than 130 kDa (middle right) and the glycogen synthase (GS) band (bottom). For simplicity, only the top 10 proteins with the highest number of peptides, is shown. Glycogenin-1 (GN1) and glycogen synthase 1 (GS1) are labelled on each graph. The other proteins identified are shown in the key.

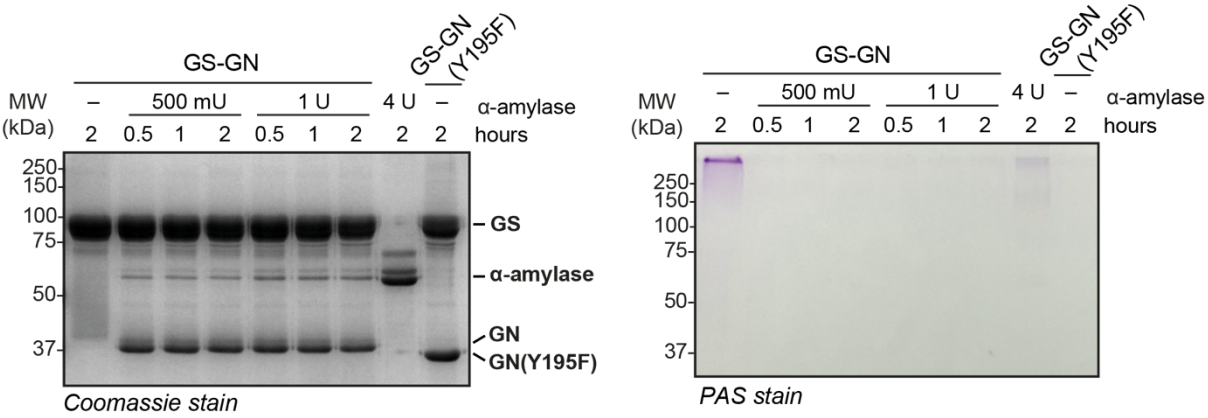

Figure 3-8: α-amylase treatment of the GS-GN complex

Disappearance of the smeared band and the appearance of a single band for GN upon α-amylase treatment can be seen by Coomassie staining (left). In addition, periodic acid-Schiff (PAS) staining shows disappearance of glucosylated species, detected by pink staining, upon α-amylase treatment.

3.3 Stoichiometry of the GS-GN complex

3.3.1 Mass photometry measurements

In order to determine the stoichiometry of the GS-GN complex, we first performed mass photometry analysis, which enables in solution mass measurements of single molecules (see section 2.11). Analysis of the GS-GN(WT) sample identified a species with an average molecular weight of 534 kDa, which is suggestive of a 4:4 stoichiometry (calculated mass of 485 kDa) (Figure 3-9a). Mass photometry analysis of the GS-GN(Y195F) complex showed a predominant molecular weight of 473 kDa, which is again indicative of a 4:4 stoichiometry (Figure 3-9b). Analysis of the non-glucosylated complex revealed a smaller distribution of the main complex peak, in comparison to the WT complex. This observed decrease in molecular mass and distribution of the Y195F mutant in comparison to the WT complex is consistent with the detected higher molecular weight of WT GN1 glucosylated species (Figure 3-6). While mass photometry lacks the resolution

to ascertain precise molecular mass of heterogenous glucosylated species, it can be speculated that the difference between the calculated and measured mass can be accounted for by phosphorylation and glucosylation associated with the GS-GN(WT) complex.

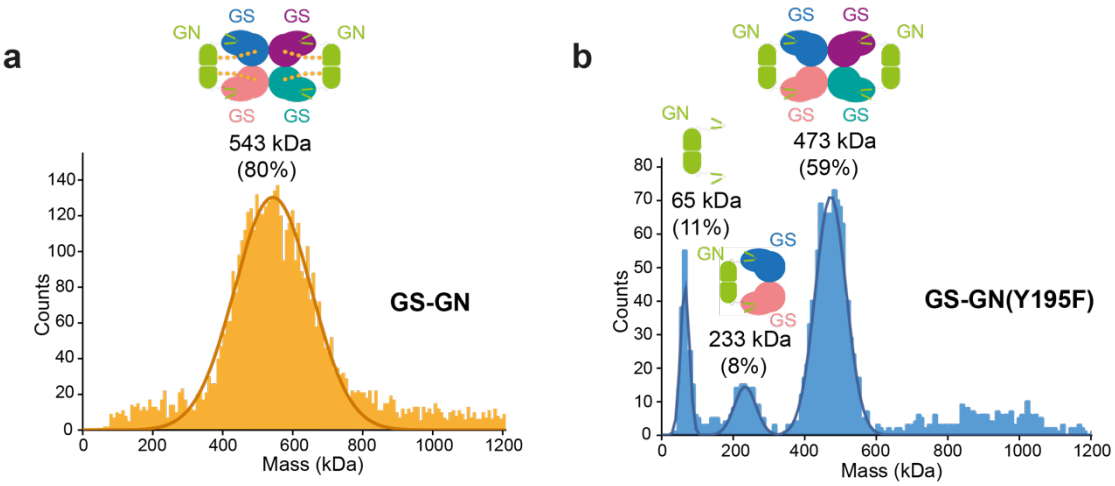

Figure 3-9: Mass photometry analysis GS-GN and GS-GN(Y195F)

a Mass photometry analysis of the GS-GN complex. Estimated molecular weight is shown and percentage of the population. Expected stoichiometry is indicated by a cartoon representation of the complex, where each GS protomer is coloured in blue, purple, pink and turquoise, the GN dimer is shown in green and glucose residues are displayed as orange hexagons. **b** Mass photometry analysis of the GS-GN(Y195F) complex.

3.3.2 Negative stain electron microscopy

In order to confirm the stoichiometry of the GS-GN complex and reveal the overall shape, we performed nsEM. For the WT complex, grids were set up with the two different peaks from the size exclusion trace, fraction 18 from peak 1 and fraction 20 from peak 2 (Figure 3-3 and Figure 3-10). The peak 1 micrographs contain more aggregated particles in comparison to the peak 2 dataset, suggesting this peak is eluting earlier than the other due to the presence of aggregated protein (Figure 3-10). Using RELION autopicking, which was optimised to avoid these aggregates, generated 2D class averages for both datasets showing the same 4:4 stoichiometry for the complex, consisting of two GN dimers, one on either side of a GS tetramer. Final 3D maps are consistent with the 2D classes, as the density can accommodate a GS tetramer and two GN dimers (Figure 3-10). The lack of GS density for the peak 2 dataset may be due to a particle number limitation. Surprisingly, the two GN dimers do not engage GS in an identical fashion, with each GN located slightly

differently on each side of the GS tetramer. In addition, the positions of each GN dimer is different between datasets. This suggests flexibility of GN and potentially each dataset has captured slightly different states of the mobile complex.

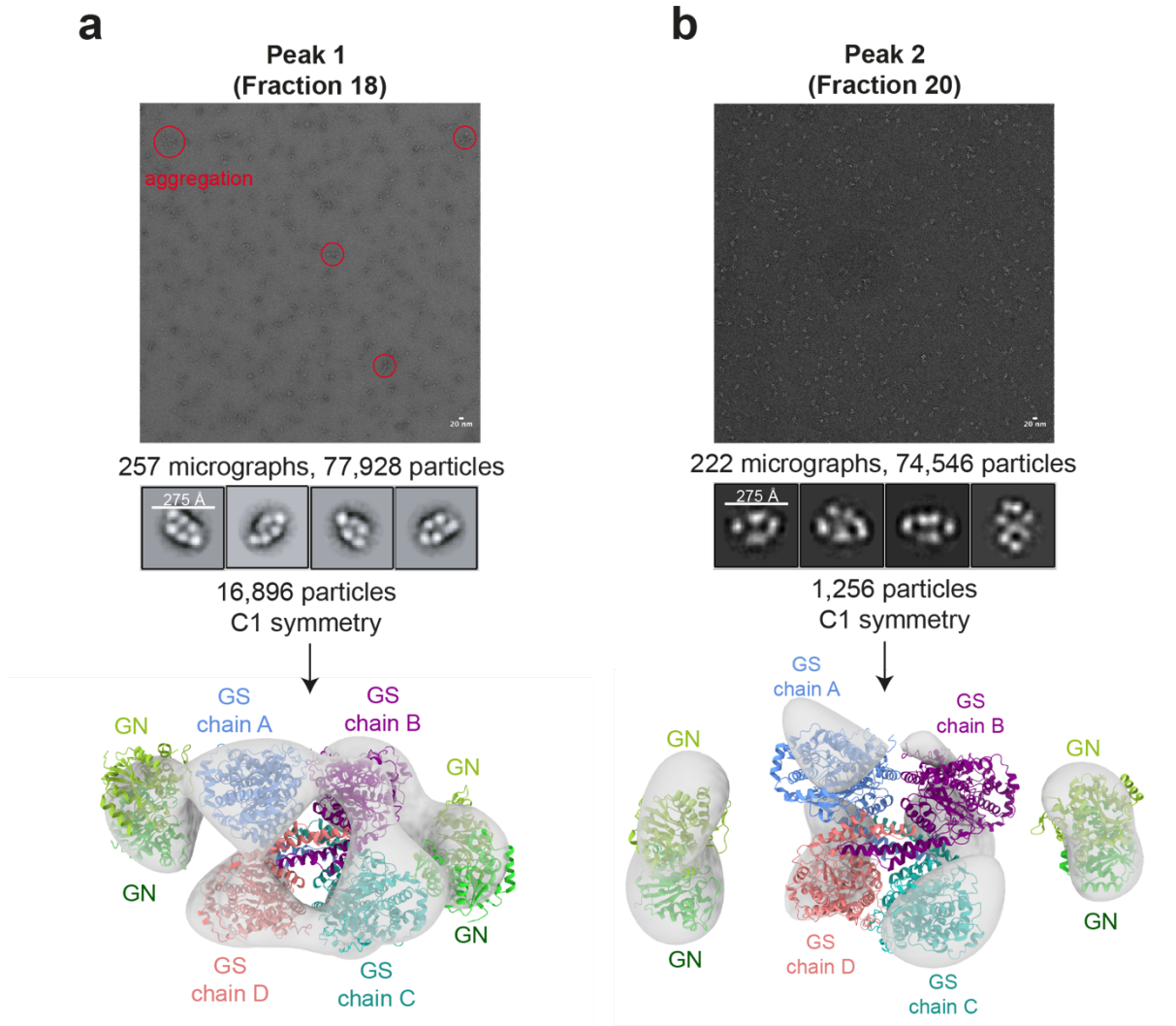

Figure 3-10: Negative stain EM analysis of the GS-GN(WT) complex

a Negative stain EM (nsEM) analysis of peak 1 (fraction 18). Aggregation visible in the micrograph is circled. 2D class averages are shown and the final 3D map. The human GS structure (reported here) and human GN (PDB ID 3T7O) are fitted. Each chain is coloured differently. **b** nsEM analysis of peak 2 (fraction 20). 2D class averages and final 3D map is displayed. The human GS structure (reported here) and human GN (PDB ID 3T7O) are fitted.

With knowledge that peak 1 in the GS-GN(Y195F) complex purification is aggregated (Figure 3-5), we only collected nsEM data for peak 2 (Figure 3-11a). Similar to the WT complex, the 2D class averages show two GN dimers, one on either side of a GS tetramer

(Figure 3-11b), and the final 3D map is consistent with this observation (Figure 3-11c). The determination of a 4:4 stoichiometry for both the WT and Y195F complexes is in agreement with previous findings that glucose chains are not essential for the GS-GN interaction, and thus a non-glucosylated GN can still form a complex with GS[69,92]. Again, GN dimers do not engage GS symmetrically, with one GN tilted slightly towards one of the GS protomers (Figure 3-11c). Collectively, these results confirm that the GS-GN complex has a 4:4 stoichiometry and these analyses offer the first glimpse of the full-length complex.

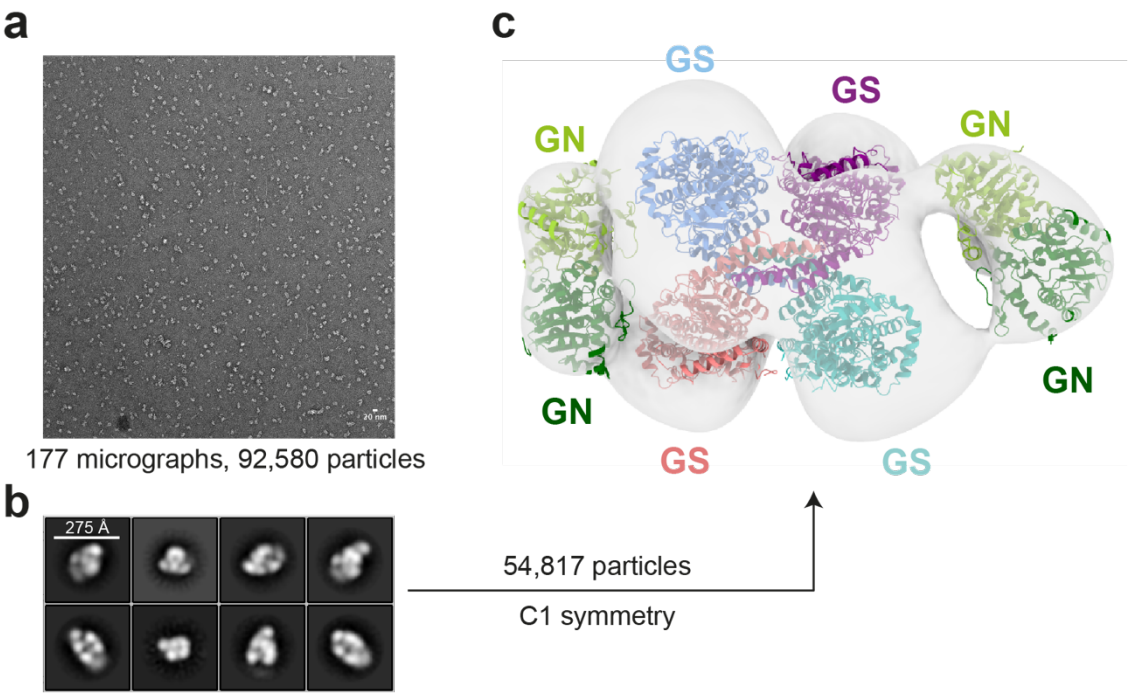

Figure 3-11: Negative stain EM analysis of the GS-GN(Y195F) complex

a Representative negative stain micrograph. **b** 2D class averages showing a 4:4 stoichiometry of the full-length GS-GN(Y195F) complex. **c** Final 3D map with human GS structure (reported here) and human GN (PDB ID 3T7O) are fitted. Density for a GS tetramer with two GN dimers, one on either side, is present.

3.4 Human GS is heavily phosphorylated when produced in insect cells

3.4.1 Dephosphorylation increases GS activity

GS structures to date were from bacterial expression systems and thus were not phosphorylated and could not provide insights into the inhibited state of GS. Our GS-GN preparations are from eukaryotic expression systems, and thus provide an opportunity to

study the phosphorylated form of GS. PP1 and lambda PP treatment resulted in faster migration of GS by SDS-PAGE and also a reduction in signal detected by specific phosphorylation site antibodies (Figure 3-12a and b). Immunoblotting confirmed that human GS is phosphorylated at S8 (site 2) and S641 (site 3a) and activity assays show that the enzyme is inactive unless stimulated by G6P or dephosphorylation (Figure 3-12b and c). A commonly used value to describe the activity state of GS is the activity ratio. The activity ratio compares the activity of GS in the absence and presence of G6P (–/+ G6P), which offers an indication of the phosphorylation level of the enzyme[238]. The value can vary between 0 (fully phosphorylated and inactive form) and 1 (fully activated state)[238]. Notably, we only see minor dephosphorylation of the GS-GN(Y195F) complex with PP1 alone, and this resulted in a 5-fold increase in basal (–G6P) activity (Figure 3-12c). Treatment with both PP1 and lambda PP results in a 15-fold increase in basal activity. All proteins were similarly active upon G6P saturation regardless of phosphorylation state (Figure 3-12c), which is consistent with studies using GS from endogenous sources[239,240].

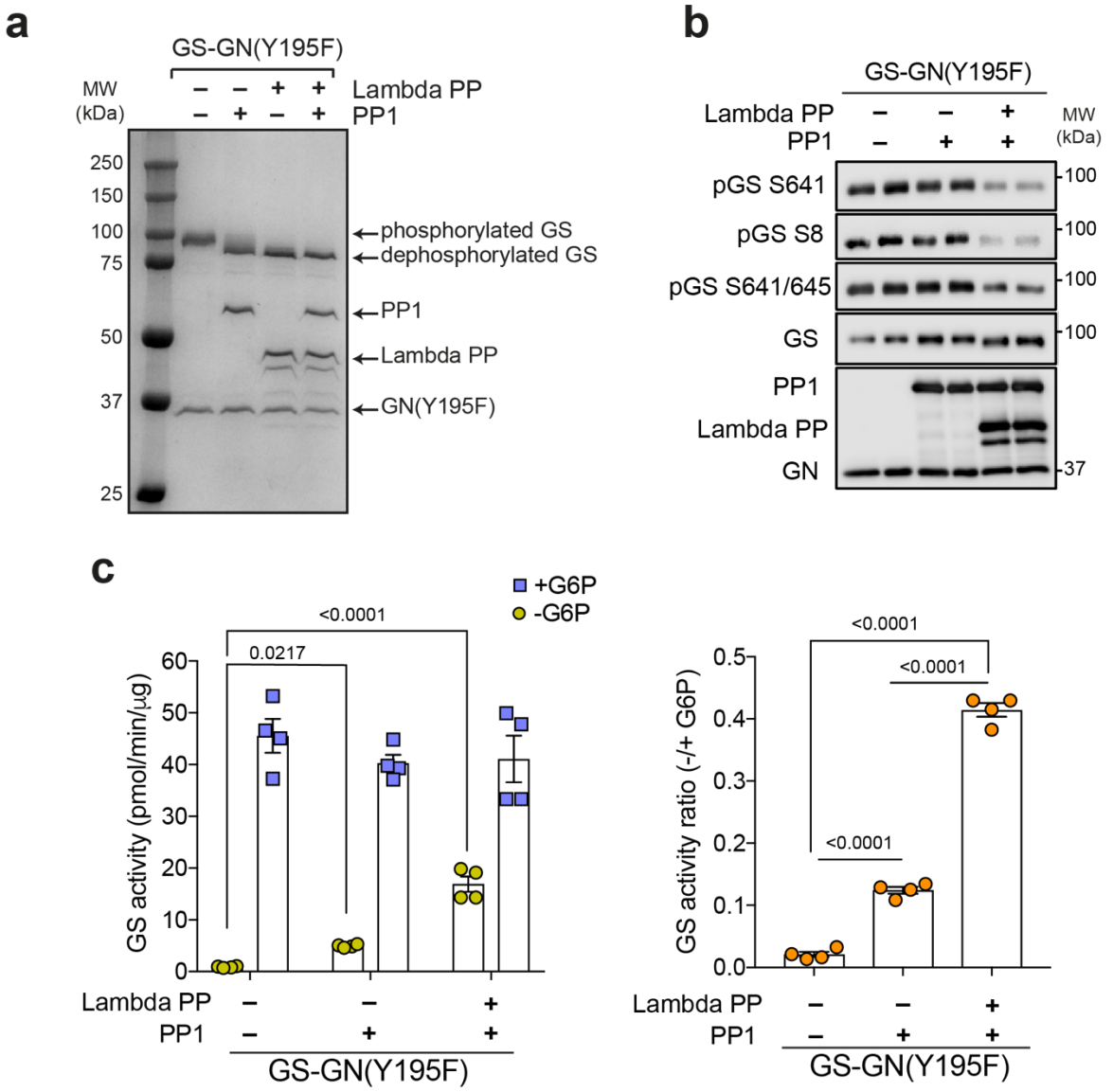

Figure 3-12: Dephosphorylation increases GS activity

a Shift in GS band by SDS-PAGE to a lower molecular weight upon dephosphorylation by PP1 and lambda PP. **b** Immunoblot for the indicated human GS phosphorylation sites, total GS and GN with and without the addition of PP1 and lambda PP. **c** Activity of GS-GN(Y195F) with and without the addition of lambda protein phosphatase (lambda PP) and protein phosphatase 1 (PP1) (left) and -/+ glucose-6-phosphate (G6P) activity ratio (right). Upon G6P saturation, GS reaches similar activity levels regardless of phosphorylation state. Data are mean +/- S.E.M. from two independent experiments performed with technical duplicates. One-way analysis of variance (Tukey's post hoc test); exact *p* values are shown.

3.4.2 Tandem MS/MS phosphorylation mapping (completed by Dr Leonard Daly)

To better understand the extent of GS phosphorylation, we mapped phosphorylation sites by tandem mass spectrometry (MS/MS) following proteolysis with either trypsin, chymotrypsin or elastase. This resulted in a sequence coverage of 97%, which is higher than the 73%[98] and 65%[99] sequence coverage achieved in earlier studies (Figure 3-13). This analysis identified canonical sites 2, 3a, 3b, 4 and 5 (S8, S641, S645, S653, S657), and also non-conventional sites (S412, S652, S727, S731) (Figure 3-13 and Table 3-1). Although these non-conventional sites have been identified before, no physiological role has been determined for these sites. In addition, HsGS site 2 (S8) phosphorylation could be detected for the first time in a recombinant enzyme preparation. All canonical phosphorylation sites identified have been previously identified in GS from endogenous sources[1,239,240], demonstrating physiological relevance of these sites. The stoichiometry of the canonical sites identified is not well understood, and will likely differ across species and cellular location. Together, these results show that expression in insect cells is sufficient to achieve a high level of phosphorylation and thus producing suitable protein preparations for understanding inactive GS.

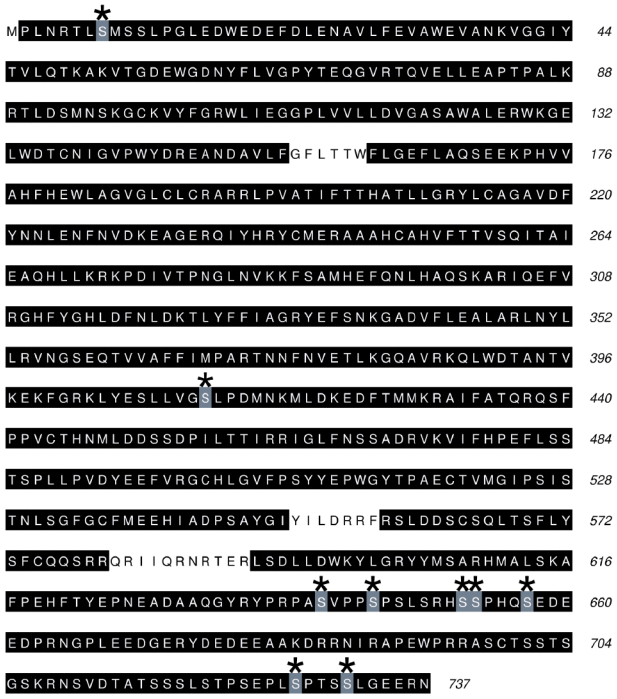

Figure 3-13: Human GS expressed in insect cells is phosphorylated

Sequence of human GS. Sequence coverage (96.7%) during phosphorylation mapping is highlighted in black. Identified phosphorylated residues are highlighted in grey and annotated with an asterix.

Table 3-1: Summary of phosphorylation site analysis.

The site of phosphorylation, peptide sequence and peptide modifications, either phosphorylation (P) or oxidation (O) are displayed. The ptmRS score indicates the confidence of where the phosphorylation is localised within the peptide. The MASCOT score is the confidence of identification. PhosphoSite Plus (web-based bioinformatics resource) was used for comparison of our results with previous literature. LTP and HTP refer to low throughput site determination (methods other than mass spectrometry) and high throughput analysis (mass spectrometry only), respectively[241].

Site	Sequence	Peptide modifications	ptmRS Score	MASCOT score	Enzyme	PhosphoSite: LTP	PhosphoSite: HTP
S8	PLNRTLsMS	S7(P)	S7: 100	40	Elastase	15	1
S8	PLNRTLsMS	S7(P); M8(O)	S7: 100	31	Elastase	15	1
S8	sMSSLPGLEDW	S1(P)	S1: 99.71	13	Chymotrypsin	15	1
S412	ESLLVGsLPDMNKMLDKEDF	S7(P)	S7: 100	31	Chymotrypsin	-	17
S412	ESLLVGsLPDMNKML	S7(P); M11(O); M14(O)	S7: 100	23	Chymotrypsin	-	17
S641	QGYRYPRPAsVPPSPS	S10(P)	S10: 99.99	20	Elastase	32	30
S641 & S645	QGYRYPRPAsVPPsPS	S10(P); S14(P)	S10: 100; S14: 100	24	Elastase	(645) 21	(645) 39
S652	RHsSPHQSEDEEDPRNGPL	S3(P)	S3: 99.78	34	Elastase	-	15
S652 & S653 & S657	RHssPHQsEDEEDPRNGPL	S3(P); S4(P); S8(P)	S3: 100; S4: 100; S8: 100	19	Elastase	(653) 13	(653) 27
S652 & S657	RHsSPHQsEDEEDPRNGPL	S3(P); S8(P)	S3: 99.67; S8: 100	53	Elastase	(657) 14	(657) 50
S727	RNSVDTATSSSLSTPSEPLsPTSSLGEER	S20(P)	S20: 100	66	Trypsin	1	22
S727	STPSEPLsPTSSL	S8(P)	S8: 99.63	20	Chymotrypsin	1	22
S731	STPSEPLSPTSsLGEERN	S12(P)	S12: 99.79	73	Chymotrypsin	-	6
S731	TPSEPLSPTSsL	S11(P)	S11: 100	22	Elastase	-	6
S727 & S731	NSVDTATSSSLSTPSEPLsPTSsLGEER	S19(P); S23(P)	S19: 100; S23: 99.52	60	Trypsin	see above	
S727 & S731	STPSEPLsPTSsLGEERN	S8(P); S12(P)	S8: 100; S12: 100	42	Chymotrypsin	see above	

3.5 Structure determination of the full-length GS-GN(Y195F) complex

3.5.1 Low resolution structure of the GS-GN(Y195F) complex

Previous attempts to crystallise full-length GS in complex with full-length GN were unsuccessful[92] and led us to pursue structural analysis using cryo-EM. NsEM analyses indicated flexibility of GN in the GS-GN complex (Figure 3-10 and Figure 3-11). Cryo-EM analysis of the GS-GN(Y195F) complex confirmed this GN flexibility, evident from the lack of GN signal in 2D class averages (Figure 3-14a) and subsequent 3D maps. Nonetheless, in one dataset collected on the Gatan K2 detector, some GN density can be seen during the 3D classification (Figure 3-14b). This class was taken forward for 3D refinement, applying D2 symmetry with the aim of reaching high-resolution, which generated a 11 Å structure of the GS-GN complex (Figure 3-14b). When fitting the human GN crystal structure (PDB ID 3T7O) into the density for GN, the density is smaller than the crystal

structure. This could be due to the lower resolution of the map, or perhaps due to the mobility of GN. Unfortunately, potentially due to particle number limitation and flexibility, the resolution was not high enough to be able to trace connecting residues between GS and GN[34], which are known to interact in *C. elegans*[92], or other unknown interactions between GS and the globular domain of GN. Nonetheless, this map is the first higher resolution representation of the full-length GS-GN complex, and consistent with the nsEM analysis, shows one GS tetramer coordinated by two GN dimers.

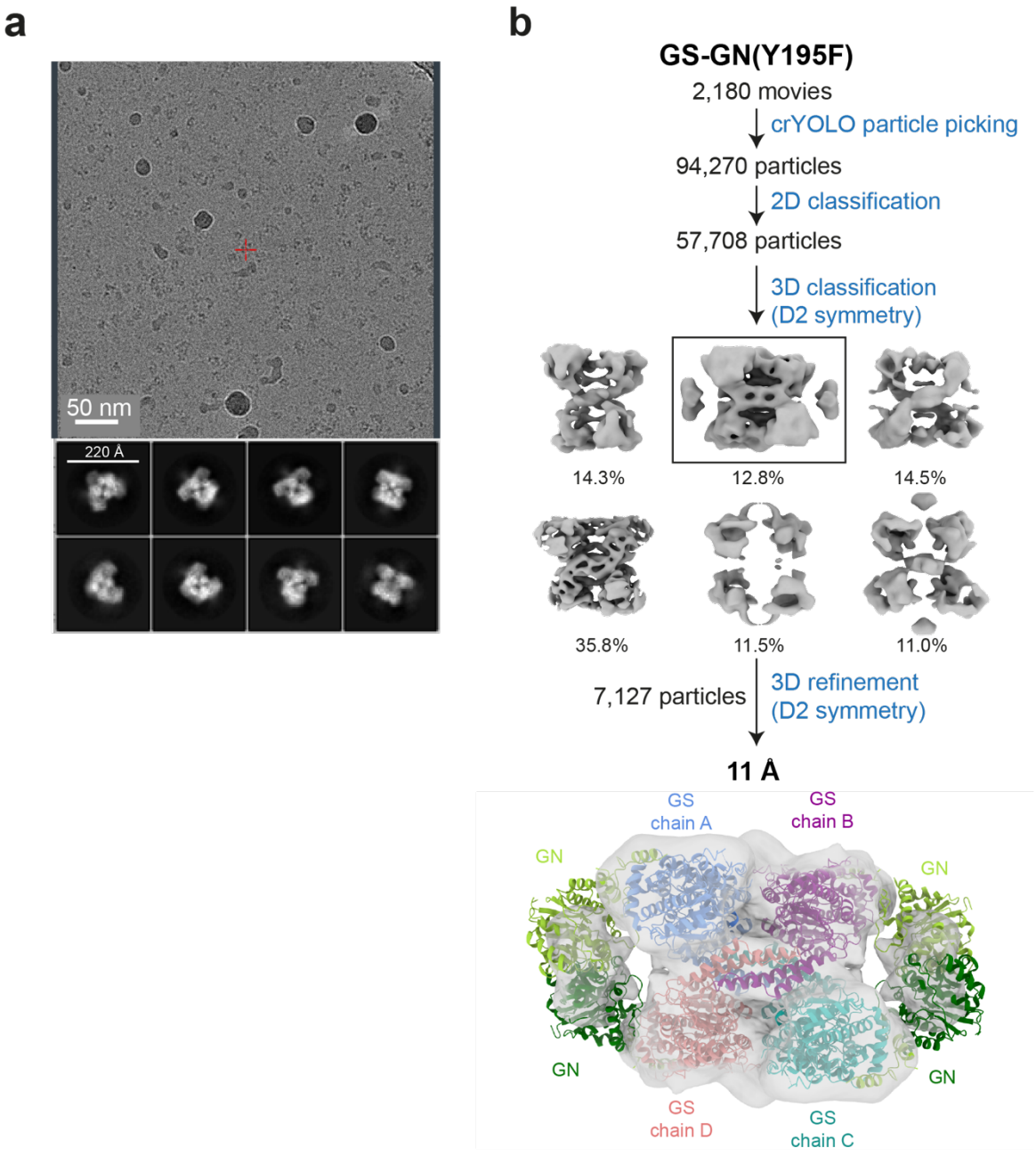

Figure 3-14: Low resolution structure of the full-length GS-GN(Y195F) complex

a Representative micrograph and 2D class averages. **b** Flow chart of data processing. The 3D class boxed was taken forward for 3D refinement to obtain a 11 Å structure of the full-length GS-

GN(Y195F) complex. The human GS structure (reported in this thesis, see chapter 5) and human GN (PDB ID 3T7O) are fitted.

3.6 Chapter summary

Although a binary complex of GS-GN was co-purified over 30 years ago[63], we have yet to confirm the stoichiometry of this complex and explain how these two enzymes assist one another in glycogen synthesis. The GS-GN complex was produced in an insect cell expression system, resulting in a glucosylated and phosphorylated form of the complex. Mutation of the auto-glucosylating Tyr to Phe produced a non-glucosylated complex that retained its phosphorylation. The extent of the phosphorylation was identified using mass spectrometry and phosphatase treatment, which showed GS is heavily phosphorylated. As expected, this complex was inactive unless stimulated by dephosphorylation or allosterically activated by adding G6P. Both glucosylated and non-glucosylated complexes form a 4:4 stoichiometry, revealing two flexible GN dimers coordinating a GS tetramer. It is exciting to speculate a model where GS could extend glucose chains from two GN dimers simultaneously, however this requires further investigation. Overall, the production of GS-GN preparations that are active and stoichiometric lay crucial foundations for studying the GS-GN complex in more detail, and specifically investigating the elusive phosphorylated and inactive state of GS.

4 Chapter 4: Structural studies of the GS-GN complex

4.1 Introduction

At the start of this project, the structure of human GS was unknown. Moreover, the aim of this work was to determine a structure of human GS to understand the complex regulation of this enzyme, particularly how phosphorylation causes inactivation and how this is coupled with allosteric activation by G6P. To achieve this, cryo-EM was used, alongside optimising various methods of sample and grid preparation, to investigate the GS-GN complex.

4.2 Overall cryo-EM structure of human GS

Due to the GN flexibility and the issues this caused in cryo-EM data processing (see section 3.5.1), we subsequently focused the data processing strategy with the goal of obtaining a high-resolution structure for human GS. Taking advantage of the prior knowledge that GS forms a tetramer, we applied D2 symmetry and achieved a global resolution of 4 Å (Figure 4-1 and Figure 4-2). The reconstruction revealed a tetrameric arrangement of HsGS and interpretable density for the C-terminal GS interacting region of GN. While the cryo-EM density was insufficient to build a *de novo* model of the human GN C-terminus, it was sufficient for a model (generated by Phyre2[227]) to be placed, and this was similar to the CeGS-GN[34] crystal structure[70] (Figure 4-1). When lowering the threshold of this map, significant density can be seen for GN on either side of the GS tetramer, however the resolution of this area is very low and noisy due to GN flexibility (Figure 4-2d).

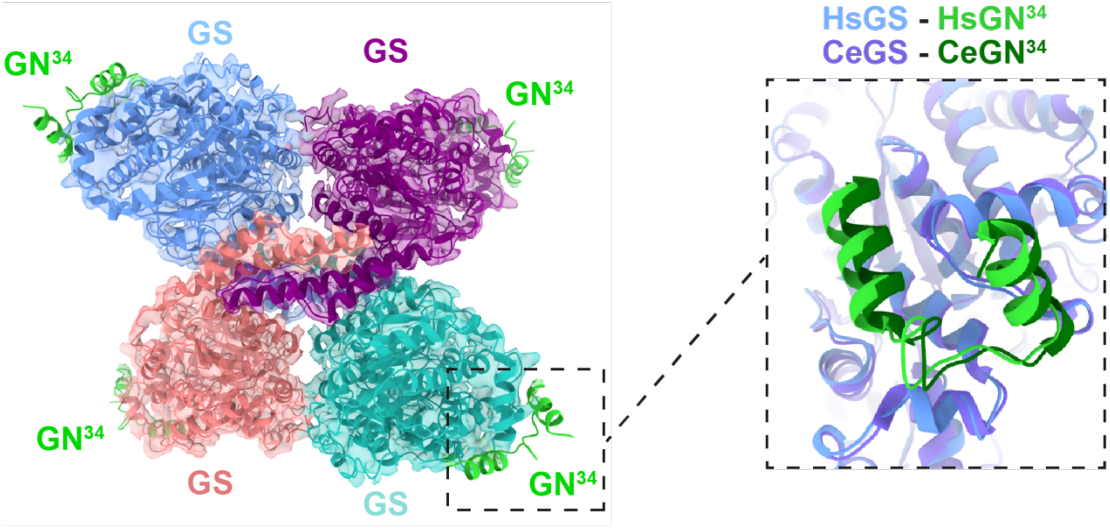

Figure 4-1: Tetrameric human GS in complex with human GN[34]

a Human GS cartoon model shown in ribbons and a 4 Å reconstructed cryo-EM map shown in transparent surface and coloured by corresponding chain. A model of the last 34 residues of the C-

terminus of human GN (GN^{34}) is displayed in green (left). The position of GN^{34} is consistent with the CeGS-GN^{34} (PDB ID 4QLB) structure.

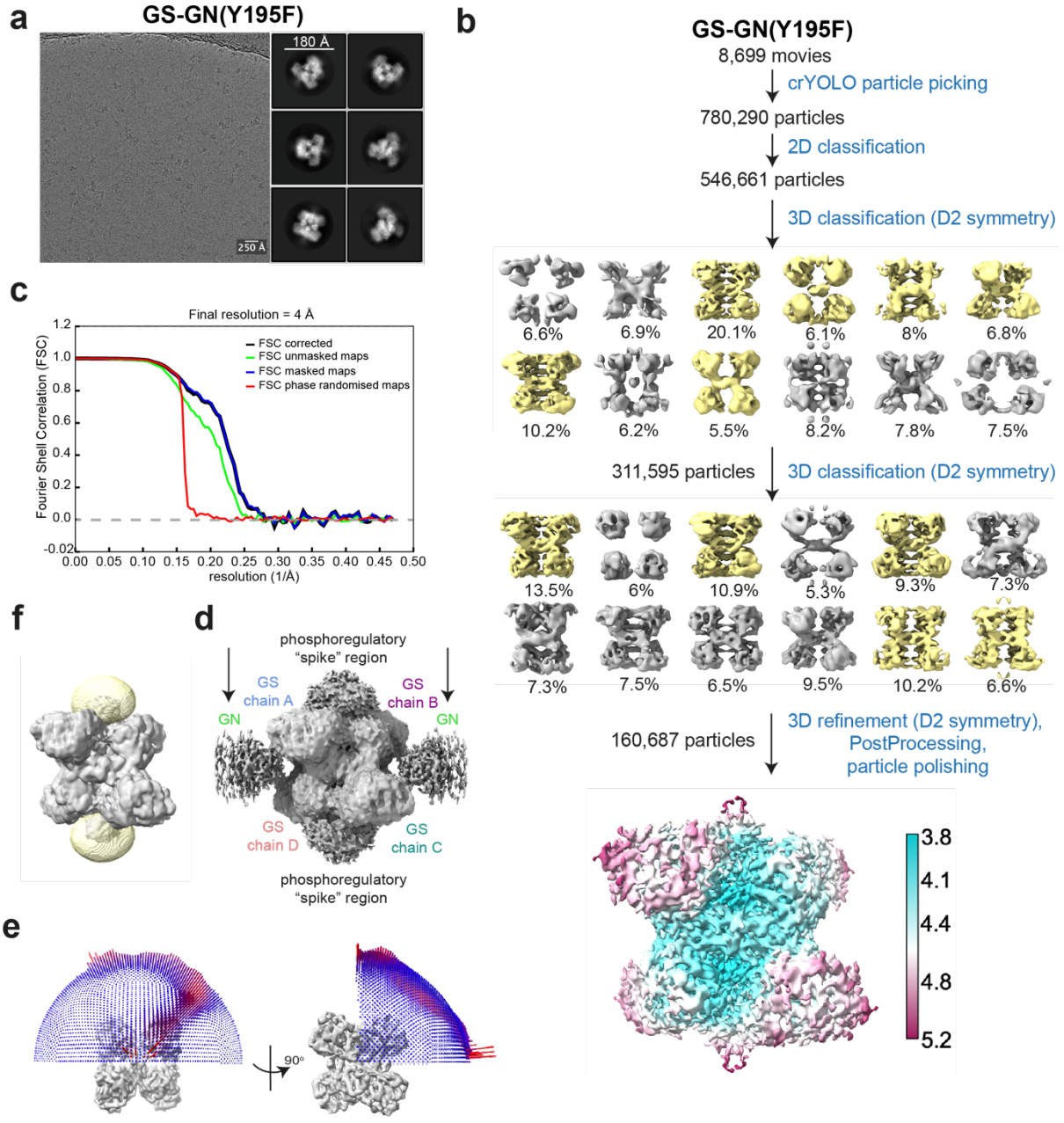

Figure 4-2: Cryo-EM analysis of GS-GN(Y195F)

a Representative micrograph and 2D class averages. **b** Flow chart of data processing. 3D classes coloured in yellow were taken forward for subsequent processing. **c** FSC curve for the final map. **d** a 3D refinement map shown at a lower threshold to visualise the weak density for the flexible GN dimers, one of either side of the GS tetramer. **e** Euler angle distribution of particles used for the 3D reconstruction. Height of rods represents number of particles, and red indicates a high number of

particles and blue indicates a low number of particles. **f** Input map and mask used in the focused 3D classification shown in Figure 4-3b.

4.3 Analysis of the human GS structure

4.3.1 Dynamic phosphoregulatory region

Notably, the EM density maps show density for an inter-subunit domain. This occupies the space where the N- and C- terminal regions, which contain the GS phosphoregulatory regions, of two adjacent GS protomers meet and then traverse away from the core (Figure 4-3a). Local resolution analysis shows that this area is lower resolution than the core of the GS tetramer (Figure 4-2b). Analysis of this ~25 kDa region by focussed 3D classification (without applying symmetry) reveals that the region is highly flexible, as seen by the various different conformations, thus explaining the lower resolution (Figure 4-3b and Figure 4-2f). Interestingly, these “spike” regions were present in all the classes generated from the focussed 3D classification. This suggests that GS exists as a continuum of structures with a core inactive tetramer and “dynamic spikes” buttressed on either side, constraining the opening of GS to an active tetramer and thus stabilising the inactive state.

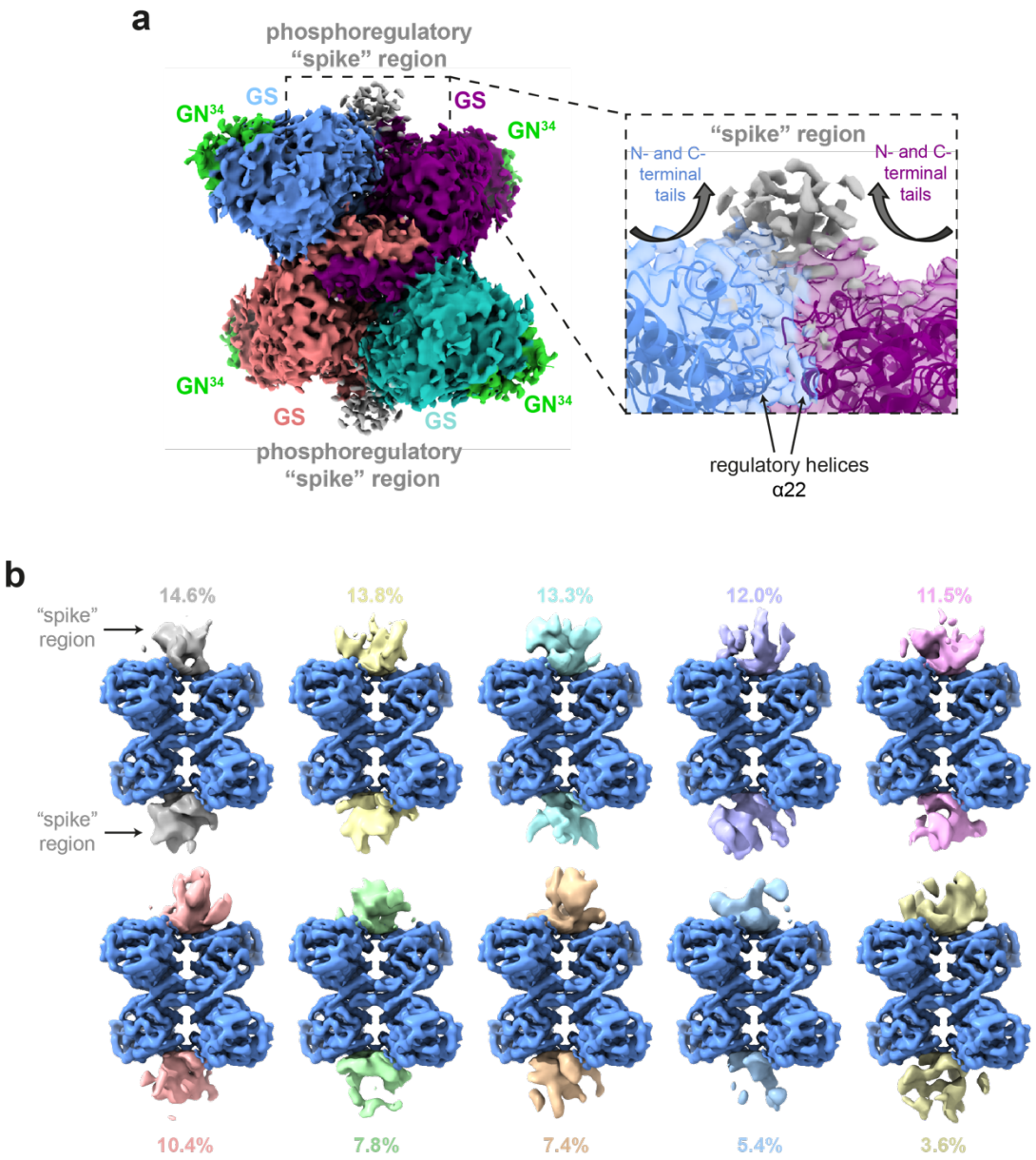

Figure 4-3: Human GS contains an inter-subunit "spike" region

a 4 Å cryo-EM density map shown at a lower threshold (left). The N- and C-terminal tails of two protomers converge and form the "spike" region, which is located above the regulatory helices (right). **b** Focussed 3D classification of the phosphoregulatory "spike" region reveals flexibility of this inter subunit domain. 3D classes, shown in different colours, and percentage population are indicated and the 3D refined map of the core region is coloured blue.

4.3.2 Treating heterogeneity in cryo-EM datasets

Both GN and the “spike” region are sources of flexibility in cryo-EM analyses of the GS-GN complex, and thus result in a heterogenous sample (as seen by multiple GS classes in 3D classification, Figure 4-2b). Recently, a number of processing tools and techniques were described which allow treatment of heterogeneity in cryo-EM datasets, facilitating the identification of multiple conformations present and to allowing generation of high-resolution structures[214]. One of these tools is particle subtraction which, briefly, can perform partial subtraction of undesired signal in particle images in order to improve the reconstruction of a further defined body[214]. This was performed to remove the signal for GN, which was potentially hindering the resolution of GS (Figure 4-4a). Focused 3D classification encompassing the GS tetramer and the spike region was then performed in order to investigate potential different conformations of GS (Figure 4-4b and c). This resulted in several different classes, and specifically when looking at the 3D slices, some classes appear more rectangular than square (Figure 4-4c). Whether this is due to various levels of phosphorylation is not known and hard to ascertain at this resolution. In addition, there is a wide variation in the conformation of the “spike” region in these classes, again highlighting the mobility in this region. Taking every 3D class forward separately for 3D refinement and postprocessing resulted in maps ranging in resolution from ~5 Å to ~11 Å (Figure 4-4c). Even the class with the most particles (class003 - 20,985 particles) resulted in a much lower resolution of 5.21 Å in comparison to the 4 Å consensus reconstruction showed earlier (Figure 4-1 and Figure 4-2). Interestingly, a class with fewer particles (class010 – 13,726 particles) reached a slightly higher resolution of 5.09 Å, showing a balance between homogeneity and particle number. Overall, removing density for GN did not result in a higher resolution map.

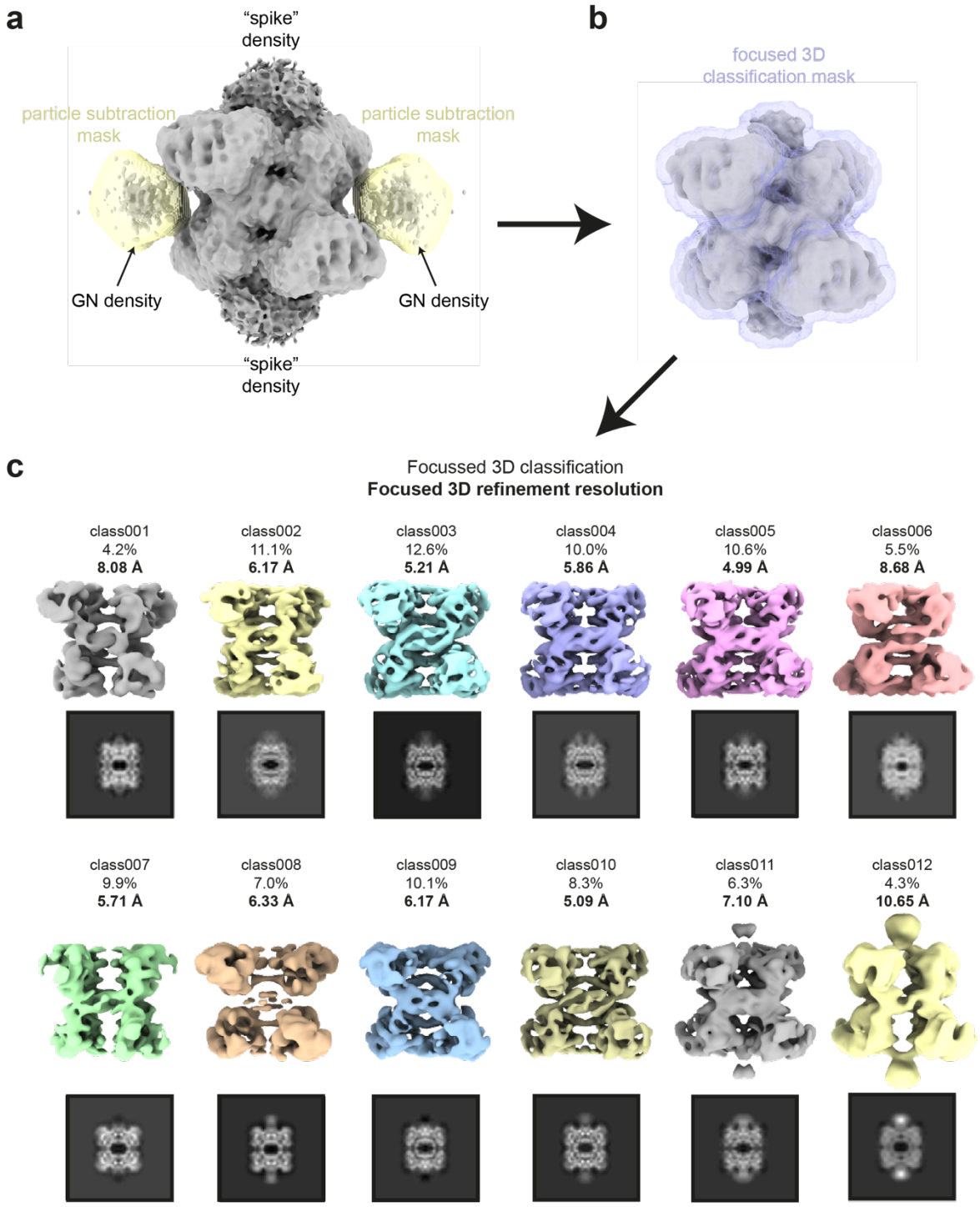

Figure 4-4: Treating heterogeneity by removing flexible glycogenin

a 3D reconstruction of GS to be used for particle subtraction (grey), using the mask shown in yellow, to remove the flexible GN region. **b** The signal subtracted map (grey) with the mask to be used in focussed 3D classification shown in purple. **c** 3D map and central slice of the twelve classes generated in the focused 3D classification (D2 symmetry) are shown. Each class was taken forward for focussed 3D refinement, and the resolution obtained is shown in bold.

Another source of heterogeneity in the dataset is the dynamic “spike” region (Figure 4-3b). Moreover, in the aim of reaching high-resolution of the core of GS, we used particle subtraction to remove both flexible regions, the “spike’’ region as well as GN (Figure 4-5). A mask encompassing both the “spike” and GN density was created (Figure 4-5a) and used during particle subtraction to remove these areas. A mask around the GS core was created and subsequently used in focussed 3D classification (Figure 4-5b). Twenty 3D classes were generated, an increase in the twelve that were generated previously (Figure 4-4), in order to investigate the extent of the various conformations. The resulting classes range in resolution as well as revealing slight changes in conformation. To further explore the flexibility and mobility of GS, we performed 3D variability analysis[221], without the application of symmetry, using cryoSPARC[211] (see section 2.5.3.9). This revealed dynamic movements of the “spike” region and concurrent flexing of the GS tetramer (Figure 4-6). This shows slight conformational changes in the GS core, similar to the classes seen in the focussed 3D classification (Figure 4-4 and Figure 4-5). Unfortunately, the lower resolution of these classes prevents identification of the precise differences between these conformations. However, the movements of the GS core can be roughly mapped, revealing slight flexing and twisting movements (Figure 4-5 and Figure 4-6). The functional significance of these conformational changes is unknown; however, it could be due to heterogenous phosphorylation levels, or perhaps differences in the GS-GN interaction. It may also be that the GS tetramer is slightly flexible in the absence of the allosteric ligand G6P or the substrate, UDP-G, although these hypotheses require further investigation. Nonetheless, these analyses reveal a role for the “spike” region in potentially constraining the opening of the GS tetramer and stabilising an inactive state.

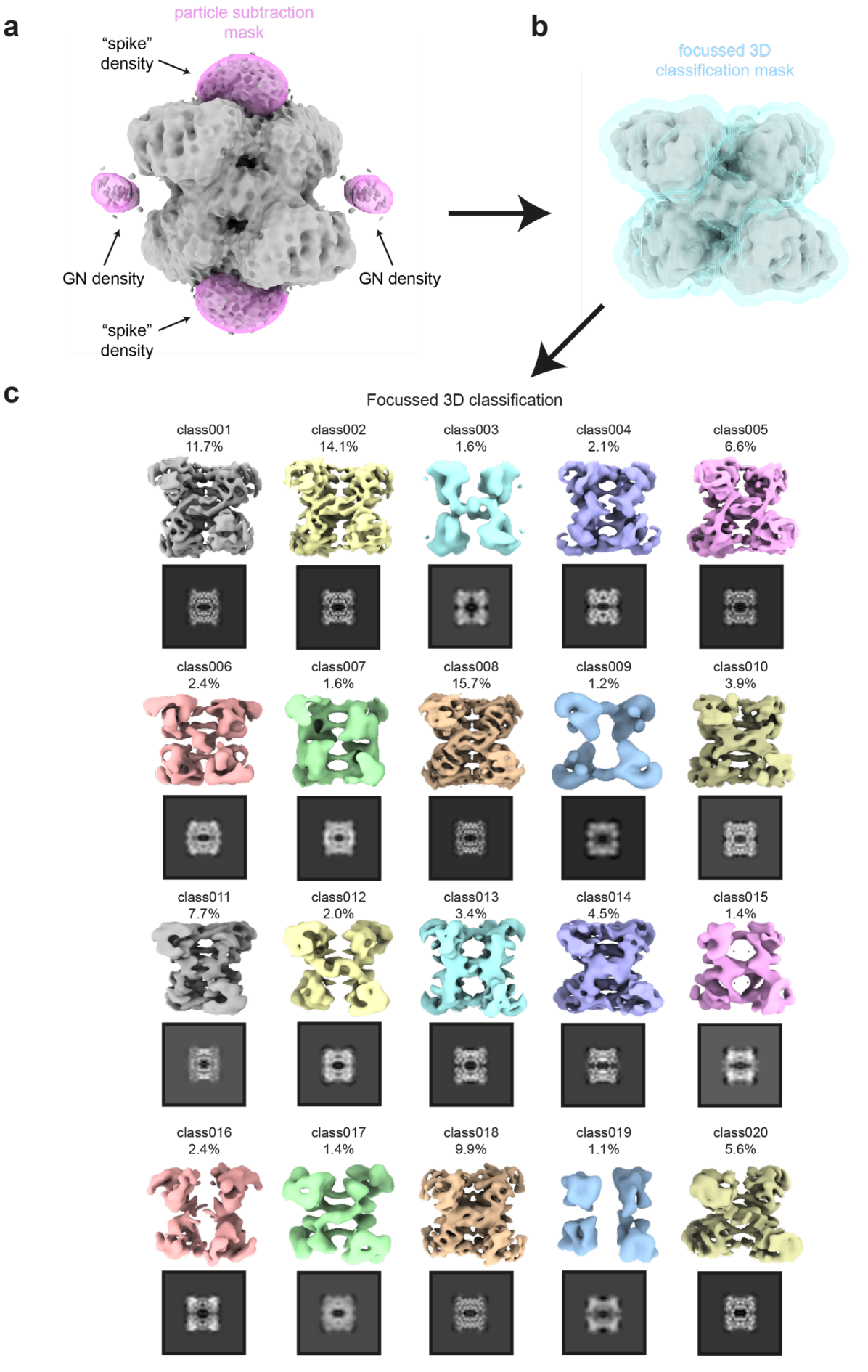
a
particle subtraction mask
"spike" density
GN density
GN density
"spike" density
b
focussed 3D classification mask
c
Focussed 3D classification
class001 11.7%
class002 14.1%
class003 1.6%
class004 2.1%
class005 6.6%
class006 2.4%
class007 1.6%
class008 15.7%
class009 1.2%
class010 3.9%
class011 7.7%
class012 2.0%
class013 3.4%
class014 4.5%
class015 1.4%
class016 2.4%
class017 1.4%
class018 9.9%
class019 1.1%
class020 5.6%

Figure 4-5: Treating heterogeneity by removing flexible domains

a 3D reconstruction of GS to be used for particle subtraction (grey), using the mask shown in pink, to remove the flexible "spike" and GN region. **b** The signal subtracted map (grey) with the mask to be used in focussed 3D classification shown in blue. **c** 3D map and central slice of the twenty classes generated in the focused 3D classification (D2 symmetry).

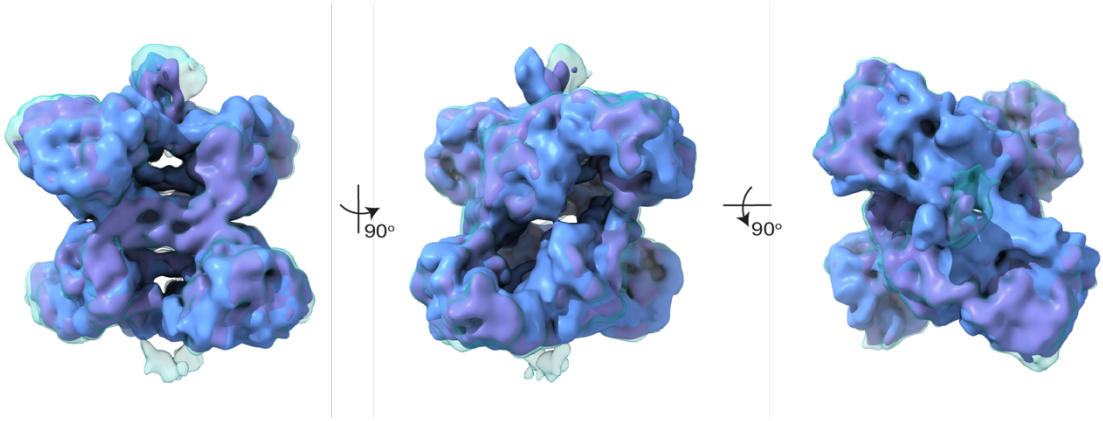

Figure 4-6: 3D variability analysis showing conformational changes of GS

Three selected frames from 3D variability analysis, shown in blue, purple and turquoise. Flexing of the GS tetramer and dynamic movement of the "spike" region can be seen.

4.4 Structural analysis of the GS-GN(WT) complex

4.4.1 Cryo-EM analysis

In order to structurally investigate the GS-GN(WT) complex, and investigate any potential differences between this glucosylated complex and the GS-GN(Y195F) complex, cryo-EM analysis was performed. Similar to GS-GN(Y195F), cryo-EM analysis of the WT complex confirmed the flexibility of GN evident from the lack of GN signal in 2D class averages (Figure 4-7a). Again, I applied D2 symmetry at the 3D classification and 3D refinement stages in the aim of achieving a high-resolution structure of GS (Figure 4-7b). From 36,972 particles, we generated a 6 Å structure of human GS (Figure 4-7). When comparing this 6 Å structure of HsGS from GS-GN dataset to the HsGS model built from the GS-GN(Y195F) dataset, no differences are observed at this resolution (Figure 4-7b), suggesting that the GN mutation does not affect the conformation of the GS tetramer. This is consistent with the fact that GN(Y195F) and GN WT interact with similar affinity to GS[92].

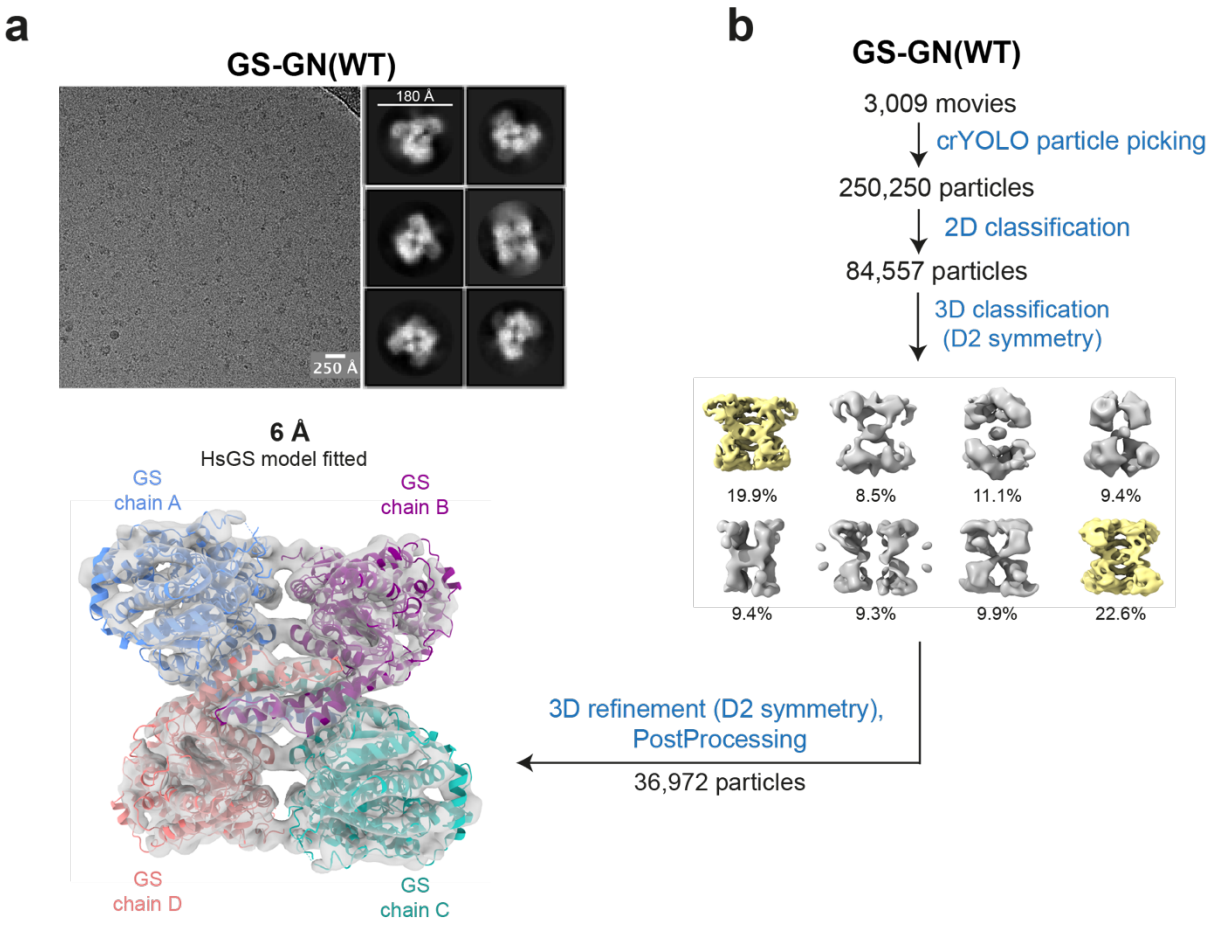

Figure 4-7: Cryo-EM analysis of GS-GN(WT)

a Representative micrograph and a selection of 2D class averages of the GS-GN(WT) dataset. **b** Flow chart of data processing. Classes coloured in yellow were taken forward for further processing. The GS-GN WT 6 Å map fits the human GS (HsGS) structure (reported here) with little differences, at this resolution.

4.5 Troubleshooting GN flexibility

From analyses thus far, it is clear that both GN dimers in the GS-GN complex are flexible. In the aim of identifying exactly how GS and GN interact, this section explores sample and grid preparation methods utilised in attempts to achieve a high-resolution GS-GN structure.

4.5.1 Chameleon

4.5.1.1 Overview

One new piece of equipment that has the aim of addressing issues caused by standard grid preparation methods is the Chameleon (SPT Labtech), which is a commercial version

of the Spotiton system[242,243]. The Spotiton system allows fast deposition of sample onto a nanowire grid[13-15]. This aims to decrease protein denaturation at the air-water interface and improve overall grid quality and reproducibility through automated grid preparation[202]. This system uses a dispensing tip to apply a stream of droplets onto a 'self-wicking' nanowire grid. The nanowires on the grid can be seen when screening the grids (Figure 4-8a). The fast speed of 10 ms that it takes for the sample to be deposited onto a grid and subsequently vitrified can reduce the effects on the sample that occur at the air-water interface[242]. Moreover, this spraying system is a potential method of increasing GN signal in cryo-EM if the issues arise from denaturation at the air-water-interface during 'standard' grid making procedures (see chapter 2.5.1).

4.5.1.2 Data processing

82,357 particles were picked using crYOLO[212], from the 1,395 movies that were acquired on a grid prepared by the Chameleon. 2D classification produced classes similar to previous datasets, showing a GS tetramer, however no obvious density can be seen for GN (Figure 4-8c). However, at this stage the 2D classes do not look as high quality as previous datasets, as the GS tetramer appears slightly blurry and no secondary structure can be seen. Following another round of 2D classification, 52,741 particles were subjected to 3D classification applying D2 symmetry. This resulted in one higher resolution class that was subsequently used for 3D refinement (Figure 4-8c). This generated a 10.2 Å map, showing a GS tetramer, which can fit the human GS cryo-EM structure. However, very little density can be seen for GN (Figure 4-8d), and the resolution at this stage is much lower than previous datasets. The high concentration of glycerol (10%) in this sample may be hindering the resolution as the undiluted sample was applied to the grid, as opposed to reducing this glycerol concentration when preparing grids using the Vitrobot. The undiluted sample was used due to the high concentration required when preparing grids using the Chameleon. Overall, using the spraying technique as opposed to traditional grid preparation methods did not solve the issue of GN flexibility in the GS-GN complex, and did not reach a higher resolution.

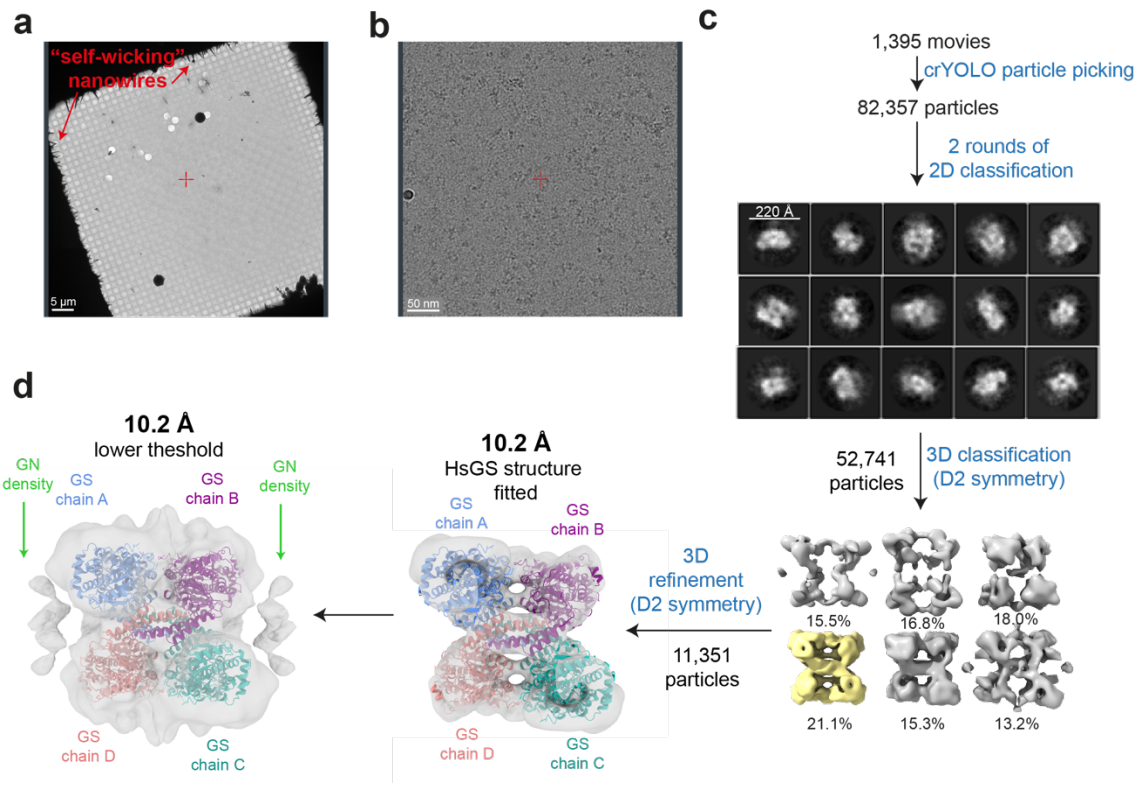

Figure 4-8: Treating GN flexibility using the Chameleon

a Image of a grid square where the "self-wicking" nanowires can be seen. **b** Representative micrograph. **c** Data processing flow chart. Final map at 10.2 Å with the human GS (HsGS) structure (reported here) rigid body fitted. The class highlighted in yellow was taken forward for 3D refinement to generate a 10.2 Å map. **d** The same 10.2 Å map show in **c** at a lower threshold. Very little density can be seen for GN in the 10.2 Å map even at a low threshold.

4.5.2 Glutaraldehyde crosslinking

Another method to reduce GN flexibility within the GS-GN complex is crosslinking the complex. There are various methods of crosslinking, one of which is using glutaraldehyde. Glutaraldehyde reacts with functional groups of proteins, including amine, thiol, imidazole and phenol[244]. Moreover, glutaraldehyde can react and crosslink proteins within a protein complex[245], and a number of successful examples using this technique for cryo-EM analysis have been published[246-248]. Albeit, this may lead to potential artefacts and may not represent the biologically relevant complex. Nonetheless, crosslinking is a potential way to reduce mobility in protein complexes, as well as preventing denaturation at the air water interface, and thus obtain structural information of problematic protein complexes.

4.5.2.1 Optimisation and cryo-EM analysis

Glutaraldehyde concentrations of 0.05% and 0.1% were used with 1.1 µM GS-GN(Y195F), and reactions were incubated for 10 or 30 min on ice or room temperature at first, based on literature at the time[246] (Figure 4-9a). Crosslinking was carried out in 9 µL reactions, and quenched by adding 4X SDS-PAGE loading dye for the purpose of optimisation, or by 500 mM Tris-HCl pH 7.5 if subsequently used and buffer exchanged into gel filtration buffer using a 0.5 mL Zeba desalting spin column. Running the samples on a Tris-Acetate 3-8% SDS-PAGE gel, two single bands can be seen in the absence of glutaraldehyde, which correspond to GS and GN(Y195F). When crosslinking reactions were performed at room temperature, the single bands for GS and GN(Y195F) disappear and a band larger than 460 kDa appears (calculated mass of 485 kDa for the GS-GN complex). However, this band cannot be seen when the crosslinking reaction is performed on ice (Figure 4-9a). In both the room temperature and ice reactions, smearing can be seen, showing that there are multiple different species in the samples. There is a slight disappearance of the single GS and GN(Y195F) bands in the ice reactions and a weak band ~117 kDa, however this is not the full GS-GN complex forming. Further optimisation, including increasing the glutaraldehyde concentration was carried out (Figure 4-9b). Again, smearing can be seen for the reactions carried out on ice. The room temperature reactions at 0.15% and 0.2% glutaraldehyde look very similar on a gel to the room temperature reactions at 0.1% (Figure 4-9b), however a higher glutaraldehyde concentration may result in non-specific crosslinking. Further optimisation involving altering the time the reaction was incubated for was performed (Figure 4-9c). This showed that using 0.1% glutaraldehyde for 20 or 30 min at 25 ºC gave very similar results on a gel. Therefore, the most optimal crosslinking reaction was determined to be 0.1% glutaraldehyde for 20 min at 25 ºC (Figure 4-9a), as this is when no un-crosslinked complex can be seen and the smearing of the crosslinking complex band is at a minimum, even though some protein that has not entered the gel suggests aggregation of the complex. These conditions were used to make cryo-EM grids.

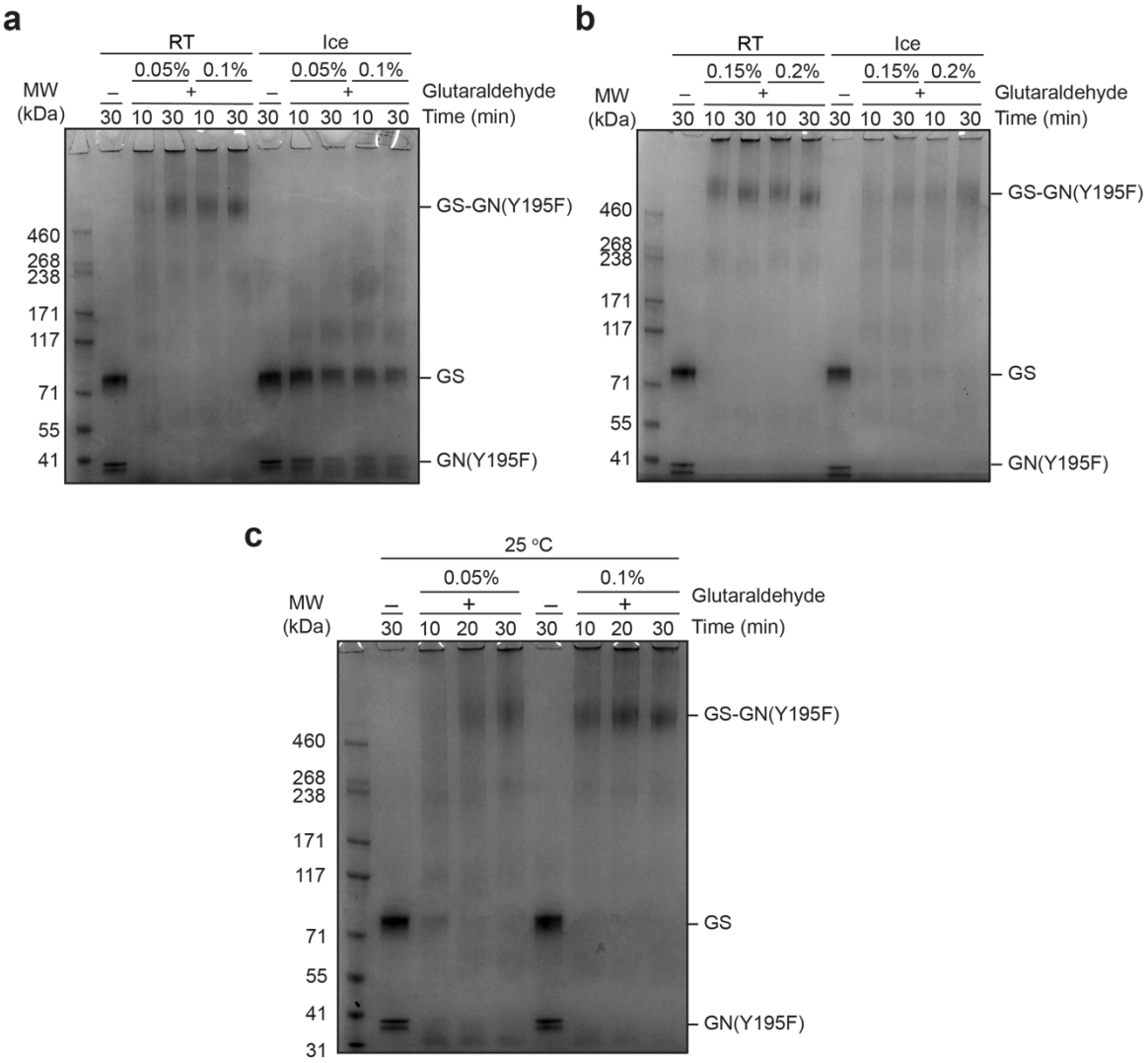

Figure 4-9: Glutaraldehyde crosslinking optimisation

a Crosslinking reactions of GS-GN(Y195F) were performed at 0.05% and 0.1% glutaraldehyde, for 10 or 30 minutes at room temperature or on ice. **b** The glutaraldehyde concentrations were increased in order to try remove the smearing seen in **a**. **c** Crosslinking reactions were performed at 0.05% and 0.1% glutaraldehyde, for 10, 20 or 30 minutes at room 25 ºC.

After the crosslinking reaction was completed with 0.1% glutaraldehyde for 20 minutes at 25 ºC, the sample was buffer exchanged as previously described (see section 2.15), and subsequently ran on a gel (Figure 4-10a). This resulted in a crosslinked sample at 0.3 mg/mL (2.5 µM), which resulted in visible particles on a grid (Figure 4-10b). Some aggregation can be seen; however, some single particles are also visible. Although there is some ice contamination, a cryo-EM dataset was collected.

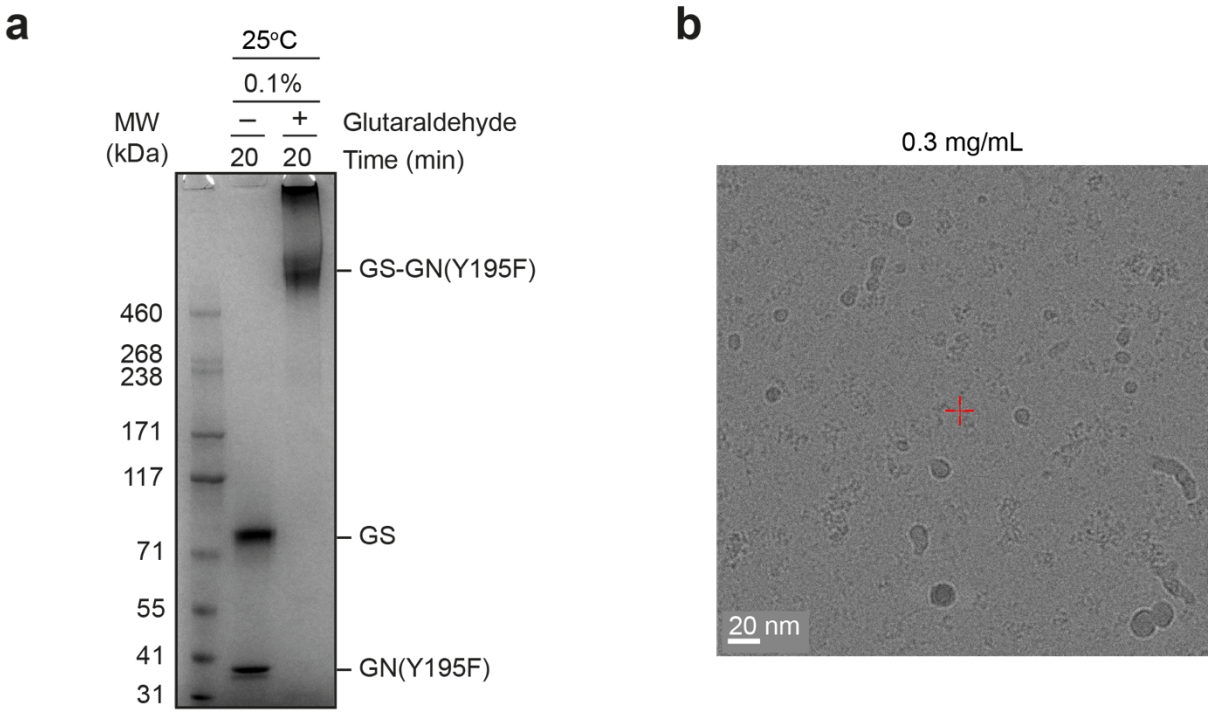

Figure 4-10: Crosslinking cryo-EM grid preparation and screening

a Crosslinking reaction using 0.1% glutaraldehyde for 20 minutes at 25 ºC were used to prepare cryo-EM grids. **b** Micrograph of the crosslinked GS-GN(Y195F) sample.

The nsEM final map for GS-GN(Y195F) (Figure 3-11c) was used to pick particles in the cryo-EM crosslinked dataset using the template picker tool in cryoSPARC[211] (see chapter 2.5.3.3). Unfortunately, 2D classification revealed poor quality classes that only showed the GS tetramer, similar to un-crosslinked 2D classes (Figure 4-11). Nonetheless, 88,011 particles were taken forward to generate *ab intio* models. However, representative of the 2D classes, some classes resembled the GS tetramer but no obvious density for GN can be seen (Figure 4-11). Due to the lack of GN density and the main aim of crosslinking being to obtain a GS-GN complex, data processing was terminated.

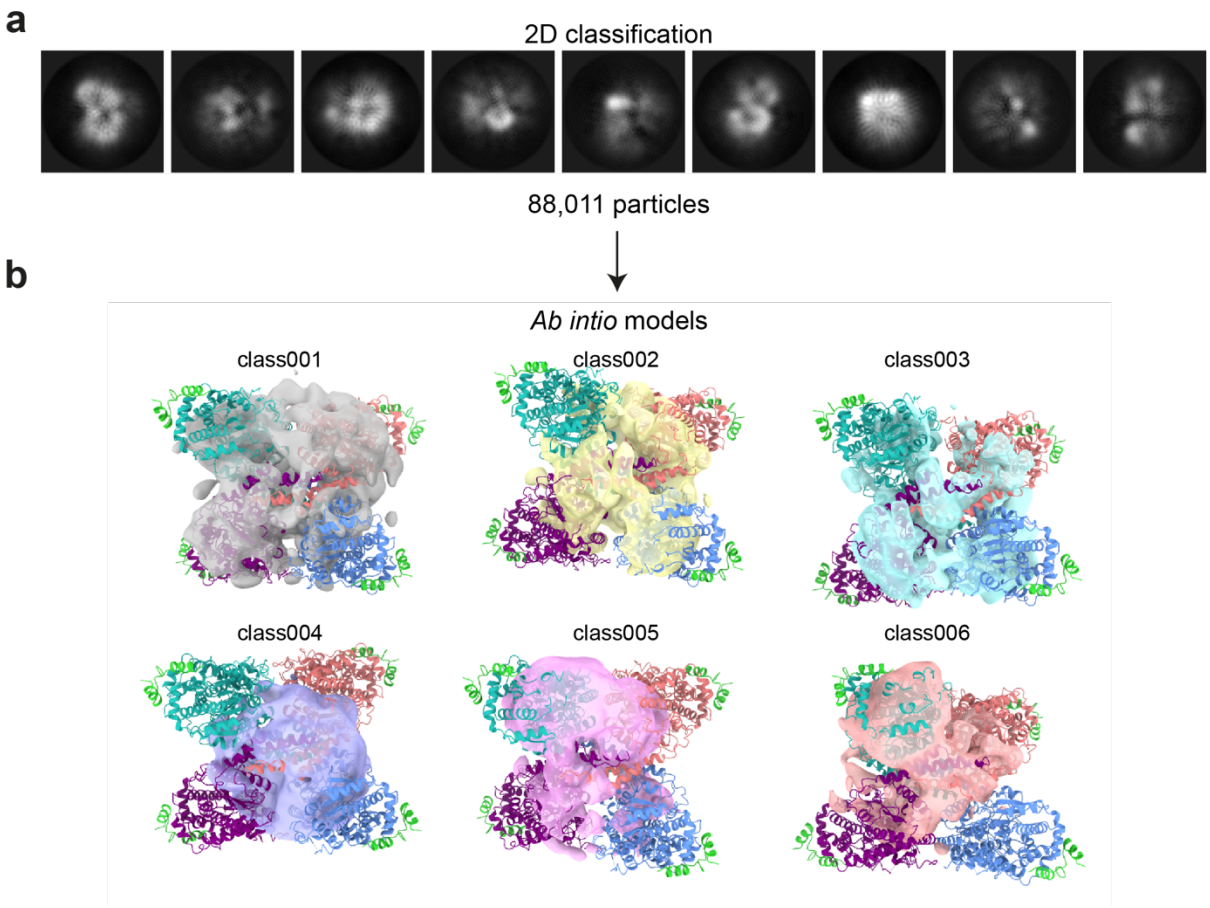

Figure 4-11: Cryo-EM analysis of crosslinked GS-GN(Y195F) complex

a 2D classification of crosslinked complex, however the classes resemble the un-crosslinked complex and only the GS tetramer can be seen. **b** *Ab initio* models from the classes shown in **a**. No density for GN can be seen, only weak density for the GS tetramer is present (human GS-GN[34] structure reported here is rigid body fitted into the maps).

4.6 Cryo-electron microscopy optimisation

4.6.1 Buffer optimisation

In order to reduce aggregation on cryo-EM grids, the GS-GN complex was purified in buffer containing 10% glycerol as this was shown to reduce aggregation during cryo-EM grid optimisation. Traditionally, it was advised that glycerol should be avoided in cryo-EM experiments as it greatly reduces contrast[249]. However, this has been debated more recently[250], and multiple examples have demonstrated that data collection in the presence of 20% (v/v) glycerol is suitable for cryo-EM analyses. To investigate whether the glycerol present in my previous datasets was hindering reaching higher than 4 Å resolution, we removed the glycerol in the GS-GN(Y195F) protein sample. To prevent aggregation of the complex we added 0.01% Tween-20, which was completed by buffer exchange using a 0.5 mL Zeba desalting spin column. Screening these cryo-EM grids showed a much better

contrast in comparison to grids with as little as 3% glycerol (Figure 4-12a and b). However, some protein aggregation can be seen when glycerol is removed even when Tween-20 is supplemented (Figure 4-12a). Nonetheless, we subsequently collected data on this grid in order to determine if glycerol is a limiting factor.

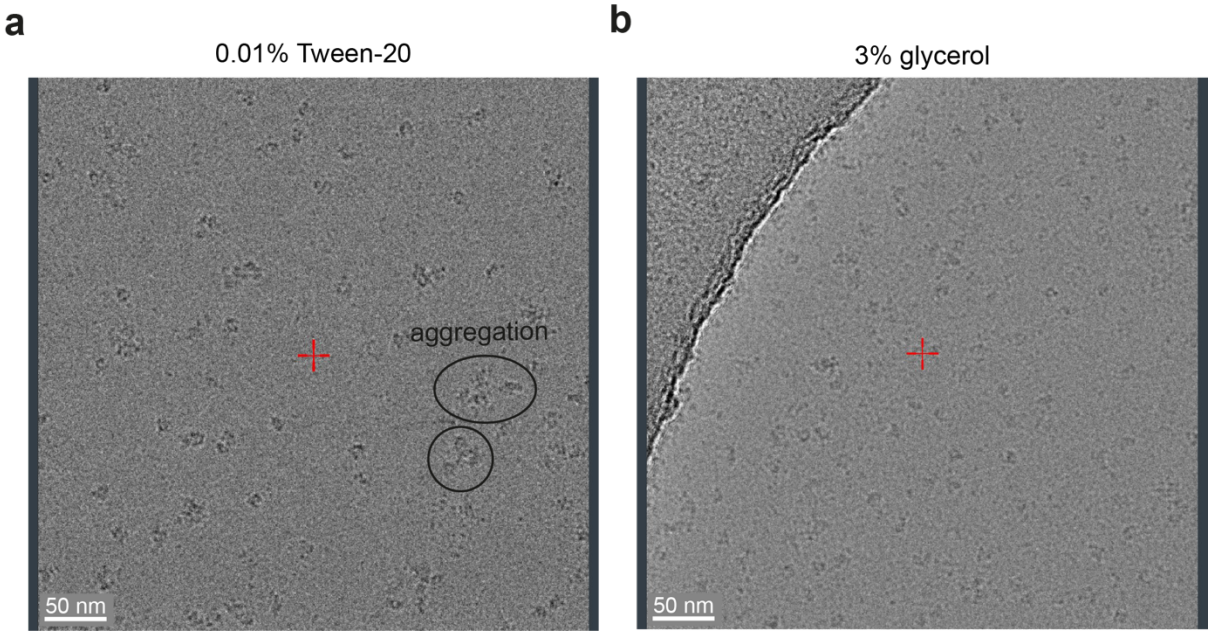

Figure 4-12: Cryo-EM buffer optimisation

a Grids prepared using buffer without glycerol and 0.01% Tween-20, much better contrast is observed in comparison to **b** containing 3% glycerol. However, some aggregation can be seen when glycerol is removed and 0.01% Tween-20 is supplemented.

4.6.2 Data collection optimisation

In addition to the buffer optimisation (discussed in 4.6.1), we also adjusted some data collection parameters in the aim of obtaining a higher resolution structure by increasing the pixel size and also reducing the electron dose. In previous data collections, a nominal magnification of 75,000 x was used, resulting in a pixel size of 1.065 Å or 1.07 Å. Also, previous data collections used an electron dose per physical pixel per second of 7.5 (see Table 2-2). However, in this new optimisation data collection we adjusted the nominal magnification to 165,000 x which resulted in an increased pixel size of 0.85 Å, and also reduced the dose per physical pixel per second to 5 (see Table 2-2).

6,071 movies were collected using a Titan Krios microscope equipped with a Gatan K2 detector, from which 157,676 particles were picked using crYOLO[212]. After 2D classification in cryoSPARC[211], 108,621 particles were subjected to 3D classification in RELION[210] (Figure 4-13b). From this 3D classification, three classes (coloured in yellow)

were taken forward for 3D refinement, PostProcessing and particle polishing, which resulted in a 4.2 Å reconstruction of GS (Figure 4-13c). This map is very similar to the previous 4 Å map obtained (Figure 4-1). To assess particle number *vs* particle quality, the highest resolution 3D class (boxed in pink) was taken forward for 3D refinement. However, this resulted in a 6.7 Å map which is lower resolution than the map generated when three classes were taken forward to increase particle number (Figure 4-13d). Another 3D class was taken forward alone for 3D refinement (boxed in green), which generated a low-resolution map of full-length GS-GN(Y195F) complex (Figure 4-13e).

Overall, both buffer optimisation by removing glycerol and replacing it with 0.01% Tween-20, as well as increasing the pixel size and reducing the electron damage during data collection, did not result in a better resolution than the previously obtained 4 Å map (Figure 4-1).

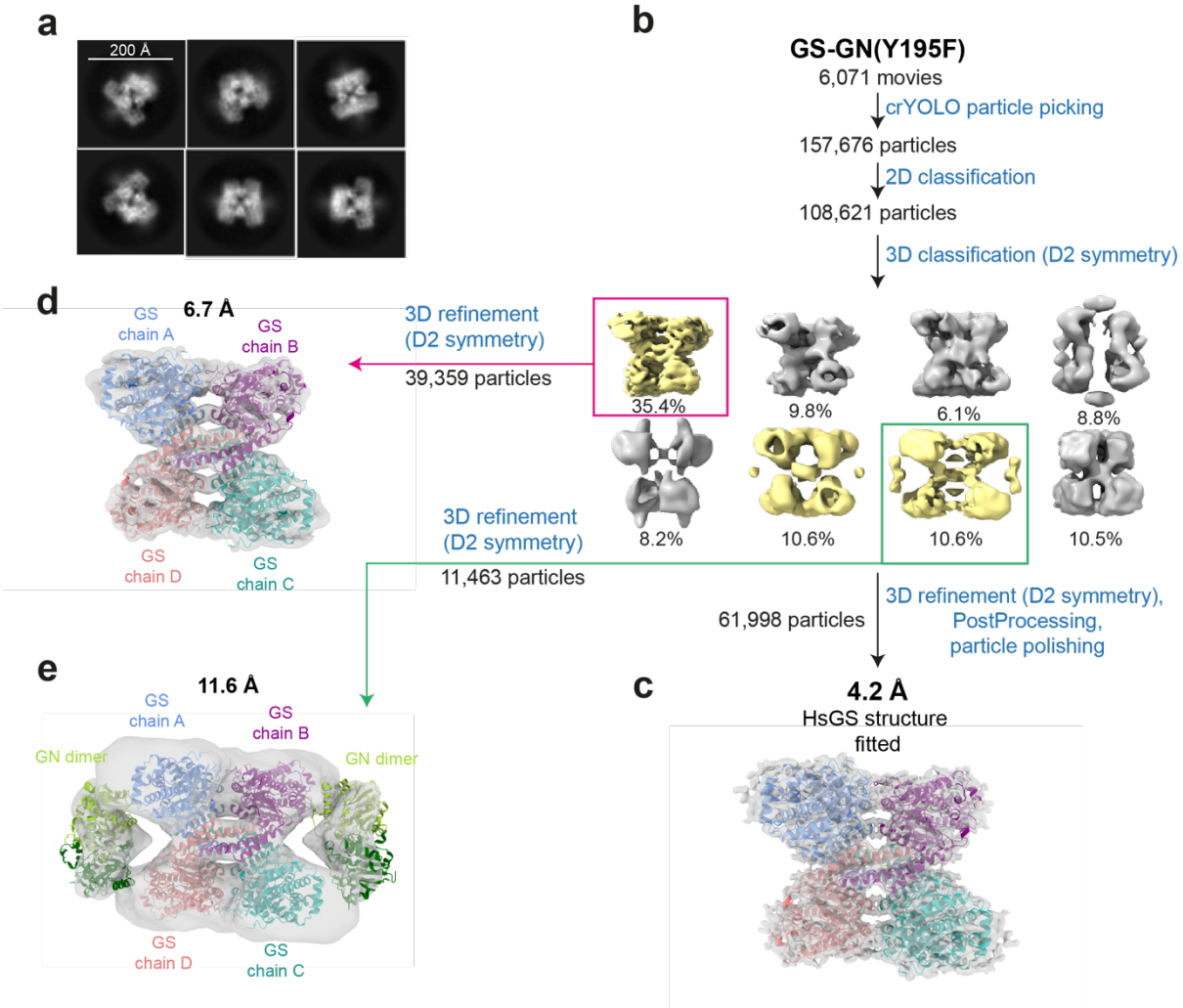

Figure 4-13: Cryo-EM optimisation by removing glycerol, increasing the pixel size and reducing electron damage

a 2D class averages. **b** Cryo-EM data processing. After the 3D classification, three classes (yellow) were taken forward to obtain a 4.2 Å structure of GS (**c**). To assess particle number *vs* particle quality, the highest resolution 3D class was taken forward for 3D refinement (pink), but the resolution was lower than the three yellow classes combined (**d**). **e** 3D refinement of the class boxed in green gave a low-resolution map of full-length GS-GN(Y195F). Human GS (reported here) and human GN (PDB ID 3T7O) structures are fitted.

4.7 Reducing sample heterogeneity

4.7.1 Allosteric activation

GS is regulated by both inhibition by phosphorylation and activation by G6P[1]. Structural and mechanistic studies of yeast GS have sophisticatedly dissected its allosteric activation by G6P, whereby G6P binding induces large translational and rotational movements[88,91]. However, it is not known if this mechanism of activation is shared with human GS. Moreover, in the aim of elucidating the mechanism of activation, nsEM and cryo-EM data was collected with the addition of G6P. Another expected benefit of adding G6P in the cryo-EM grids was to reduce heterogeneity. As purified, GS is phosphorylated at multiple sites, resulting in a heterogenous sample. However, the binding of G6P is known to override this phosphorylation[5], also potentially inducing a conformational change as seen in crystal structures of GS from yeast[88,91], thus resulting in a more homogenous sample.

NsEM analysis of GS-GN(Y195F) in the presence of 25-fold molar excess of G6P (0.14 μM GS-GN(Y195F) and 3.5 μM G6P) revealed 2D classes that resemble the classes seen for inhibited GS-GN(Y195F) (Figure 4-14). However, 3D reconstruction reveals a map that, at this resolution, seems to be in the active state, as the yGS+G6P crystal structure fits slightly better than the HsGS inhibited structure, although it is difficult to make conclusions at this resolution (Figure 4-14c and d). Moreover, cryo-EM data was collected with GS-GN(Y195F) plus 25-fold excess G6P (1.8 μM GS-GN(Y195F) and 46 μM G6P) (Figure 4-15).

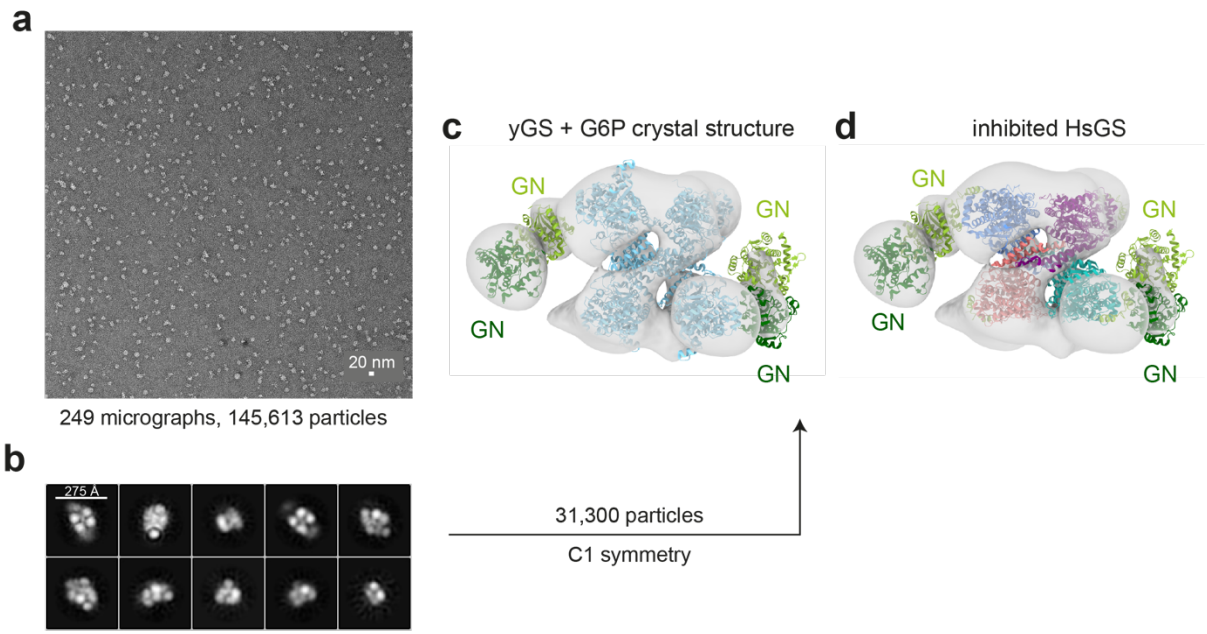

Figure 4-14: Negative stain EM analysis of GS-GN(Y195F) + G6P

a Representative negative stain micrograph. **b** 2D class averages showing a 4:4 stoichiometry of the full-length GS-GN(Y195F) complex. **c** Final 3D map with the yeast GS + glucose-6-phosphate (G6P) crystal structure (PDB ID 5SUK) and human GN (PDB ID 3T7O) fitted. Density for a GS tetramer with two GN dimers one on either side is present, and the activated yeast GS (yGS) structure fits the density slightly better than the inhibited human GS (HsGS) rigid body fitted into the same map (shown in **d**).

2,180 movies were collected using a Titan Krios microscope equipped with a Gatan K2 detector, from which 94,270 particles were picked and used for 2D classification. 57,708 particles were selected after 2D classification and subsequently used in a 3D classification. During the 3D classification, two different conformations of GS can be seen (which are highlighted in yellow) which resemble the inhibited human GS state previously seen (Figure 4-1) and also the activated state of GS seen by the yeast crystal structure (Figure 4-15b). Taking the activated state forward for 3D refinement (applying D2 symmetry) generated a map at 13 Å (Figure 4-15b). Fitting the crystal structure of yGS + G6P (PDB ID 5SUK) fits the map better than the inhibited cryo-EM structure. However, it was a small percentage of the particles that are in this conformation (11%), owing the low resolution to limited particle number (Figure 4-15b). The 3D class with the most particles was taken forward for further processing, generating a 4.1 Å map of inhibited human GS (Figure 4-15c). Moreover, this dataset with 25-fold excess G6P did not result in a high-resolution structure of activated human GS. Albeit, this dataset generated a high resolution inhibited, phosphorylated HsGS structure, which is very similar to the map previously described

(Figure 4-1). Perhaps 25-fold molar excess of G6P is not sufficient to saturate GS, and thus form a homogenous sample.

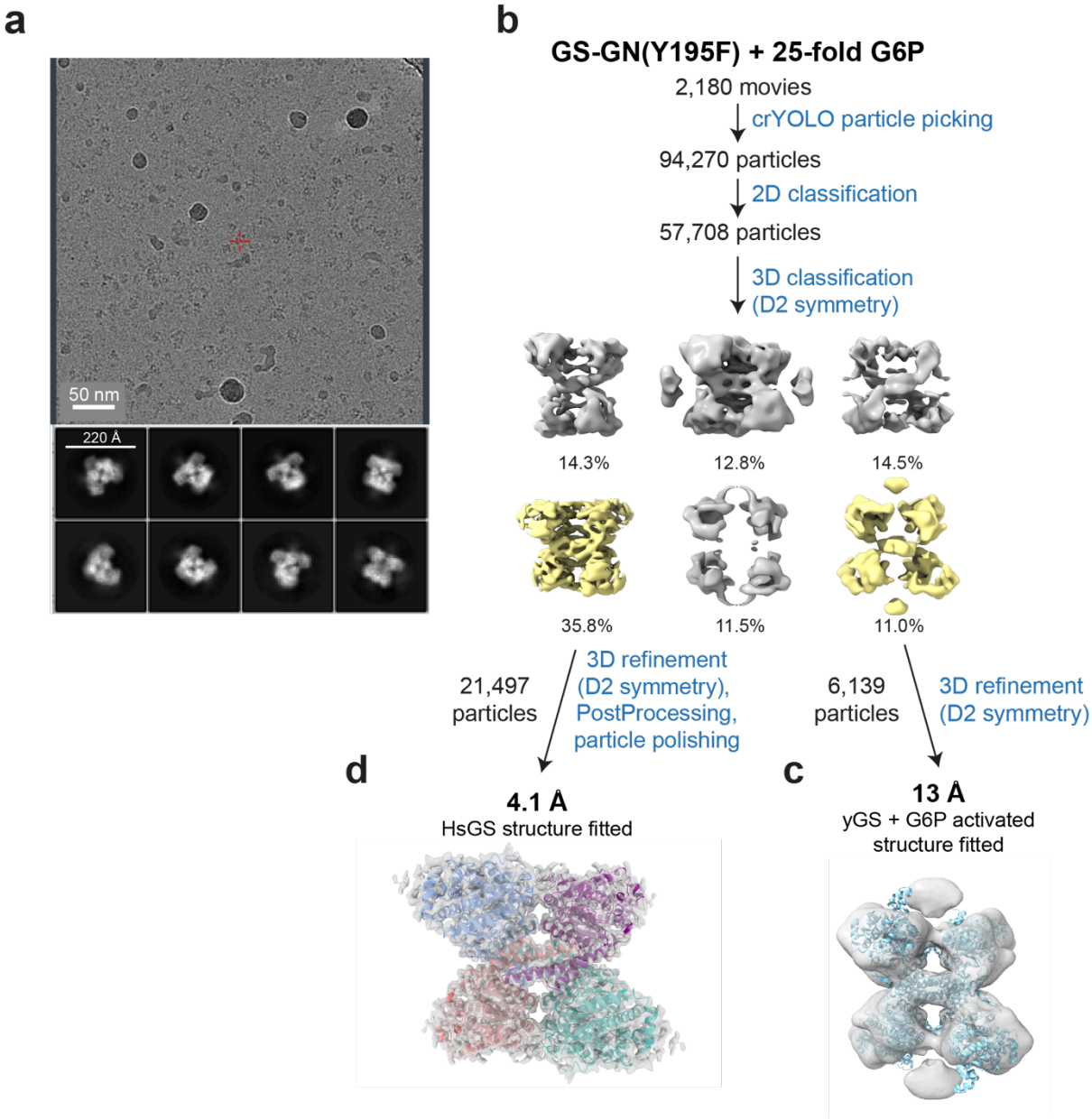

Figure 4-15: Cryo-EM analysis of GS-GN(Y195F) + 25-fold molar excess G6P

a Representative micrograph and 2D class averages. **b** Flow chart of data processing. Each class coloured in yellow was taken forward separately. **c** 13 Å map of human GS, with the activated yeast GS (yGS) crystal structure (PDB ID 5SUK) fitted. This map is a different conformation to the human GS (HsGS) inhibited structure. **d** 4.1 Å map of human with the inhibited state structure fitted (reported here).

In light of the results of the 25-fold G6P dataset, we set up cryo-EM grids with 100-fold molar excess G6P (1.5 μM GS-GN(Y195F) and 150 μM G6P) (Figure 4-16). 100-fold molar

excess was chosen with the aim of keeping the G6P concentration as low as possible in order to prevent high concentrations of G6P effecting ice quality. 2,223 movies were collected using a Titan Krios microscope equipped with a Falcon III detector, from which 231,003 particles were picked and used for 2D classification. 85,102 particles were selected after 2D classification and used for 3D classification. Similar to the +25-fold G6P dataset (Figure 4-15), two conformations can be seen (coloured in yellow). The majority of the particles are in the inhibited, phosphorylated conformation, as seen by the map generated by 3D refinement (Figure 4-16d). 20.8% of the particles are in the G6P bound, activated conformation, as seen by fitting the yGS + G6P crystal structure into the density (Figure 4-16c). In comparison to the +25-fold dataset, this +100-fold dataset contains more particles that are in the activated state. However, this number is still relatively low at 17,619 particles (20.8% of the total particles), and thus has not resulted in a high-resolution structure. The addition of 100-fold excess G6P to the GS-GN(Y195F) has not resulted in saturation of the enzyme, suggesting that G6P has a low affinity to phosphorylated GS.

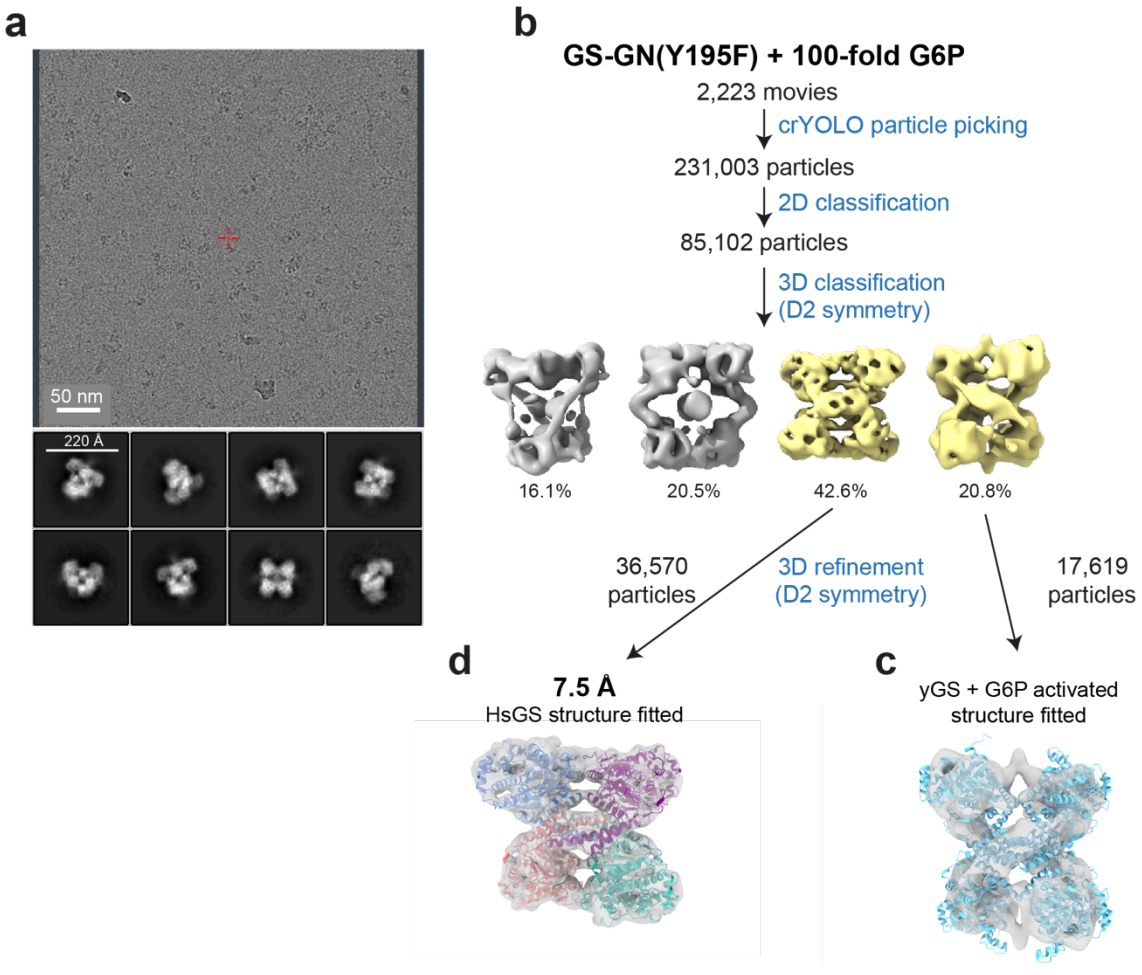

Figure 4-16: Cryo-EM analysis of GS-GN(Y195F) + 100-fold molar excess G6P
a Representative micrograph and 2D class averages. **b** Flow chart of data processing. Each class coloured in yellow was taken forward separately. **c** 3D class map with the activated yeast GS (yGS) crystal structure (PDB ID 5SUK) fitted. This map is a different conformation to the human GS (HsGS) inhibited structure. **d** 7.5 Å map of HsGS with the inhibited state structure fitted (reported here).

Perhaps the reason for the G6P activated state of GS containing a small percentage of the particles is that the phosphorylation state of GS is having an impact on the affinity for G6P. Microscale thermophoresis (MST) was performed to investigate the affinity of G6P to the GS-GN complex, which determined a K_d of 12.3 ± 1.9 mM (Figure 4-17). This low affinity indicates that the amount of G6P added during the cryo-EM analyses did not saturate GS, and describes why only a small percentage of the particles were in the activated conformation (Figure 4-16). Perhaps this heavily phosphorylated enzyme preparation is the reason for the high K_d, and it is still unclear if the dephosphorylated state of the GS-GN complex binds G6P more easily.

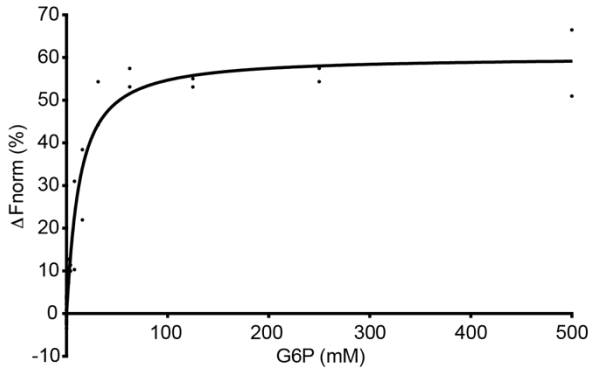

Figure 4-17: G6P binding affinity to GS-GN(Y195F)
Microscale thermophoresis (MST) analysis of glucose-6-phoshate (G6P) binding to GS-GN(Y195F). Data shows mean from two independent experiments carried out in technical duplicates, where the Hill plot was used to generate a K_d value of 12.3 ± 1.9 mM.

4.7.2 Dephosphorylation by lambda protein phosphatase and protein phosphatase 1

The phosphorylation of GS at multiple sites results in a heterogenous sample. One method to reduce this heterogeneity is to dephosphorylate GS-GN(Y195F). Lambda PP and PP1

can dephosphorylate the GS-GN complex, as observed by reduction in signal for antibodies for specific phosphorylation sites and faster migration of the dephosphorylated sample on an SDS-PAGE gel (Figure 3-12).

The creation of dephosphorylated GS-GN(Y195F) could potentially produce a more homogenous sample, which is ideal for cryo-EM studies. Structural studies of the dephosphorylated enzyme will also further our knowledge of glycogen synthase regulation, as well as allowing investigation into how phosphorylation is affecting G6P binding. It may also reduce the concentration of G6P required to saturate GS, which would be more suitable for cryo-EM analyses. With these aims in mind, I set out to optimise the preparation of dephosphorylated GS-GN(Y195F) complex.

4.7.2.1 Expression and purification of lambda protein phosphatase

The lambda PP construct, which has an N-terminal GST tag and a C-terminal His tag, was purchased from MRC PPU reagents and services. The expression of lambda PP was successful, and it was purified using nickel affinity chromatography and size exclusion chromatography (see section 2.2.5 for purification details) (Figure 4-18a and b). This lambda PP was active, as a faster migration can be seen for the GS band on an SDS-PAGE gel, owing to a reduction in molecular weight through dephosphorylation (Figure 4-18c).

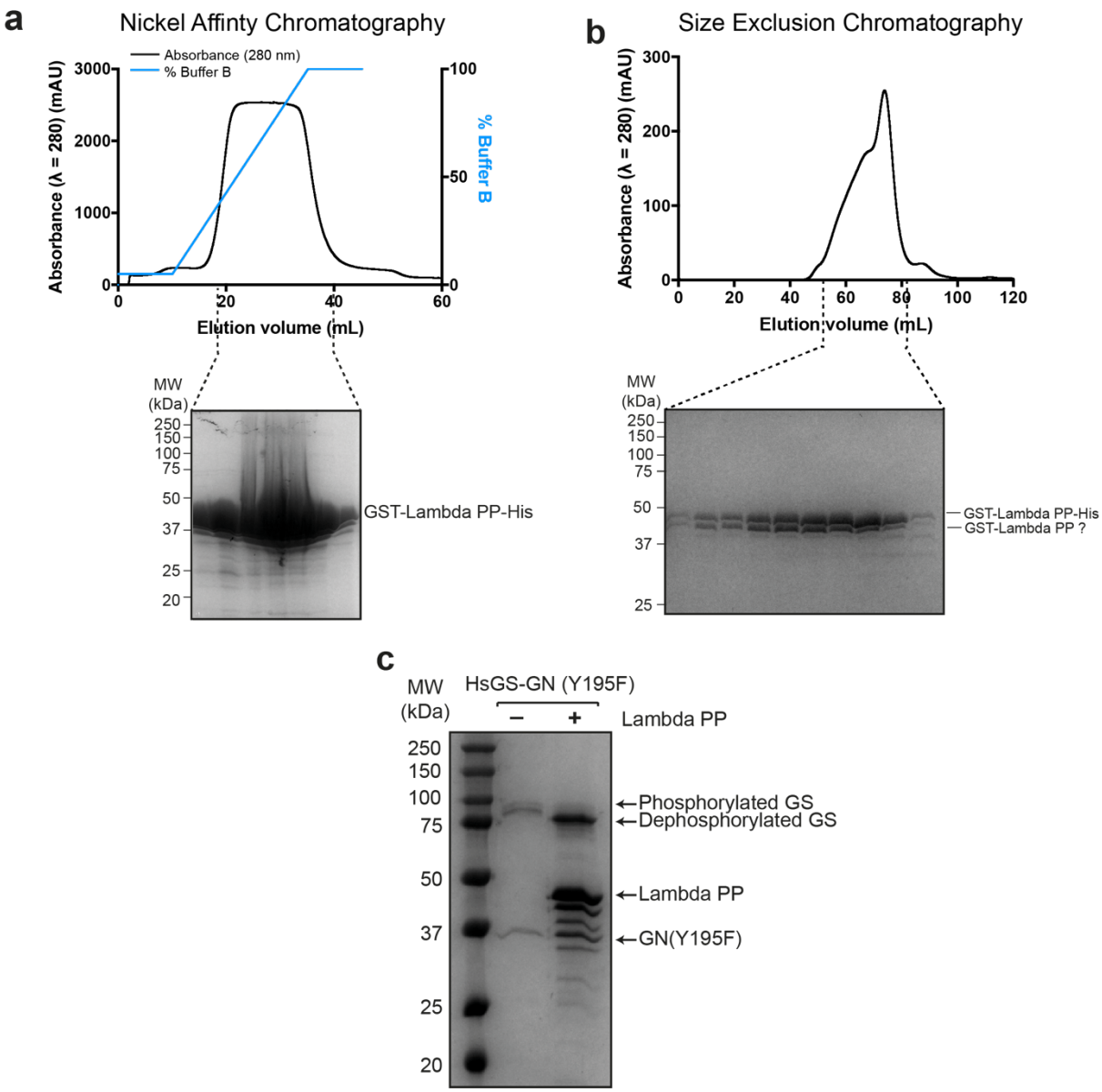

Figure 4-18: Purification of lambda protein phosphatase

a Nickel affinity chromatography trace and SDS-PAGE gel of peak fractions. **b** Size exclusion chromatography trace and SDS-PAGE gel of peak fractions. **c** Purified lambda phosphatase is active and can dephosphorylate GS.

4.7.3 Dephosphorylation optimisation

After obtaining active lambda PP and PP1 (purified PP1 protein was purchased from MRC PPU Reagents and Services), the dephosphorylation reaction requires optimisation. Initially there are two possibilities: dephosphorylation during protein purification or dephosphorylation of purified protein. We explored dephosphorylation during protein purification by adding both lambda PP and PP1 during the overnight dialysis stage after nickel affinity chromatography. Lambda PP and PP1 were added to the protein with a final

concentration of 1 mM $MnCl_2$ in the buffer, as well as TEV protease, in a 1:100 ratio (1 mg protease/phosphatase:100 mg GS-GN(Y195F)). However, following overnight dialysis a white precipitate was observed upon visual inspection, which is indicative of precipitated protein. Following centrifugation to separate this precipitate and running samples of the supernatant and precipitated protein on an SDS-PAGE gel, it seemed that it was the dephosphorylated form of GS that has precipitated. As only the lower molecular weight band for GS was present in the precipitate sample, as opposed to the doublet band routinely seen for phosphorylated GS, which can be seen in the supernatant sample (Figure 4-19b), this suggested that dephosphorylated GS was unstable, in the current buffer, at 4 ºC overnight. Nonetheless, the purification was continued with size exclusion chromatography. SDS-PAGE analysis shows GS-GN(Y195F) complex in the void volume and peak 2, and excess GN(Y195F) in peak 3 (Figure 4-19c).

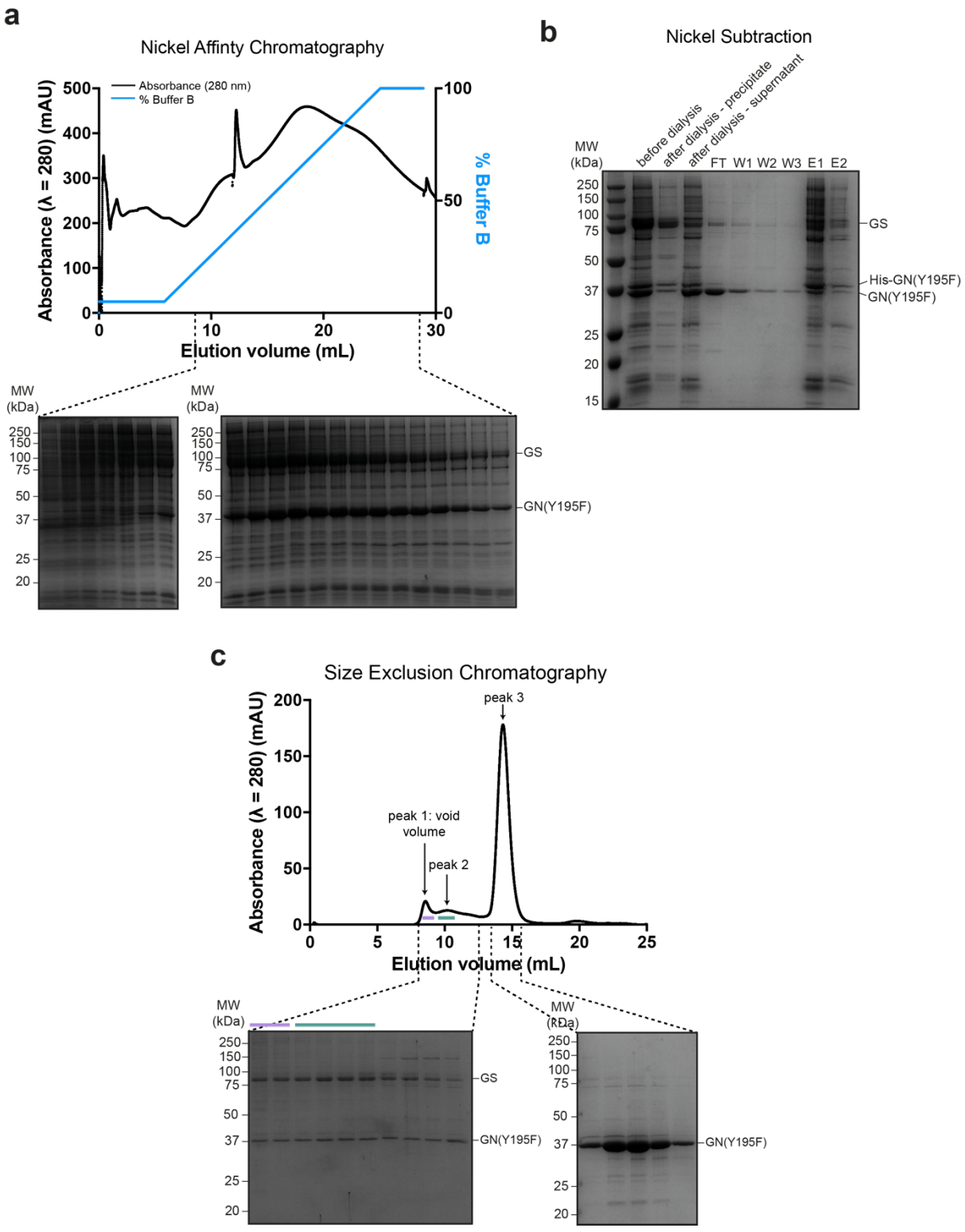

Figure 4-19: Dephosphorylation of the GS-GN complex

a Nickel affinity chromatography trace and SDS-PAGE gel of peak fractions. **b** Nickel subtraction step to remove His-tagged TEV protease. After dephosphorylation during overnight dialysis, the GS-GN(Y195F) precipitated, and the majority of the precipitate contained dephosphorylated GS. **c** Size exclusion chromatography trace and SDS-PAGE gel of peak fractions.

Protein from peak 2 was concentrated and used to make nsEM grids. A negative stain dataset was collected, however the 2D class averages generated do not resemble the GS-GN(Y195F) complex usually attained (Figure 4-20a, b and c). Initial models generated from these 2D classes do not fit the structure of the GS-GN complex well, and thus data processing was terminated (Figure 4-20d).

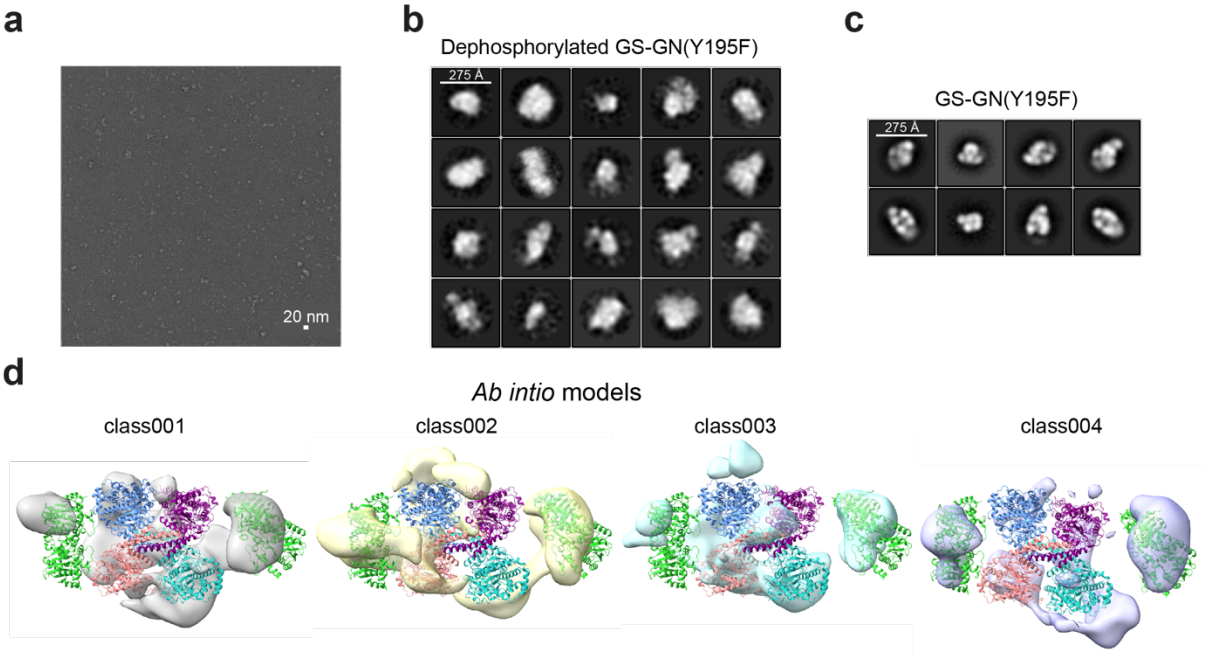

Figure 4-20: Negative stain EM analysis of dephosphorylated GS-GN(Y195F)
a Representative micrograph. **b** 2D class averages, which do not resemble the 2D classes usually seen for the phosphorylated GS-GN complex, shown in **c**. **d** *Ab initio* models of dephosphorylated GS-GN(Y195F). The human GS structure (reported here) and human GN dimer (PDB ID 3T7O) is fitted.

Due to the little success of dephosphorylating the GS-GN complex during protein purification, the method of dephosphorylation post purification was explored. Firstly, different molar ratios of GS-GN(Y195F) to lambda PP and PP1 were tested, at room temperature and 4 ºC both for 2 hours (Figure 4-21a). This showed that the most migration to a lower molecular weight species was achieved with 10-fold less phosphatases than GS-GN(Y195F) complex at room temperature, therefore, these were the conditions subsequently used. Firstly, this optimised dephosphorylation reaction was performed, and the phosphatases were removed by incubating the reaction with glutathione Sepharose resin and subsequently removing the resin using a spin-X column to elute the dephosphorylated complex (E1) (Figure 4-21b). Dephosphorylated GS-GN(Y195F) was applied to cryo-EM grids, however the protein was aggregated (Figure 4-21c). This

assessment of dephosphorylation of purified protein agrees with the first method of dephosphorylation during protein purification, as the aggregation observed on a cryo-EM grid as well as the protein precipitation suggests that dephosphorylated GS is unstable.

To assess and confirm the destabilisation effect of GS dephosphorylation that is suggested by aggregation on a cryo-EM grid and precipitation during protein purification, thermostability analysis was performed. Differential scanning fluorimetry (DSF) was used which revealed GS-GN(Y195F) as purified (i.e. phosphorylated) had a melting temperature (T_m) of 52 ºC. However, dephosphorylation reduces the T_m to 48 ºC. This 4 ºC change in T_m confirms that the dephosphorylated state of GS is less stable than the phosphorylated state (Figure 4-21d), thus suggesting a role for phosphorylation of the “spike” regions in strengthening the inter-subunit interactions within the tetramer and holding the enzyme in the “tense” conformation.

Upon addition of 12.5 mM G6P, the phosphorylated state of GS-GN(Y195F) is stabilised to 55 ºC, with a ΔT_m of 3 ºC. G6P can also stabilise the dephosphorylated state of GS-GN(Y195F) to 57 ºC, with a ΔT_m of 9 ºC in comparison to the no G6P sample (Figure 4-21d). Furthermore, in an attempt to aid stabilisation of the dephosphorylated complex, we performed the dephosphorylation reaction in the presence of G6P (Figure 4-21e). However, this did not abolish aggregation of the dephosphorylated sample on a cryo-EM grid (Figure 4-21f). This is surprising, as thermostability analysis shows that dephosphorylated GS + G6P is slightly more stable than phosphorylated GS with and without G6P, neither of which appear aggregated on a cryo-EM grid (Figure 4-2, Figure 4-15 and Figure 4-16). However, perhaps the dephosphorylation has resulted in the complex being more sensitive to the air-water interface, thus causing aggregation. In summary, optimisation of a dephosphorylated GS-GN(Y195F) sample that is suitable for cryo-EM analysis needs further investigation.

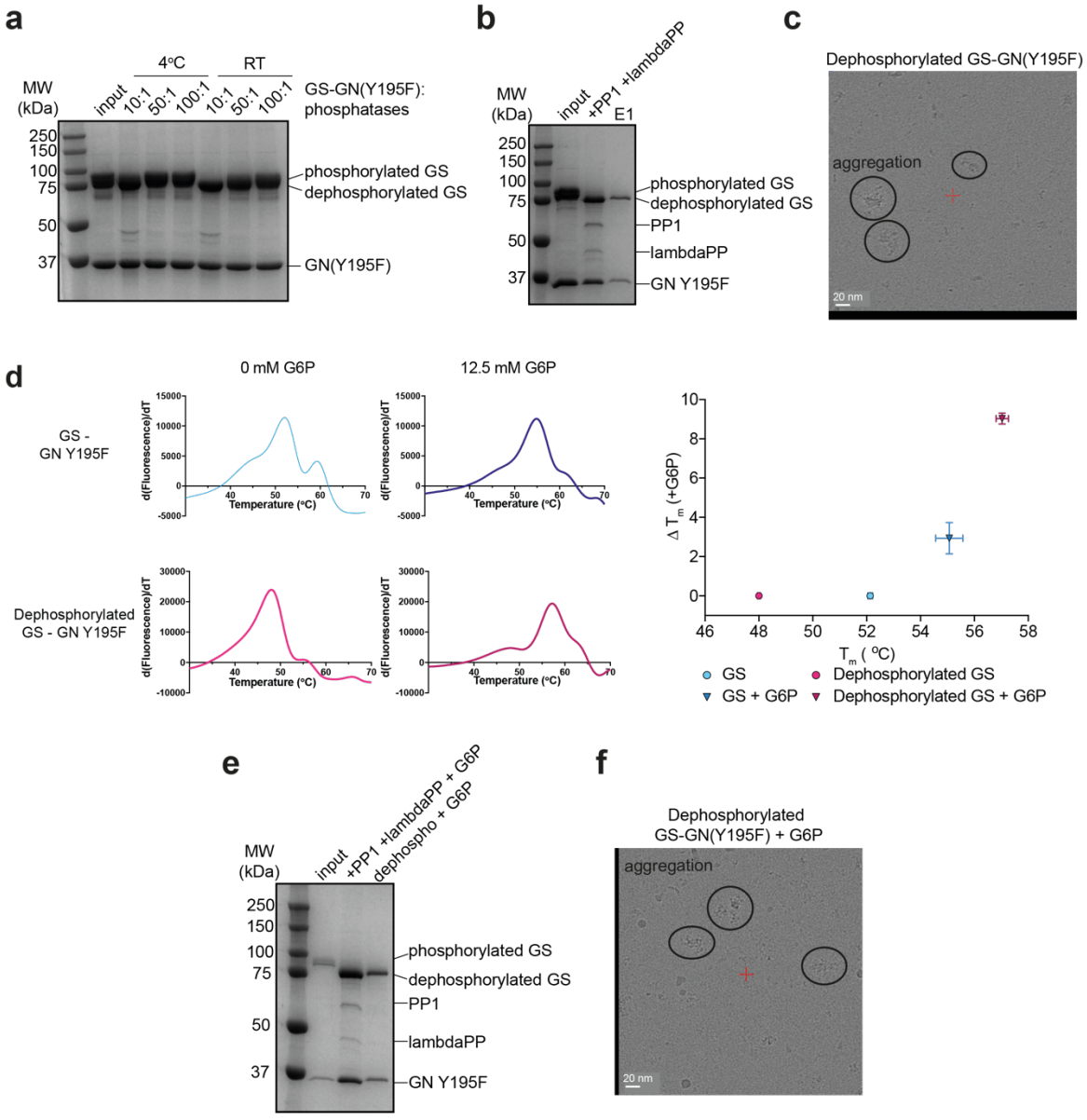

Figure 4-21: Dephosphorylation destabilises the GS-GN(Y195F) complex

a SDS-PAGE gel showing dephosphorylation via migration at different ratios of GS-GN(Y195F): phosphatases (lambda PP and PP1). The most migration is seen with 10-fold less phosphatases than GS-GN(Y195F). **b** Dephosphorylation and subsequent removal of phosphatases, resulting in a pure dephosphorylated sample (E1). **c** The dephosphorylated complex appears aggregated on a cryo-EM grid. **d** Melting curves from thermostability analysis of phosphorylated and dephosphorylated GS-GN(Y195F) (left) and corresponding melting temperatures (right). Data are mean +/- S.E.M from three independent experiments carried out in technical duplicates (dephosphorylated GS) and triplicates (GS). **e** Dephosphorylation of the GS-GN(Y195F) complex in the presence of glucose-6-phoshate (G6P). **f** Even with the addition of G6P, dephosphorylated GS-GN(Y195F) appears aggregated on a cryo-EM grid.

4.8 Chapter summary

At the start of this project, the structure of human GS was unknown and the mechanism of inactivation remained elusive. This chapter reveals the first high-resolution structure of human GS, at a global resolution of 4 Å. Two interdomain subunits, termed the “spike” regions, which house the GS N- and C-termini that contains phosphorylation sites appear to constrain GS in an inactive conformation. Cryo-EM data processing methods to treat heterogeneity were performed in an attempt to increase the resolution, however, despite these attempts the resolution was limited to 4 Å. Following from this, sample preparation techniques including an alternative cryo-EM grid making method, protein crosslinking and buffer optimisation were used to overcome GN flexibility and the limiting resolution. However, these were unsuccessful in obtaining higher resolution. Exploring GS regulation through G6P activation and dephosphorylation did not lead to a high-resolution structure, but revealed insights into the affinity of G6P to the phosphorylated state of GS and the destabilising effect of dephosphorylation. Future work may establish the mechanism of human GS activation and how phosphorylation seemingly stabilises GS.

5 Chapter 5: High resolution structure of inhibited GS in complex with GN

5.1 Introduction

The 4 Å cryo-EM structure of human GS described in chapter 4 revealed some insights into the inhibited state of GS, however the map lacked precise details of this mechanism and cryo-EM density corresponding to residues of the phospho-regulated tails was limited to mainchain atoms. Therefore, the aim of this work was to further investigate the detailed mechanism in which phosphorylation at the N- and C-terminal tails causes inhibition, through obtaining a higher resolution structure of the enzyme. Building an atomic model of human GS using a higher resolution map would hopefully allow investigation into the relationship between allosteric activation and covalent inhibition.

5.2 Overall cryo-EM structure of human GS

Following on from the 4 Å cryo-EM structure human GS (see section 4.2), we set out to reach higher resolution in the aim of gaining more insight into the mechanism of inactivation. Upgrades in the EM facility provided an opportunity to push the current 4 Å resolution further using new detectors. The Falcon 4 detector with an energy filter, in comparison to Falcon III and K2, offers a performance boost in various ways. This new detector increases the signal to noise ratio as well as allowing more images to be collected per hour, both of which allow for higher resolution reconstructions[251]. In addition, new data format electron-event representation (EER) offers benefits. In comparison to the existing MRC format in which camera frames are summed and represented as images, the EER format saves the time and location of recording of each detected electron event, thus preserving the full temporal and spatial resolution of data, allowing 3D reconstructions to reach higher resolution[207].

The collection of a new dataset, using a Titan Krios microscope equipped with a Falcon 4 detector with a 10 eV energy filter, generated a structure at a global resolution of 2.6 Å using D2 symmetry (Figure 5-1 and Figure 5-2). Consistent with the 4 Å structure, we see tetrameric arrangement of GS in complex with the C-terminal GS interacting region of GN. The oligomeric state of human GS is in agreement with the crystal structures of the yeast GS and *C. elegans* GS enzymes, with root mean square deviation (RMSD) values of 0.9 Å (between 522 Cα atom pairs) and 1.1 Å (between 484 Cα atom pairs) respectively

(Figure 5-3a and b). Consistent with the previous 4 Å cryo-EM structure human GS, we also see the inter-subunit domain that houses the phosphorylation sites (Figure 5-2e).

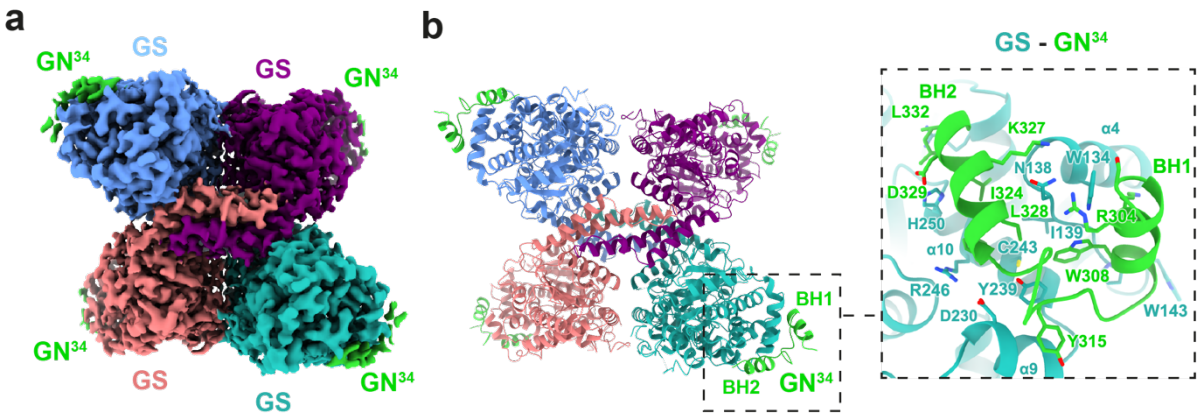

Figure 5-1: Tetrameric human GS in complex with human GN^{34}

a 2.6 Å cryo-EM map of the GS tetramer coloured by corresponding chain. Density corresponding to the GN^{34} C-terminal region is shown in green. **b** Human GS-GN^{34} cartoon model shown in ribbons coloured by corresponding chain (left). Interaction between GS and GN^{34} (right).

This higher resolution structure now allows for *de novo* model building of residues 300-332 of the GN C-terminus (GN^{34}). Four GN peptides bind to the GS tetramer, and these residues form a helix-turn helix, where the first helix is denoted BH1 and the second as BH2 (Figure 5-1b). This structural organisation is similar to the *C. elegans* GS-GN^{34} crystal structure[92], with an RMSD value of 0.8 Å (between 30 Cα atom pairs) (Figure 5-3c). The binding of this helix-turn-helix is mediated by combination of hydrophobic and hydrogen bonding interactions involving α4, α9 and α10 from human GS, and is consistent with the interactions observed for GS-GN^{34} from *C. elegans*[92] (Figure 5-3c).

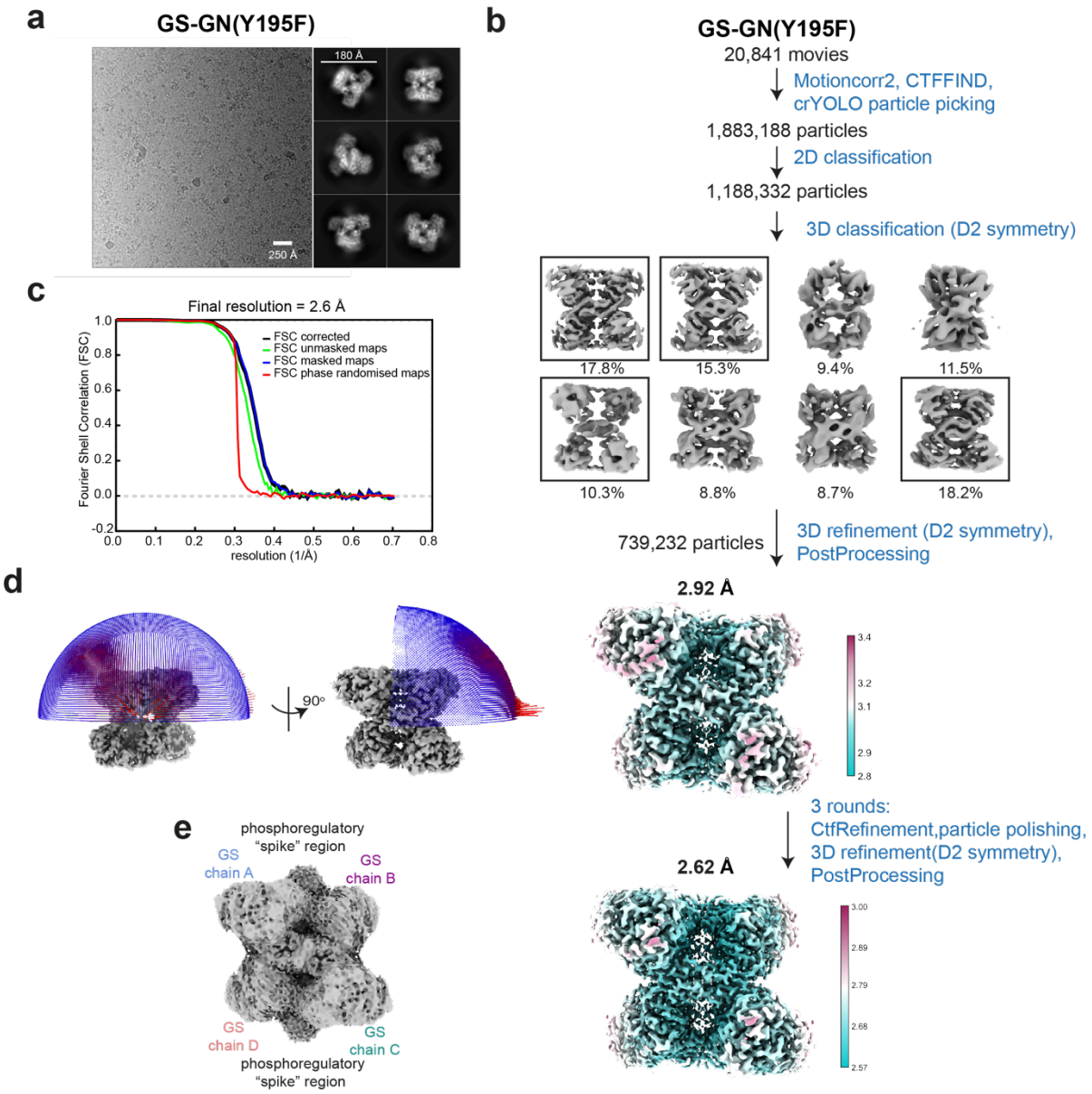

Figure 5-2: Cryo-EM analysis of GS-GN(Y195F)

a Representative micrograph and 2D class averages. **b** Flow chart of data processing. 3D classes boxed were taken forward for subsequent processing. **c** FSC curve for the final map. **d** Euler angle distribution of particles used for the 3D reconstruction. Height of rods represents number of particles, and red indicates a high number of particles and blue indicates a low number of particles. **e** A 3D refinement map shown at a lower threshold to visualise the phosphoregulatory "spike" region.

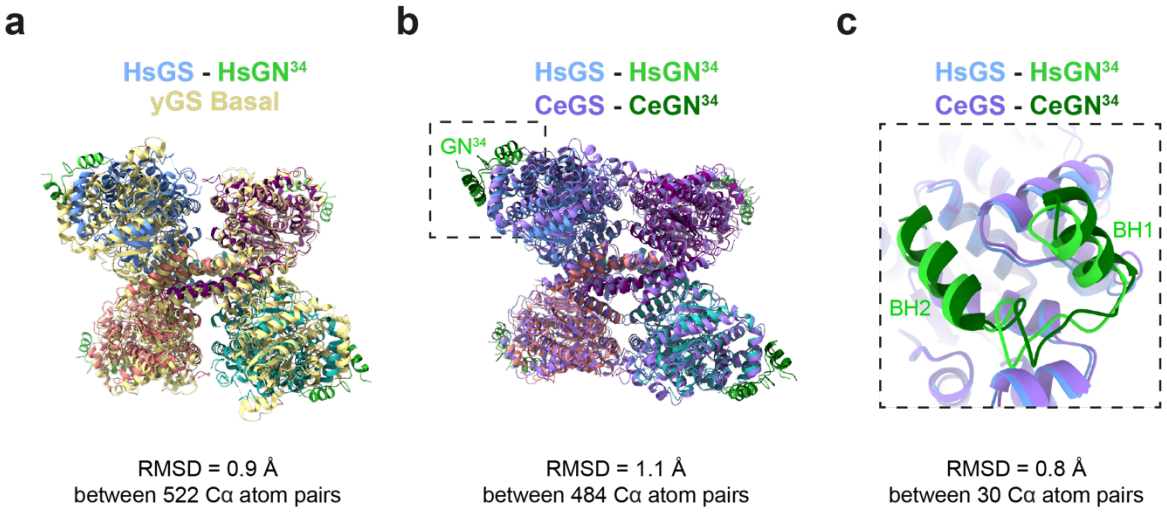

Figure 5-3: Comparison of human GS to previous crystal structures

a Human GS (HsGS) has a tetrameric arrangement, agreeing with crystal structures of yeast GS (yGS) (yellow, PDB ID 3NAZ). **b** HsGS tetrameric arrangement agrees with the crystal structure of *C. elegans* GS (CeGS) (PDB ID 4QLB). The area boxed is shown in **c. c** $HsGN^{34}$ is consistent with the CeGS-GN^{34} structure (PDB ID 4QLB).

5.3 Analysis of the human GS structure

5.3.1 Mechanism of GS inactivation

A unique feature of metazoan GS is that both N- and C-terminal tails are phosphorylated. However, previous crystal structures of yeast and *C. elegans* GS have lacked an understanding of the inhibited state of GS, due to a lack of sequence conservation in these phosphoregulatory regions (Figure 1-15) and expression in bacteria, meaning GS was not phosphorylated. Thus, the mechanism by which the N- and C-terminal tails participate in enzyme inactivation has remained elusive.

We were able to build a model for the GS N-terminus starting from residue 13, and of the C-terminus up to residue 625, and then from 630-639 (chain A/C) and 630-642 (chain B/D). The N- and C-terminal tails of each GS protomer lie almost parallel to each other, and travel side by side along the GS tetrameric core to reach the centre where they meet other C-terminal and N-terminal tails from adjacent GS protomers (Figure 5-4). A 2.8 Å cryo-EM map of GS generated without the application of D2 symmetry averaging (Figure 5-5), suggests that these two C-terminal tails do not engage GS in a symmetrical manner. One chain (chain A) disengages with the GS core earlier than the other C-terminal tail from the

adjacent protomer (chain B), and begins to traverse away from the core. However, the C-terminal tail from chain B continues to travel further across the regulatory helices than chain A. This allows chain B to interact with the regulatory helices α22, specifically S641 engaging with four arginine residues, R588 and R591 from both chain A and chain B, which we have termed the "arginine cradle" (Figure 5-4). From our immunoblotting data and phosphorylation mapping analysis (Figure 3-12, Figure 3-13 and Table 3-1), we know our GS enzyme is phosphorylated at S641. Moreover, in the chain that interacts with the arginine cradle we have modelled S641 as phosphorylated (pS641) (Figure 5-4). Interestingly, there is only room for one pS641 residue, and the other chain is relocated into the "spike" region. S641 (site 3a), the first C-terminal phosphorylation site, was shown to have a dominant role in the regulation of GS[121,240], and interaction of pS641 with the arginine cradle in helix α22 shows the mechanism of inactivation which constrains the GS tetramer in a "tense state".

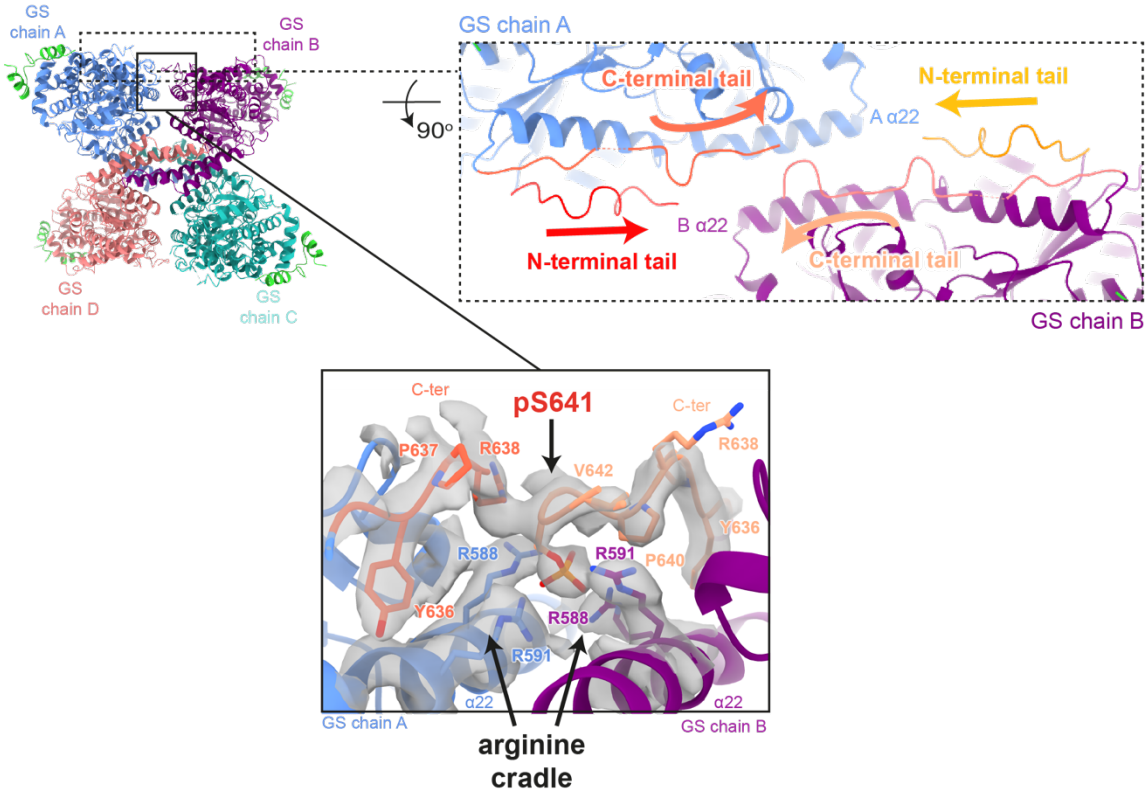

Figure 5-4: Autoinhibition of human GS

Human (Hs)GS-GN[34] structure shown in ribbons (top left). The N- and C- terminal tails of one GS protomer (chain A) lie next to one another and move towards the adjacent protomer, meeting the N- and C-terminal tails from chain B. Arrows indicate continuation of cryo-EM density (top right). Electron density (C1 symmetry) for phosphorylated S641 (pS641) interacting with R588 and R591 on the regulatory helices α22 (bottom).

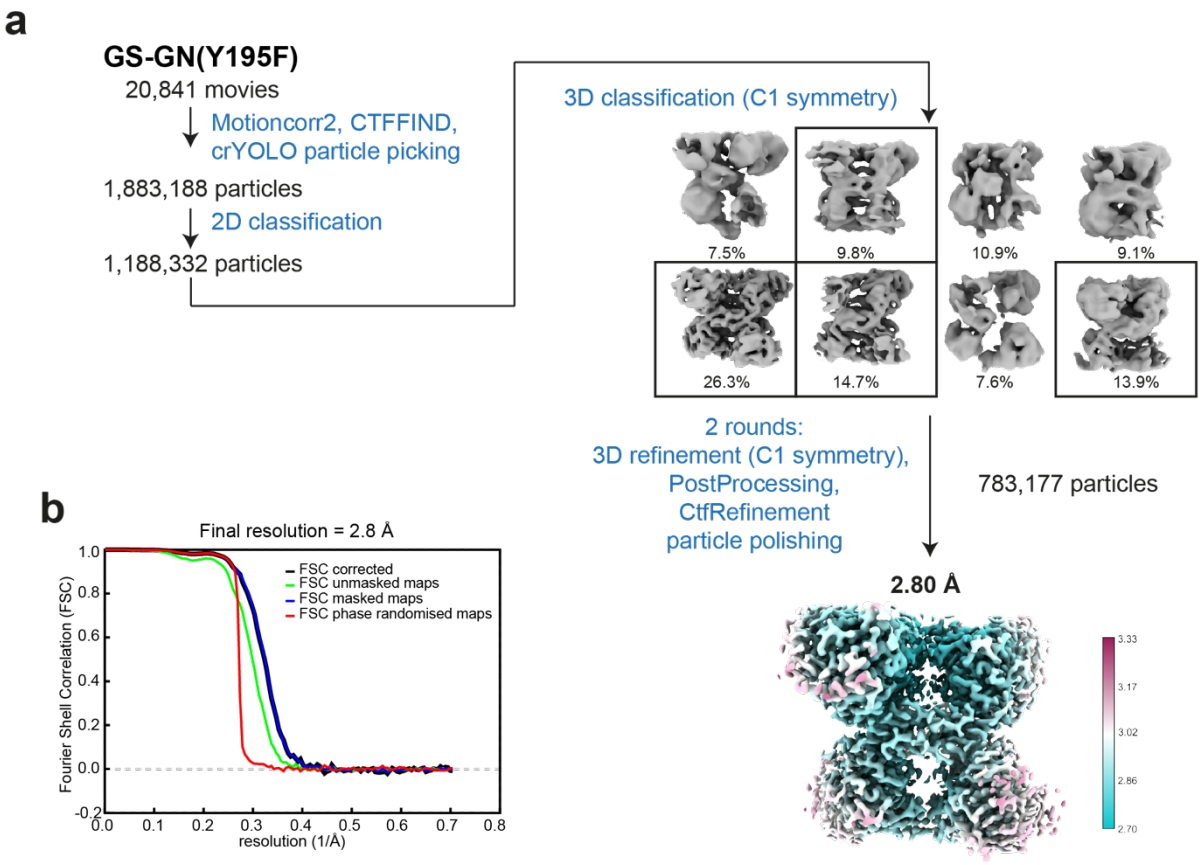

Figure 5-5: Cryo-EM analysis of human GS-GN(Y195F) complex without the application of symmetry averaging

a Flow chart of data processing strategy used. 3D classes boxed were selected for subsequent processing. Final map at a global resolution of 2.8 Å is shown. **b** FSC curve for the final map. The resolution was calculated using the gold-standard FSC cut-off at 0.143 frequency.

5.3.2 Relationship between allosteric activation and inhibition by phosphorylation

The regulatory helices α22 which interact with pS641 are also known to interact with the allosteric activator G6P (Figure 5-6a). The arginine residues R582 and R586[88] on α22, along with other residues in the G6P binding pocket, stabilise G6P binding and induce conformational changes in the GS tetramer leading to a more open active site and a higher activity state[88]. This reveals that interactions involving the regulatory helices α22 are vital in responding to both activating and inhibitory stimuli. Previous studies hypothesised that the phosphoregulatory regions may be able to engage with the allosteric binding site and thus directly compete with G6P binding. However, this human GS structure in the inhibited state does not support this hypothesis as we see no additional density for the

phosphorylated tails in the G6P binding site (Figure 5-6a). Instead, we speculate that the phosphoregulatory regions indirectly affect G6P binding by constraining the opening and closing of the GS tetramer.

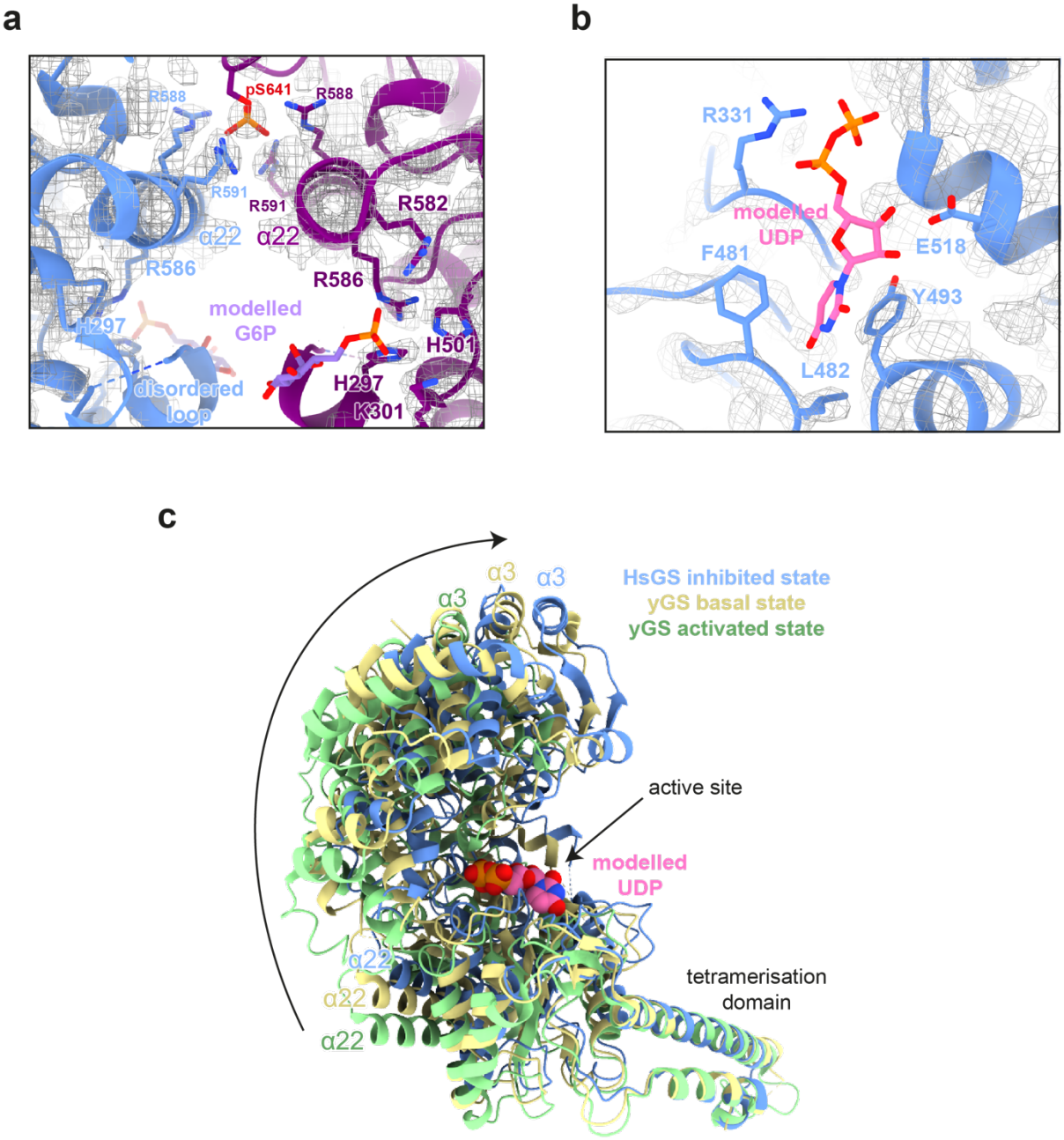

Figure 5-6: The active and allosteric binding site in inhibited human GS

a Close up view of the allosteric binding site and pS641 binding. Helices α22 interacts with both glucose-6-phoshate (G6P) and pS641. G6P is modelled using the yeast GS + G6P structure (using PDB ID 5SUK). The residues involved in the G6P and phosphate interaction and are shown as sticks. **b** Close up view of the closed active site. UDP is modelled in the active site (using PDB ID 3O3C) and the residues involved in the interaction are shown as sticks and labelled. **c** GS

monomers from the human (HsGS) inhibited state, the yeast (yGS) activated (PDB ID 5SUK) and yGS basal in complex with UDP (PDB ID 3O3C). GS monomers were superimposed using the tetramerisation domain, revealing a tightening of the monomer in order to close the active site, as indicated by the arrow.

In comparison to the G6P bound state, the active site in inhibited human GS appears closed (Figure 5-6b). This suggests that accessibility to the catalytic site may dictate the activity state of the GS enzyme. In addition, when comparing the conformation of the GS monomer across activity states, by superimposing on the tetramerisation domain, a much tighter conformation is observed in the inhibited state (Figure 5-6c). This closure involves both the N-terminal and C-terminal domains, as highlighted by the movement of both α3 and α22 helices. We postulate that this tightening and closure of the GS protomer results in a closure of the active site that results in a less active enzyme. Collectively, our structural analyses support a model by which phosphorylated N- and C-terminal tails inhibit the GS tetramer by constraining a tense conformation and blocking substrate access to the catalytic site.

As discussed in the introduction (see section 1.4.1), the regulatory helices have been widely investigated and they are thought to have a role in determining the activity state of GS. Comparing the distance between the regulatory helices (α22) in adjacent monomers, the distance changes according to the activity state of GS (Figure 5-7). In the phosphorylated human GS structure, helices α22 lie 7.9 Å apart when measuring Cα-Cα distances from Arg591 on chain A and Arg580 on chain B (Figure 5-7). The similar measurement of the corresponding residues in yeast proteins reveals that in a structure of a mimic of the inhibited state, where residues R589 and R592 were mutated to Ala and GS was produced in bacteria, the helices are close together at 8 Å (Figure 5-7). This is comparable to the phospho-human GS, where phosphorylation appears to contribute to the closing of the regulatory helices constraining the GS tetramer and thus locking it in a tense, inactive state. In the basal state of yeast GS, these helices are positioned further apart at 11 Å, and in the G6P activated state they are furthest apart, at 16 Å (Figure 5-7). This suggests a role of these helices in altering the conformation of GS, as well as altering the accessibility for substrate binding, moving GS into an inhibited “tense” state or an activated “relaxed” state. Moreover, a model where phospho-S641 leads to GS tetrameric closure into a “tense” state, where substrate access to the active site is blocked is proposed. And this autoinhibition can be overcome by G6P, which in turn drives the helices away from each other, releasing this “tense” state and forming a “relaxed” activated state.

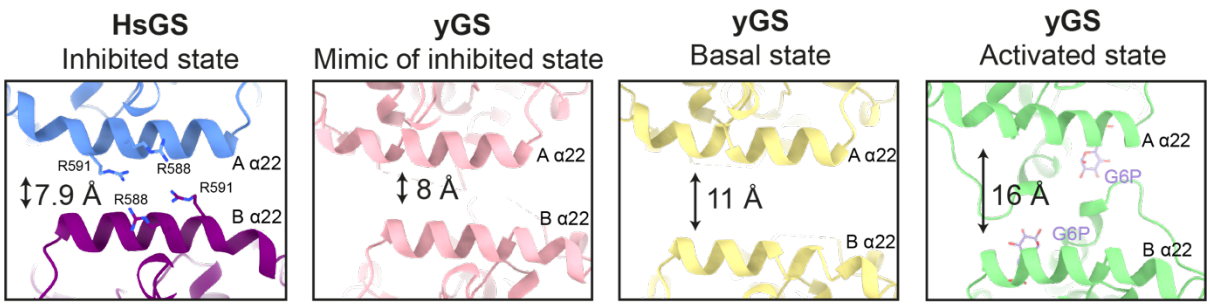

Figure 5-7: Positioning of the regulatory helices across various activity states

Comparison of distances between regulatory helices of adjacent monomers of human GS (HsGS) (reported here), low activity inhibited mimic (PDB ID 5SUL), basal state (PDB ID 3NAZ) and glucose-6-phoshate (G6P) activated (PDB ID 5SUK) yeast GS (yGS) crystal structures. Quoted distances were measured from Cα of Arg591 (chain A) and -Cα of Arg580 (chain B) of HsGS and corresponding yeast residues.

5.3.3 Role of the N-terminal tail

The position of the extreme N-terminus is markedly different in human and *C. elegans* GS structures compared to yeast. The majority of the first β-strand in all 3 species is in a similar position. In human and *C. elegans* GS the residues preceding β1 move towards the regulatory helices, however in yeast, the extreme N-terminus is in the opposing direction (Figure 5-8a). The structural analysis of *C. elegans* GS proposed a hypothesis where phosphorylation could enable the N-terminal tail to engage with the regulatory helices[70]. Our structure of the human, phosphorylated enzyme supports this hypothesis, although the current density does not allow model building before residue 13. To aid model building and electron density interpretation, the program LAFTER[252] was used. LAFTER uses a local denoising filter to recover features across a wide range of signal to noise ratios and produces a denoised map. This denoised map reveals additional density present next to the regulatory helices (Figure 5-8b). This suggests that the N-terminal phosphorylation sites, S8 and S11, are positioned near to the regulatory helices, specifically adjacent to R579 and R580, thus also playing a role in the inhibition of GS (Figure 5-8b).

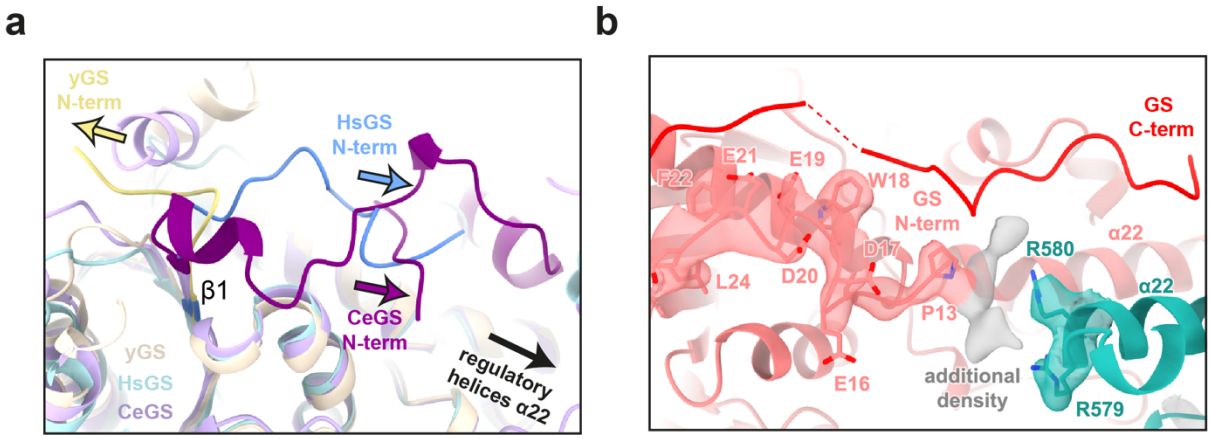

Figure 5-8: N-terminal tail position in HsGS, CeGS and yGS

a Human (HsGS) and C. elegans (CeGS) GS N-terminal tails are positioned towards the regulatory helices. However, the N-terminus of yeast GS (yGS) is facing away from the regulatory helices. **b** LAFTER was used to produce a denoised map, which reveals additional density for the N-terminus of HsGS near the regulatory helices, however, the resolution does not allow model building.

5.3.4 Comparison to the AlphaFold predicted model

The GS core containing the two Rossmann fold domains were predicted to a high level of accuracy by AlphaFold[225] in comparison to the human GS structure (RMSD 1.0 Å between 575 Cα atoms) (Figure 5-9a). However, the position of the N- and C- terminal tails do not agree entirely with our inhibited structure (Figure 5-9a). Nevertheless, the position of S641 in monomeric GS predicted by AlphaFold is consistent with our cryo-EM structure where phospho-S641 is modelled (Figure 5-9b). This indicates that an interaction between S641 and the arginine cradle may also be possible in the non-phosphorylated state, although the negative charge on the phosphate group would naturally strengthen the interaction with the positively charged arginine cradle.

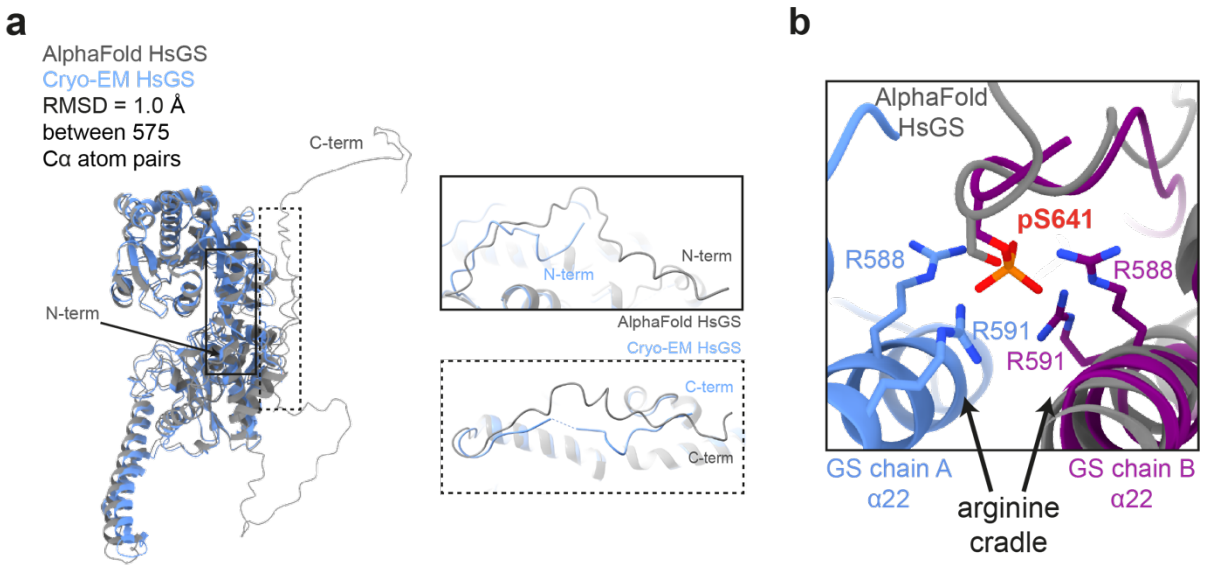

Figure 5-9: Comparison between cryo-EM human GS and AlphaFold GS

a Comparison of human GS monomer model (reported here) with the AlphaFold model (accessed on 1 October 2021) and position of the N- and C- terminal tails. **b** Comparison of human GS pS641 cryo-EM model with the AlphaFold model.

5.4 Dislodging the phosphoregulatory region

Due to the flexibility of the N- and C-terminal tails, we were unable to build phosphorylated residues other than phospho-S641. However, we can see the beginning of the flexible phosphoregulatory "spike" region and residues from the GS "core tetramer" which interact with this regulatory region. To investigate the mechanism of GS inactivation, we mutated residues in GS that contact the beginning of the phosphoregulatory region. We selected residues which are not involved in G6P binding and mutated these in order to "dislodge" the regulatory tails (Figure 5-10). If the "spike" region which houses the remaining phosphorylation sites are indeed maintaining GS in an inactive state, then weakening the interaction between the core tetramer and the N- and C-termini inhibitory regions should create an enzyme with higher basal activity in comparison to the WT. Therefore, the structure-guided mutagenesis resulted in the generation of the following GS mutants: R588A+R591A, Y600A, R603A, H610E, S641R, W18A.

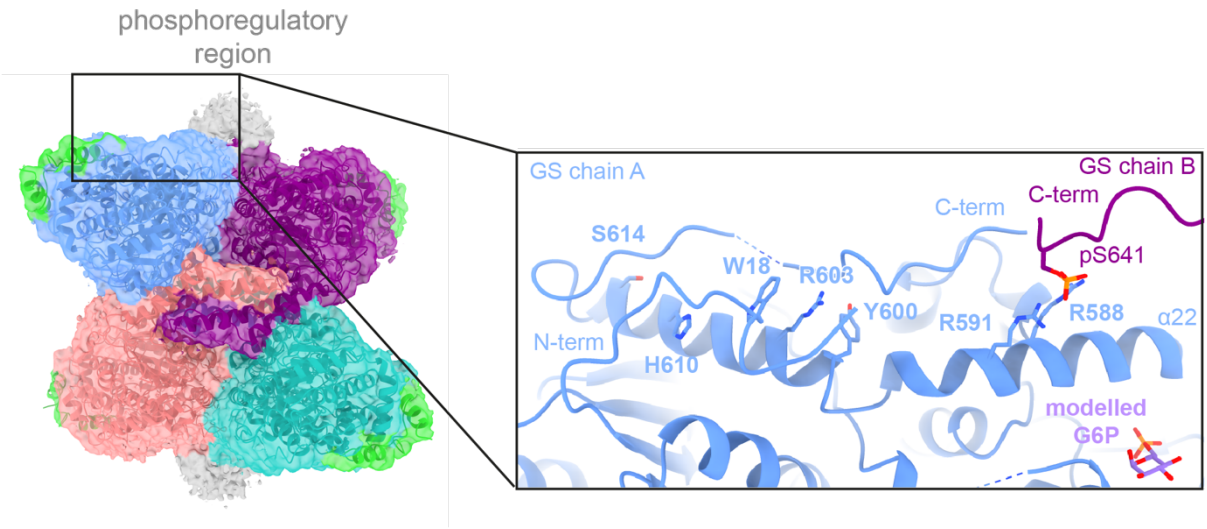

Figure 5-10: GS mutations created in order to dislodge the phosphoregulatory regions
Model of human GS in complex with GN C-terminus and cryo-EM map shown in transparent surface and coloured by chain. The phosphoregulatory region that houses the majority of the phosphorylation sites is shown in grey and labelled (left). The N- and C-terminus of one GS protomer (blue) and the C-terminal of the adjacent GS protomer (purple). The residues which are mutated in this study are show as sticks. Glucose-6-phoshate (G6P) is modelled using the yeast GS + G6P crystal structure (PDB ID 5SUK) (right).

5.4.1 Production of GS mutants

All mutants were generated via site directed mutagenesis and confirmed by DNA sequencing. Similar to GS WT, mutants were co-expressed with GN(Y195F) in insect cells and subsequently purified via nickel affinity chromatography and size exclusion (Figure 5-11).

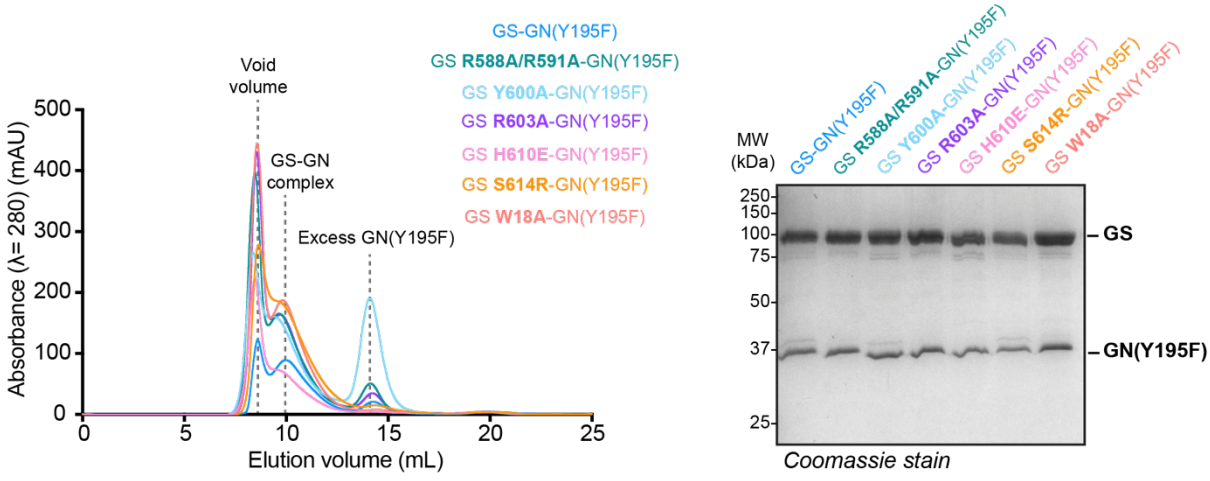

Figure 5-11: Purification of GS mutants in the GS-GN(Y195F) complex

Size-exclusion chromatography profile of GS WT and mutants in the GS-GN Y195F complex (left), and SDS-PAGE analysis of the corresponding mutants (right).

5.4.2 Dislodging the phosphoregulatory region increases basal activity

Consistent with our hypothesis, we observed a marginal increase in basal (-G6P) GS activity in R588A+R591A, Y600A, R603A, H610E and W18A mutants, that was reflective of the phosphorylated state at residues S8, S641 and S645 (Figure 5-12a and b). No increase in basal activity was observed for S641R, potentially as this mutation is furthest away from the base of the "spike region" (Figure 5-10 and Figure 5-12a). All mutants were unaffected in terms of co-purification with GN, and with the exception of R588A+R591A mutant, they had similar melting (T_m) profiles and oligomeric state to the WT GS complexes (Figure 5-11 and Figure 5-12b and c and Figure 5-14). As expected, most mutants could still be activated to similar levels to the WT upon addition of G6P (Figure 5-12a), which is consistent with previous studies that demonstrate that G6P can override inhibition by phosphorylation and restore full activity[1]. With the exception of Y600A which did not recover to full activity following G6P activation, and although Y600 does not directly bind to G6P, UDP or sugars[88-90], it is possible that this hydrophobic residue is important for interdomain movements which are required for full GS activation.

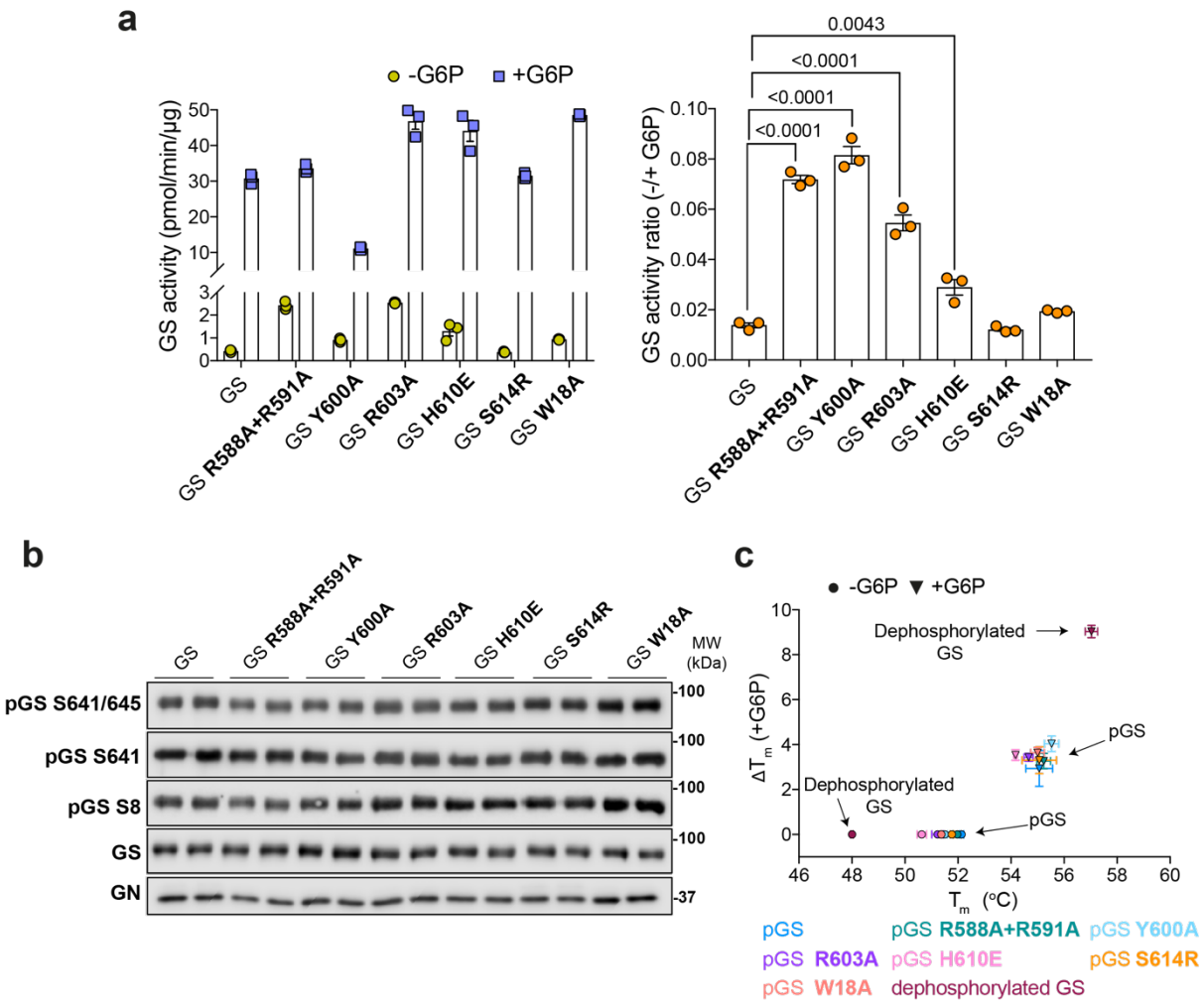

Figure 5-12: GS mutants increase basal activity

a Activity of GS WT and indicated mutants in the GS-GN(Y195F) complex in the presence and absence of glucose-6-phoshate (G6P) (left) and -/+ G6P activity ratio (right). Data are mean +/- S.E.M representative of two independent experiments performed in technical triplicates. One-way analysis of variance, (Tukey's post hoc test); exact *p* values are shown. **b** Western blot for human GS phosphorylation sites S641/645, S641, S8, and total GS and GN. **c** Melting temperature (T_m) of GS WT and mutants in the GS-GN(Y195F) complex. Changes in melting temperature upon addition of 12.5 mM G6P ($\Delta T_m = T_m^{+G6P} - T_m^{-G6P}$). Data are mean +/- S.E.M from three independent experiments carried out in technical duplicates (dephosphorylated GS) and triplicates (WT and mutant GS).

5.4.3 Dislodging the phosphoregulatory region increases accessibility to phosphatases

To further investigate the effect of dislodging the phosphoregulatory region we examined selected mutants, R588A+R591A, R603A and W18A in their response to

dephosphorylation. The C-terminal mutants R588A+R591A and R603A were chosen as the largest difference in basal activity was observed (Figure 5-12a), and W18A was selected to investigate a potential role of the N-terminus. Upon addition of PP1 and lambda PP, these GS mutants were more robustly dephosphorylated at S641 and S8 than WT GS (Figure 5-13a). This suggests that the dislodging increases exposure of the “spike” region to phosphatases. For the W18A mutant, this dephosphorylation by both lambda-PP and PP1 resulted in over a 20-fold increase in basal activity, and also, an approximately 3-fold increase in comparison to WT GS (Figure 5-13b). The GS R603A-GN(Y195F) mutant has a basal activity similar to WT GS upon dephosphorylation. Surprisingly, the robust dephosphorylation at S641 and S8 in GS R588A+R591A was not associated with an increase in activity (Figure 5-13).

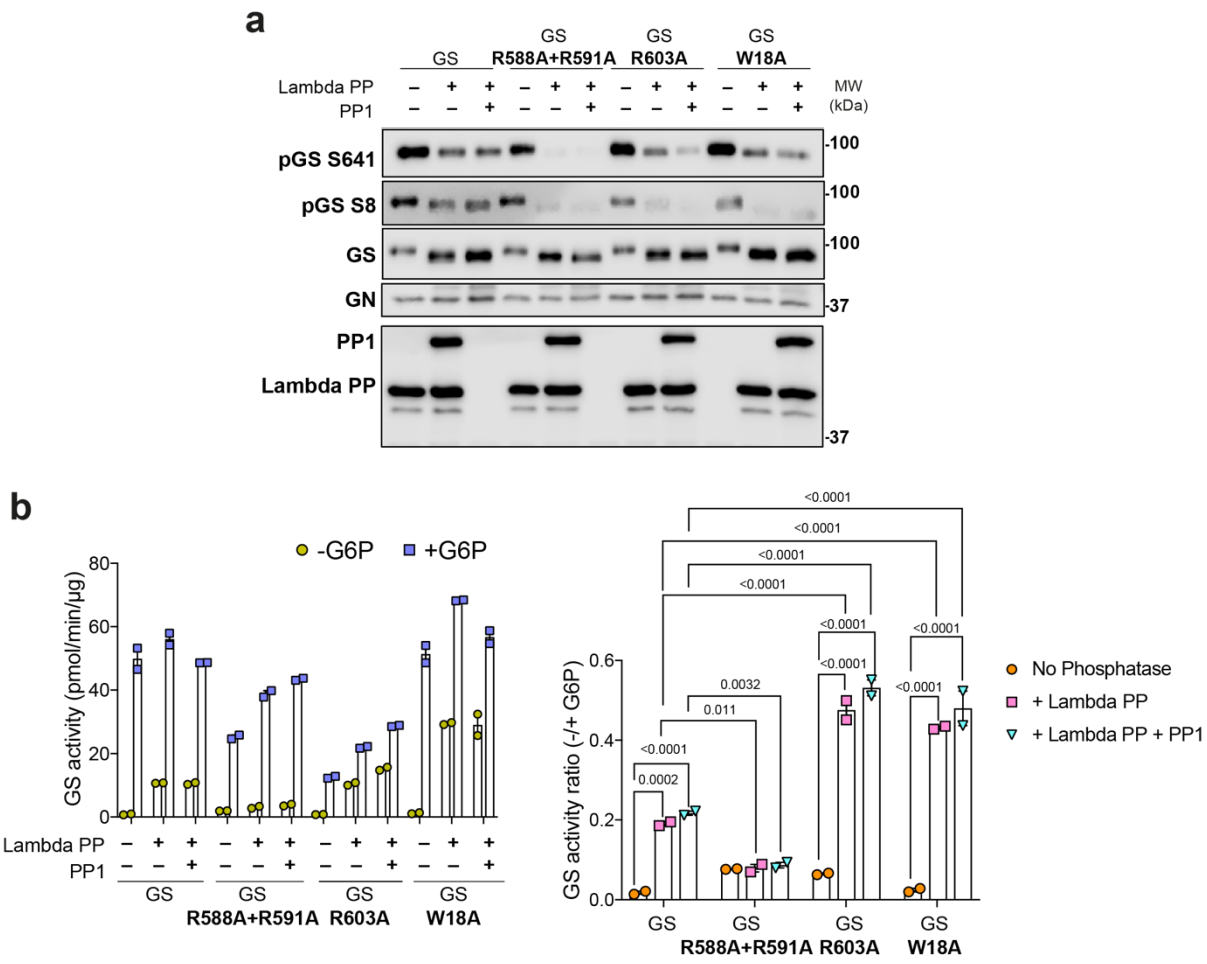

Figure 5-13: GS mutants increase accessibility to phosphatases

a Immunoblots of GS WT and mutants in the GS-GN(Y195F) complex after dephosphorylation with protein phosphatase 1 (PP1) and/or lambda protein phosphatase (lambda PP). **b** Activity of phosphorylated and dephosphorylated GS WT and indicated mutants (left) and -/+ glucose-6-phoshate (G6P) activity ratio (right). Data are mean +/- S.E.M representative of two independent

experiments performed in technical duplicates. Two-way analysis of variance (Tukey's post hoc test); exact *p* values are shown.

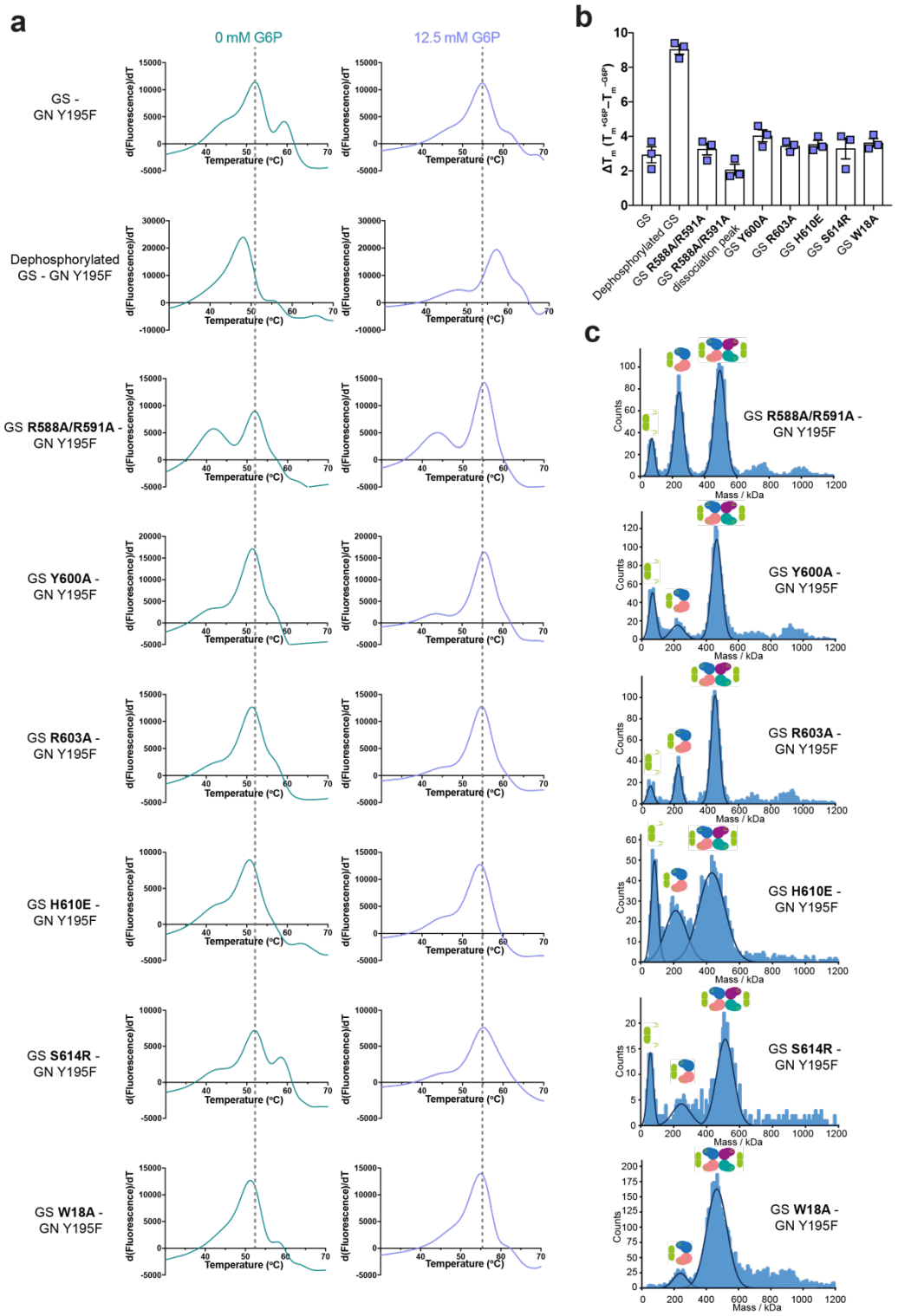

Figure 5-14: Stability of GS WT and mutants in the GS-GN(Y195F) complex

a Differential scanning fluorimetry (DSF) analysis of GS WT, various mutants and dephosphorylated GS in the GS-GN(Y195F) complex. Addition of glucose-6-phosphate (G6P) leads to stabilisation of

the complex, seen by an increase in melting temperature. **b** Changes in melting temperature (ΔT_m) was calculated by subtracting the T_m of 0 mM G6P of the corresponding protein. Data in **b** is shown as mean +/- S.E.M. Data in **a** and **b** are the mean from three independent experiments carried out in technical duplicates (dephosphorylated GS) and triplicates (WT and mutant GS). **c** Mass photometry analysis of GS mutants in the GS-GN(Y195F) complex. Expected stoichiometry is indicated by a cartoon representation of the complex, where each GS protomer is coloured in blue, purple, pink and turquoise and the GN dimer is shown in green.

As described previously, R588 and R591 lie on the regulatory helices and are also involved in inter-subunit interactions and form the arginine cradle that interacts with phospho-S641 (Figure 5-4). We observed some dissociation of the GS R588A+R591A double mutant complex in mass photometry (Figure 5-14c). This dissociation was also seen in cryo-EM analysis of the GS(R588A+R591A) – GN(Y195F) complex. In 2D classification some GS tetramer can be seen, however some classes do not resemble the tetramer (Figure 5-15). Although it is difficult to distinguish between 2D projection views and dissociation, the *ab initio* models generated have poor density that does not fit the GS tetramer well (Figure 5-15). Due to the poor *ab initio* classes obtained from this dataset, image processing was terminated. These results are suggestive of dissociation of the complex, in agreement with mass photometry analysis and multiple melting peaks in thermostability assays (Figure 5-14). Therefore, the role of R588 and R591 in stabilising the GS tetramer may be the cause for the lack of rescue of activity upon dephosphorylation (Figure 5-13).

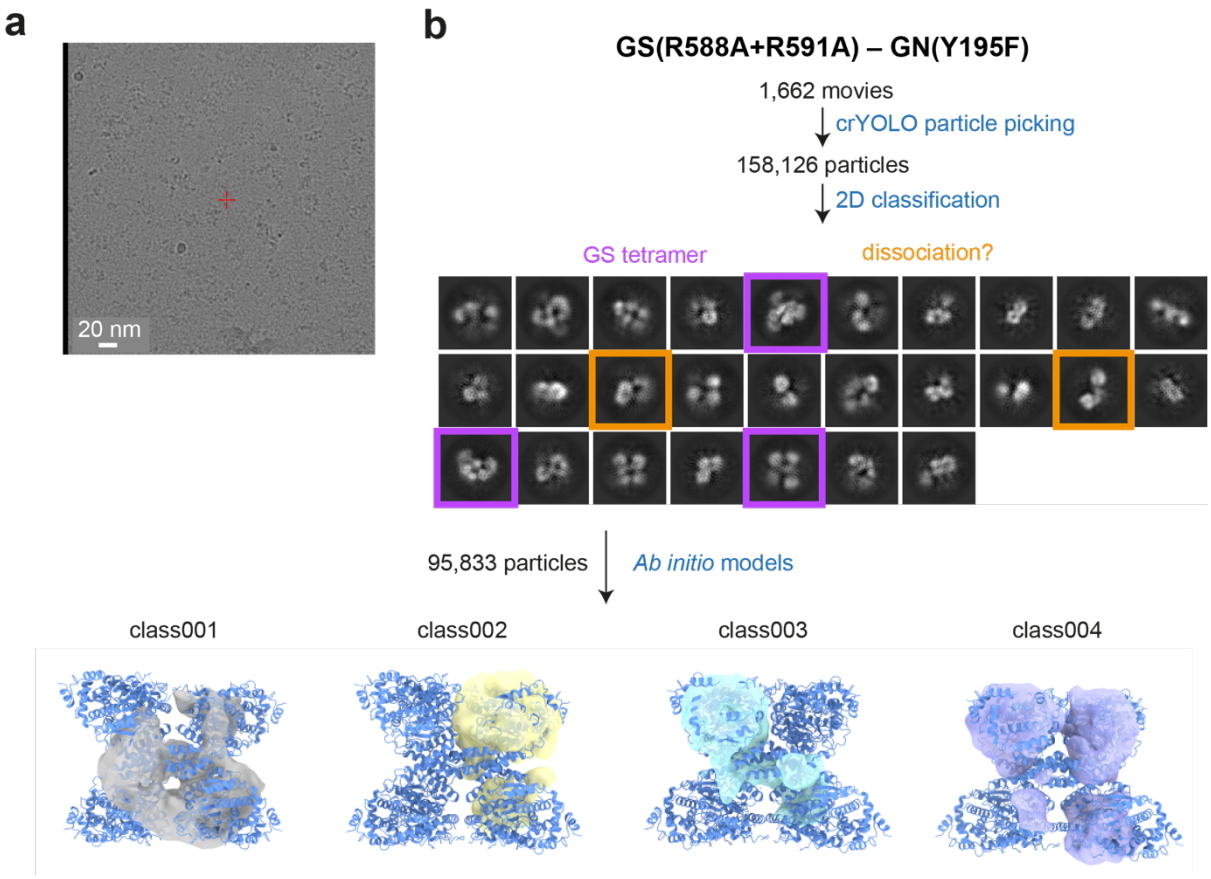

Figure 5-15: Cryo-EM analysis of GS(R588A+R591A)-GN(Y195F)

a Representative micrograph. **b** Data processing pipeline. 2D classification revealed some classes which resemble the GS tetramer, however some classes do not and these are potentially dissociated complex. The *ab initio* models generated do not fit the GS tetramer (reported here) well.

5.5 Chapter summary

This chapter reveals a higher resolution structure of human GS at 2.6 Å. A key discovery is the interaction of phosphorylated S641 with the arginine cradle situated on the regulatory helices α22. This autoinhibition mechanism indicates a vital role of helix α22 in both allosteric activation and covalent inhibition. Structure guided mutagenesis revealed that dislodging the phosphoregulatory region results in an increase in basal activity, as well as leaving the phosphorylation sites more accessible to phosphatases. The structure reveals inter- and intra-protomer interactions of the N- and C-terminal tails within the GS tetramer, thus highlighting a crucial role of these regulatory regions in holding GS in a “tense” and inactive state.

6 Chapter 6: Discussion

6.1 Overview

GS is a metabolic enzyme involved in glycogen synthesis, and is the rate-limiting enzyme in this process. Despite the extensive investigation into GS enzyme kinetics and structural determination of bacterial[86,87], yeast[88,90,91] and *C. elegans*[70] GS, the regulation and structure of human GS remained elusive. This has hindered the development of GS inhibitors[236,253], which is a potential therapeutic to treat GSDs.

The first results chapter in this thesis (see chapter 3) shows optimisation of the expression and purification of the full-length GS-GN complex. In addition, the stoichiometry of the complex was measured by mass photometry and nsEM, revealing two flexible GN dimers, one on either side of a GS tetramer.

The second results chapter in this thesis (see chapter 4) aims to address the current lack of structural information for human GS. A global resolution of 4 Å was achieved through a combination of sample and data processing optimisation steps. These efforts resulted in the first insights of the inhibited state of GS. In addition, efforts towards producing and analysing an activated and a basal state enzyme are shown.

Finally, the third results chapter in this thesis (see chapter 5), shows determination of a high-resolution structure of human GS, at a global resolution of 2.6 Å. This structure revealed the mechanism of inactivation through phosphorylation at the N- and C-terminal tails in unprecedented detail. This enabled the design of structure-guided mutants to elucidate the role of the inter-subunit domain named the “spike” region and elucidate new details of GS biological mechanism of action.

This discussion will reiterate some of the key findings of this thesis, as well as describing and explaining results in the context of previous studies and suggesting future directions.

6.2 Cooperation of GS and GN during glycogen synthesis

The co-expression of full-length human GS and GN successfully produces a stoichiometric complex, and nsEM analysis shows two GN dimers co-ordinating a GS tetramer (Figure 3-11). The different positions of the GN dimers, with one being more tilted towards GS, show that these proteins are flexible. The 4:4 stoichiometry of this complex suggests that

GS can potentially elongate four glucose chains simultaneously, one from each GN protomer, however further work is required to confirm this.

The high-resolution structure of human GS has shown the mechanism of inactivation through auto-inhibition; however, it is still not clear why human GS requires stabilising interactions with GN to be expressed, unlike its yeast and *C. elegans* counterparts. Especially considering the GS-GN^{34} interaction is very similar in both humans and *C. elegans* (Figure 5-3). Recently, another study has also solved the structure of human GS in complex with the C-terminus of GN (GN^{34})[254]. The authors identified three GS cysteine residues (C137, C189 and C251) at the GS-GN^{34} interface that they propose are responsible for the need to co-express human GS with GN. The distances between these residues are within disulphide bonding distance, and visualising the electron density map at lower threshold suggests possible disulphide bonding between two of these cysteines. However, in both the 4 Å and 2.6 Å described in this thesis, no evidence of disulphide bonding was observed. The lack of conservation of this cysteine pocket in *C. elegans* (substituted by C154, L207 and T269) and yeast GS (substituted by V126, P177 and S240) (Figure 1-15) may explain this unique requirement of human GS, as in the absence of GN this cysteine pocket would be solvent-exposed and may lead to aberrant disulphide bonding[254].

It is evident from our negative stain and cryo-EM analyses that GN is flexible in the GS-GN complex, and the two GN dimers neighbouring a GS tetramer do not interact in an identical manner, with one GS dimer tilted closer towards GS in comparison to the other (Figure 3-11). This observed mobility of GN may be aided by the variable length linker, that joins the catalytic domain and the C-terminal region that anchors GS. The linker length was shown to govern glycogen particle size and molecular weight distribution in vitro[92], however further investigation is required to understand the precise functional relevance of this movement. It is exciting to speculate that the ability for GN to interact flexibly with GS may facilitate the wide range of size and distribution of glycogen particles seen in multiple species and tissues[97].

6.3 Mechanism of GS inactivation

Overall, our structure of human GS in the phosphorylated state reveals auto-inhibition of the enzyme. Phosphorylated S641 (site 3a) from one protomer interacts with the arginine cradle, R588 and R591 from two GS protomers, which lie on the regulatory helices α22 (Figure 5-4). Interestingly, only one C-terminal tail can interact with α22 and the other tail

from the adjacent protomer is excluded to the flexible “spike” region (Figure 5-4). This interaction closes the regulatory helices and constrains GS in a “tense” state where the active site is occluded, preventing substrate access and thus explaining why the enzyme is in a low activity state (Figure 6-1).

This structure highlights an important role of phosphorylated S641 (site 3a) in the inhibition of GS, which is consistent with previous mutagenesis studies in rabbit GS. This study showed that upon phosphorylation by GSK3, a Ser to Ala mutation at site 3a resulted in an increase of the activity ratio by around 6-fold in comparison to WT, meaning this mutant enzyme is more active than the WT phosphorylated form[121]. This suggests that phosphorylation at site 3a is critical in GS inactivation[121]. In addition, phosphorylation at the other C-terminal sites, sites 5, 4, 3c and 3b did not have effect on G6P sensitivity in comparison to unphosphorylated WT. Whereas introduction of the fifth phosphate at site 3a resulted in a 10-fold increase in the G6P concentration required for half-maximal activity[121]. However, this G6P sensitivity is not necessarily due to site 3a only, as the other four C-terminal sites of rabbit GS were also phosphorylated, in the nature of the sequential phosphorylation[121]. Nonetheless, this study highlighted a dominant role for site 3a in the inactivation of rabbit GS by GSK3.

The structure of human GS reported in this thesis revealed extra density nearby the regulatory helices, where S8 and S11 (sites 2 and 2a respectively) may be located. This suggests a role for the N-terminus in the mechanism of GS inactivation, consistent with previous studies that revealed that phosphorylation at both the N- and C-terminus is required for inhibition of rabbit GS[75]. It would be interesting to further investigate the role of the human GS N-terminal phosphorylation sites by measuring activity in a delta C-terminal GS construct or with the C-terminal phosphorylation sites mutated from Ser to Ala. Although preliminary analysis of the protein expression would be required as this majorly dephosphorylated enzyme may be unstable or insoluble.

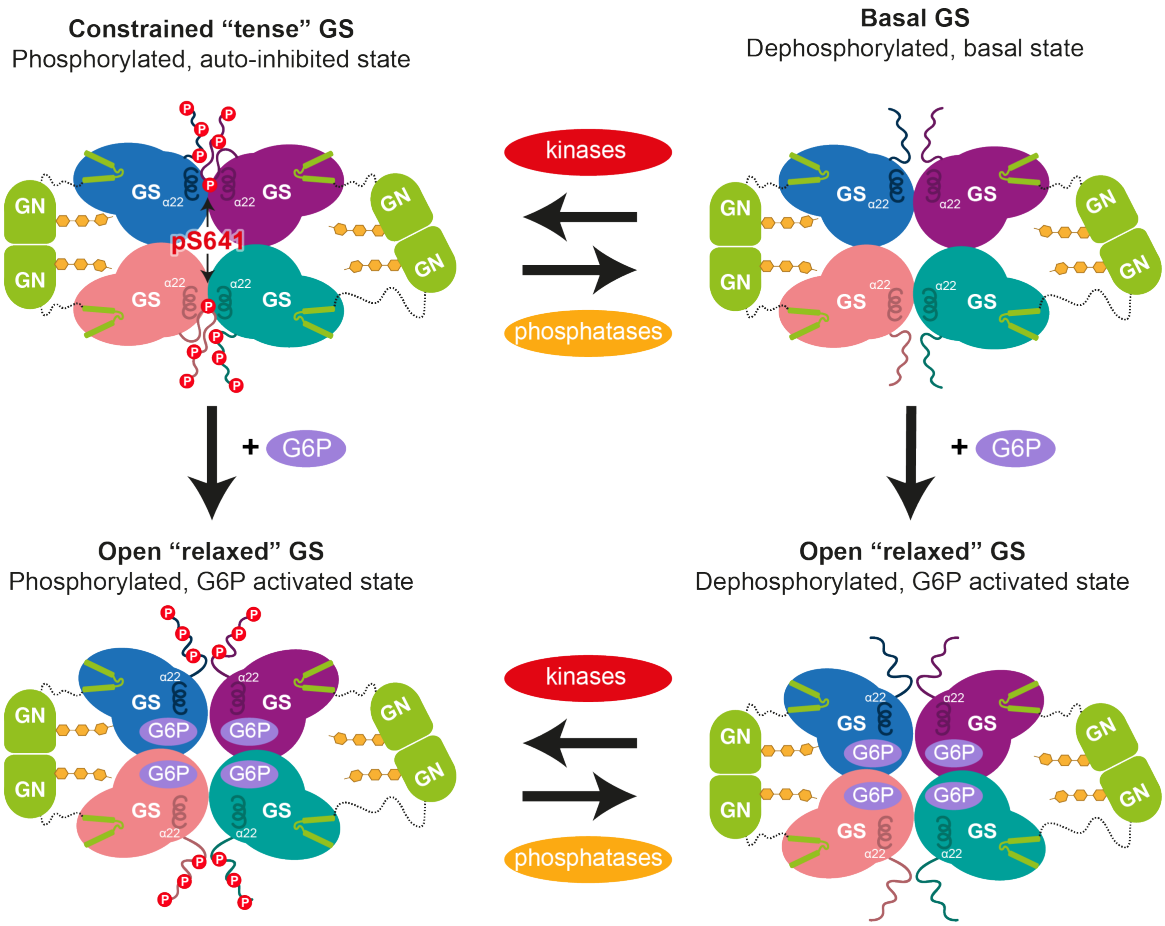

Figure 6-1: Regulation of the GS-GN complex

Glucose is converted into glycogen through the action of glycogenin (GN), glycogen synthase (GS) and glycogen branching enzyme (GBE). GN interacts with GS to feed the initial glucose chain into the GS active site for elongation. GS is regulated by allosteric activation and inhibitory phosphorylation. Phospho-S641 (pS641) from one C- terminal tail interacts with the regulatory helices α22 to cause enzyme inhibition. This can be relieved by glucose-6-phoshate (G6P), with or without phosphatases, to reach a high activity state. Kinases can phosphorylate GS to inhibit the enzyme.

The multiple phosphorylation sites of GS indicate a high level of redundancy, perhaps to ensure GS remains in the inhibited state where a phosphorylated residue is interacting with the arginine cradle. In addition, each phosphorylation site is regulated by different kinases, meaning GS sits at the nexus of multiple signalling pathways. It is not known whether only S641 can interact with the regulatory helices, and in a state of GS where S641 is not phosphorylated, an alternative phosphorylation site can interact with the regulatory helices, ensuring GS remains in its auto-inhibited state.

Interestingly, R588 and R591 which form the arginine cradle and S641 are not conserved in *C. elegans* GS (Figure 1-15). This suggests an evolutionary divergence and speculates additional mechanisms for CeGS inactivation. The difference in the position of the N-terminus in CeGS and HsGS (Figure 5-8) as well as the absence of the arginine cradle suggests an important role of the N-terminal phosphorylation sites in the inactivation of CeGS, perhaps interacting with the other arginine residues in the regulatory helices.

In this thesis, structure guided mutagenesis to elucidate the role of the GS phosphoregulatory region was performed, in context of the GS-GN(Y195F) complex. Mutating GS residues at the interface between the core and the “spike” region mostly resulted in enzymes that had higher basal activity (-G6P), confirming the hypothesis that the phosphoregulatory region holds GS in a constrained, inhibited state (Figure 5-12). Three mutants were further investigated, and two of these mutants (R603A and W18A) were more easily dephosphorylated and had higher basal activity upon dephosphorylation compared to WT. This suggests that dislodging the “spike” region lead to the phosphorylation sites being more accessible to phosphatases (Figure 5-13). Surprisingly, mutating the arginine cradle (R588A+R591A), which interacts with phosphorylated S641, showed robust dephosphorylation by immunoblotting, but did not have increased basal activity upon dephosphorylation in comparison to phosphorylated GS R588A+R591A and WT (Figure 5-13). Further analysis by mass photometry and cryo-EM showed dissociation of this mutant complex, and so this may be the reason for the lack of rescue in activity upon dephosphorylation. Further work to investigate the effect of these GS mutations in the complex with GN WT would be interesting to evaluate any impact of glucosylation on the regulation of GS.

Interestingly, arginine cradle residues have been mutated previously in yeast GS, by mutating the corresponding residues R589A+592A[91]. The authors refer to this double mutant as a ‘mimic of the phosphorylated state’, as they found yeast GS(R589A+592A) to have an activity ratio similar to the yeast GS enzyme when it has been phosphorylated *in vitro*. However, our observation that GS(R588A+R591A)-GN(Y195F) has higher basal activity and activity ratio than GS-GN(Y195F) is not consistent with these findings in the yeast enzyme. However, the activity levels of phosphorylated human GS and phosphorylated yeast GS are yet to be directly compared, and due to the increased number of phosphorylation sites in human GS in comparison to yeast GS, they are unlikely to be comparable/similar. The authors did not report any dissociation of this yeast mutant enzyme, however, as this protein was produced in bacteria it was not phosphorylated.

Thus, perhaps dissociation only occurs when these arginine residues are essential for binding to a phosphorylated residue to hold GS in a tense, inactive state. Overall, the comparison across species and different expression systems may account for the discrepancies.

Structural analysis of yeast GS(R589A+R592A) is similar to phosphorylated human GS, with an RMSD of 0.94 Å between 537 Cα atoms. The authors revealed that the regulatory helices (α22) in the yeast GS(R589A+592A) are closer together than in the yeast basal state, at 8 Å apart in comparison to 11 Å (Figure 5-7). This is consistent with the human GS inactive state where the regulatory helices lie 7.9 Å apart (Figure 5-7), overall highlighting a role for helices α22 in regulating GS activity across orthologues, and with the fact that this yeast mutant is indeed inactive.

6.4 Regulation of GS

The structure of human GS highlights an essential role for the regulatory helices (α22) in GS inhibition. These helices are also involved in GS activation, as R582 and R586, are involved in G6P binding, moreover, α22 plays a vital role in both allosteric activation and covalent inhibition of GS (Figure 6-1). Recently, a study solved the structure of human GS in complex with G6P[254]. This revealed a binding mode similar to the previous structure determined of yeast GS in the activated state[88], whereby G6P induces conformational changes where the regulatory helices are pushed apart. In the human enzyme, this activation dislodges pS641 interacting with the regulatory helices, leaving the active site open. Interestingly, upon UDP-G binding to the activated state, each GS subunit is rotated in order to close the active site, leaving GS in a catalytically competent state[254]. In this structure, UDP-G was hydrolysed, where UDP remains bound in the active site and the glucose is found adjacent to this pocket. The location of this glucose (+0) is adjacent to the glucose binding sites identified in *E.coli* GS (+1, +2, +3)[90] (Figure 6-2), which suggests a mechanism in which the growing glucose chain is fed into and out of the active site through a cleft[254]. However, further investigation into exactly how the primer glucose chain is fed from GN into GS is needed to confirm this mechanism.

In addition, the structure of the basal, dephosphorylated state of human GS has not been solved, leaving parts of the regulatory mechanism of GS elusive. Due to the instability of dephosphorylated GS *in vitro* (Figure 4-21), it will be difficult to solve this structure. However, it would be interesting to structurally investigate the observation that dephosphorylation increases the affinity for G6P[239]. In addition, the GS affinity for UDP-G

is complicated and still unclear[255]. It has been shown that high glycogen content decreases the affinity for UDP-G *in vivo*[256], and also that dephosphorylation increases affinity for UDP-G[257]. However, the underlying structural mechanism of this remains unclear. Moreover, further efforts into solving structures for these various states, as well as biochemical studies, may shed light into the allosteric and covalent regulation of the kinetic properties of human GS.

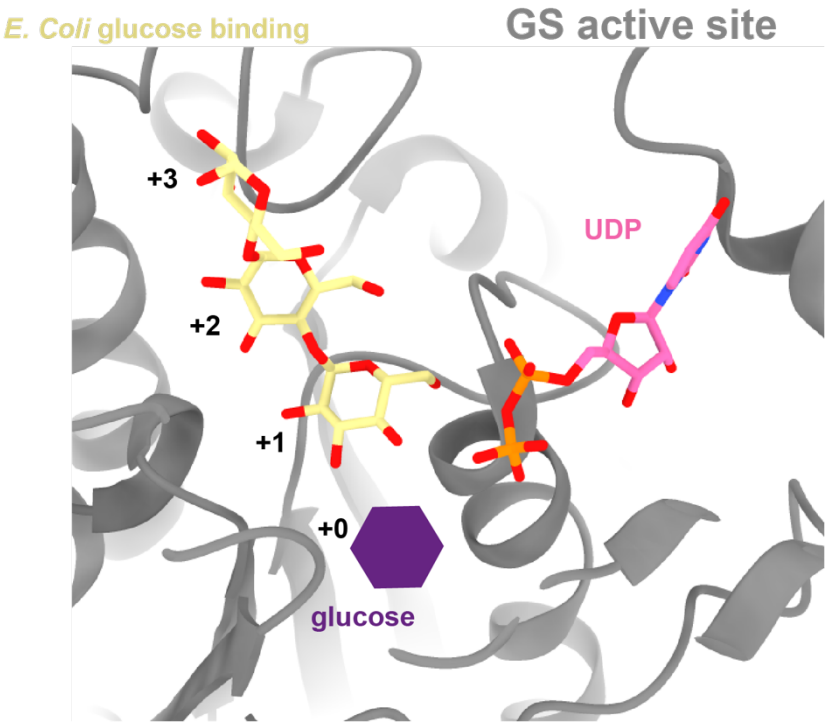

Figure 6-2: The UDP and glucose binding sites in GS

At the time of writing, co-ordinates of the human GS structure in complex with UDP-glucose were not available, moreover the yeast GS + UDP structure (PDB 3O3C) was used to place UDP and the glucose residue (shown as a purple hexagon) was approximately placed based on a figure shown in the pre-print of McCorvie *et al*, 2022[254]. The glucose binding sites of *E. coli* GS (PDB ID 3CX4) are named positions +1, +2 and +3, which are adjacent to the human +0 glucose binding site.

The extensive investigation into yeast GS identified six arginine residues (human residues: R579, R580, R582, R586, R588 and R591) which are known in the literature as the arginine cluster (Figure 1-15). Mutagenesis, kinetic and structural work has shown roles for R580, R582 and R586, in the activation of GS, by interacting with G6P, in both yeast[88]

and human[254]. The first two residues, R579 and R580, were proposed to be responsible for allowing yeast GS to respond to phosphorylation and potentially involved in binding to a phosphorylation site. While the final two residues, R588 and R591, were also thought to be involved in the response to phosphorylation as well as setting the basal activity level by keeping the regulatory helices apart when GS is not phosphorylated. The new structure of human GS reported here has now confirmed the roles of the final two arginines (R588 and R591) in the cluster, now termed the arginine cradle, in directly binding to a phosphorylated residue. While the work described in this thesis resolves decades of investigation and speculation about the arginine cluster, further work into fully understanding the roles of R579 and R580 is required, particularly for human GS.

6.5 Future perspectives

Although the high-resolution structure of human GS described in this thesis has greatly improved our knowledge of GS regulation, the cooperation of GS and GN in the synthesis of glycogen still remains elusive. A potential method to investigate how GN feeds the glucose primer into GS for elongation is to perform cryo-EM analysis of a catalytically inactive GS in complex with WT GN. This analysis could hopefully characterise the cooperation of GS and GN, and examine whether or not GS can extend glucose primers from two GN dimers simultaneously.

A potential therapeutic for GSDs is to target enzymes involved in glycogen synthesis. In the drug discovery process, cryo-EM is a tool that can be utilised to perform structural analysis of small molecule inhibitors. The achievement of solving human GS at 2.6 Å shows that cryo-EM is a viable technique that can be used, as the high resolution allows for structure activity relationship studies that are essential in the drug discovery process. The structure will also allow for *in silico* structure-based drug discovery attempts, to identify binders of GS in various pockets, including the active site and allosteric G6P binding site. However, the flexibility of GN caused issues in the data processing and potentially hinders the achievable resolution. One method to streamline the structure-based drug design approach to allow more reliable and faster structure solving is to remove the globular domain of GN. This could be completed by inserting a protease site in GN after the minimal binding region to GS, moreover expressing the full-length complex and using protease digestion to form a final product of GS in complex with GN^{34}. Alternatively, instead of co-expressing GS and full-length GN, a construct that only contains the C-terminus of GN could be designed. With the use of fusion tags to aid solubility, GN^{34} could be co-expressed

with GS, resulting in a less flexible complex than the full-length proteins which may be more suitable for drug discovery programs using cryo-EM.

To investigate the purpose of multiple GS phosphorylation sites, a series of point mutations altering each phosphorylation site to an alanine one by one, could answer whether it is possible for other phosphorylation sites apart from pS641 to interact with the arginine cradle. For example, if S641 is mutated to alanine, does the next C-terminal phosphorylation site, S645, interact with the arginine cradle instead, or perhaps one of the nearby N-terminal phosphorylation sites. Knowing GS is destabilised upon dephosphorylation (Figure 4-21), ensuring only one phosphorylation site is mutated to alanine at any one time should ensure GS remains soluble, allowing successfully expression, purification and structural analysis. Creation and successful expression of delta N-terminal or delta C-terminal GS constructs may be able to investigate this further, and identify roles for each phosphorylated tail by structural analysis and also activity assays.

In addition, molecular dynamic simulations could be performed using the structure of human GS, which may be able to generate models of how other phosphorylation sites may interact with the arginine cradle and may give insight into the purpose of the “spike region” being flexible.

6.6 Therapeutic potential of targeting GS

Mutations of GS1 and GS2 are common in GSDs, and are often found within pockets of GS, affecting G6P, sugar and UDP-G binding (Appendix 2). Some mutations also affect the GS-GN[34] interaction (Appendix 2), which is consistent with the observation that this interaction is essential for glycogen synthesis *in vitro*[70]. More recently, GS1 has emerged as a therapeutic target for several GSDs, including Lafora disease[159], Pompe disease[236] and Andersen disease[253]. In these examples, pathogenicity is driven by excess glycogen, therefore, inhibiting GS and in turn reducing glycogen levels could be beneficial[178,258-261]. However, the process has thus far been hindered by the lack of structural information of human GS. Moreover, the ability to gain high resolution information of human GS will certainly be beneficial in efforts to design GS inhibitors to target specific pockets. GS is an attractive enzyme to target as it has several druggable pockets, including the G6P allosteric site, the active site and also potentially the GS-GN[34] interface. Although the allosteric site may offer a potential pocket, it may be difficult to mimic G6P with a small molecule that is suitable as a therapeutic treatment due to its negative charge. It is

important to consider the differences in the sequences of GS1 (predominantly in the muscle) and GS2 (exclusively in the liver) (Figure 1-15). Moreover, significant efforts must be made to ensure a small molecule is selective for GS1, successfully lowering glycogen levels in the abnormal areas yet not impacting overall blood glucose homeostasis. Further work involving structural analysis of GS2 would aid this process.

Another potential therapeutic avenue for GSDs is to inhibit GS dephosphorylation, and thus holding GS in an inactive state, by targeting PTG, the subunit of PP1 which targets the phosphatase to GS. A mutation of PP1R3A (R_{GL}), the regulatory subunit of PP1, directly impairs glycogen synthesis and decreases glycogen levels in human skeletal muscle through decreasing GS activity and increasing GP activity[130]. It has been suggested that this mutation may be involved in the development of type 2 diabetes[130], however it also shows that dysfunctional PP1 is not associated with detrimental health consequences that cannot be treated. Moreover, reduction in muscle glycogen is tolerated and thus this supports the safety and viability of this therapeutic approach.

6.7 Conclusions and final remarks

The structure of GS has been solved at 2.6 Å resolution by cryo-EM. This represents the first structural analysis of human GS and the first structure of the phosphorylated, inhibited state. This revealed that the phosphorylated N- and C-terminal tails form “spike” regions, and we propose that its dynamic nature provides a functional redundancy mechanism. This region holds GS in a constrained “tense” state, allowing pS641 to engage with the regulatory helices. As well as exposing phosphorylated residues to phosphatases, N- and C-termini regulatory regions allow a “tuneable rheostat” rather than an on/off switch for regulating GS activity. The structure of human GS described in this thesis represents an invaluable resource for understanding GS disease mutations, as well as for discovery and investigation of small molecule inhibitors of GS.

References

1. Roach, P. J. Glycogen and its metabolism. *Curr. Mol. Med.* **2,** 101–120 (2002).
2. Young, F. G. *Claude Bernard and the discovery of glycogen; a century of retrospect. British medical journal* **1,** 1431–1437 (BMJ Publishing Group, 1957).
3. Ellingwood, S. S. & Cheng, A. Biochemical and clinical aspects of glycogen storage diseases. *Journal of Endocrinology* **238,** R131–R141 (2018).
4. Ros, S. & Schulze, A. Linking Glycogen and Senescence in Cancer Cells. *Cell Metabolism* **16,** 687–688 (2012).
5. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* **441,** 763–787 (2012).
6. Gunja-Smith, Z., Marshall, J. J., Mercier, C., Smith, E. E. & Whelan, W. J. A revision of the Meyer-Bernfeld model of glycogen and amylopectin. *FEBS Lett* **12,** 101–104 (1970).
7. Goldsmith, E., Sprang, S. & Fletterick, R. Structure of maltoheptaose by difference Fourier methods and a model for glycogen. *Journal of Molecular Biology* **156,** 411–427 (1982).
8. Meléndez, R., Meléndez-Hevia, E. & Cascante, M. How did glycogen structure evolve to satisfy the requirement for rapid mobilization of glucose? A problem of physical constraints in structure building. *J Mol Evol* **45,** 446–455 (1997).
9. Marchand, I. *et al.* Quantification of subcellular glycogen in resting human muscle: granule size, number, and location. *Journal of Applied Physiology* **93,** 1598–1607 (2002).
10. Rybicka, K. K. Glycosomes--the organelles of glycogen metabolism. *Tissue and Cell* **28,** 253–265 (1996).
11. Meyer, F., Heilmeyer, L. M., Haschke, R. H. & Fischer, E. H. Control of phosphorylase activity in a muscle glycogen particle. I. Isolation and characterization of the protein-glycogen complex. *J. Biol. Chem.* **245,** 6642–6648 (1970).
12. Greenberg, C. C., Jurczak, M. J., Danos, A. M. & Brady, M. J. Glycogen branches out: new perspectives on the role of glycogen metabolism in the integration of metabolic pathways. *Am J Physiol Endocrinol Metab* **291,** E1–8 (2006).
13. Liu, Q.-H. *et al.* From Prokaryotes to Eukaryotes: Insights Into the Molecular Structure of Glycogen Particles. *Front Mol Biosci* **8,** 673315 (2021).

14. DROCHMANS, P. [Morphology of glycogen. Electron microscopic study of the negative stains of particulate glycogen]. *J Ultrastruct Res* **6,** 141–163 (1962).
15. Stapleton, D. *et al.* Analysis of hepatic glycogen-associated proteins. *Proteomics* **10,** 2320–2329 (2010).
16. Stapleton, D. *et al.* The 3T3-L1 adipocyte glycogen proteome. *Proteome Sci* **11,** 11–13 (2013).
17. Chee, N. P. & Geddes, R. The structure of liver glycogen. *FEBS Lett* **73,** 164–166 (1977).
18. Besford, Q. A. *et al.* Protein Component of Oyster Glycogen Nanoparticles: An Anchor Point for Functionalization. *ACS Appl Mater Interfaces* **12,** 38976–38988 (2020).
19. Powell, P. O. *et al.* Acid hydrolysis and molecular density of phytoglycogen and liver glycogen helps understand the bonding in glycogen α (composite) particles. *PLoS One* **10,** e0121337 (2015).
20. Sullivan, M. A. *et al.* Molecular insights into glycogen α-particle formation. *Biomacromolecules* **13,** 3805–3813 (2012).
21. Torija, M.-J. *et al.* Glycogen synthesis in the absence of glycogenin in the yeast Saccharomyces cerevisiae. *FEBS Lett* **579,** 3999–4004 (2005).
22. Testoni, G. *et al.* Lack of Glycogenin Causes Glycogen Accumulation and Muscle Function Impairment. *Cell Metabolism* **26,** 256–266.e4 (2017).
23. Liu, Q.-H. *et al.* From Prokaryotes to Eukaryotes: Insights Into the Molecular Structure of Glycogen Particles. *Front Mol Biosci* **8,** 673315 (2021).
24. Prats, C., Gómez-Cabello, A. & Hansen, A. V. Intracellular compartmentalization of skeletal muscle glycogen metabolism and insulin signalling. *Experimental Physiology* **96,** 385–390 (2011).
25. Ball, S. *et al.* From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. *Cell* **86,** 349–352 (1996).
26. Fridén, J., Seger, J. & EKBLOM, B. Topographical localization of muscle glycogen: an ultrahistochemical study in the human vastus lateralis. *Acta Physiologica Scandinavica* **135,** 381–391 (1989).
27. Marchand, I. *et al.* Quantitative assessment of human muscle glycogen granules size and number in subcellular locations during recovery from prolonged exercise. *The Journal of Physiology* **580,** 617–628 (2007).
28. Lees, S. J., Franks, P. D., Spangenburg, E. E. & Williams, J. H. Glycogen and glycogen phosphorylase associated with sarcoplasmic reticulum: effects of fatiguing activity. *Journal of Applied Physiology* **91,** 1638–1644 (2001).

29. Nielsen, J., Schrøder, H. D., Rix, C. G. & Ørtenblad, N. Distinct effects of subcellular glycogen localization on tetanic relaxation time and endurance in mechanically skinned rat skeletal muscle fibres. *The Journal of Physiology* **587,** 3679–3690 (2009).
30. Thorens, B. & Mueckler, M. Glucose transporters in the 21st Century. *Am J Physiol Endocrinol Metab* **298,** E141–5 (2010).
31. Navale, A. M. & Paranjape, A. N. Glucose transporters: physiological and pathological roles. *Biophys Rev* **8,** 5–9 (2016).
32. Agius, L. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem. J.* **414,** 1–18 (2008).
33. Thorens, B. GLUT2, glucose sensing and glucose homeostasis. *Diabetologia* **58,** 221–232 (2015).
34. Simpson, I. A. *et al.* The facilitative glucose transporter GLUT3: 20 years of distinction. *Am J Physiol Endocrinol Metab* **295,** E242–53 (2008).
35. Bryant, N. J., Govers, R. & James, D. E. Regulated transport of the glucose transporter GLUT4. *Nat Rev Mol Cell Biol* **3,** 267–277 (2002).
36. Di Mauro, S. Muscle glycogenoses: an overview. *Acta Myol* **26,** 35–41 (2007).
37. Nogales-Gadea, G. *et al.* The pathogenomics of McArdle disease--genes, enzymes, models, and therapeutic implications. *J Inherit Metab Dis* **38,** 221–230 (2015).
38. Nelson, T. E., Kolb, E. & LARNER, J. Purification and properties of rabbit muscle amylo-1,6-glucosidase-oligo-1,4-1,4-transferase. *Biochemistry* **8,** 1419–1428 (1969).
39. BROWN, D. H. & ILLINGWORTH, B. The properties of an oligo-1,4--1,4-glucantransferase from animal tissues. *Proceedings of the National Academy of Sciences* **48,** 1783–1787 (1962).
40. Raben, N., Plotz, P. & Byrne, B. J. Acid alpha-glucosidase deficiency (glycogenosis type II, Pompe disease). *Curr. Mol. Med.* **2,** 145–166 (2002).
41. Campbell, J., Davies, G., Bulone, V. & Henrissat, B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **329 (Pt 3),** 719 (1998).
42. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *Journal of Molecular Biology* **328,** 307–317 (2003).
43. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50,** D571–D577 (2022).

44. Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A. & Imberty, A. Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *J Biochem* **123,** 1000–1009 (1998).
45. NAD-binding domains of dehydrogenases. *Current Opinion in Structural Biology* **5,** 775–783 (1995).
46. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77,** 521–555 (2008).
47. Structure/function studies of glycosyltransferases. *Current Opinion in Structural Biology* **9,** 563–571 (1999).
48. Taujale, R. *et al.* Mapping the glycosyltransferase fold landscape using interpretable deep learning. *Nature Communications* **12,** 5656–12 (2021).
49. An Evolving Hierarchical Family Classification for Glycosyltransferases. *Journal of Molecular Biology* **328,** 307–317 (2003).
50. Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R. & Davies, G. J. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem Biol* **9,** 1337–1346 (2002).
51. Sommer, N., Depping, R., Piotrowski, M. & Rüger, W. Bacteriophage T4 alpha-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. *Biochemical and Biophysical Research Communications* **323,** 809–815 (2004).
52. Cid, E., Gomis, R. R., Geremia, R. A., Guinovart, J. J. & Ferrer, J. C. Identification of two essential glutamic acid residues in glycogen synthase. *J. Biol. Chem.* **275,** 33614–33621 (2000).
53. McNamara, J. T., Morgan, J. L. W. & Zimmer, J. A molecular description of cellulose biosynthesis. *Annu. Rev. Biochem.* **84,** 895–921 (2015).
54. Cori, C. F., Schmidt, G. & Cori, G. T. THE SYNTHESIS OF A POLYSACCHARIDE FROM GLUCOSE-1-PHOSPHATE IN MUSCLE EXTRACT. *Science* **89,** 464–465 (1939).
55. LELOIR, L. F., Society, C. C. J. O. T. A. C.1957. BIOSYNTHESIS OF GLYCOGEN FROM URIDINE DIPHOSPHATE GLUCOSE1. *ACS Publications*
56. Krisman, C. R. & Barengo, R. A precursor of glycogen biosynthesis: alpha-1,4-glucan-protein. *European Journal of Biochemistry* **52,** 117–123 (1975).
57. Kennedy, L. D., muscle, B. K. A.1985. *The biogenesis of rabbit-muscle glycogen*. *ICSU Press/IRL Press Oxford*

58. Aon, M. A. & Curtino, J. A. Evidence for the glycoprotein nature of retina glycogen. *European Journal of Biochemistry* **140,** 557–566 (1984).
59. AON, M. A. & Curtino, J. A. Protein-bound glycogen is linked to tyrosine residues. *Biochem. J.* **229,** 269–272 (1985).
60. Rodriguez, I. R. & Whelan, W. J. A novel glycosyl-amino acid linkage: rabbit-muscle glycogen is covalently linked to a protein via tyrosine. *Biochemical and Biophysical Research Communications* **132,** 829–836 (1985).
61. Campbell, D. G. & Cohen, P. The amino acid sequence of rabbit skeletal muscle glycogenin. *European Journal of Biochemistry* **185,** 119–125 (1989).
62. Nimmo, H. G., Proud, C. G. & Cohen, P. The purification and properties of rabbit skeletal muscle glycogen synthase. *European Journal of Biochemistry* **68,** 21–30 (1976).
63. Pitcher, J., Smythe, C., Campbell, D. G. & Cohen, P. Identification of the 38-kDa subunit of rabbit skeletal muscle glycogen synthase as glycogenin. *European Journal of Biochemistry* **169,** 497–502 (1987).
64. Pitcher, J., of, C. S. E. J.1988. Glycogenin is the priming glucosyltransferase required for the initiation of glycogen biogenesis in rabbit skeletal muscle. *European Journal of Biochemistry* **176,** 391–395 (1988).
65. Mu, J., Skurat, A. V. & Roach, P. J. Glycogenin-2, a Novel Self-glucosylating Protein Involved in Liver Glycogen Biosynthesis. *J. Biol. Chem.* **272,** 27589–27597 (1997).
66. Curtino, J. A. & Aon, M. A. From the seminal discovery of proteoglycogen and glycogenin to emerging knowledge and research on glycogen biology. *Biochem. J.* **476,** 3109–3124 (2019).
67. Zhai, L., Schroeder, J., Skurat, A. V. & Roach, P. J. Do Rodents Have a Gene Encoding Glycogenin-2, the Liver Isoform of the Self-Glucosylating Initiator of Glycogen Synthesis? *IUBMB Life* **51,** 87–91 (2001).
68. Gibbons, B. J., Roach, P. J. & Hurley, T. D. Crystal Structure of the Autocatalytic Initiator of Glycogen Biosynthesis, Glycogenin. *Journal of Molecular Biology* **319,** 463–477 (2002).
69. Skurat, A. V., Dietrich, A. D. & Roach, P. J. Interaction between glycogenin and glycogen synthase. *Archives of Biochemistry and Biophysics* **456,** 93–97 (2006).
70. Zeqiraj, E. *et al.* Structural basis for the recruitment of glycogen synthase by glycogenin. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E2831–40 (2014).
71. Baqué, S., Guinovart, J. J. & Ferrer, J. C. Glycogenin, the primer of glycogen synthesis, binds to actin. *FEBS Lett* **417,** 355–359 (1997).

72. Prats, C., Graham, T. E. & Shearer, J. The dynamic life of the glycogen granule. *J. Biol. Chem.* **293,** 7089–7098 (2018).

73. Zhai, L., Mu, J., Zong, H., Depaoli-Roach, A. A. & Roach, P. J. Structure and chromosomal localization of the human glycogenin-2 gene GYG2. *Gene* **242,** 229–235 (2000).

74. Hansen, B. F., Derave, W., Jensen, P. & Richter, E. A. No limiting role for glycogenin in determining maximal attainable glycogen levels in rat skeletal muscle. *Am J Physiol Endocrinol Metab* **278,** E398–404 (2000).

75. Skurat, A. V., Wang, Y. & Roach, P. J. Rabbit skeletal muscle glycogen synthase expressed in COS cells. Identification of regulatory phosphorylation sites. *J. Biol. Chem.* **269,** 25534–25542 (1994).

76. Skurat, A. V., Lim, S. S. & Roach, P. J. Glycogen biogenesis in rat 1 fibroblasts expressing rabbit muscle glycogenin. *European Journal of Biochemistry* **245,** 147–155 (1997).

77. Moslemi, A.-R. *et al.* Glycogenin-1 Deficiency and Inactivated Priming of Glycogen Synthesis. *http://dx.doi.org/10.1056/NEJMoa0900661* (2010). doi:10.1056/NEJMoa0900661

78. Visuttijai, K. *et al.* Glycogenin is Dispensable for Glycogen Synthesis in Human Muscle, and Glycogenin Deficiency Causes Polyglucosan Storage. *J Clin Endocrinol Metab* **105,** (2020).

79. Issoglio, F. M., Carrizo, M. E., Romero, J. M. & Curtino, J. A. Mechanisms of monomeric and dimeric glycogenin autoglucosylation. *J. Biol. Chem.* **287,** 1955–1961 (2012).

80. Chaikuad, A. *et al.* Conformational plasticity of glycogenin and its maltosaccharide substrate during glycogen biogenesis. *Proceedings of the National Academy of Sciences* **108,** 21028–21033 (2011).

81. Bilyard, M. K. *et al.* Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* 1–24 (2018). doi:10.1038/s41586-018-0644-7

82. Buschiazzo, A. *et al.* Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. *The EMBO Journal* **23,** 3196–3205 (2004).

83. Wilamowitz-Moellendorff, von, A. *et al.* Glucose-6-phosphate-mediated activation of liver glycogen synthase plays a key role in hepatic glycogen synthesis. *Diabetes* **62,** 4070–4082 (2013).

84. Browner, M. F., Nakano, K., Bang, A. G. & Fletterick, R. J. Human muscle glycogen synthase cDNA sequence: a negatively charged protein with an

asymmetric charge distribution. *Proceedings of the National Academy of Sciences* **86,** 1443–1447 (1989).

85. Nuttall, F. Q., Gannon, M. C., Bai, G. & Lee, E. Y. Primary structure of human liver glycogen synthase deduced by cDNA cloning. *Archives of Biochemistry and Biophysics* **311,** 443–449 (1994).
86. Sheng, F., Jia, X., Yep, A., Preiss, J. & Geiger, J. H. The crystal structures of the open and catalytically competent closed conformation of Escherichia coli glycogen synthase. *J. Biol. Chem.* **284,** 17796–17807 (2009).
87. Horcajada, C., Guinovart, J. J., Fita, I. & Ferrer, J. C. Crystal structure of an archaeal glycogen synthase: insights into oligomerization and substrate binding of eukaryotic glycogen synthases. *J. Biol. Chem.* **281,** 2923–2931 (2006).
88. Baskaran, S., Roach, P. J., Depaoli-Roach, A. A. & Hurley, T. D. Structural basis for glucose-6-phosphate activation of glycogen synthase. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17563–17568 (2010).
89. Chikwana, V. M. *et al.* Structural basis for 2'-phosphate incorporation into glycogen by glycogen synthase. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 20976–20981 (2013).
90. Baskaran, S. *et al.* Multiple glycogen-binding sites in eukaryotic glycogen synthase are required for high catalytic efficiency toward glycogen. *J. Biol. Chem.* **286,** 33999–34006 (2011).
91. Mahalingan, K. K., Baskaran, S., Depaoli-Roach, A. A., Roach, P. J. & Hurley, T. D. Redox Switch for the Inhibited State of Yeast Glycogen Synthase Mimics Regulation by Phosphorylation. *Biochemistry* **56,** 179–188 (2017).
92. Zeqiraj, E. *et al.* Structural basis for the recruitment of glycogen synthase by glycogenin. *Proceedings of the National Academy of Sciences* **111,** E2831–E2840 (2014).
93. LARNER, J., Takeda, Y. & Hizukuri, S. The influence of chain size and molecular weight on the kinetic constants for the span glucose to polysaccharide for rabbit muscle glycogen synthase. *Mol Cell Biochem* **12,** 131–136 (1976).
94. Sheng, F., Yep, A., Feng, L., Preiss, J. & Geiger, J. H. Oligosaccharide Binding in Escherichia coli Glycogen Synthase. *Biochemistry* **48,** 10089–10097 (2009).
95. Díaz, A., Martínez-Pons, C., Fita, I., Ferrer, J. C. & Guinovart, J. J. Processivity and Subcellular Localization of Glycogen Synthase Depend on a Non-catalytic High Affinity Glycogen-binding Site. *J. Biol. Chem.* **286,** 18505–18514 (2011).

96. Nimmo, H. G., Proud, C. G. & Cohen, P. The purification and properties of rabbit skeletal muscle glycogen synthase. *European Journal of Biochemistry* **68,** 21–30 (1976).

97. Zeqiraj, E. & Sicheri, F. Getting a handle on glycogen synthase – Its interaction with glycogenin. *Molecular Aspects of Medicine* **46,** 63–69 (2015).

98. Hunter, R. W., Zeqiraj, E., Morrice, N., Sicheri, F. & Sakamoto, K. Expression and purification of functional human glycogen synthase-1:glycogenin-1 complex in insect cells. *PROTEIN EXPRESSION AND PURIFICATION* **108,** 23–29 (2015).

99. Khanna, M. *et al.* Expression and purification of functional human glycogen synthase-1 (hGYS1) in insect cells. *PROTEIN EXPRESSION AND PURIFICATION* **90,** 78–83 (2013).

100. Gibson, W. B., Illingsworth, B. & BROWN, D. H. Studies of glycogen branching enzyme. Preparation and properties of -1,4-glucan- -1,4-glucan 6-glycosyltransferase and its action on the characteristic polysaccharide of the liver of children with Type IV glycogen storage disease. *Biochemistry* **10,** 4253–4262 (1971).

101. Moses, S. W. & Parvari, R. The variable presentations of glycogen storage disease type IV: a review of clinical, enzymatic and molecular studies. *Curr. Mol. Med.* **2,** 177–188 (2002).

102. Pederson, B. A., Cheng, C., Wilson, W. A. & Roach, P. J. Regulation of Glycogen Synthase IDENTIFICATION OF RESIDUES INVOLVED IN REGULATION BY THE ALLOSTERIC LIGAND GLUCOSE-6-P AND BY PHOSPHORYLATION. *J. Biol. Chem.* **275,** 27753–27761 (2000).

103. Yeast Glycogen Synthetase in the Glucose 6-Phosphate-dependent Form: II. THE EFFECT OF PROTEOLYSIS. *J. Biol. Chem.* **249,** 3858–3861 (1974).

104. Wilson, W. A. *et al.* Regulation of glycogen metabolism in yeast and bacteria. *FEMS Microbiol Rev* **34,** 952–985 (2010).

105. Smith, C. H., Brown, N. E. & LARNER, J. Molecular characteristics of the totally dependent and independent forms of glycogen synthase of rabbit skeletal muscle. II. Some chemical characteristics of the enzyme protein and of its change on interconversion. *Biochim Biophys Acta* **242,** 81–88 (1971).

106. Carling, D. & Hardie, D. G. The substrate and sequence specificity of the AMP-activated protein kinase. Phosphorylation of glycogen synthase and phosphorylase kinase. *Biochim Biophys Acta* **1012,** 81–86 (1989).

107. Roach, P. J., DePaoli-Roach, A. A. & LARNER, J. Ca2+-stimulated phosphorylation of muscle glycogen synthase by phosphorylase b kinase. *J Cyclic Nucleotide Res* **4,** 245–257 (1978).

108. Ahmad, Z., DePaoli-Roach, A. A. & Roach, P. J. Purification and characterization of a rabbit liver calmodulin-dependent protein kinase able to phosphorylate glycogen synthase. *J. Biol. Chem.* **257,** 8348–8355 (1982).

109. Huang, T. S. & Krebs, E. G. Amino acid sequence of a phosphorylation site in skeletal muscle glycogen synthetase. *Biochemical and Biophysical Research Communications* **75,** 643–650 (1977).

110. Ahmad, Z., Lee, F. T., DePaoli-Roach, A. & Roach, P. J. Phosphorylation of glycogen synthase by the Ca2+- and phospholipid-activated protein kinase (protein kinase C). *J. Biol. Chem.* **259,** 8743–8747 (1984).

111. Flotow, H. & Roach, P. J. Synergistic phosphorylation of rabbit muscle glycogen synthase by cyclic AMP-dependent protein kinase and casein kinase I. Implications for hormonal regulation of glycogen synthase. *J. Biol. Chem.* **264,** 9126–9128 (1989).

112. Flotow, H. *et al.* Phosphate groups as substrate determinants for casein kinase I action. *J. Biol. Chem.* **265,** 14264–14269 (1990).

113. Picton, C., Woodgett, J., Hemmings, B. & Cohen, P. Multisite phosphorylation of glycogen synthase from rabbit skeletal muscle. Phosphorylation of site 5 by glycogen synthase kinase-5 (casein kinase-II) is a prerequisite for phosphorylation of sites 3 by glycogen synthase kinase-3. *FEBS Lett* **150,** 191–196 (1982).

114. DePaoli-Roach, A. A., Ahmad, Z., Camici, M., Lawrence, J. C. & Roach, P. J. Multiple phosphorylation of rabbit skeletal muscle glycogen synthase. Evidence for interactions among phosphorylation sites and the resolution of electrophoretically distinct forms of the subunit. *J. Biol. Chem.* **258,** 10702–10709 (1983).

115. Wilson, W. A. *et al.* Control of mammalian glycogen synthase by PAS kinase. *Proceedings of the National Academy of Sciences* **102,** 16596–16601 (2005).

116. Skurat, A. V. & Dietrich, A. D. Phosphorylation of Ser640 in muscle glycogen synthase by DYRK family protein kinases. *J. Biol. Chem.* **279,** 2490–2498 (2004).

117. Fiol, C. J., Mahrenholz, A. M., Wang, Y., Roeske, R. W. & Roach, P. J. Formation of protein kinase recognition sites by covalent modification of the substrate. Molecular mechanism for the synergistic action of casein kinase II and glycogen synthase kinase 3. *J. Biol. Chem.* **262,** 14042–14048 (1987).

118. Roach, P. J. Multisite and hierarchal protein phosphorylation. *J. Biol. Chem.* **266,** 14139–14142 (1991).

119. Zhang, W. M., Depaoliroach, A. A. & Roach, P. J. Mechanisms of Multisite Phosphorylation and Inactivation of Rabbit Muscle Glycogen Synthase. *Archives of Biochemistry and Biophysics* **304,** 219–225 (1993).

120. Skurat, A. V. & Roach, P. J. Phosphorylation of Sites 3a and 3b (Ser640 and Ser644) in the Control of Rabbit Muscle Glycogen Synthase *. *J. Biol. Chem.* **270,** 12491–12497 (1995).

121. Wang, Y. & Roach, P. J. Inactivation of rabbit muscle glycogen synthase by glycogen synthase kinase-3. Dominant role of the phosphorylation of Ser-640 (site-3a). *J. Biol. Chem.* **268,** 23876–23880 (1993).

122. Cohen, P. T. & Cohen, P. Discovery of a protein phosphatase activity encoded in the genome of bacteriophage lambda. Probable identity with open reading frame 221. *Biochem. J.* **260,** 931–934 (1989).

123. Kumar, G. S. *et al.* Identification of the substrate recruitment mechanism of the muscle glycogen protein phosphatase 1 holoenzyme. *Sci Adv* **4,** eaau6044 (2018).

124. Christiansen, C. *et al.* The carbohydrate-binding module family 20--diversity, structure, and function. *FEBS J* **276,** 5006–5029 (2009).

125. Fong, N. M. *et al.* Identification of binding sites on protein targeting to glycogen for enzymes of glycogen metabolism. *J. Biol. Chem.* **275,** 35034–35039 (2000).

126. Printen, J. A., Brady, M. J. & Saltiel, A. R. PTG, a protein phosphatase 1-binding protein with a role in glycogen metabolism. *Science* **275,** 1475–1478 (1997).

127. Berman, H. K., O'Doherty, R. M., Anderson, P. & Newgard, C. B. Overexpression of protein targeting to glycogen (PTG) in rat hepatocytes causes profound activation of glycogen synthesis independent of normal hormone- and substrate-mediated regulatory mechanisms. *J. Biol. Chem.* **273,** 26421–26425 (1998).

128. Strålfors, P., Hiraga, A. & Cohen, P. The protein phosphatases involved in cellular regulation. Purification and characterisation of the glycogen-bound form of protein phosphatase-1 from rabbit skeletal muscle. *European Journal of Biochemistry* **149,** 295–303 (1985).

129. Aschenbach, W. G. *et al.* The muscle-specific protein phosphatase PP1G/R(GL)(G(M))is essential for activation of glycogen synthase by exercise. *J. Biol. Chem.* **276,** 39959–39967 (2001).

130. Savage, D. B. *et al.* A prevalent variant in PPP1R3A impairs glycogen synthesis and reduces muscle glycogen content in humans and mice. *PLoS Med* **5,** e27 (2008).

131. Savage, D. B. *et al.* Digenic inheritance of severe insulin resistance in a human pedigree. *Nat Genet* **31,** 379–384 (2002).

132. VILLAR-PALASI, C. & LARNER, J. Insulin-mediated effect on the activity of UDPG-glycogen transglucosylase of muscle. *Biochim Biophys Acta* **39,** 171–173 (1960).

133. Valentine, R. J., Coughlan, K. A., Ruderman, N. B. & Saha, A. K. Insulin inhibits AMPK activity and phosphorylates AMPK Ser$^{485/491}$ through Akt in hepatocytes, myotubes and incubated rat skeletal muscle. *Archives of Biochemistry and Biophysics* **562,** 62–69 (2014).

134. Newgard, C. B., Brady, M. J., O'Doherty, R. M. & Saltiel, A. R. Organizing glucose disposal: emerging roles of the glycogen targeting subunits of protein phosphatase-1. *Diabetes* **49,** 1967–1977 (2000).

135. Han, H.-S., Kang, G., Kim, J. S., Choi, B. H. & Koo, S.-H. Regulation of glucose metabolism from a liver-centric perspective. *Exp Mol Med* **48,** e218–e218 (2016).

136. Bouskila, M., Hirshman, M. F., Jensen, J., Goodyear, L. J. & Sakamoto, K. Insulin promotes glycogen synthesis in the absence of GSK3 phosphorylation in skeletal muscle. *Am J Physiol Endocrinol Metab* **294,** E28–35 (2008).

137. Saltiel, A. R. Insulin signaling in health and disease. *J Clin Invest* **131,** (2021).

138. Jiang, G. & Zhang, B. B. Glucagon and regulation of glucose metabolism. *Am J Physiol Endocrinol Metab* **284,** E671–8 (2003).

139. Dietz, M. R., Chiasson, J. L., Soderling, T. R. & Exton, J. H. Epinephrine regulation of skeletal muscle glycogen metabolism. Studies utilizing the perfused rat hindlimb preparation. *J. Biol. Chem.* **255,** 2301–2307 (1980).

140. MacKintosh, C., Campbell, D. G., Hiraga, A. & Cohen, P. Phosphorylation of the glycogen-binding subunit of protein phosphatase-1G in response to adrenalin. *FEBS Lett* **234,** 189–194 (1988).

141. Foulkes, J. G. & Cohen, P. The hormonal control of glycogen metabolism. Phosphorylation of protein phosphatase inhibitor-1 in vivo in response to adrenaline. *European Journal of Biochemistry* **97,** 251–256 (1979).

142. Nielsen, J. N. & Wojtaszewski, J. F. P. Regulation of glycogen synthase activity and phosphorylation by exercise. *Proc Nutr Soc* **63,** 233–237 (2004).

143. Roach, P. J., Depaoli-Roach, A. A., Hurley, T. D. & Tagliabracci, V. S. Glycogen and its metabolism: some new developments and old themes. *Biochem. J.* **441,** 763–787 (2012).

144. Katz, A. A century of exercise physiology: key concepts in regulation of glycogen metabolism in skeletal muscle. *Eur J Appl Physiol* 1–22 (2022). doi:10.1007/s00421-022-04935-1

145. Kraniou, Y., Cameron-Smith, D., Misso, M., Collier, G. & Hargreaves, M. Effects of exercise on GLUT-4 and glycogenin gene expression in human skeletal muscle. *Journal of Applied Physiology* **88,** 794–796 (2000).

146. Kohler, L., Puertollano, R. & Raben, N. Pompe Disease: From Basic Science to Therapy. *Neurotherapeutics* **15,** 928–942 (2018).

147. Leslie, N. & Bailey, L. Pompe disease. (2017).

148. Peruzzo, P., Pavan, E. & Dardis, A. Molecular genetics of Pompe disease: a comprehensive overview. *Ann Transl Med* **7,** 278–4 (2019).

149. Byrne, B. J. *et al.* Pompe disease: design, methodology, and early findings from the Pompe Registry. *Molecular Genetics and Metabolism* **103,** 1–11 (2011).

150. Schoser, B. Pompe disease: what are we missing? *Ann Transl Med* **7,** 292–18 (2019).

151. Meena, N. K., Ralston, E., Raben, N. & Puertollano, R. Enzyme Replacement Therapy Can Reverse Pathogenic Cascade in Pompe Disease. *Mol Ther Methods Clin Dev* **18,** 199–214 (2020).

152. Douillard-Guilloux, G. *et al.* Restoration of muscle functionality by genetic suppression of glycogen synthesis in a murine model of Pompe disease. *Hum. Mol. Genet.* **19,** 684–696 (2010).

153. Clayton, N. P. *et al.* Antisense Oligonucleotide-mediated Suppression of Muscle Glycogen Synthase 1 Synthesis as an Approach for Substrate Reduction Therapy of Pompe Disease. *Mol Ther Nucleic Acids* **3,** e206 (2014).

154. Heinicke, K. *et al.* Reproducibility and absolute quantification of muscle glycogen in patients with glycogen storage disease by 13C NMR spectroscopy at 7 Tesla. *PLoS One* **9,** e108706 (2014).

155. Yi, H. *et al.* Characterization of a canine model of glycogen storage disease type IIIa. *Dis Model Mech* **5,** 804–811 (2012).

156. Ribeiro, L. F. P. *et al.* Effects of swimming training on tissue glycogen content in experimental thyrotoxic rats. *Can J Physiol Pharmacol* **90,** 587–593 (2012).

157. López-Soldado, I. *et al.* Liver glycogen reduces food intake and attenuates obesity in a high-fat diet-fed mouse model. *Diabetes* **64,** 796–807 (2015).

158. Delgado-Escueta, A. V. Advances in lafora progressive myoclonus epilepsy. *Curr Neurol Neurosci Rep* **7,** 428–433 (2007).

159. Minassian, B. A. Lafora's disease: towards a clinical, pathologic, and molecular synthesis. *Pediatr Neurol* **25,** 21–29 (2001).

160. Striano, P. *et al.* Typical progression of myoclonic epilepsy of the Lafora type: a case report. *Nat Clin Pract Neurol* **4,** 106–111 (2008).

161. Chan, E. M., Andrade, D. M., Franceschetti, S. & Minassian, B. Progressive myoclonus epilepsies: EPM1, EPM2A, EPM2B. *Adv Neurol* **95,** 47–57 (2005).

162. Ganesh, S., Puri, R., Singh, S., Mittal, S. & Dubey, D. Recent advances in the molecular basis of Lafora's progressive myoclonus epilepsy. *J Hum Genet* **51,** 1–8 (2006).

163. Minassian, B. A. *et al.* Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. *Nat Genet* **20,** 171–174 (1998).

164. Chan, E. M. *et al.* Laforin preferentially binds the neurotoxic starch-like polyglucosans, which form in its absence in progressive myoclonus epilepsy. *Hum. Mol. Genet.* **13,** 1117–1129 (2004).

165. Wang, J., Stuckey, J. A., Wishart, M. J. & Dixon, J. E. A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen. *J. Biol. Chem.* **277,** 2377–2380 (2002).

166. Chan, E. M. *et al.* Mutations in NHLRC1 cause progressive myoclonus epilepsy. *Nat Genet* **35,** 125–127 (2003).

167. Gentry, M. S., Worby, C. A. & Dixon, J. E. Insights into Lafora disease: Malin is an E3 ubiquitin ligase that ubiquitinates and promotes the degradation of laforin. *Proceedings of the National Academy of Sciences* **102,** 8501–8506 (2005).

168. Vilchez, D. *et al.* Mechanism suppressing glycogen synthesis in neurons and its demise in progressive myoclonus epilepsy. *Nat Neurosci* **10,** 1407–1413 (2007).

169. Singh, P. K., Singh, S. & Ganesh, S. The laforin-malin complex negatively regulates glycogen synthesis by modulating cellular glucose uptake via glucose transporters. *Molecular and Cellular Biology* **32,** 652–663 (2012).

170. Roach, P. J. Are there errors in glycogen biosynthesis and is laforin a repair enzyme? *FEBS Lett* **585,** 3216–3218 (2011).

171. Tagliabracci, V. S. *et al.* Laforin is a glycogen phosphatase, deficiency of which leads to elevated phosphorylation of glycogen in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 19262–19266 (2007).

172. Tagliabracci, V. S. *et al.* Phosphate incorporation during glycogen synthesis and Lafora disease. *Cell Metabolism* **13,** 274–282 (2011).

173. Romá-Mateo, C., Sanz, P. & Gentry, M. S. Deciphering the role of malin in the lafora progressive myoclonus epilepsy. *IUBMB Life* **64,** 801–808 (2012).

174. Donohue, K. J., Gentry, M. S. & Sun, R. C. The E3 ligase malin plays a pivotal role in promoting nuclear glycogenolysis and histone acetylation. *Ann Transl Med* **8,** 254–254 (2020).

175. García-Gimeno, M., Knecht, E. & Sanz, P. Lafora Disease: A Ubiquitination-Related Pathology. *Cells* **7,** 87–17 (2018).

176. Worby, C. A., Gentry, M. S. & Dixon, J. E. Malin decreases glycogen accumulation by promoting the degradation of protein targeting to glycogen (PTG). *J. Biol. Chem.* **283,** 4069–4076 (2008).

177. Ahonen, S. *et al.* Gys1 antisense therapy rescues neuropathological bases of murine Lafora disease. *Brain* (2021). doi:10.1093/brain/awab194

178. Varea, O., Duran, J., Aguilera, M., Prats, N. & Guinovart, J. J. Suppression of glycogen synthesis as a treatment for Lafora disease: Establishing the window of opportunity. *Neurobiol Dis* **147,** 105173 (2021).

179. Nitschke, S. *et al.* An inducible glycogen synthase-1 knockout halts but does not reverse Lafora disease progression in mice. *J. Biol. Chem.* **296,** 100150 (2021).

180. Gertz, H. J., Cervos-Navarro, J., Frydl, V. & Schultz, F. Glycogen accumulation of the aging human brain. *Mech Ageing Dev* **31,** 25–35 (1985).

181. Duran, J. *et al.* Deleterious effects of neuronal accumulation of glycogen in flies and mice. *EMBO Mol Med.* **4,** 719–729 (2012).

182. Sinadinos, C. *et al.* Neuronal glycogen synthesis contributes to physiological aging. *Aging Cell* **13,** 935–945 (2014).

183. Nielsen, J. *et al.* Subcellular localization-dependent decrements in skeletal muscle glycogen and mitochondria content following short-term disuse in young and old men. *Am J Physiol Endocrinol Metab* **299,** E1053–60 (2010).

184. Seo, Y.-H. *et al.* Enhanced glycogenesis is involved in cellular senescence via GSK3/GS modulation. *Aging Cell* **7,** 894–907 (2008).

185. Khandelwal, R. L., Enno, T. L. & Narayanan, N. Effects of age on glycogen synthase and phosphorylase activities in rat liver. *Mech Ageing Dev* **28,** 13–22 (1984).

186. Dall'Aglio, E., Chang, H., Reaven, G. M. & Azhar, S. Age-related changes in rat muscle glycogen synthase activity. *J Gerontol* **42,** 168–172 (1987).

187. Moore, S. A., Peterson, R. G., Felten, D. L. & O'Connor, B. L. Glycogen accumulation in tibial nerves of experimentally diabetic and aging control rats. *J Neurol Sci* **52,** 289–303 (1981).

188. Cheong, J.-W. *et al.* Constitutive phosphorylation of FKHR transcription factor as a prognostic variable in acute myeloid leukemia. *Leuk Res* **27,** 1159–1162 (2003).

189. Eldar-Finkelman, H. & Ilouz, R. Challenges and opportunities with glycogen synthase kinase-3 inhibitors for insulin resistance and Type 2 diabetes treatment. *Expert Opin Investig Drugs* **12,** 1511–1519 (2003).

190. Minhas, P. S. *et al.* Restoring metabolism of myeloid cells reverses cognitive decline in ageing. *Nature* **590,** 122–128 (2021).

191. Kelsall, I. R. *et al.* HOIL-1 ubiquitin ligase activity targets unbranched glucosaccharides and is required to prevent polyglucosan accumulation. *The EMBO Journal* e109700 (2022). doi:10.15252/embj.2021109700

192. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nat Methods* **3,** 1021–1032 (2006).

193. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112,** 531–552 (1999).

194. Doerr, A. Single-particle cryo-electron microscopy. *Nat Methods* **13,** 23–23 (2016).

195. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online* **6,** 23–34 (2004).

196. Scarff, C. A., Fuller, M. J. G., Thompson, R. F. & Iadaza, M. G. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. *JoVE* 1–8 (2018). doi:10.3791/57199

197. Passmore, L. A. & Russo, C. J. Specimen Preparation for High-Resolution Cryo-EM. *Methods Enzymol* **579,** 51–86 (2016).

198. Grassucci, R. A., Taylor, D. J. & Frank, J. Preparation of macromolecular complexes for cryo-electron microscopy. *Nat Protoc* **2,** 3239–3246 (2007).

199. Bhella, D. Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev* **11,** 515–519 (2019).

200. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* **21,** 129–228 (1988).

201. Dubochet, J., Microscopy, A. M. J. O.1981. Vitrification of pure water for electron microscopy. *serval.unil.ch*

202. Levitz, T. S., Brignole, E. J., Fong, I., Darrow, M. C. & Drennan, C. L. Effects of chameleon dispense-to-plunge speed on particle concentration, complex formation, and final resolution: A case study using the Neisseria gonorrhoeae ribonucleotide reductase inactive complex. *Journal of Structural Biology* **214,** 107825 (2022).

203. Orlova, E. V. & Saibil, H. R. Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev* **111,** 7710–7748 (2011).

204. Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat Protoc* **14,** 100–118 (2018).

205. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14,** 331–332 (2017).

206. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6,** 5–17 (2019).

207. Guo, H. *et al.* Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. *IUCrJ* **7,** 860–869 (2020).

208. Zhang, K. Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* **193,** 1–12 (2016).

209. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192,** 216–221 (2015).

210. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLIFE* **7,** (2018).

211. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14,** 290–296 (2017).

212. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2,** 218–13 (2019).

213. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180,** 519–530 (2012).

214. Scheres, S. H. W. *Processing of Structurally Heterogeneous Cryo-EM Data in RELION*. *Processing of structurally heterogeneous cryo-EM data in RELION* **579,** 125–157 (Elsevier Inc., 2016).

215. Gilbert, P. F. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. II. Direct methods. *Proc R Soc Lond B Biol Sci* **182,** 89–102 (1972).

216. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology* **333,** 721–745 (2003).

217. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. in **20,** 205–214 (2012).

218. van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *Journal of Structural Biology* **151,** 250–262 (2005).

219. Saibil, H. R. Macromolecular structure determination by cryo-electron microscopy. *Acta Crystallogr D Biol Crystallogr* **56,** 1215–1222 (2000).

220. Fernández, J. J., Luque, D., Castón, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *Journal of Structural Biology* **164,** 170–175 (2008).

221. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213,** 107702 (2021).

222. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66,** 486–501 (2010).

223. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66,** 213–221 (2010).

224. Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* **60,** 2184–2195 (2004).

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

226. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

227. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10,** 845–858 (2015).

228. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30,** 70–82 (2021).

229. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47,** W636–W641 (2019).

230. Bond, C. S. & Schüttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr D Biol Crystallogr* **65,** 510–512 (2009).

231. Ferries, S. *et al.* Evaluation of Parameters for Confident Phosphorylation Site Localization Using an Orbitrap Fusion Tribrid Mass Spectrometer. *J Proteome Res* **16,** 3448–3459 (2017).

232. Daly, L. A. *et al.* Oxygen-dependent changes in binding partners and post-translational modifications regulate the abundance and activity of HIF-1α/2α. *Sci Signal* **14,** (2021).

233. Dephoure, N., Gould, K. L., Gygi, S. P. & Kellogg, D. R. Mapping and analysis of phosphorylation sites: a quick guide for cell biologists. *Mol Biol Cell* **24,** 535–542 (2013).

234. Candiano, G. *et al.* Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25,** 1327–1333 (2004).

235. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science* **360,** 423–427 (2018).

236. Tang, B. *et al.* Discovery and Development of Small-Molecule Inhibitors of Glycogen Synthase. *J. Med. Chem.* **63,** 3538–3551 (2020).

237. Smythe, C., Caudwell, F. B., Ferguson, M. & Cohen, P. Isolation and structural analysis of a peptide containing the novel tyrosyl-glucose linkage in glycogenin. *The EMBO Journal* **7,** 2681–2686 (1988).

238. Roach, R. J. & LARNER, J. Covalent phosphorylation in the regulation glycogen synthase activity. *Mol Cell Biochem* **15,** 179–200 (1977).

239. Bouskila, M. *et al.* Allosteric Regulation of Glycogen Synthase Controls Glycogen Synthesis in Muscle. *Cell Metabolism* **12,** 456–466 (2010).

240. McManus, E. J. *et al.* Role that phosphorylation of GSK3 plays in insulin and Wnt signalling defined by knockin analysis. *The EMBO Journal* **24,** 1571–1583 (2005).

241. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43,** D512–20 (2015).

242. Dandey, V. P. *et al.* Time-resolved cryo-EM using Spotiton. *Nat Methods* **17,** 897–900 (2020).

243. Budell, W. C., Allegri, L., Dandey, V., Potter, C. S. & Carragher, B. Cryo-Electron Microscopic Grid Preparation for Time-Resolved Studies using a Novel Robotic System, Spotiton. *JoVE* e62271 (2021). doi:10.3791/62271

244. Habeeb, A. J. & Hiramoto, R. Reaction of proteins with glutaraldehyde. *Archives of Biochemistry and Biophysics* **126,** 16–26 (1968).

245. Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques* **37,** 790–6– 798–802 (2004).

246. Witus, S. R. *et al.* BRCA1/BARD1 site-specific ubiquitylation of nucleosomal H2A is directed by BARD1. *Nat Struct Mol Biol* **28,** 268–277 (2021).

247. Hu, Q. *et al.* Mechanisms of BRCA1-BARD1 nucleosome recognition and ubiquitylation. *Nature* **596,** 438–443 (2021).

248. Bertram, K. *et al.* Structural Insights into the Roles of Metazoan-Specific Splicing Factors in the Human Step 1 Spliceosome. *Mol Cell* **80,** 127–139.e6 (2020).

249. Drulyte, I. *et al.* Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr D Struct Biol* **74,** 560–571 (2018).

250. Basanta, B., Hirschi, M. M., Grotjahn, D. A. & Lander, G. C. A case for glycerol as an acceptable additive for single-particle cryoEM samples. *Acta Crystallogr D Struct Biol* **78,** 124–135 (2022).

251. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587,** 152–156 (2020).

252. Ramlaul, K., Palmer, C. M. & Aylett, C. H. S. A Local Agreement Filtering Algorithm for Transmission EM Reconstructions. *Journal of Structural Biology* **205,** 30–40 (2019).

253. Kakhlon, O. *et al.* Guaiacol as a drug candidate for treating adult polyglucosan body disease. *JCI Insight* **3,** (2018).

254. McCorvie, T. J. *et al.* Molecular basis for the regulation of human glycogen synthase by phosphorylation and glucose-6-phosphate. *bioRxiv* 2021.11.12.468446 (2022). doi:10.1101/2021.11.12.468446

255. Palm, D. C., Rohwer, J. M. & Hofmeyr, J.-H. S. Regulation of glycogen synthase from mammalian skeletal muscle - a unifying view of allosteric and covalent regulation. *FEBS J* **280,** 2–27 (2012).

256. Lai, Y.-C., Stuenaes, J. T., Kuo, C.-H. & Jensen, J. Glycogen content and contraction regulate glycogen synthase phosphorylation and affinity for UDP-glucose in rat skeletal muscles. *Am J Physiol Endocrinol Metab* **293,** E1622–9 (2007).

257. Roach, P. J., Takeda, Y. & LARNER, J. Rabbit skeletal muscle glycogen synthase. I. Relationship between phosphorylation state and kinetic properties. *J. Biol. Chem.* **251,** 1913–1919 (1976).

258. Pursell, N. *et al.* Inhibition of Glycogen Synthase II with RNAi Prevents Liver Injury in Mouse Models of Glycogen Storage Diseases. *Mol Ther* **26,** 1771–1782 (2018).

259. Ashe, K. M. *et al.* Inhibition of glycogen biosynthesis via mTORC1 suppression as an adjunct therapy for Pompe disease. *Molecular Genetics and Metabolism* **100,** 309–315 (2010).

260. Turnbull, J. *et al.* PTG depletion removes Lafora bodies and rescues the fatal epilepsy of Lafora disease. *PLoS Genet* **7,** e1002037 (2011).

261. Pederson, B. A. *et al.* Inhibiting glycogen synthesis prevents Lafora disease in a mouse model. *Ann Neurol.* **74,** 297–300 (2013).

Appendix 1. Table of primers

PCR primers used in this study for site directed mutagenesis. The mutation created is shown in red. All primers read 5' to 3' and both the forward (Fwd) and reverse (Rev) primers are shown.

Mutagenesis primers	
GS:	
W18A Fwd	CCTGCCTGGCCTGGAAGATGCGGAGGATGAGTTCGACCTGGAGAAC
W18A Rev	CAGGTCGAACTCATCCTCCGCATCTTCCAGGCCAGGCAGGGAG
R588A+R591A Fwd	CGCCAACGTATCATCCAGCGCAACGCTACCGAGGCCCTGAGCGATCTGC TGGACTGG
R588A+R591A Rev	CTTCCAGTCCAGCAGATCGCTCAGGGCCTCGGTAGCGTTGCGCTGGATG ATACGTTG
Y600A Fwd	CTGAGCGATCTGCTGGACTGGAAGGCCCTGGGTCGCTACTACATGTCC
Y600A Rev	CGGACATGTAGTAGCGACCCAGGGCCTTCCAGTCCAGCAGATCGC
R603A Fwd	CTGGACTGGAAGTACCTGGGTGCCTACTACATGTCCGCTCGCCACATGG
R603A Rev	GTGGCGAGCGGACATGTAGTAGGCACCCAGGTACTTCCAGTCCAGCAG
H610E Fwd	GGTCGCTACTACATGTCCGCTCGCGAAATGGCTCTCAGCAAAGCCTTC
H610E Rev	GGGGAAGGCTTTGCTGAGAGCCATTTCGCGAGCGGACATGTAGTAGCG
S614A Fwd	CATGTCCGCTCGCCACATGGCTCTCAGAAAAGCCTTCCCCGAACATTTC
S614A Rev	GGTGAAATGTTCGGGGAAGGCTTTTCTGAGAGCCATGTGGCGAGCGGAC

Appendix 2. GS mutations

Mutations of GS1 and GS2 genes. Source: clinvar, last accessed 4 October 2021

GS1 mutations					
CDS mutation	**AA mutation**	**Type**	**Accession code (clinvar)**	**Condition(s)**	**Residue interaction**
c.1A>C	M1L	missense	VCV000430320	Glycogen storage disease 0, muscle	
c.3G>T	M1I	missense	VCV001039847	Glycogen storage disease 0, muscle	
c.20T>C	L7S	missense	VCV000999389	Glycogen storage disease 0, muscle	
c.40G>C	G14R	missense	VCV000964057	Glycogen storage disease 0, muscle	
c.56A>T	E19V	missense	VCV001063670	not provided	
c.80C>T	A27V	missense	VCV001064089	Glycogen storage disease 0, muscle	
c.101G>T	W34L	missense	VCV000429776	Glycogen storage disease 0, muscle	
c.114C>G	N38K	missense	VCV000946451	Glycogen storage disease 0, muscle	
c.134C>T	T45M	missense	VCV000945011	Glycogen storage disease 0, muscle	
c.178G>A	D60N	missense	VCV001002555	Glycogen storage disease 0, muscle	
c.213G>C	Q71H	missense	VCV000642436	not provided	
c.214G>T	G72C	missense	VCV001014411	Glycogen storage disease 0, muscle	
c.217G>A	V73M	missense	VCV000970915	Glycogen storage disease 0, muscle	
c.221G>C	R74T	missense	VCV000645191	Glycogen storage disease 0, muscle	
c.254C>T	P85L	missense	VCV000960357	Glycogen storage disease 0, muscle	
c.278C>T	S93F	missense	VCV000854343	Glycogen storage disease 0, muscle	
c.310G>A	G104R	missense	VCV000450024	not provided	
c.314G>A	R105H	missense	VCV000654802	Glycogen storage disease 0, muscle	
c.324C>G	I108M	missense	VCV000998960	Glycogen storage disease 0, muscle	
c.362C>T	A121V	missense	VCV001040534	Glycogen storage disease 0, muscle	
c.395A>G	E132G	missense	VCV000567434	Glycogen storage disease 0, muscle	
c.418G>A	G140R	missense	VCV001004862	Glycogen storage disease 0, muscle	GN^{34} interaction

c.425C>T	P142L	missense	VCV000941325	Glycogen storage disease 0, muscle	GN^{34} interaction
c.448G>A	D150N	missense	VCV000942370	Glycogen storage disease 0, muscle	
c.500C>A	A167E	missense	VCV000652496	Glycogen storage disease 0, muscle	
c.505A>G	S169G	missense	VCV000894589	Glycogen storage disease 0, muscle	
c.524T>C	V175A	missense	VCV000946427	Glycogen storage disease 0, muscle	
c.553G>A	G185S	missense	VCV000862647	Glycogen storage disease 0, muscle	
c.556G>A	V186I	missense	VCV000329823	Glycogen storage disease 0, muscle	
c.578C>T	A193V	missense	VCV000847954	Glycogen storage disease 0, muscle	GN^{34} interaction
c.584G>A	R195Q	missense	VCV000967367	Glycogen storage disease 0, muscle	GN^{34} interaction
c.631C>T	R211C	missense	VCV000970295	Glycogen storage disease 0, muscle	
c.646G>A	G216S	missense	VCV000939691	Glycogen storage disease 0, muscle	
c.646G>T	G216C	missense	VCV000966878	Glycogen storage disease 0, muscle	
c.650C>T	A217V	missense	VCV000949974	Glycogen storage disease 0, muscle	
c.652G>A	V218M	missense	VCV000835130	Glycogen storage disease 0, muscle	
c.652G>C	V218L	missense	VCV001054282	Glycogen storage disease 0, muscle	
c.666C>G	N222K	missense	VCV000836150	Glycogen storage disease 0, muscle	sugar binding
c.666C>A	N222K	missense	VCV001037659	Glycogen storage disease 0, muscle	sugar binding
c.683A>G	N228S	missense	VCV001007054	Glycogen storage disease 0, muscle	GN^{34} interaction
c.704A>T	E235V	missense	VCV000856687	Glycogen storage disease 0, muscle	
c.722G>A	R241Q	missense	VCV001018886	Glycogen storage disease 0, muscle	
c.754G>A	A252T	missense	VCV000391854	Glycogen storage disease 0, muscle	
c.760G>A	V254I	missense	VCV000940671	Glycogen storage disease 0, muscle	
c.793G>A	E265K	missense	VCV000653921	Glycogen storage disease 0, muscle	
c.803A>T	H268L	missense	VCV000894180	Glycogen storage disease 0, muscle	

c.872A>G	H291R	missense	VCV001023691	Glycogen storage disease 0, muscle	G6P binding
c.944A>T	H315L	missense	VCV001015336	Glycogen storage disease 0, muscle	
c.1012G>C	G338R	missense	VCV000945364	Glycogen storage disease 0, muscle	
c.1015G>T	A339S	missense	VCV001013781	Glycogen storage disease 0, muscle	
c.1020C>A	D340E	missense	VCV000955642	Glycogen storage disease 0, muscle	
c.1021G>A	V341I	missense	VCV000329819	Glycogen storage disease 0, muscle	
c.1049A>G	N350S	missense	VCV001024950	Glycogen storage disease 0, muscle	
c.1052A>G	Y351C	missense	VCV000659273	Glycogen storage disease 0, muscle	sugar binding
c.1075G>A	E359K	missense	VCV001056519	Glycogen storage disease 0, muscle	
c.1077_1078delinsAA	Q360K	missense	VCV000858591	Glycogen storage disease 0, muscle	
c.1078C>A	Q360K	missense	VCV000894178	Glycogen storage disease 0, muscle	
c.1112G>A	R371Q	missense	VCV000844943	Glycogen storage disease 0, muscle	
c.1121A>G	N374S	missense	VCV000585159	Glycogen storage disease 0, muscle	
c.1127A>G	N376S	missense	VCV001007393	Glycogen storage disease 0, muscle	sugar binding
c.1144G>A	G382S	missense	VCV000944904	Glycogen storage disease 0, muscle	
c.1145G>A	G382D	missense	VCV000329816	Glycogen storage disease 0, muscle	tetramerisation domain
c.1156C>T	R386C	missense	VCV000893350	Glycogen storage disease 0, muscle	tetramerisation domain
c.1184C>T	T395M	missense	VCV000893348	Glycogen storage disease 0, muscle	tetramerisation domain
c.1190A>G	K397R	missense	VCV000641050	Glycogen storage disease 0, muscle	tetramerisation domain
c.1236C>G	S412R	missense	VCV000662709	Glycogen storage disease 0, muscle	phosphorylation site (identified by phosphorylation mapping)
c.1243G>A	D415N	missense	VCV000808611	Glycogen storage disease 0, muscle	
c.1246A>G	M416V	missense	VCV000329814	Glycogen storage disease 0, muscle	tetramerisation domain
c.1279A>G	M427V	missense	VCV001017549	Glycogen storage disease 0, muscle	

c.1303A>G	T435A	missense	VCV000893132	Glycogen storage disease 0, muscle	
c.1315_1316delinsAA	S439N	missense	VCV000649510	Glycogen storage disease 0, muscle	
c.1318T>C	F440L	missense	VCV000498163	not provided	sugar binding
c.1324C>G	P442A	missense	VCV000498167	Glycogen storage disease 0, muscle	sugar binding
c.1324C>A	P442T	missense	VCV000936512	Glycogen storage disease 0, muscle	sugar binding
c.1363C>T	P455S	missense	VCV001004987	Glycogen storage disease 0, muscle	
c.1407T>A	S469R	missense	VCV000893129	Glycogen storage disease 0, muscle	
c.1412A>G	D471G	missense	VCV000947297	Glycogen storage disease 0, muscle	
c.1432C>T	H478Y	missense	VCV000942790	Glycogen storage disease 0, muscle	
c.1436C>T	P479L	missense	VCV000643030	Glycogen storage disease 0, muscle	
c.1475A>T	D492V	missense	VCV000938769	Glycogen storage disease 0, muscle	
c.1492C>T	R498C	missense	VCV001014236	Glycogen storage disease 0, muscle	
c.1520C>T	S507F	missense	VCV000546850	not provided	
c.1559C>T	T520M	missense	VCV001026554	Glycogen storage disease 0, muscle	
c.1615G>A	E539K	missense	VCV000329810	Neuroferritinopathy, Hereditary hyperferritinemia with congenital cataracts, Glycogen storage disease 0, muscle	sugar binding
c.1754A>G	Q585R	missense	VCV000859569	Glycogen storage disease 0, muscle	regulatory helix, α22
c.1822G>A	A608T	missense	VCV001016164	Glycogen storage disease 0, muscle	
c.1823C>G	A608G	missense	VCV001052744	Glycogen storage disease 0, muscle	
c.1834G>C	A612P	missense	VCV001051757	Glycogen storage disease 0, muscle	
c.1835C>T	A612V	missense	VCV001061438	Glycogen storage disease 0, muscle	
c.1859A>G	H620R	missense	VCV001043805	Glycogen storage disease 0, muscle	
c.1879G>A	E627K	missense	VCV000576955	Glycogen storage disease 0, muscle	
c.1898G>T	G633V	missense	VCV001021949	Glycogen storage disease 0, muscle	

c.1909C>G	P637A	missense	VCV000651479	Glycogen storage disease 0, muscle	
c.1922C>T	S641L	missense	VCV000916245	Glycogen storage disease 0, muscle	phosphorylation site
c.1961C>T	P654L	missense	VCV000329807	Glycogen storage disease 0, muscle	
c.1985A>G	D662G	missense	VCV000934446	Glycogen storage disease 0, muscle	
c.1990C>T	R664W	missense	VCV000329806	Glycogen storage disease 0, muscle	
c.2008G>A	E670K	missense	VCV001009397	Glycogen storage disease 0, muscle	
c.2009A>G	E670G	missense	VCV000999215	Glycogen storage disease 0, muscle	
c.2014G>A	G672S	missense	VCV000945494	Glycogen storage disease 0, muscle	
c.2017G>A	E673K	missense	VCV000957665	Glycogen storage disease 0, muscle	
c.2021G>A	R674H	missense	VCV001021262	Glycogen storage disease 0, muscle	
c.2032G>A	D678N	missense	VCV001024883	Glycogen storage disease 0, muscle	
c.2035G>A	E679K	missense	VCV000894150	Glycogen storage disease 0, muscle	
c.2042C>T	A681V	missense	VCV000985784	Inborn genetic diseases	
c.2044G>A	A682T	missense	VCV001018906	Glycogen storage disease 0, muscle	
c.2065C>T	R689C	missense	VCV001006088	Glycogen storage disease 0, muscle	
c.2096G>T	C699F	missense	VCV001010299	Glycogen storage disease 0, muscle	
c.2105C>T	S702F	missense	VCV000836504	Glycogen storage disease 0, muscle	
c.2120A>C	K707T	missense	VCV000894149	Glycogen storage disease 0, muscle	
c.2122C>G	R708G	missense	VCV000844469	Glycogen storage disease 0, muscle	
c.2206C>T	R736C	missense	VCV000837003	Glycogen storage disease 0, muscle	
c.2207G>A	R736H	missense	VCV000329804	Glycogen storage disease 0, muscle, Neuroferritinopathy, Hereditary hyperferritinemia with congenital cataracts	
c.907C>T	R303*	nonsense	VCV000965769	Glycogen storage disease 0, muscle	
c.913C>T	Q305*	nonsense	VCV001074189	Glycogen storage disease 0, muscle	

c.1384C>T	R462*	nonsense	VCV000016057	Glycogen storage disease 0, muscle	sugar binding
c.1615G>T	E539*	nonsense	VCV000570876	Glycogen storage disease 0, muscle	sugar binding
c.160dup	T54fs	frameshift - duplication	VCV000422359	not provided	
c.162_163del	D56fs	frameshift - deletion	VCV000128236	Glycogen storage disease 0, muscle	
c.699_700del	R236fs	frameshift - deletion	VCV001069149	Glycogen storage disease 0, muscle	
c.1204del	R402fs	frameshift - deletion	VCV000567037	Glycogen storage disease 0, muscle	
c.2207del	R736fs	frameshift - deletion	VCV000861694	Glycogen storage disease 0, muscle	
GS2 mutations					
CDS mutation	**AA mutation**	**Type**	**Accession code (clinvar)**	**Condition(s)**	**Residue interaction**
c.50A>G	Q17R	missense	VCV000665936	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.116A>G	N39S	missense	VCV000016054	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.154G>A	A52T	missense	VCV000308011	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.215A>G	H72R	missense	VCV000884167	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.279C>G	D93E	missense	VCV000214522	not specified	
c.280G>A	A94T	missense	VCV000137522	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.289A>G	K97E	missense	VCV000308009	Glycogen storage disease due to hepatic glycogen synthase deficiency	sugar binding
c.299G>A	C100Y	missense	VCV000214527	not provided	
c.395G>A	G132D	missense	VCV000214523	not specified	
c.421G>A	G141S	missense	VCV000137523	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.427C>T	P143S	missense	VCV000861394	Glycogen storage disease due to hepatic glycogen synthase deficiency	GN^{34} interaction
c.470C>T	S157F	missense	VCV000999043	Glycogen storage disease due to hepatic glycogen synthase deficiency	

c.520T>C	Y174H	missense	VCV000380333	not specified	
c.526G>A	V176I	missense	VCV000736468	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.556A>T	I186F	missense	VCV000214524	not specified	
c.577G>A	A193T	missense	VCV000261471	Glycogen storage disease due to hepatic glycogen synthase deficiency	GN^{34} interaction
c.653T>C	I218T	missense	VCV000964160	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.753C>G	C251W	missense	VCV001036478	Glycogen storage disease due to hepatic glycogen synthase deficiency	GN^{34} interaction
c.755C>G	A252G	missense	VCV000308003	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.799G>A	E267K	missense	VCV000882213	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.956A>T	D319V	missense	VCV000376836	not provided	
c.1015G>C	A339P	missense	VCV000016052	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1087A>G	M363V	missense	VCV000261462	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1129G>A	V377M	missense	VCV000880824	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1171G>T	D391Y	missense	VCV000884120	Glycogen storage disease due to hepatic glycogen synthase deficiency	tetramerisation domain
c.1245C>G	D415E	missense	VCV000137524	Glycogen storage disease due to hepatic glycogen synthase deficiency	tetramerisation domain
c.1277C>T	T426I	missense	VCV000214531	not provided	tetramerisation domain
c.1279A>C	I427L	missense	VCV001058720	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1334C>T	T445M	missense	VCV000952317	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1336C>G	H446D	missense	VCV000016056	Glycogen storage disease due to hepatic glycogen synthase deficiency	

c.1418T>G	V473G	missense	VCV000307993	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1427T>A	I476N	missense	VCV000214532	not provided	
c.1436C>A	P479Q	missense	VCV000016051	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1447T>C	S483P	missense	VCV000016055	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1472T>G	M491R	missense	VCV000016053	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1475A>T	D492V	missense	VCV000418239	not provided	
c.1477T>A	Y493N	missense	VCV000521873	Inborn genetic diseases	active site
c.1522T>G	Y508D	missense	VCV000418240	not provided	sugar binding
c.1549G>C	A517P	missense	VCV000307992	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1553A>C	E518A	missense	VCV000214533	Glycogen storage disease due to hepatic glycogen synthase deficiency	active site
c.1636A>G	T546A	missense	VCV000137526	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1657G>A	V553I	missense	VCV000214525	not specified	
c.1672C>T	R558C	missense	VCV000648650	Glycogen storage disease due to hepatic glycogen synthase deficiency	sugar binding
c.1745G>A	R582K	missense	VCV000534624	Glycogen storage disease due to hepatic glycogen synthase deficiency	G6P binding
c.1774C>G	L592V	missense	VCV000653240	Glycogen storage disease due to hepatic glycogen synthase deficiency	regulatory helix, α22
c.1789G>A	D597N	missense	VCV000214535	not provided	
c.1790A>G	D597G	missense	VCV000307990	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1820A>G	H607R	missense	VCV000307989	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1829A>G	H610R	missense	VCV000936980	Glycogen storage disease due to hepatic glycogen synthase deficiency	

c.1880C>T	S627T	missense	VCV000766741	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1889C>T	T630M	missense	VCV000719047	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1906T>C	Y636H	missense	VCV000658693	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1965G>C	Q655H	missense	VCV000137528	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.2005G>A	D669N	missense	VCV000214526	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.2054T>C	F685S	missense	VCV000137529	Glycogen storage disease due to hepatic glycogen synthase deficiency, not provided	
c.2068G>T	V690F	missense	VCV000720229	not provided	
c.2072C>T	P691L	missense	VCV000798247	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.547C>T	Q183*	nonsense	VCV000214529	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.574C>T	R192*	nonsense	VCV000214530	Glycogen storage disease due to hepatic glycogen synthase deficiency	GN^{34} interaction
c.736C>T	R246*	nonsense	VCV000016049	Glycogen storage disease due to hepatic glycogen synthase deficiency	GN^{34} interaction
c.925C>T	R309*	nonsense	VCV000078954	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1156C>T	R386*	nonsense	VCV000569452	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.465del	F155fs	frameshift - deletion	VCV000937499	Glycogen storage disease due to hepatic glycogen synthase deficiency	
c.1081del	T361fs	frameshift - deletion	VCV000667422	Glycogen storage disease	
c.1974dup	V659fs	frameshift - duplication	VCV000631679	Glycogen storage disease due to hepatic glycogen synthase deficiency	