Structural characterisation of calnexin cycle components and assessment as antiviral targets



Johan C. Hill

Magdalen College, University of Oxford

Trinity term, 2018

A thesis submitted for the degree of
Doctor of Philosophy

Structural characterisation of calnexin cycle components and assessment as antiviral targets

Johan C. Hill

Magdalen College, University of Oxford

Trinity term, 2018

A thesis submitted for the degree of

Doctor of Philosophy

N-glycosylated proteins that traverse the endoplasmic reticulum (ER) can make use of the calnexin cycle to attain their correct fold. The calnexin cycle modifies the N-glycan structure and allows for association of the glycoprotein with the ER lectins calnexin and calreticulin, which in turn recruit further chaperones that assist folding. Most enveloped viruses encode glycoproteins, which, upon infection of a host cell, crucially depend on the calnexin cycle to aid their folding. This includes diverse families such as *Flaviviridae*, *Retroviridae* and *Orthomyxoviridae*. We studied the calnexin cycle components with the ultimate aim of developing broad-spectrum antivirals.

X-ray crystallography was used to structurally characterise the murine ER α-glucosidase I, which controls entry into the calnexin cycle, with a number of inhibitory antiviral iminosugars. These data reveal flexibility in the ligands' alkyl tails and may act as a basis for the discovery of enzyme specific inhibitors.

UDP-glucose: glycoprotein glucosyltransferase (UGGT) is the quality control checkpoint of the calnexin cycle whose full-length structure from the thermophilic fungus *Chaetomium thermophilum* was recently determined. Presented here are a higher resolution structure in addition to SAXS studies of UGGT's interaction with Sep15, a protein that enhances UGGT activity.

UGGT's reaction releases into the ER lumen UDP, which is the only known small molecule inhibitor of UGGT. An ER-resident UDPase, ENTPD5, breaks down UDP into UMP. Enzymatic characterisation of ENTPD5 reveals its substrate specificities; in addition we show a paralog, ENTPD6, possesses similar activities. Presented here is work towards crystallisation of these two proteins and a test of the anti-Zika activity of ENTPD5 inhibitors.

Finally, CRISPR/Cas9 knock-out cells were generated to test, in principle, whether modulation of the activity of proteins involved in the calnexin cycle could be antiviral. The data confirm that the ER glucosidases are likely the best targets of those studied.

Acknowledgements

I've really come to think of the whole lab as a big, extended family all working together. Nicole obviously takes her place at the head of the family, attempting to get everyone organised and working cohesively whilst everyone else tries their best to stop this from happening. Pietro is the fun uncle who will happily give you his attention when you need it but is also a bit of a trouble-maker (but will always deny this fact). I think of both Dom and Alex as two big brothers: despite vastly different personalities, they have been there to give advice when needed but also the odd bit of teasing! Whilst it might be a little bit rude to say, I've always thought of George as the family dog, who's always happy to go for a walk or take 5 minutes just to mess around. Then of course everyone else fits in to the busy milieu of the bustling family-dynamic that has made the lab an absolute pleasure to work in over the past four years.

From day one, Nicole has been a caring supervisor who, despite being ridiculously busy, has always been happy for me to just drop in and chat about my work or anything else. She has been there to give me the confidence to go ahead and attempt things I've never tried before and learn from the outcomes. I think I have been incredibly lucky to have a supervisor who is so supportive and has a great sense of humour, as well as having the foresight to hire Pietro.

Pietro really is an amazing personality to work with, who can always make us laugh with either his stories or what he gets up to. In the lab, he has been the one that I have gone to whenever I had problems, whenever I had questions and whenever I had results that I wanted to show someone. I would like to thank him for always being interested in my work wherever it ended up and for imparting some of his endless crystallographic knowledge to me. I wish him the best of luck for his fellowship up in Leicester and I know that he'll come to be relied upon for his skills and knowledge by many over there. He always makes sure that credit is given where due.

In my first year in the lab, I don't know how well I would have gotten by without Alex to guide me, teaching me all of his little tips and tricks as well as leaving me a legacy in the form of glucosidase I. Dom has been there for many a walk around the University Parks, to encourage me to finally pop the question and he's also a dab-hand at the ol' HPLC. I'd

like to thank the "Virus Club" lot for welcoming me into the world of cell-based assays, especially to Michelle for the original introduction with Zika virus and all of her help and then also to Beatrice for an introduction to dengue. Abhinav, on the mass spectrometry side-of-things, to help answer any scientific queries. Mario, for introducing me to climbing, for being a great squash partner and for the crucial role of carrying on the p7 legacy. And then a final thanks to everyone else in the Zitzmann group, as well as the Oxford Protein Production Facility, Ed Lowe (for SAXS and crystallisation expertise) and David Staunton.

My parents have always been a big inspiration to me. It's to them I owe my "just-get-on-with-it" attitude and being content with doing pretty much anything as long as I'm being kept busy. Without that, I don't think I would have done quite so well; having to pipette marginally different volumes of slightly different liquids into hundreds or thousands of different wells. They have always supported anything I've done and have told me to only ever do a job if you enjoy what you're doing and who you're doing it with.

A huge chunk of the credit for my D.Phil has to go to Em, my fiancée, to whom I'll probably be married by the time anyone actually reads this finished thesis. She's been by my side from the application process all the way to the finished thesis, and she'll hopefully continue to put up with me for many, many years to come. I need to thank Em for pretending to be interested in my work and my day to day experiments but also for pushing me to be the best. But most important of all, a thank you to Em for her amazing brownies, which have helped me through the darkest of days when all the chocolate in the house ran out.

Finally, I would like to thank Warren Harding for keeping me on the straight and narrow and reminding me that there's more to life than the Structural Characterisation of Calnexin Cycle Components and Assessment as Antiviral Targets.

Statement of Authorship

The work presented here was completed by myself in Nicole Zitzmann's laboratory at the Oxford Glycobiology Institute, with the following exceptions. As acknowledged in section 2.1.1, the initial protein expression trials of *Mus musculus* ER α-glucosidase I (glu I), were performed by Dr Alessandro Caputo, under the supervision of Dr Louise Bird, at the Oxford protein production facility (OPPF). The initial cloning and expression trials of *Homo sapiens* UGGT and ENTPD5 were carried out by myself at OPPF with the assistance of Dr Jo Nettleship. As acknowledged in sections 2.4.2 and 4.2.5, Dr Abhinav Kumar (of Nicole Zitzmann's laboratory) performed the mass-spectrometry experiments. Dr David Staunton, of the Biochemistry Department's biophysics suite, carried out the SEC-MALS analysis of proteins, as stated in 2.6.3. Dr Alessandro Caputo set up the first crystal screen of glu I to yield a crystal (3.4.1). Cloning, expression and set up of crystallisation screens for the double cysteine mutants were carried out together with an Erasmus student, Andrea Lia, as acknowledged in chapter 4.1. As acknowledged in 6.3.3, Dr Dominic Alonzi performed the free oligosaccharide analysis.

Contents

Abstract **i**

Acknowledgements **iii**

List of figures **xiii**

List of tables **xv**

List of abbreviations and acronyms **xvi**

1. Introduction **1**
1.1. The calnexin cycle 2
1.1.1. An overview 2
1.1.2. ER α-glucosidase I 5
1.1.3. ER α-glucosidase II 7
1.1.4. UDP-glucose: glycoprotein glucosyltransferase 10
1.1.5. Selenoprotein 15 14
1.1.6. The UDP-Glc cycle 15
1.1.7. Malectin 18
1.2. Enveloped Viruses 19
1.2.1. Dengue virus 20
1.2.2. Zika virus 22
1.2.3. Other viruses rely on glycosylation 24
1.3. Iminosugars 26
1.3.1. A history 26
1.3.2. Use in disease 28
1.3.3. Iminosugars as antivirals 30
1.4. Project aims 33

2. Materials and methods **35**
2.1. Cloning strategies 36
2.1.1. Oxford protein production facility 36
2.1.2. Cloning and expression vectors 36
2.1.3. Ligate dependent cloning 37
2.1.4. Gibson Assembly 37
2.1.5. Site-directed mutagenesis 37

2.1.6. CRISPR cloning 38
2.2. Protein Expression 41
2.2.1. Protein expression in *E.coli* 41
2.2.2. Protein expression in HEK293F cells 42
2.3. Protein purification 42
2.3.1. *E.coli* lysate purification 42
2.3.2. Mammalian sample preparation 43
2.3.3. Affinity purification 44
2.3.4. Size exclusion chromatography 44
2.3.5. Anion exchange 45
2.3.6. Protein concentration 45
2.4. Protein modification 46
2.4.1. Refolding *in vitro* 46
2.4.2. Methylation 46
2.4.3. Proteolysis 47
2.5. Gel electrophoresis 48
2.5.1. SDS-PAGE 48
2.5.2. Western blotting 48
2.5.3. DNA gels 49
2.6. Protein characterisation 50
2.6.1. Dynamic scanning fluorimetry 50
2.6.2. Circular dichroism 50
2.6.3. SEC-MALS 51
2.6.4. Small angle X-ray scattering 51
2.6.5. The MESG activity assay 53
2.6.6. Isothermal titration calorimetry 54
2.7. Crystallisation 55
2.7.1. Set-up of crystallisation screens 55
2.7.2. Data collection 55
2.7.3. Data processing 56
2.8. Cell culture 56
2.8.1. HEK293T cells 57
2.8.2. Vero cells 58
2.8.3. LLCMK2 cells 58
2.8.4. CRISPR transfection and screening 58
2.9. Cellular assays 62
2.9.1. Toxicity assay 62
2.9.2. Antiviral assays with Zika virus 62
2.9.3. Antiviral assays with dengue virus 63
2.9.4. Plaque assay 64
2.9.5. QRT-PCR 66
2.9.6. Immunofluorescence 67

3. ER α-glucosidase I **69**
3.1. Cloning of *Mm* glu I 70
3.2. Expression of *Mm* glu I 70
3.3. Purification of *Mm* glu I 71
3.3.1. Purification of *Mm* glu I 71
3.3.2. Methylation of *Mm* glu I 71
3.3.3. Mass-spectrometry analysis of *Mm* glu I 72
3.4. *Mm* glu I – the 3.26 Å dataset 74
3.4.1. Crystallisation and data collection 74
3.4.2. Data processing 74
3.4.3. Phase determination 75
3.4.4. Model building and refinement 77
3.5. The 2.1 Å resolution structure of *Mm* glu I 78
3.5.1. Obtaining higher resolution crystals 78
3.5.2. Data collection and processing 78
3.5.3. Model building, refinement and validation 79
3.5.4. Analysis of the structure 82
3.6. The binding of other iminosugars 87
3.6.1. Crystallisation 87
3.6.2. Data collection and processing 87
3.6.3. Model building, refinement and validation 87
3.6.4. Iminosugars in the active site 90
3.6.5. An additional iminosugar binding site? 93
3.6.6. Likely binding conformation of glucose in the +1 subsite 95
3.7. Ligand binding by ITC 96
3.8. Discussion 101

4. UGGT and Sep15 **106**
4.1. Cloning 107
4.1.1. Cloning of a double cysteine *Ct* UGGT mutant 107
4.1.2. Cloning of *Hs* UGGT1 with OPPF 109
4.1.3. Cloning of the *Hs* Sep15-interacting loop from *Hs* UGGT1 into *Ct* UGGT 109
4.1.4. Cloning of *Hs* Sep15 111
4.2. Purification and expression 111
4.2.1. Of the UGGT chimera 111
4.2.2. Of Sep15 from pET303 112
4.2.3. Of Sep15 in pET/M-43 113
4.2.4. Of the UGGT chimera-Sep15 complexes 114
4.2.5. Of the double cysteine *Ct* UGGT mutant 114
4.3. Crystallisation trials 114
4.3.1. The *Ct* UGGT chimera and its complexes 114

4.3.2. Crystallisation of *Ct* UGGT D611C:G1050C 116
4.4. Data processing 116
4.4.1. Data processing 117
4.4.2. Model building and refinement 118
4.5. The structure of the *Ct* UGGT chimera 119
4.6. The structure of *Ct* UGGT D611C:G1050C 121
4.7. SAXS studies of the *Ct* UGGT chimera 122
4.7.1. Collection of data and initial data processing 122
4.7.2. *Ab initio* modelling 124
4.7.3. Model of the *Ct* UGGT and MBP-Sep15 complex 129
4.7.4. Analysis of the conformation of UGGT in solution 130
4.8. Discussion 131

5. ENTPD5 136
5.1. Design of constructs 137
5.2. Cloning of constructs 140
5.3. Expression trials 141
5.4. Purification, refolding and methylation 143
5.4.1. *Hs* MBP-ENTPD5 143
5.4.2. *Ct* ENTPD6 144
5.5. Characterisation and activity assays 145
5.5.1. SEC-MALS 145
5.5.2. Dynamic scanning fluorimetry 146
5.5.3. Activity assay optimisation 149
5.5.4. Michaelis-Menten curves 150
5.5.5. E172Q: The catalytic residue mutations 153
5.5.6. Inhibition of the ER-UDPase 154
5.6. Attempts at crystallisation 156
5.6.1. Crystallisation of *Hs* MBP-ENTPD5 156
5.6.2. Crystallisation of *Ct* ENTPD6 158
5.7. Antiviral properties of ENTPD5 inhibitors 159
5.7.1. Toxicity assay 159
5.7.2. 6403111 in the Zika antiviral assay 160
5.8. Discussion 163

6. CRISPR/Cas9 knock-out cell lines 166
6.1. Bioinformatics and construct design 167
6.2. Cloning, transfection and clonal selection 170
6.3. Screening of clones 170
6.3.1. PCR screening 170
6.3.2. Sequencing to confirm knock-outs 172
6.3.3. Free oligosaccharide analysis 176
6.4. Immunofluorescence experiments 179

6.5. Zika antiviral assays 182
6.6. Dengue virus – antiviral assays 186
6.7. Discussion 189

7. Conclusions **194**

Appendix A: Primers used in this thesis **198**

Appendix B: X-ray diffraction data collection parameters **202**

Appendix C: Sequences used for molecular biology **203**

References **213**

List of figures

1.1. Overview of the calnexin cycle 3
1.2. ER α-glucosidase I domain organisation and structure 6
1.3. ER α-glucosidase II domain organisation and structure 9
1.4. Predicted domain organisation of UGGT and structure 12
1.5. The UDP-glucose cycle 17
1.6. Structure of example iminosugars 27

2.1. DNA gel of PCR screening for CRISPR knock-outs 61

3.1. Mass spectrometry coverage of methylated *Mm* glu I 72
3.2. Native Patterson and self-rotation function of *Mm* glu I diffraction data 76
3.3. Ramachandran plot of *Mm* glu I structure 81
3.4. Overall structure of *Mm* glu I 82
3.5. Active site of *Mm* glu I with M*O*N-DNJ 83
3.6. Vacuum electrostatics and surface conservation of *Mm* glu I 84
3.7. Naturally occurring mutations mapped onto *Mm* glu I structure 86
3.8. Electron density of a panel of iminosugars in the glu I active site 91
3.9. Active site, visualisation of a flexible loop 92
3.10. Structure of *Mm* glu I in complex with SR1 94
3.11. Superimposed structures of sugars in the glu I active site 96
3.12. ITC data of *Mm* glu I titrated against DNJ 98
3.13. ITC data of *Mm* glu I titrated against *N*B-DNJ 99
3.14. ITC data of *Mm* glu I titrated against M*O*N-DNJ 100

4.1. Design of *Hs* UGGT1 and UGGT2 constructs 108
4.2. UV traces of UGGT chimera and Sep15 complexes from SEC 113
4.3. Morphology of the *Ct* UGGT "transformer" crystals 115
4.4. Crystal packing of the *Ct* UGGT chimera 120
4.5. Structure of the *Ct* UGGT double mutant 122
4.6. Background-subtracted SAXS diffraction data of the *Ct* UGGT chimera with constructs of Sep15 123
4.7. *Ab initio* models of the *Ct* UGGT chimera 125
4.8. *Ab initio* models of the *Ct* UGGT chimera with Sep16.5 126
4.9. *Ab initio* models of the *Ct* UGGT chimera with MBP-Sep15 127
4.10. Fit of the UGGT and MBP-Sep15 complex into a SAXS envelope 129
4.11. SREFLEX analysis of the *Ct* UGGT chimera 130

5.1. Predicted binding of nucleotides into the active site of ENTPD5 139

5.2. Elution profile of proteolysed *Ct* UGGT from a MonoQ column 145
5.3. Accuracy of buffer system used for DSF 147
5.4. Experimental results of DSF experiments 148
5.5. Activity of *Hs* ENTPD5 and *Ct* ENTPD6 at different pH 149
5.6. Specificity constants of *Hs* ENTPD5 and *Ct* ENTPD6 152
5.7. Michaelis-Menten curves of the ENTPD5 E172Q mutant 153
5.8. IC50 curves of potential *Hs* ENTPD5 inhibitors 154
5.9. X-ray diffraction of a potential *Hs* ENTPD5 crystal 157
5.10. MTS cytotoxicity assay of ENTPD5 inhibitors 160
5.11. Antiviral effect of 6403111 against Zika virus 161

6.1. PCR screen of glu II knock-out clonal populations 171
6.2. DNA sequence of an UGGT1 knock-out 172
6.3. DNA sequence of a glu I knock-out 173
6.4. DNA sequence of an UGGT2 knock-out 174
6.5. DNA sequence of a malectin knock-out 175
6.6. DNA sequence of an ENTPD5 knock-out 176
6.7. Free oligosaccharide analysis of glucosidase knock-outs 178
6.8. Immunofluorescence experiments to assess percentage of cells infected by Zika virus 180
6.9. Immunofluorescence experiments to assess percentage of cells infected by dengue virus 181
6.10. Plaque and qRT-PCR results of the Zika antiviral assay against knock-out cell lines 183
6.11. Plaque and qRT-PCR results of the dengue antiviral assay against knock-out cell lines 187

List of tables

2.1. Volumes of reagents required for cell culture in different flasks 56
2.2. Reagents required for different plaque overlay mixes 65

3.1. Data processing statistics of the 3.26 Å dataset of *Mm* glu I 73
3.2. Data processing statistics of the 2.1 Å dataset of *Mm* glu I 78
3.3. Model building and refinement statistics of *Mm* glu I 80
3.4. Data processing statistics of the panel of iminosugars in *Mm* glu I 88
3.5. Model building and refinement statistics of the panel of iminosugars in *Mm* glu I 89

4.1. Data processing statistics of the *Ct* UGGT crystal data 117
4.2. Model building and refinement statistics of the UGGT structures 118

5.1. Showing which constructs of ENTPD5 were successfully cloned 141
5.2. Expression levels of the different ENTPD5 constructs 142
5.3. Michaelis-Menten constants obtained for *Hs* ENTPD5 and *Ct* ENTPD6 151
5.4. Michaelis-Menten constants for the *Hs* ENTPD5 E172Q mutant 153
5.5. IC50s of the *Hs* ENTPD5 inhibitors 155

List of abbreviations

AMP-DNJ	*N*-(5-adamantane-1-yl-methoxypentyl)-deoxynojirimycin
ATPase	Adenosine triphosphate phosphatase
CBM	Carbohydrate binding module
CDG-IIb	Congenital disorder of glycosylation, type IIb
CI	Confidence interval
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CST	Castanospermine
CV	Column volumes
D	Deletion screening primers
DENV (1-4)	Dengue virus (serotypes 1 – 4)
DEPC	Diethylpyrocarbonate
DGJ	Deoxygalactonojirimycin
DHF	Dengue haemorrhagic fever
DLS	Diamond Light Source
DMAB	Dimethylamine borane
DMSO	Dimethyl sulfoxide
DNJ	Deoxynojirimycin
Dol	Dolichol
DSF	Dynamic scanning fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ENTPD	Ectonucleoside triphosphate diphosphohydrolase family
Env	Envelope protein of a virus
ER	Endoplasmic reticulum
ERAD	ER associated degradation

ESRF	European Synchrotron Radiation Facility
FBS	Foetal bovine serum
FOS	Free oligosaccharide
FPLC	Fast protein liquid chromatography
Fru	Fructose
Gal	Galactose
GH	Glycosyl hydrolase
Glc	Glucose
GlcNAc	*N*-Acetyl-d-glucosamine
Glu I	ER α-glucosidase I
Glu II	ER α-glucosidase II
GMPPNP	Non-hydrolysable analogue of GTP
GPCP	Non-hydrolysable analogue of GDP
GP	Glycoprotein (of viral origin)
GT	Glycosyl transferase
HA	Hemagglutinin
HBV	Hepatitis B virus
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HRV	Human rhinovirus
Hsp70	Heat shock protein family 70
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_M	Michaelis constant
M	'Membrane' protein of a virus
Man	Mannose
MBP	Maltose binding protein
MESG	7-methyl-6-thioguanosine
MEM	Minimal essential medium

MGAM	Intestinal maltase-glucoamylase
MOGS	Mannosyl-oligosaccharide transferase
M*O*N-DNJ	methoxy,9-nonyl-deoxynojirimycin
MOI	Multiplicity of infection
MS	Mass spectrometry
MW (CO)	Molecular weight (cut-off)
NA	Neuraminidase
*N*B-DNJ	*N*-butyl deoxynojirimycin
ND	Non-deletion screening primers
NDPase	Nucleoside diphosphate phosphatase
NMR	Nuclear magnetic resonance
NS	Non-structural protein
NSD	Normalised spacial discrepancy
OPPF	Oxford protein production facility
OST	Oligosaccharyltransferase
PBS	Phosphate buffered saline
PDB	Protein data bank
PEG	Polyethylene glycol
PenStrep	Mix of penicillin and streptomycin
pfu	Plaque forming units
P_I	Free phosphate
PP	Diphosphate
prM	Pre-cursor of membrane protein (of viral origin)
PVDF	Polyvinylidene difluoride
qRT-PCR	Quantitative reverse transcriptase real time PCR
RMSD	Root mean squared deviation
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography

SEC-MALS	SEC-multi angle laser light scattering
Sep15	Selenoprotein-15
Sep16.5	Selenoprotein-15 made in the pET303 vector
sgRNA	Single guide RNA
SI	Sucrase isomaltase
SLC35B1	Solute carrier family 35, member B1
SOP	Standard operating procedures
TAE	Tris-acetate-EDTA buffer
TBS	Tris buffered saline
TCEP	Tris(2-carboxyehtyl)phosphine
T_M	Melting temperature
TRXL	Thioredoxin-like domain
UDPase	Uridine diphosphate phosphatase
UGGT	UDP-glucose: glycoprotein glucosyltransferase
UPR	Unfolded protein response
V_{MAX}	Maximal rate of reaction
WHO	World Health Organisation
WNV	West Nile virus
ZIKV	Zika virus

1. Introduction

Due to the lack of effective, broad-spectrum antivirals, treatments of many emerging viruses are limited to those just targeting the symptoms, as was the case in the recent epidemic in 2015 of Zika virus and the 2014 West African Ebola virus outbreak[2-4]. The evolutionary strategies of viruses rely on hijacking the pathways and processes of the host cells they infect[5]. From the perspective of antiviral drug development, the host cellular processes on which many diverse viruses depend could be targeted by a single

compound to effectively hinder replication of many viruses at once[6-8]. Such a broad-spectrum antiviral strategy could lead to a "golden pill" that could effectively treat a vast spectrum of current and emerging viral infections.

1.1. The calnexin cycle

The controlled folding of N-glycosylated proteins in the endoplasmic reticulum (ER) is one of these pathways that a diverse array of viruses depend upon. N-glycosylation can be found on proteins of many enveloped viruses such as *Flaviviridae*[9], *Herpesviridae*[10], *Filoviridae*[11] and *Retroviridae,* amongst others[12]. N-glycosylated host cell receptors can be required for cell entry by both enveloped and non-enveloped viruses[13, 14].

Known as the "calnexin cycle", a number of ER lectins, enzymes and chaperones make up the N-linked glycoprotein folding quality control machinery*. Misfolded glycoproteins are selectively retained in the ER whilst correctly folded glycoproteins are able to leave and reach their destinations. It has been estimated that over half of all eukaryotic proteins are glycosylated[15]. Such a diversity in potential substrates requires flexibility in the ER folding quality control systems to ensure that each glycoprotein can be monitored and retained if appropriate.

1.1.1. An overview of glycoprotein folding in the ER

The majority of proteins destined for the secretory pathway are co-translationally or sometimes post-translationally translocated into the ER via the SEC61 translocon complex[16]. Nascent N-glycans originate on the cytoplasmic side of the ER membrane,

* Throughout this thesis I shall refer to N-linked glycoproteins simply as "glycoproteins", a term which would otherwise include other forms of glycosylated proteins.

conjugated to a dolichol lipid anchor. Through a series of transferases first on the cytoplasmic and subsequently on the lumenal side of the ER membrane, the glycan is incrementally synthesised to attain its full composition: Dol-PP-$GlcNAc_2Man_9Glc_3$[17]. Oligosaccharyltransferase (OST), tightly associated with the eukaryotic translocon, can add a glycan to an asparagine side chain in the consensus sequence Asn - X – Ser/Thr (and less commonly, NXC); wherein X is any amino acid except proline[18, 19].

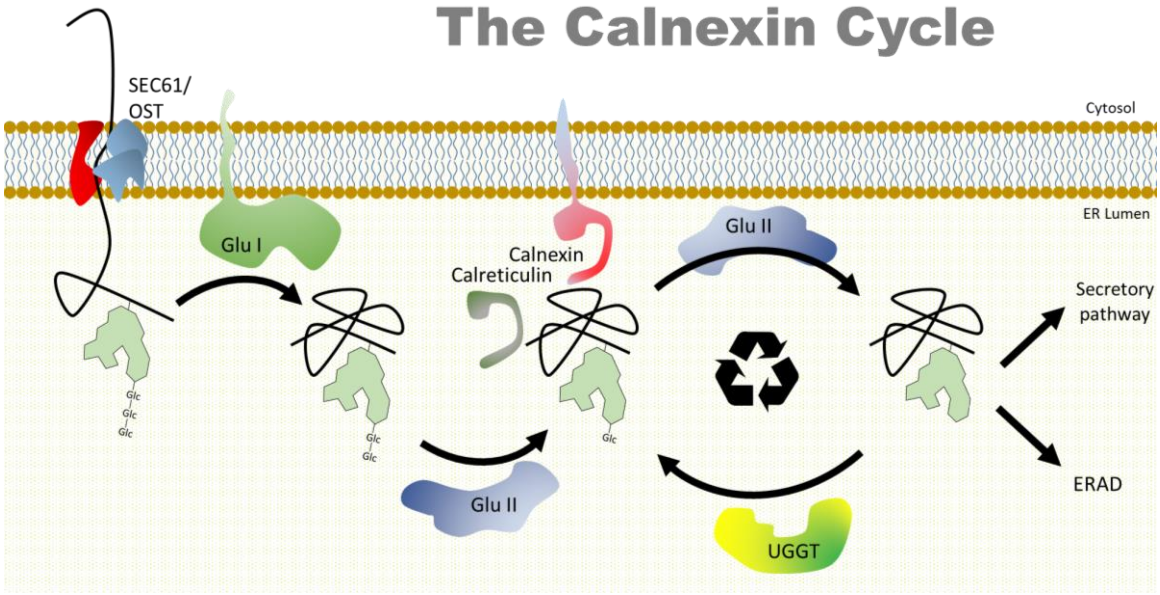


Figure 1.1. Overview of the calnexin cycle. Polypeptide is shown as a squiggly black line, glycan as a green polygon with glucose individually labelled. Proteins involved in the calnexin cycle are labelled and illustrated as colourful shapes.

Regardless of the folding state of the protein, the ER α-glucosidase I (glu I or MOGS) removes the terminal α(1,2) conjugated glucose moiety from the glycan[20]. ER α-glucosidase II (glu II) subsequently removes the middle α(1,3) conjugated glucose to yield a glycoprotein with a $GlcNAc_2Man_9Glc_1$ glycan. This is selectively recognised by the ER lectins, calnexin and calreticulin, which in turn recruit co-chaperones to the glycoprotein in order to aid its folding. Eventually, glu II removes the innermost α(1,3) conjugated glucose, freeing the glycoprotein from further interaction with either ER lectin[21, 22].

As shown in figure 1.1, one of three outcomes can now occur:

1. A misfolded protein may be recognised by the ER folding sensor, UDP-glucose: glycoprotein glucosyltransferase (UGGT). UGGT adds a single α(1,3) linked glucose back onto the glycan in the same position as the original innermost one. The resultant glycoprotein can re-associate with the ER lectins and is retained within the ER[23, 24].
2. Correctly folded glycoproteins are not recognised by UGGT and thus continue travel along the secretory pathway.
3. A terminally misfolded protein will eventually be dispatched to ER associated degradation (ERAD). This occurs via slowly-acting mannosidases, such as EDEM-1, which remove mannoses from the glycan until UGGT can no longer recognise the protein as a substrate, eventually targeting the misfolded glycoprotein to ERAD receptors and retro-translocation into the cytoplasm for degradation[25, 26].

Once a protein moves onwards through the secretory pathway, the N-glycans can be heavily modified to become "complex glycans". Modifications that are commonly seen include fucosylation, mannosylation or the addition of galactose or *N*-acetylneuraminic acid[27]. These glycans play crucial roles in cellular processes ranging from E-cadherin-mediated cell-cell interactions[28], the orientation of the immunological synapse[29] and synaptic transport of key neurotransmitters[30, 31]. In these cases, changes to the glycan composition affect the function of the glycoprotein.

As so many processes depend on correct glycoprotein fold and glycosylation, it is barely a surprise that mutations in the enzymes of the calnexin cycle cause disease. For example, inherited mutations in the glu I gene lead to the congenital disorder of glycosylation IIb

(CDG-IIb) whilst those in the glu II α and β genes can lead to polycystic kidney disease[32]. Despite this, the targeting of these enzymes for short-term treatment of acute disease in adult organisms has therapeutic potential[33].

1.1.2. ER α-Glucosidase I

ER α-glucosidase I was first discovered in the 1970s and subsequently determined to be the first ER enzyme involved in N-glycan modification by removing the terminal glucose from a glycan, using an inverting mechanism specific for α(1,2) glucosidic bonds[20, 34, 35]. Glu I apparently acts in a co-translational manner, as measured by glycans on nascent polypeptides that were partially translocated into the ER, which led to the speculation that it could be associated with OST and therefore the translocon complex[36].

Glu I is a singly N-glycosylated, type II membrane protein of 837 amino acids in humans[37] (Uniprot entry Q13724) (Figure 1.2.A). The cytoplasmic sequence is short and is likely responsible for the ER localisation of the protein whilst the lumenal portion is comprised of two domains: an N-terminal domain of unknown function and the C-terminal catalytic domain belonging to the GH63 family of protein folds[38, 39].

The first crystal structure of a member of the GH63 family appeared in 2008. The *Escherichia coli* glucosidase hydrolase, YgjK, was crystallised a number of times in complex with a range of substrates such as glucose, galactose and mannose[40]. However, as a bacterially derived protein, its natural substrate is not likely to be a eukaryotic N-glycan. Work suggested YgjK has broad substrate specificity to allow for the cleavage of a range of polysaccharides, but the Glc-α(1,2)-Gal bond is likely to be the major

substrate[40, 41]. Hence, this bacterial protein may not be the best model for modelling the structure and function of human glu I.

In 2013 the crystal structure of the glu I yeast homologue was determined[42]. The yeast glu I structure confirmed the protein's subdomain boundaries predicted from sequence (Figure 1.2.B). However, in this structure the active site was occluded by the expression tag and all attempts to dislodge it or crystallise the enzyme without the hexahistidine expression tag failed, as it formed part of an important crystal contact. With the ultimate

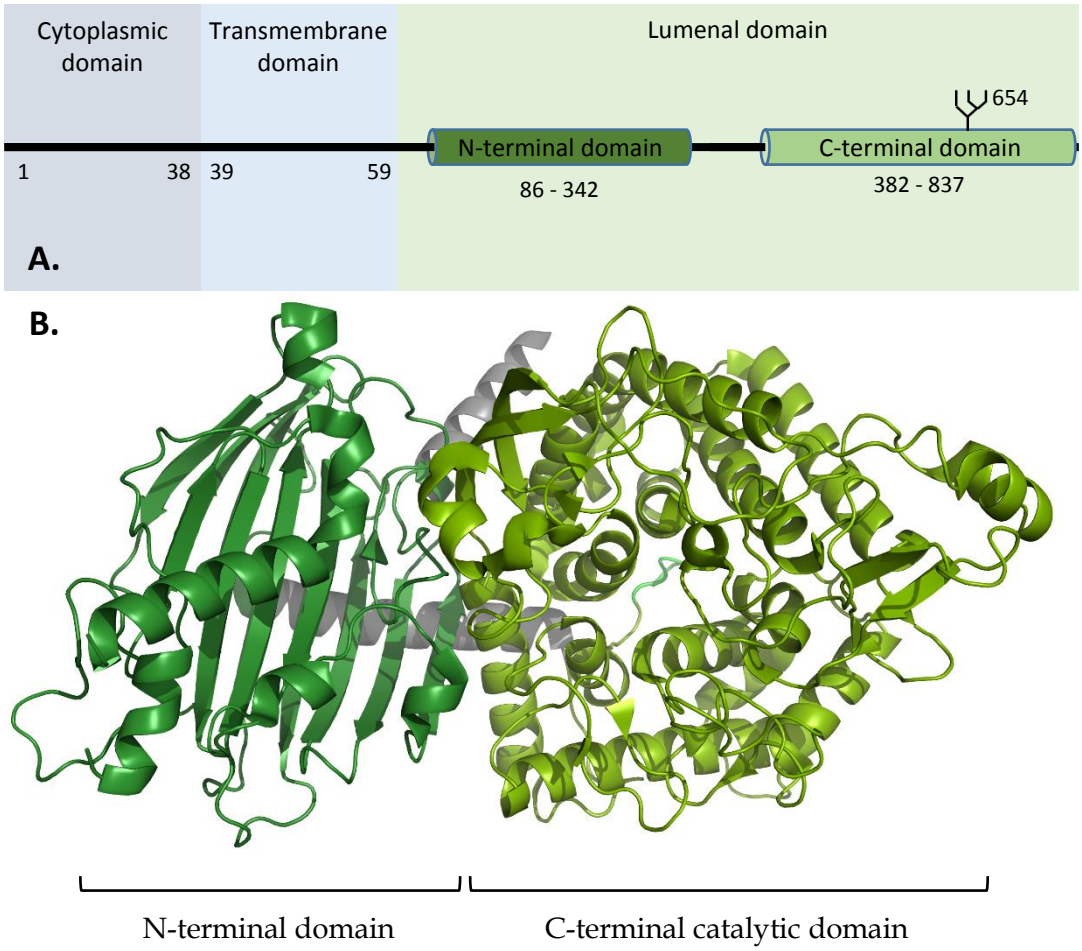


Figure 1.2. (A) Sequence-based domain boundary of human glu I. The predicted N-glycosylation site is shown decorated with a branched design indicating the glycan. (B) Structure of *S.cerevisiae* processing α-glucosidase I, with the N-terminal domain in dark green, the C-terminal domain in a lighter green, and linker elements in grey.

aim of producing specific inhibitors of the protein based on high-resolution information of the active site, this particular construct was not suitable.

Glu I appears to be a promising target for broad-spectrum antivirals on the basis of clinical observations in patients carrying naturally occurring mutations within the gene. In 2000, the first two glu I mutations in patients suffering from the disease MOGS-CDG (CDG-IIb), were described: both R486T and F652L mutations of glu I lead to an inactive protein[43, 44]. Other glu I mutations in CDG-IIb patients have since been reported, including A22E and R110H, which features incorrect splicing of the mRNA and early degradation, respectively; as well as a Q124 to stop codon causing the early termination of the protein[45]. The prognosis for those suffering from this disorder is not promising, as there are many associated developmental issues and complications such as liver failure and seizures among others, with the sufferers not expecting to live past their twenties[44]. Interestingly, all CDG-IIb patients are fairly resistant to viral infection, while ERAD and the unfolded protein response (UPR) appear to occur at normal levels in their cells. Whilst a lifetime of altered glu I activity, especially during development, has severe consequences, it is conceivable that inhibition of its activity for a short period of time to deal with acute viral infection could be tolerated by the adult body.

1.1.3. ER α-Glucosidase II

ER α-glucosidase II (glu II) is the second glycan-processing enzyme involved in the entry of glycoproteins into the calnexin cycle. Its role is to cleave, via a retaining mechanism, the α-(1,3) glycosidic bonds that connect the two innermost glucose moieties to the body of the glycan[35, 46]. The enzyme is a heterodimer; its GH31-family catalytic α-subunit is

encoded by the *GANAB* gene whilst the β-subunit is encoded by *PRKCSH*, which carries the ER retrieval sequence and likely plays a regulatory role[47]. The C-terminal mannose-6-phosphate receptor homology (MRH) domain of the β-subunit has been structurally characterised by both x-ray crystallography and NMR[48]. The domain is proposed to help recognise the glycan substrate by detecting terminal mannose residues in the glycan[49]. A terminal mannose binding to the MRH domain would lead to conformational alterations and allosterically enable the catalytic domain to cleave its substrate, whereas the LDLRα domains help maintain the α-subunit in a "catalytically competent conformation"[46, 50].

Whilst the removal of the first glucose from a nascent glycan by glu I is near instantaneous, glu II acts more slowly. First, it may have to compete with an ER lectin, malectin, for access to the glycan. After cleavage of the middle glucose, the glycoprotein glycan needs to reorient itself in the glu II active site for the second cleavage to occur[36]. In the period between the two cleavage events catalysed by glu II, the monoglucosylated glycoprotein can be sequestered by the ER lectins calnexin and calreticulin, slowing the apparent enzymatic activity of the enzyme.

The structure of a trypsinolytic fragment of *Mus musculus* glu II was recently determined by Caputo *et al.* in the Zitzmann lab (Figure 1.3). The majority of the catalytic α-subunit was ordered in the crystals, with just two large loops either removed by trypsin-mediated proteolysis or disordered. From the β-subunit, only the two LDLRα domains survived trypsinolysis, their structure reveals for the first time part of the interface between the subunits[46].

A substrate analogue, a covalent catalytic intermediate mimic, the product D-glucose and a number of iminosugar-based inhibitors were co-crystallised or soaked into the crystals, allowing for the elucidation of the reaction mechanism and revealing how the existing inhibitors bind at a molecular level. Glucosidases exist throughout the body and belong to a number of different protein families that cleave different glycosidic bonds; the closest glu II analogues are the GH31-family enzymes, intestinal maltase-glucoamylase (MGAM)

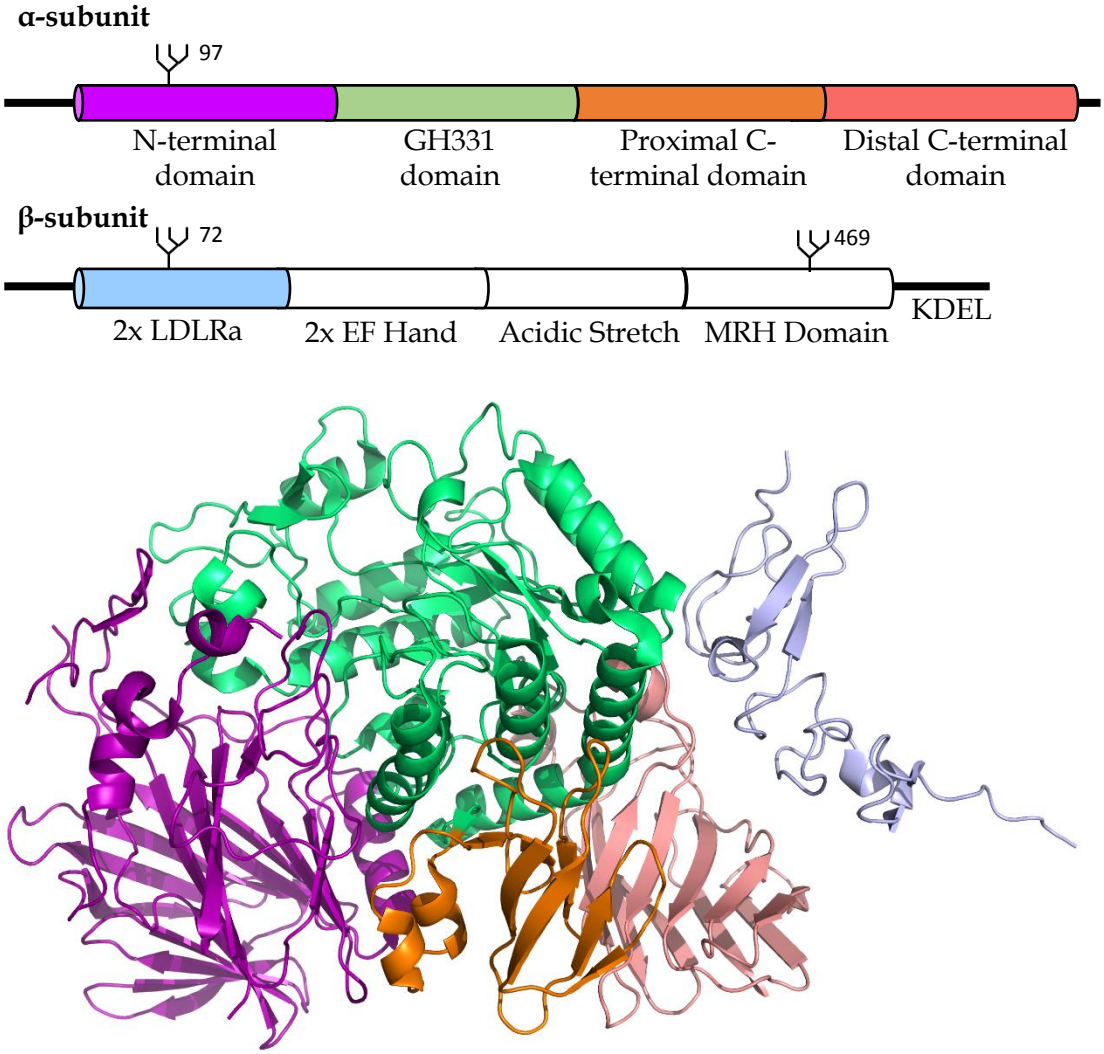


Figure 1.3. Overall domain architecture (top diagram) and structure (bottom diagram) of the glu II heterodimer, colour coded against one another. In white are domains not defined in the crystal structure. Figure adapted from Caputo, Alonzi et al. 2016.

and sucrase-isomaltase (SI). The glu II crystal structures are paving the way for the design of active site inhibitors that are specific for glu II over those intestinal glucosidases, the latter being the cause for the most pronounced side effects reported during clinical trials of iminosugars against HIV infection[51]. Moreover, the interface between the two glu II subunits is unique to this particular glucosidase: molecules targeting this region would disrupt the ER localisation of the catalytic subunit and possibly constitute a new class of specific glu II inhibitors.

In contrast to glu I, naturally occurring mutations in the genes encoding glu II are not reported to lead to a resistance to viral infections. However, mutations in both *PRKCSH* and *GANAB* genes have been reported to lead to the same disease: polycystic kidney or liver disease[32, 52-54]. It has been suggested that disruption of the calnexin cycle via these mutations leads to the incorrect folding and/or localisation of polycystin and fibrocystin families of glycoproteins, precipitating the disease[55, 56].

For preventing entry of glycoproteins into the calnexin cycle, glu I may eventually turn out to be the better target of the two ER glucosidases. Firstly, endo-α-mannosidases within the Golgi can act on the $Glc_2Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ glycans that result from glu II inhibition and consequently bypass glu II this way (although this also involves bypassing calnexin cycle mediated folding)[57]. Secondly, glu II is a fairly promiscuous enzyme and can cleave a range of glycosidic bonds, albeit with lower efficiency than the Glc α(1,3)-Glc bond it evolved to cleave, possibly making it more difficult to develop a specific inhibitor[22].

1.1.4. UDP-glucose: glycoprotein glucosyltransferase

UDP-glucose: glycoprotein glucosyltransferase acts as the glycoprotein misfold sensor of the calnexin cycle. It is able to recognise a broad range of misfolded glycoproteins and targets them for ER retention via the re-glucosylation of $GlcNAc_2Man_{7-9}$ glycans and their subsequent re-association with the ER-resident lectins, calnexin and calreticulin[23, 25, 58, 59]. UGGT will not retain folded proteins in the ER as it cannot recognise them as substrates, nor will it retain terminally misfolded proteins for long, because ER-resident mannosidases will eventually cleave the exposed mannose arms of the glycan to the extent that UGGT is unable to recognise these glycoproteins[25, 60]. Lower eukaryotes only encode a single copy of UGGT in their genome, yet in many higher eukaryotes, two isoforms are present: UGGT1 and UGGT2, sharing as much as 50% overall sequence homology[61]. Up until as recently as 2013, it was believed that just UGGT1 was the active isoform whilst the role of UGGT2 was unknown[62, 63]. However, a number of recent studies have shown that UGGT2 does have the ability to bind and re-glucosylate misfolded proteins, just as UGGT1 does[64-66].

Structural data on UGGT was non-existent until recently, when the first structure of a single thioredoxin domain from the thermophilic fungus *Chaetomium thermophilum*, was published in 2014[67]. Sequence-based analysis of the approximately 1500 residue polypeptide chain erroneously predicted three thioredoxin-like (TRXL) domains towards the N-terminal side of the protein and a catalytic domain of approximately 300 residues at the far C-terminus (Figure 1.4.A)[67].

In 2017, the Zitzmann group published three crystal structures of the full-length *C.thermophilum* protein from different crystal forms and in different conformations[68]. Figure 1.4.B highlights the newly discovered domain architecture: unexpectedly, four

TRXL domains were found in the N-terminal portion of the protein followed by two β-sandwiches, which in turn clasp the catalytic domain. The protein folds in a non-sequential order, which is especially unusual due to its size and complexity. Figure 1.4(C) shows the three-dimensional structure of the protein. TRXL2, on the right-hand side of figure 1.4 C, showed the greatest conformational flexibility out of all of the domains in the published structures and it was postulated that it was this interdomain

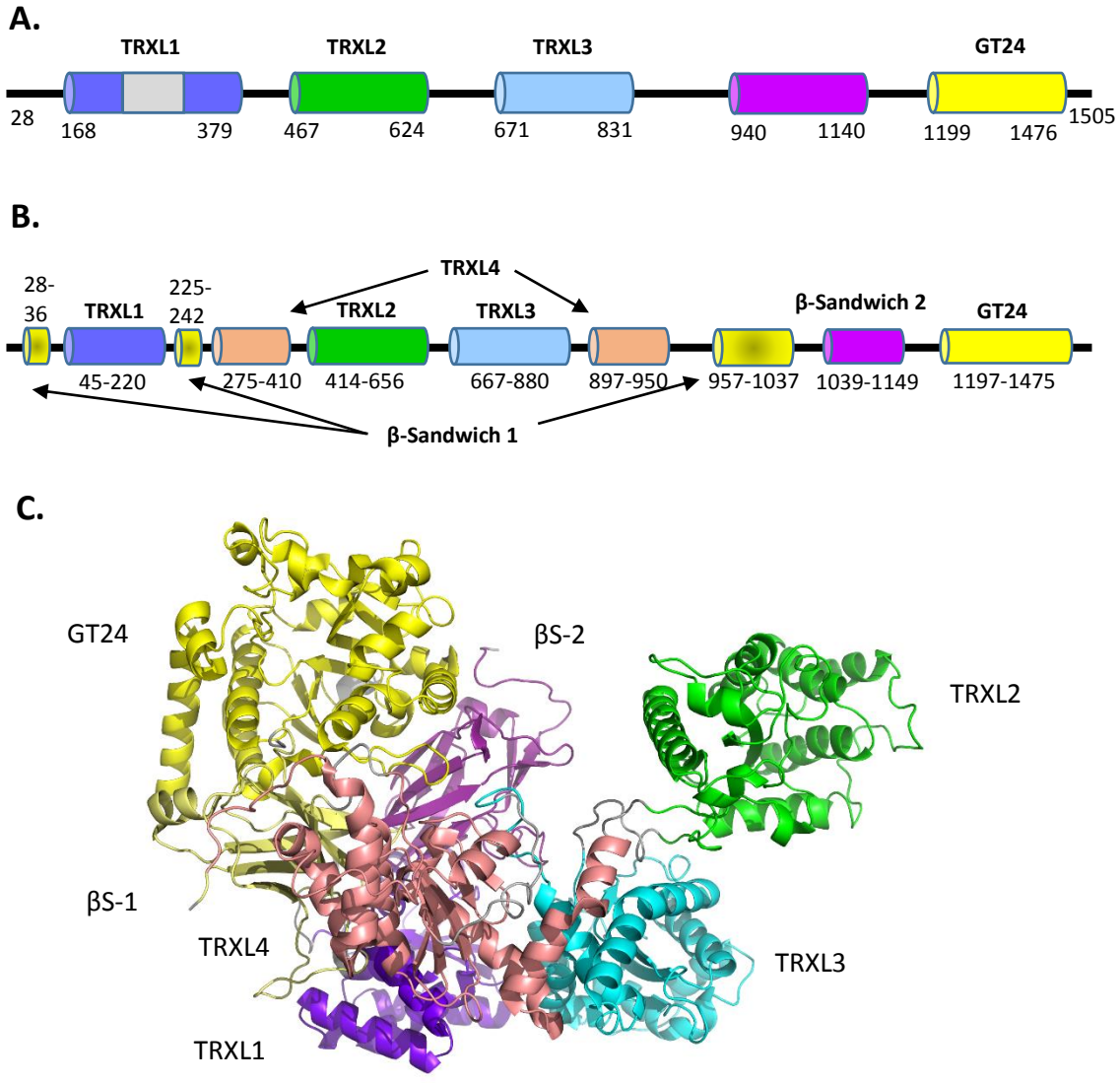


Figure 1.4. (A) Adapted figure of the original predicted domain architecture of UGGT by Koichi Kato *et al* (2014). (B) Adapted figure of domain architecture once the full-length structure was solved by Roversi *et al* (2017). (C) Three-dimensional structure of *Ct* UGGT, PDB ID 5MU1.

conformational flexibility that aids the ability of UGGT to recognise a large range of substrates.

TRXL domains are found in a number of ER localised proteins that include protein disulphide isomerase, ERDJ5 (an Hsp40 chaperone) and Selenoprotein 15[69-71]. This fold has been described to have patches that can recognise misfolded proteins, and so it was not unreasonable to hypothesise that the TRXL domains of UGGT may play the same role, to help recognise misfolded glycoproteins[72]. In fact, the activity of the catalytic domain of UGGT2 is established against a non-substrate client glycoprotein when a chimera of UGGT2's catalytic domain is fused with the N-terminal domains of UGGT1[62]. Knowledge of the misfold sensing epitopes would allow for the targeted design of ligands to affect the interaction interface between UGGT and its client glycoproteins.

The latest structures of UGGT to be published are from the organism *Thermomyces dupontii* (*Td*UGGT). The structures of the N-terminal domains and the catalytic domain were determined separately to a resolution of 3.1 Å and 1.5 Å, respectively[73]. In the active site of three distinct catalytic domain crystal structures, Ca^{2+}, a buffer molecule and UDP-glucose or UDP are found; the latter ligands are one of the substrates and one of the products of the enzyme's catalytic activity. The catalytic domains of fungal UGGTs bear high sequence similarity to their human homologues, so that their high-resolution structures could allow for identification of the first UGGT inhibitors through fragment-based screening by protein crystallography.

Whilst naturally occurring mutations of glu I and glu II in humans eventually cause chronic disease, no such mutations of UGGT are known. Homozygous knock-outs of UGGT1 in mice are embryonically lethal, however heterozygous knock-out mice were

found to possess ~50% of the wild-type levels of the enzyme and displayed no obvious phenotype other than a reduced susceptibility to RNA viruses and reduction in the resultant morbidity[74, 75]. These results raise the possibility that UGGT1 inhibition could be an antiviral therapy and that transient treatment of acute viral disease in a mature individual may be viable.

1.1.5. Selenoprotein 15

Selenoprotein 15 (Sep15 or SELENOF) is a small, 15 kDa protein shown to bind to UGGT with nanomolar affinity which likely allows for its ER retention[76, 77]. NMR structures confirmed a predicted TRXL fold and gave rise to a hypothesis that Sep15 plays a role as an oxidoreductase, although the structure of the N-terminal cysteine rich region was not studied[71]. This latter region has 15-20% sequence homology with Zinc-finger family proteins and mediates its interaction with UGGT: if this Sep15 N-terminus is grafted onto an unrelated protein, the new chimera will gain the ability to associate with UGGT with a stoichiometry of 1:1[78].

As for its potential role in the calnexin cycle, Sep15 has been shown to increase the activity of both UGGT1 and UGGT2[64]. Sep15 can block secretion of IgM and IgG immunoglobulins if they contain mis-formed disulphide bonds, thereby supporting redox quality control of these proteins[79]. ER stress and the unfolding protein response (UPR), induced by the presence of tunicamycin or brefeldin A, lead to the transcriptional upregulation of Sep15 levels within the ER; however, Sep15 knock-down with small interfering RNA did not lead to ER stress, indicating that there are other redundant

systems in place or that Sep15 only plays a role in the folding of a specific subset of glycoproteins[80].

Many mutations in the gene encoding Sep15 exist, which has led to speculation on the role of Sep15 in both cancer risk and progression of certain viruses such as HIV[81-88]. Knock-out mice are fully fertile and viable but develop early cataracts[89]. Were viral proteins to fold under the control of Sep15-assisted glycoprotein folding and quality control, targeting of Sep15 could constitute an antiviral or anticancer strategy with minimal risk of side effects.

1.1.6. The ER UDP-Glc Cycle

For UGGT to carry out its role in re-glucosylating misfolded glycoproteins, a plentiful supply of the glucose-donor, UDP-glucose, must be available[22, 23]. Along the same lines, the only known small molecule inhibitor of UGGT is UDP, a product of the reaction that UGGT catalyses[90]. These two facts would be problematic for the cell, were it not for the presence of two proteins that, together with UGGT, form the ER UDP-Glc cycle:

1. An ER-localised UDPase that breaks down UDP into UMP and monophosphate.
2. An ER-localised antiporter which pumps UMP into the cytoplasm in return for a molecule of UDP-Glc, synthesised in the cytoplasm[91].

One proposed candidate for the human ER-UDPase is a member of the ecto-nucleoside triphosphate diphosphohydrolase (ENTPD) family, known as ENTPD5[90, 92]. ENTPD5, otherwise known as CD39L4 or PCPH, was first discovered in 1998 and the first few papers on its enzymatic activity described its apparent ADP hydrolysis activity[93-95]. It was eventually revealed that all its given names referred to the same protein and

nomenclature was standardised and the family given its "ENTPD" based names[96]. Other well-known members of this family include ENTPD1 (CD39) that hydrolyses ligands such as ATP, ADP, UTP and UDP to their monophosphate derivatives and also ENTPD2, a cell-surface ATPase[97].

Further characterisation of the protein ENTPD5 revealed both an ER and an extracellular localisation, consistent with the lack of any obvious retention signals but raising the possibility that the protein interacts with ER-localised partners for its own ER localisation[92, 98]. After further studies, the enzymatic activity was primarily described as a UDPase, not an ATPase as originally described, with significant GDP hydrolysis activity as well[99]. These data supported ER-localisation, as unwarranted ATP hydrolysis would be detrimental to the function of other ER-localised proteins such as BiP. Separate studies have started to build up a body of evidence that imply a role of ENTPD5 in the Warburg effect (enhanced glycolysis found in cancer cells)[92, 100]. Many studies have correlated dysregulation of the expression of the protein with a range of cancers, emphasizing its important role in human disease and justifying the need for further study[96, 101-112].

Further supporting evidence for the enzyme's role in the calnexin cycle comes from many observations that knock-down of the protein causes ER stress whilst induction of ENTPD5 expression alleviates ER stress[100].

Another candidate for the ER UDPase is a closely related family member, ENTPD6. This protein has been reported as a Golgi-localised GDPase, yet a look at the sequence shows that this type-II membrane protein has a possible ER-localisation signal[113-115]. Many NDPases have broad specificity towards other nucleosidase diphosphates and so it is

possible that ENTPD6, or even other family members, play a redundant role in the folding of glycoproteins in the ER, under certain conditions[100].

As for the possible candidate for the UDP-Glc: UMP antiporter, the identity of this transporter is currently unknown in humans[116, 117]. However, studies in *Arabidopsis thaliana* have identified a protein that plays this role, *At*UTr1 is upregulated by the UPR and is absolutely required for UDP-Glc incorporation into the ER; its human homologue has been identified as solute carrier family 35, member B1 (SLC35B1 or UGTrel1)[118-120]. Whilst annotated as a UDP-galactose antiporter in UniProt (ID P78383), a few sources do

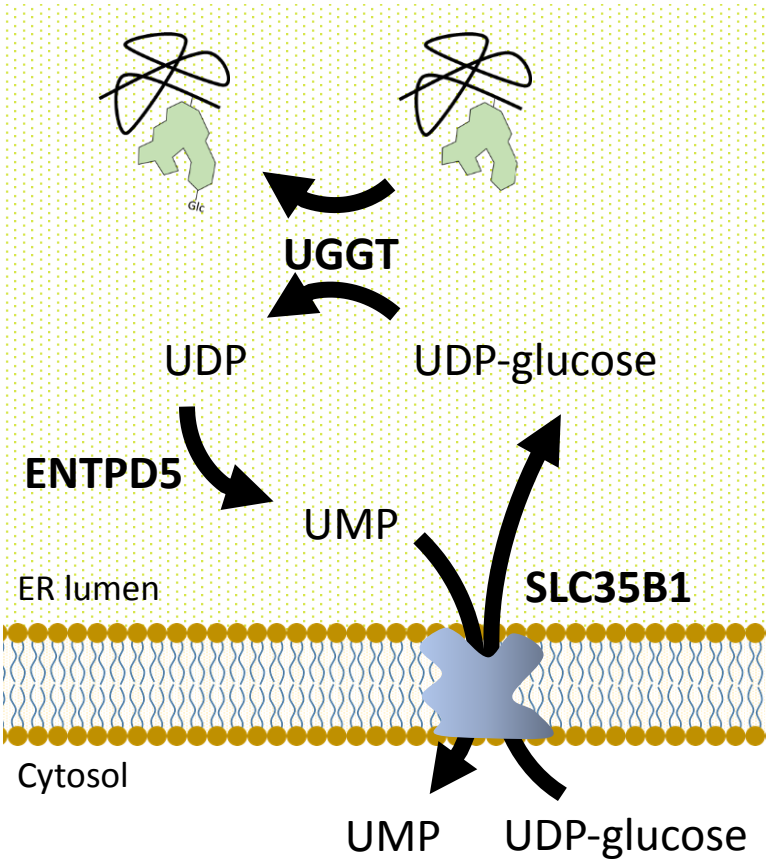


Figure 1.5. The UDP-glucose cycle. UGGT requires UDP-glucose to add glucose to misfolded glycoprotein. UDP is hydrolysed by ENTPD5 to UMP + P_I UMP is pumped out of the ER by an antiporter, likely SLC35B1, in return for a molecule of UDP-glucose, completing the cycle.

refer to UDP-Glc antiporter activity from SLC35B1, although it is hard to trace this information back to a peer-reviewed publication[117, 121].

Figure 1.5 highlights the proposed general architecture of this auxiliary cycle to the calnexin cycle, of which a good review was written by Hadley *et al*. in 2014[122]. Similar nucleotide-sugar cycles are repeated within the different sub-cellular compartments of cells and so there is a strong expectation that one must exist for UDP in the endoplasmic reticulum. Besides UGGT, if ENTPD5 or SLC35B1 can be inhibited with small molecule ligands, these proteins too have promise as targets for broad-spectrum antivirals, as well as potential applications in many cancers[112].

1.1.7. Malectin

Whilst the calnexin cycle can play an important role in the folding of many N-glycosylated proteins, there is a large variety of other enzymes and chaperones that also aid the folding environment within the ER.

Malectin is a newly discovered player to the ER glycoprotein folding field and accompanies calnexin and calreticulin as an ER-resident lectin. Only discovered in 2008, it is a type-I transmembrane protein with highest ligand specificity for glycans terminating in α(1,3)-linked diglucose moieties, the product of the glu I mediated cleavage[123]. Malectin is, however, slightly promiscuous and can bind glycans in the ER such as $Glc_3Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$, as well as the iminosugar miglitol[124].

It is therefore a distinct possibility that malectin can bind glycoproteins that are a product of glu I, before glu II has had a chance to act. Both glu I and malectin have been proposed to associate with ER-resident complexes such as the OST complex for their ER

localisation[125]. Malectin has, as of the time of writing, not been designated a functional role. It has been suggested that it acts in a similar manner to calnexin or calreticulin by recruiting co-chaperones to aid folding of glycoproteins, although the only non-native conformers that have been proven to associate with malectin are influenza hemagglutinin (HA) and a folding-defective variant of the secretory protein α-1-antitrypsin[126].

It could be a possibility that malectin-mediated folding allows for the bypass of the calnexin cycle by the ensuing actions of Golgi mannosidases allowing a client's subsequent continuation down the secretory pathway before glu II and calnexin/calreticulin can act. Much more work on this protein needs to be done before any conclusions on its role and possible effects can be established.

1.2. Enveloped viruses

Viruses can be very broadly classified into two groups, enveloped and non-enveloped viruses. The former differ from the latter by the fact that enveloped viruses contain a host-derived lipid bilayer in addition to the viral capsid encasing the viral genome (be it single-stranded, double-stranded, DNA or RNA).

Most enveloped viruses infect animals, are much less resistant to the environment outside of the host than non-enveloped viruses, and generally require transmission between hosts. Enveloped viruses also do not necessarily follow a lytic life cycle: they often can leave the cells by budding from the host membrane, thus leaving the cell alive for future viral replication and paving the way for persistent or latent infection of the host[127-129].

Studded throughout this host derived viral envelope membrane, virus-encoded proteins, often N-glycosylated, can frequently be found and are important for viral structure, target

cell recognition and infection. Viruses that encode glycoproteins include some of the most deadly, some of the most widespread and some of the most recent emerging pathogens on the planet. These viruses are amongst the potential targets for broad-spectrum antivirals that are the aim of this thesis.

1.2.1. Dengue virus

The first serotype of dengue virus (DENV), a member of the *flavivirus* genus, was isolated in 1943 and within 10 years, the four major circulating serotypes had been identified: DENV1-4. In 2013, a fifth serotype was identified with non-human primates as the major hosts[130]. Altogether, 2.5 billion people are at risk of infection, with an estimated 100 million cases every year that lead to 25 000 deaths.

Transmission of DENV across human populations tends to be mediated by an *Aedes* genus mosquito vector, of which *Aedes aegypti* is the principal vector[131]. Primary infection with any DENV serotype causes dengue fever but confers long-term immunity to that particular serotype and does not tend to lead to severe disease. However, subsequent infection with a different serotype can lead to antibody-dependent enhancement and other factors that contribute to the development of the more deadly dengue haemorrhagic fever (DHF)[132].

Dengue virus is one of the most heavily studied viruses from a structural perspective. As a positive-sense, single-stranded RNA virus, direct translation of the ~11 kb genome into a single polypeptide is possible. This is subsequently cleaved into 10 proteins, three of which are structural: Core, Envelope (Env) and (pre)Membrane (prM; M); and seven that are non-structural: NS1, 2A, 2B, 3, 4A, 4B and 5[133]. The mature virus was first visualised

by cryo-electron microscopy in 2002[9]. A poorly defined inner nucleocapsid was observed, surrounded by a protected lipid bilayer studded with copies of the M protein, likely to be structurally relevant to the positioning of the Env dimer. These dimers form the outer shell, with three Env dimers per icosahedral asymmetric unit[9]. Env is responsible for the recognition of host cell receptors, internalisation into the host cell and the subsequent pH dependent fusion of host cell and virus membranes[134, 135].

Of the ten dengue proteins, only three, M, Env and NS1, are identified as N-glycosylated. NS1 has two N-glycosylation sites, one of which carries a complex glycan and the other a high-mannose type glycan[136]. Glycosylation of NS1 is likely involved in viral RNA replication and is critical for the replication of the virus[137, 138]. NS1 is also secreted as a hexamer and interacts with the host's complement system. The stability of this hexamer and its interaction with complement is hindered by mutations that remove the glycosylation sites[139, 140].

Similarly, the M protein has a number of potential N-glycosylation sites which are also crucial for the production of infectious virus[136]. Inhibition of the endoplasmic reticulum α-glucosidases by the iminosugars deoxynojirimycin (DNJ) or castanospermine (CST) leads to incomplete M protein folding and formation of an unstable complex with Env, affecting the assembly of dengue virions[141].

Env has a number of glycosylation sites that may include N153/4 and N67, although its glycosylation pattern is much more variable between the different serotypes than those of the other two glycoproteins[136]. The glycosylation site at N153/4 is not essential for the virus, however there have been signs that mutation of this residue can affect membrane fusion and neurovirulence[142]. The N-glycosylation site at N67 mediates the interaction

with the host receptor, DC-SIGN[143]. If this glycosylation site were removed, it would severely affect virus growth and propagation[144, 145].

At present, there is no specific antiviral treatment against dengue nor is there an efficacious, licensed vaccine[136]. Dengue vaccines have, until recently, proved elusive due to the need to target all four serotypes to avoid antibody dependent enhancement[146]. Dengvaxia, a tetravalent dengue vaccine, had been approved for use in some endemic countries, however its efficacy and safety has been hotly debated and may increase the incidence of severe disease in children under 9 years of age, generally the target group for vaccinations[147-150]. It led the company that developed it, Sanofi, to admit that the vaccine poses higher risks to those without prior exposure to DENV. Whilst most primary infection does not require treatment, the more severe DHF requires hospitalisation and thus new treatments are still needed to combat this disease.

A host-protein targeting approach would potentially be efficacious against the virus' multiple serotypes and work around antiviral resistance: for example, the Hsp70 family proteins have been suggested as valid antiviral targets[151]. The need for correct glycosylation on the three DENV glycoproteins makes a strong case for an anti-dengue therapy based on inhibitors of the ER α-glucosidases that control entry to the calnexin cycle[141, 152, 153].

1.2.2. Zika virus

Zika virus (ZIKV) is another *flavivirus* and causes much of the same flu-like symptoms as dengue. ZIKV is transmitted by the same mosquito vectors yet does not have the multiple serotypes that allow for repeated infection as with DENV. One major difference is its

ability to affect the nervous system, leading to the neurological conditions of Guillain-Barré syndrome and microcephaly[154, 155].

Although first isolated in humans in 1952[156], ZIKV did not reach the public's attention until the 2015 outbreak centred in Brazil because of the association with microcephaly in newborns and Guillain-Barré syndrome in adults[157-160]. It is generally agreed that the Latin America epidemic is now over, yet the virus may still become endemic in many countries, just as dengue virus has[161]. Endemic ZIKV is predicted to lead to millions of cases every year and cause possible new epidemics in areas which have so far been little exposed to the virus. Whilst infection with ZIKV is likely to lead to lifelong immunity against the virus (not yet fully confirmed), there are no antiviral therapies nor vaccines against it[162].

ZIKV is a 10.8 kb long, single-stranded RNA virus and encodes homologues to the same three structural and seven non-structural proteins in dengue virus[163]. In 2016, two groups determined 3.7 and 3.8 Å resolution structures of ZIKV, discovering that it had a very similar structure to other *flaviviruses* but was more thermally stable than dengue, possibly due to an extra hydrogen bond between Env domain IIIs at a five-fold symmetry axis[164, 165]. Whilst closely related to DENV, both groups noted the fact that only the Asn^{154} site of the Env protein was glycosylated, despite glycosylation at Asn^{67} being required for proper folding in DENV. It was shown that the loop on which Env is glycosylated in ZIKV varies between isolates but is also distinct to dengue and much more similar to Env from West Nile virus (WNV), another closely related *flavivirus*. In WNV, the glycosylation site has been linked to neurotropism, which could explain the neuropathological conditions seen in ZIKV infection[166].

Asn^{154} is varies between ZIKV strains, occurring in nearly all sequenced cases from Latin America, in most cases in Asian strains and only in African strains which had not been passaged extensively in cell culture or mice brains[167, 168]. It is thought the process of this passaging may have led to the deletions in this region, as observed with the closely related WNV[169]. Nonetheless, the importance in the glycosylation of ZIKV Env at Asn^{154} has been made clear. ZIKV mutants lacking this glycosylation site give rise to a lower viremia and cause no mortality in mice, whilst in mosquito cells they show much lower infectivity[170]. Absence of glycan at Asn^{154} leads to deficiencies in packaging and impaired infectivity in a human leukaemia cell line, attenuated neuroinvasion in mice and prevents infection of mosquitos via the oral route[171-173]. Glycosylation sites are also present in prM, at position Asn^{69} in African, Asian and Latin American strains[167]. Overall, ZIKV glycoproteins may also rely on the calnexin cycle in order to obtain their correct fold. The development of a broad-spectrum antiviral that could be used against emerging threats would therefore be of great use.

1.2.3. Other enveloped viruses rely on glycosylation

Both dengue virus and ZIKV were used in this thesis, but the array of enveloped viruses which utilise and rely on N-glycosylation is not restricted to these *flaviviruses*.

Ebola virus is a *filovirus* that causes episodical epidemics in sub-Saharan Africa and leads to a high fatality rate due to the severity of the hemorrhagic fever that it induces. Ebola glycoprotein (GP) is a heterodimer of GP1 and GP2: the former has 15 N-glycosylation sites whilst the latter has only 2 extremely well-conserved sites. The separate removal of every single N-glycan site on GP1 does not affect the stability of the virus and actually

improves viral entry efficiency; however, these mutants showed a much greater sensitivity to antibody neutralisation, highlighting how glycosylation is important to shield Ebola virus from the immune system[174]. The removal of the two GP2 N-glycosylation sites does not affect the expression of the viral protein but greatly hinders oligomerisation and the ability of the monomers to be incorporated into virions. Knock-outs of calnexin and calreticulin showed lower expression of GP, underscoring the importance of the calnexin cycle in the life cycle of Ebola[175-177].

Influenza, or flu, circulates in seasonal epidemics as one of four types, A-D, and is estimated to cause 3-5 million severe cases and 290 000 – 650 000 deaths annually, according to the World Health Organisation (WHO)[178]. Seasonal vaccinations can prove efficacious but require the circulating strain to be correctly predicted[179]. Influenza viruses are members of the *Orthomyxoviridae* family and types A and B encode two major surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA)[180]. It has been suggested that multiple glycosylation sites on HA, whilst not affecting the virus structure or fusion activities of the protein, may be used to evade antibody-mediated neutralisation[181]. Glycosylation of HA at specific sites also appear to have evolved to aid in the folding and maturation of the protein[182]: the removal of some of these sites severely reduces viral growth and replication, in a neuraminidase-subtype specific manner[183]. Influenza glycoproteins interact with calnexin and calreticulin[184] and blocking access to the calnexin cycle, such as in cells lacking glu I (such as in CDG-IIb), leads to much lower hemagglutinin titres and lower susceptibility to infection[45].

In these same cells derived from sufferers of CDG-IIb, a much lower susceptibility to HIV infection was also observed. HIV is from the *Retroviridae* family and its genome encodes

fifteen proteins[185]. The HIV pandemic started in the 1980s and has led to at least 70 million infections and 35 million deaths (World Health Organisation, 2017). Once infected, individuals cannot be cured, but combination antiretroviral therapy can prevent transmission and prolong survival[186]. The GP120 (envelope) protein of HIV is highly glycosylated, with up to 50% of its mass made up of sugars. This protein is the only known target for broadly-neutralising antibodies against HIV, due to the virus' high mutation rate[187]. Loss of the GP120 glycans at many sites reduces binding of HIV to the host cell receptor, CD4[188] and studies of inhibitors that interfere with glycan processing show that glycosylation is important for viral infectivity and cytopathogenicity[189, 190]. The glycan shield surrounding GP120 is also important in immune evasion, with mutations of many single glycosylation sites sufficient to increase antibody mediated neutralisation[191-193].

Ebola, influenza and HIV are just three examples of the large subset of enveloped viruses whose infectivity and/or replication rely on proper folding of their glycoproteins. It is therefore of interest to develop a drug that could interfere with these processes and act as a broadly-acting antiviral. Fortunately, some possible candidates for this role have already been identified: iminosugars.

1.3. Iminosugars

1.3.1. A history

Iminosugars are carbohydrate mimetics where the oxygen in the ring is replaced with nitrogen. The first iminosugar, nojirimycin, was isolated from *Streptomyces roseochromogenes* and *Streptomyces lavendulae* in 1966: the molecule only differed from D-

glucose by the substitution of the endocyclic oxygen with a nitrogen atom[194]. Compounds with the stereochemistry of galactose and mannose were discovered shortly after in the 1980s, yet all three were fairly unstable due to the hydroxyl group on C_1.

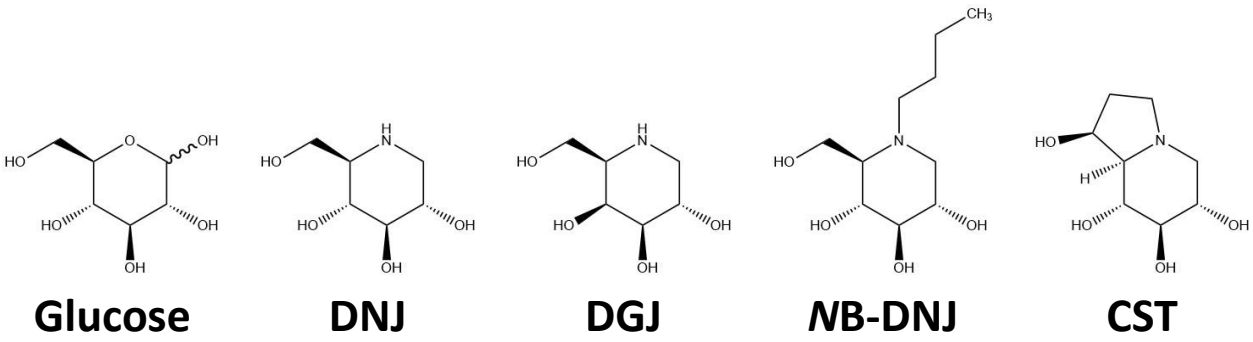


Figure 1.6. Structure of glucose, deoxynojirimycin (DNJ), deoxygalactonojirimycin (DGJ), *N*-butyl DNJ (*N*B-DNJ) and castanospermine (CST).

1-deoxynojirimycin (DNJ), a much more stable glucose mimetic (Figure 1.6) was isolated from the root bark of the mulberry tree in 1976[195]. It has been shown since to be produced by a number of bacterial species as well as by chemical synthesis[196]. *N*-alkylation of iminosugars, specifically of DNJ, occurs naturally; the synthesis of longer chain alkylated iminosugars was first reported in 1984 and was used to help improve the specificity for certain enzymes[196, 197]. Bicyclic iminosugars also occur naturally: castanospermine (CST), was found in *Castanospermum australe,* a plant that causes severe gastrointestinal distress when ingested by livestock (and can sometimes prove fatal)[198].

Naturally occurring iminosugars are all inhibitors of certain glycosidases, due to the structural similarities to the natural substrates of the enzymes[199]. For example, DNJ has been shown to inhibit α-glucosidases such as the intestinal disaccharidases and the ER glucosidases I and II, over β-glucosidases[200-202]. Slight modifications to the DNJ ring can affect specificity of the compounds. In the search for more specific inhibitors of the

ceramide glucosyltransferases involved in glycolipid biosynthesis, the stereo-chemistry was changed from glucose-like (*N*B-DNJ) to galactose-like (*N*B-DGJ) which was sufficient to produce an inhibitor which no longer targets the ER α-glucosidases nor many intestinal glucosidases such as sucrase or maltase[203]. Modification of CST with a glucoside can increase the specificity for sucrase over the other gut glucosidases[204], while alkyl groups added to DNJ can make inhibitors specific for the non-lysosomal glucocerebrosidase, over other cellular α-glucosidases[205].

Greater specificity for their target enzymes means that these iminosugars can be better suited to treat disease than their non-specific counterparts.

1.3.2. Use in disease

Mulberry leaves were traditionally used in ancient Chinese medicine to treat diabetes: the iminosugar they contain was later found to act on the intestinal glucosidases, which break down complex carbohydrates into monosaccharides[196]. DNJ isolated from these plants decreases blood glucose levels, whilst other polysaccharides from the plant could help reduce damage to pancreatic islets[206]. Derivatives of DNJ have been developed to increase potency and specifically inhibit the intestinal α-glycosidases[207, 208], eventually leading to the development of Miglitol (Glyset, *N*-hydroxyethyl DNJ), which is now used clinically against type II diabetes mellitus[209-211]. Miglitol is thought to act via the inhibition of α(1,4) glucosidases, which in turn reduces glucose uptake from the gut and therefore reduces levels of glucose, insulin and glycated haemoglobin in the blood, although it does come with side effects of gastrointestinal discomfort[212, 213].

*N*B-DNJ, otherwise known as Miglustat or under the trade name Zavesca, has a slightly longer alkyl tail than Miglitol and was approved for the treatment of type I Gaucher's disease in 2002[214]. This lysosomal storage disease is caused by the lack of glucocerebrosidase activity and a build-up of glucosylceramide in the lysosome. The first available treatment of the disease was by enzyme replacement therapy yet costs hundreds of thousands of pounds per year of treatment. The approval of *N*B-DNJ helped make treatment more accessible and cheaper for many Gaucher patients[215, 216].

*N*B-DNJ has also been approved for the treatment of Niemann-Pick disease type C, a lipid storage disorder, in 2009[217, 218]. The molecule works via a similar mechanism to that of the treatment of Gaucher's disease, by inhibiting a glucosyltransferase in the primary step of glycosphingolipid synthesis, thus preventing the toxic accumulation of glycosphingolipids in lysosomes[219]. Whilst not curing the disease, which is caused by inherited mutations in the *Niemann-Pick C1* or *C2* genes, *N*B-DNJ stabilises or improves a number of pathogenic manifestations of the disease[220]. In general, iminosugars have broad applicability to many lysosomal storage diseases and iminosugar treatment has been suggested for Fabry's disease, Pompe disease, Tays-Sachs disease and Sandhoff disease[221-224].

Studies based on obese rodent models with insulin resistance found that some iminosugar derivatives have the potential to be used as drugs to fight obesity. For example, AMP-DNJ[205] decreases blood glucose levels and food uptake and increases fat metabolism via a combined mechanisms of inhibition of gut glycosidases and of glycosphingolipid synthesis[225-227].

Iminosugars have even been suggested as possible anti-cancer drugs. CST and swainsonine are both characterised as interfering with protein N-glycosylation: the former as an inhibitor of the ER α-glucosidases and the latter as an inhibitor of Golgi mannosidase II[228, 229]. The two compounds have both shown good potential for use alone or in combination as anti-metastatics[230-235]. Swainsonine, particularly, has passed Phase I and Ib clinical trials, although it later showed little therapeutic effect in a Phase II trial[236, 237].

1.3.3. Iminosugars as antivirals

The broad range of potential applications of iminosugars is not limited to those described so far: the inhibition of ER α-glucosidases has been implicated in the antiviral activity of iminosugars.

Schlesinger *et al.* were among the first to study the effects of iminosugars on the replication of viruses, finding that inhibitors of the ER α-glucosidases caused lower growth of vesicular stomatitis virus and Sindbis virus, evidence that showed the reliance of some viral glycoproteins on simply the removal of glucose from their glycans[238, 239]. Over the next two decades, antiviral activity of iminosugars has been demonstrated against influenza (albeit possibly via interference with lipid oligosaccharide synthesis)[240-242], mouse hepatitis virus[243], a murine leukaemia virus[244], human cytomegalovirus[245], human immunodeficiency virus[246, 247] and herpes simplex virus[248] among many others.

Just these few examples represent diverse families of viruses such as *Orthomyxoviridae*, *Coronaviridae*, *Herpesviridae*, *Retroviridae*, *Rhabdoviridae* and *Togaviridae*, emphasising the scope of iminosugars as broad-spectrum antivirals.

Mechanistic studies have found that the way in which virus replication, formation or infectivity is affected by iminosugars can be quite varied. In a study of dengue virus, antiviral activity of iminosugars such as CST and DNJ was originally found to affect the folding and formation of complexes between prM and Env proteins with unprocessed glycans[249]. *N*-nonyl DNJ, an iminosugar with a long alkyl tail, was shown to affect the interaction of dengue glycoproteins with calnexin, adding credence to the hypothesis that ER α-glucosidases were targeted[250]. As suggested for influenza virus, the inhibition of enzymes involved in glycosphingolipid biosynthesis was a possible iminosugar antiviral mechanism but was recently discounted[153]. Phase I clinical trials of *N*-9-methoxy,nonyl-DNJ (M*O*N-DNJ) against dengue virus, as well as celgosivir have already taken place, with a phase II trial planned for the latter[251].

DNJ derivatives targeting the ER glucosidases were shown to have an antiviral effect against human immunodeficiency virus (HIV). The exact nature of the effect is under debate, but it is thought that GP120 folding is relatively unaffected as it can still recognise and bind CD4 in virions normally even when cells are treated with the compound. Yet DNJ-based compounds appear to change the conformations of three GP120 loops and affect the ability of the virion to fuse with the target cell[252, 253]. In clinical trials, both CST and *N*B-DNJ have made it to Phase II against HIV, however, therapeutic concentrations of the iminosugars in serum were hard to maintain, whilst causing intestinal distress[251]. An ER α-glucosidase inhibitor that does not also target intestinal glucosidases is highly desirable.

Replication of hepatitis B virus (HBV), a member of the *Hepadnaviradae* family of viruses, is also inhibited by iminosugar derivatives. *N*B-DNJ and *N*-nonyl DNJ are ER α-

glucosidase inhibitors and have been suggested to affect the folding of HBV's glycoproteins; some of the antiviral effects were independent of the calnexin cycle though, as DGJ derivatives maintained a similar effect, despite being unable to inhibit ER α-glucosidases[254]. *N*N-DGJ is proposed to affect the virus via glycosphingolipid biosynthesis and processing only, in stark contrast to the mechanism of action of iminosugars against dengue virus where DGJ based compounds showed no antiviral effect[153].

Whilst iminosugars do hold promise as broad-spectrum antivirals, they are not necessarily beneficial for all enveloped viruses tested so far. Initial tests of iminosugar lead compounds against Ebola virus found that they protected mice in a lethal mouse model and inhibited the ER α-glucosidases[255]. However, more recent work failed to find efficacy in other rodent models, although it could be due to the host organism used or that inhibitory concentrations of iminosugars were not reached[256].

Encouraging as antiviral iminosugar therapy may be, further work is needed to check their efficacy against additional target viruses, at the same time ideally improving their specificity towards the ER α-glucosidases and away from the intestinal glycosidases which cause unpleasant side effects.

1.4. Project aims

One important requirement for the development of new, broad-spectrum antivirals is the need for greater specificity towards the target enzymes, the ER α-glucosidases. Potent drugs whose therapeutic dose is too low to inhibit the intestinal glucosidases and/or

selective drugs that are no longer able to inhibit them in the first place, will be essential to avoid the gastrointestinal distress caused by many iminosugars.

Work within the Zitzmann laboratory is already progressing on the front of more potent and specific ER glu II inhibitors, with the *Mm* trypsinised glu II crystal structure determined by itself and in complex with many active site ligands; and fragment-based lead development having revealed novel ligand binding sites on the enzyme.

On the glu I front, development of potent and/or specific inhibitors will require a deeper understanding of the mechanisms of substrate and ligand recognition. In Chapter 3, we aim to obtain structural data on a mammalian glu I in order to better understand how iminosugars bind to it and to identify interesting binding sites which could be exploited by specific ligands in future.

Widening the search for antiviral calnexin cycle modulators beyond ER α-glucosidase inhibitors has been the second aim of this project: many other ER enzymes and proteins are involved in surveillance of enveloped viral glycoprotein folding and associated processes. These proteins too have the potential to be targets for antivirals. By targeting proteins other than ER α-glucosidases, some of the known side-effects that are associated with glucosidase inhibitors could be avoided.

A second important enzyme within the calnexin cycle is its checkpoint enzyme, UGGT, which can be modulated by Sep15, a small TRXL-containing selenoprotein. The UGGT: Sep15 interaction constitutes a unique target within the human body for small molecule ligands that may lower UGGT's activity under stress and act as novel antivirals. The characterisation of this interaction is presented in chapter 4.

ENTPD5, the ER-UDPase, is another potential alternative target for antivirals in the ER. Chapter 5 describes the work to further characterise this enzyme and the potential use of UDPase inhibitors as antivirals.

Finally, in Chapter 6, CRISPR/Cas9 knock-outs of proteins associated with the calnexin cycle were generated to assess the effect that these proteins have on the replication and infectivity of two topical viruses, dengue and Zika virus.

2. Materials and methods

All chemicals were obtained from Sigma-Aldrich and all enzymes were obtained from New England BioLabs, unless otherwise stated. All solutions were made up using ultrapure reagent-grade type 1 water from a MilliQ Academic apparatus (Millipore) except those for RNA/DNA, which were made up in nuclease-free H_2O, not diethylpyrocarbonate (DEPC) treated (Ambion). All 0.22µm filtering of solutions was performed with Millipore Express PLUS 0.22µm PES membranes (Millipore). All

molecular biology cloning utilised NEB® 5-alpha Competent *E. coli* (high efficiency) cells (Cat #C2987). All graphs were plotted and statistics calculated using GraphPad Prism 7.0 (GraphPad Software Inc.).

2.1. Cloning strategies

2.1.1. Oxford protein production facility

High-throughput cloning and screening for expression of *Mus musculus* (*Mm)* ER α-glucosidase I (Glu I) was carried out by Dr Alessandro Caputo under the supervision of Dr Louise Bird at the Oxford Protein Production Facility (OPPF)[257]. Primers were designed by Dr Bird and expression carried out under OPPF standard operating procedures (SOPs) (see OPPF standard protocols: Cloning and Expression Screening, https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/).

High-throughput screening of *Homo sapiens (Hs)* UDP-glucose: glycoprotein glucosyltransferase 1/2 (UGGT1/ UGGT2) and *Hs* ectonucleoside triphosphate diphosphohydrolase 5 (ENTPD5) were carried out by myself under the supervision of Dr Jo Nettleship at OPPF according to the same SOPs.

2.1.2. Cloning and expression vectors

Cloning vector Litmus28i was obtained from New England BioLabs. Vectors pETM-43, pETM-33 and pET303 (Thermo Fisher Scientific) were used for bacterial expression of protein constructs. The mammalian expression vector pHLsec was a gift from Prof Radu Aricescu[258].

2.1.3. Ligase dependent cloning

For sticky-end cloning of constructs, the DNA construct and target vector were digested with an identical pair of restriction enzymes. For cloning between pHLsec and Litmus28i, 20 units of both AgeI-HF and KpnI-HF were incubated with 1x CutSmart Buffer (New England BioLabs) and 500ng of DNA and digested at 37°C overnight. DNA was run on a 0.8% agarose gel and the correctly-sized fragment excised and purified using the QIAquick Gel Extraction Kit (QIAgen).

Ligation of linearised vector and construct was achieved by mixing the two in a 1:1 ratio with 1 µl T4 DNA ligase in 1x T4 DNA ligase buffer (New England BioLabs). The mix was incubated at room temperature for 30 minutes followed by 10 minutes at 65°C. From the reaction mix, 5 µl was used to transform competent cells.

2.1.4. Gibson assembly

Gibson assembly was carried out using the Gibson Assembly® Master Mix and NEBuilder® HiFi DNA Assembly Master Mix from New England BioLabs according to the manufacturer's protocols. For Gibson assembly from an existing DNA construct within the laboratory, PCR primers were designed with a 5′ 15-25bp overlap with the target DNA vector. For Gibson assembly of new DNA constructs into a vector, constructs were designed with the overlap already added to both 5′ ends of the target sequence.

2.1.5. Site-directed mutagenesis

For mutagenesis of constructs in pHLsec, the constructs were first moved into the Litmus28i vector using ligase-dependent cloning (2.1.3.). For mutagenesis of constructs

in bacterial expression plasmids, mutagenesis was done *in situ*. Primers for mutagenesis were designed using the NEBaseChanger® tool from New England BioLabs (see Appendix).

Cloning was carried out with the Q5® Site Directed Mutagenesis Kit (New England BioLabs) according to the manufacturer's protocol. In short, PCR was first used to amplify the DNA and introduce the mutation(s): In a 20 µl reaction, 1 µl of both 10 µM forward and 10 µM reverse primers were mixed with < 1 ng of template DNA, 10 µl of 2x Q5 Hot Start Master Mix and made up to the final volume with nuclease-free water.

Thermocycling was carried out following the manufacturer's protocols, and by varying the annealing temperature and extension time to suit the amplification primers and length of the target DNA.

Circularisation of the vector was carried out as follows: 1 µl PCR product, 5 µl KLD 2x reaction buffer, 1 µl 10x KLD enzyme mix and 3 µl nuclease-free water were incubated for 10 minutes at room temperatures before 5 µl was used to transform competent *E.coli*.

2.1.6. CRISPR cloning

The pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid was a gift from Feng Zhang (Addgene plasmid # 62988). It contains puromycin and ampicillin resistance genes, a Cas9 expression cassette and template single guide (sg) RNA for insertion of 20bp of DNA complementary to the targeted region[259]. Cloning was based on the Feng Zhang laboratory general target sequence cloning protocol and was carried out as follows:

85 µl of PX459 (400 ng/µl) was mixed with 5 µl BpiI (Thermo Fisher Scientific) and 10 µl of 10x buffer G (Thermo Fisher Scientific). This mix was incubated overnight at 37°C and

then run on a 0.8% agarose (a4-0700, geneflow) in Tris-Acetate-EDTA (TAE) buffer. The DNA band corresponding to the linearised vector was extracted using the QIAquick gel extraction kit (QIAgen).

Annealing step: 1 µl of forward primer (100µM) and 1 µl of reverse primer (100µM) were mixed with 0.5 µl T4 polynucleotide kinase (M0201S, New England Biolabs), 1 µl T4 DNA ligase buffer (New England Biolabs) and 6.5 µl nuclease-free water. The mix was incubated at 37°C for 1 hour, at 95°C for 5 minutes and then the temperature was subsequently ramped down to 25°C at a rate of 5°C per minute, all before chilling the reaction on ice.

Ligation reaction: 2 µl of linearised PX459 (0.07pMol) plasmid was mixed with 1 µl annealed primers, 2 µl T4 DNA ligase buffer, 0.5 µl T4 DNA ligase (New England Biolabs) and 14.5 µl nuclease-free water. This mix was incubated at room temperature for 3 hours, at 65°C for 10 minutes and then stored on ice.

An aliquot of competent cells was transformed with 5 µl of ligation reaction and then plated on LB Agar + 0.1 mg/ml carbenicillin plates and incubated overnight at 37°C. Single colonies were picked and grown in 5 ml LB overnight in a shaking incubator at 37°C. Plasmids were purified using the QIAprep Spin Miniprep Kit (QIAgen, #27106) and sent for sequencing using primer 26_BbsI_fwd (see Appendix). Results were compared with the expected sequence to confirm the correct insertion.

Design of primers

The Uniprot entry for the protein to be knocked out was used as a source of sequence information. NCBI tblastn was used to search an organism's mRNA library for the

translated protein sequence. The sequence of the mRNA from that organism was then inserted into NCBI Blast Genomes, targeting the search to the same organism, to get the DNA sequence of the gene of interest, partitioned into its exons.

Two exons for each gene to be knocked out were chosen for targeting when between 2-10 kb apart. These stretches of DNA were inserted into the fly CRISPR tool: http://tools.flycrispr.molbio.wisc.edu/targetFinder/[260] to find potential CRISPR sites with the "NGG" protospacer adjacent motif (PAM) sequence. Potential targets were checked for off-target effects by searching NCBI Blast Genomes with the sequences and assessing whether this was likely; any hit with 3 or fewer mismatches led to the rejection of the target sequence. Other well-known CRISPR design tools were not used as none could search the *Chlorocebus sabaeus* genome, required for CRISPR of Vero cells. Each target was modified as follows to produce two primers for cloning into PX459 and ordered as standard desalted oligonucleotides:

5′ **CACCG**NNNNNNNNNNNNNNNNNNN 3′

3′ **C**NNNNNNNNNNNNNNNNNNNN**CAAA**5′

Non-Deletion (ND) and Deletion (D) primers (used to screen for genomic deletion by PCR), as well as 'scarring primers', used to amplify the region around each targeted loci and allow for sequencing, were designed as described by Orkin *et al*[261].

2.2. Protein expression

2.2.1. Protein expression in *E.coli*

E. coli BL21 (DE3) were transformed with vector containing an inducible bacterial expression cassette: 10 ng of plasmid DNA was mixed with 50 µl of chemically competent of cells and incubated on ice for 30 minutes, in a 42°C water bath for 30 seconds and returned to ice for 5 minutes before growing the cells in SOC outgrowth medium for 1 hour at 37°C and plating them on appropriate LB agar + antibiotic plates and incubation at 37°C overnight. Kanamycin (pET/M vectors) was used at a final concentration of 0.05 mg/ml and carbenicillin (all other vectors) was used at 0.1 mg/ml respectively.

A single colony was picked from the plate and grown in 5 ml LB + antibiotic (at the same concentrations as the agar plates) overnight at 37°C. The following morning, 2.5 l baffled flasks containing 500 ml LB + antibiotic were seeded with 1 ml of the overnight culture and incubated at 37°C in a shaking incubator at 150 rpm. Once the absorbance (600 nm) of the culture reached 0.6, flasks were rapidly cooled to 16°C before addition of a final concentration 0.05 mM isopropyl-β-D-thiogalactoside (IPTG). Cultures were incubated at 16°C overnight at 150 rpm before harvesting.

2.2.2. Protein expression in HEK293F cells

Freestyle™ 293-F cells (Thermo Fisher Scientific) were grown and passaged in Freestyle™ 293 expression media (Thermo Fisher Scientific) according to the manufacturer's protocols, ensuring that cell density stayed within the limits of 0.1-2.0x10^6 cells/ml and the passage number did not exceed 40. Cells were cultured in sterile Erlenmeyer flasks,

restricting the volume of culture to 40% of the maximum capacity of the flask, in a shaking incubator at 150 rpm, 37°C and 8% CO_2.

Transfection of cells were carried out with Freestyle™ MAX reagent (ThermoFisher) according to the manufacturer's protocol

Expression conditions were optimised by determining when the expression of the desired protein was at a peak. An aliquot of the culture was harvested every day on days 1-5 after transfection, cells spun down and supernatants analysed on an SDS-PAGE gel in order to establish when maximal protein expression occurred.

2.3. Protein purification

2.3.1. *E.coli* lysate purification

Cells were spun at 5000 xg for 30 minutes at 4°C. The supernatant was discarded and cells resuspended in 25 ml of 2x phosphate buffered saline (PBS), 225 mM NaCl, 0.5 mM ethylenediaminetetraacetic acid (EDTA) and 0.05 mM Tris(2-carboxyethyl)phosphine (TCEP) with the pH adjusted to 7.4 and supplemented with lysozyme (Sigma Aldrich #L6876), DNaseI (Roche #10104159001) and protease inhibitors (Sigma Aldrich #05892953001), per 5g of cell pellet. Cells were hand homogenised using a Potter-Elvehjem style mortar and pestle before sonication: 12 cycles of 15 second, 85 Watt bursts separated by 30 second rests on ice. The lysate was spun at 35,000 xg for 30 minutes at 4°C. The resultant supernatant was 0.22 µm filtered.

2.3.2. Mammalian sample preparation

Once protein expression was at its optimal level, the expression culture was harvested using a 1500 xg spin for 10 minutes. 10x PBS was added to the supernatant so that a final concentration of 1x PBS was achieved, the pH was adjusted to 7.4 with NaOH, and it was filtered with a 0.22 µm filter.

2.3.3. Affinity purification

Three different pre-packed columns were used for affinity purification of proteins: HisTrap™ excel, MBPTrap™ HP or GSTrap™ HP (GE Healthcare Life Sciences). Either a 1 ml or 5 ml column was used depending on the level of expression achieved: the smaller column was used preferentially to reduce non-specific binding. All columns were stored in 20% ethanol.

HisTrap™

Binding Buffer: 1x PBS, pH 7.4 (0.22 µm filter-sterilised)

Elution Buffer: 1x PBS, 500 mM imidazole, pH 7.4 (0.22 µm filter-sterilised)

Buffers were supplemented with 5% glycerol and 0.1 mM TCEP for purification of *Mus musculus (Mm)* glu I constructs.

The HisTrap™ column was prepared by flowing 5 column volumes (CV) MilliQ water through the column, followed by 5CV of each binding, elution and again binding buffers. Samples were prepared as described and loaded onto the column using an ÄKTA Start chromatography system in a 4°C cabinet at a flow-rate of 1 ml/minute. Once loaded, the

column was washed with 20CV of binding buffer and eluted using a gradient elution of 0-100% elution buffer over 20CV. Fractions of all steps were collected.

GSTrap™

Binding Buffer: 1x PBS

Elution Buffer: 50 mM Tris pH 7.6, 10 mM glutathione and 5 mM dithiothreitol (DTT)

The GSTrap™ column was prepared with 5CV of MilliQ followed by 5CV of each binding, elution and again binding buffers. Sample was loaded and washed in the same manner as for the HisTrap™ and eluted with a block step of 100% elution buffer. Fractions were collected at all stages of the purification.

MBPTrap™

Binding Buffer: 20 mM Tris, 200 mM NaCl, pH 8.0

Elution Buffer: 20 mM Tris, 200 mM NaCl, 10 mM maltose, pH 8.0

The MBPTrap™ column was prepared in the same manner as the HisTrap™ column. Sample was loaded at 1 ml/minute and eluted with a block step of 100% elution buffer. Fractions from all steps were collected.

2.3.4. Size exclusion chromatography

All size exclusion chromatography (SEC) columns were obtained from GE Life Sciences. Columns were chosen depending on the size of the protein in solution and the scale of the preparation done. For analytical scale SEC runs, 10/300 GL column sizes were used; for preparation scale runs, 16/600 PG size columns were used.

For proteins between 10 kDa and 50 kDa, a Superdex 75 was used. For proteins between 40 kDa and 300 kDa, a Superdex 200 was used. All columns were stored in 20% ethanol and run using an ÄKTA pure FPLC chromatography system at room temperature. Before equilibrating into running buffer, columns were equilibrated with 1.5CV MilliQ. All buffers used for SEC were degassed and 0.22 µm filter-sterilised. The exact buffer used depended on the protein analysed All samples were 0.22 µm filter-sterilised upon loading onto the injection loop.

Mm Glu I	20 mM HEPES, 100 mM NaCl, 0.1 mM TCEP, 5% Glycerol, pH 7.2
Hs Sep15-6xHis	20 mM HEPES, 100 mM NaCl, 0.1 mM TCEP pH 7.4
Hs MBP-ENTPD5	20 mM Tris, 150 mM NaCl, 2 mM $CaCl_2$, pH 8.0

For all other constructs, 20 mM HEPES, 100 mM NaCl, pH 7.4 was used. 1.2CV of buffer was used for each run at a flow rate of 1 ml/min and fractions were collected after 0.2CV.

2.3.5. Anion exchange

A MonoQ® 5/50 GL column (GE life sciences) was used for anion exchange using an ÄKTA pure FPLC chromatography system at room temperature. Both buffers were degassed and 0.22 µm filtered before use. The column was stored in 20% ethanol.

Binding Buffer: 20 mM Tris, pH 8.0

Elution Buffer: 20 mM Tris, 1 M NaCl, pH 8.0

The column was equilibrated with 8CV MilliQ before 8CV of each, binding, elution and binding buffer again. Sample was diluted at least 1 in 20 using binding buffer and loaded onto the column sing a 50 ml superloop, at 0.5 ml/minute. Elution was performed with a

100CV gradient from 0-100% elution buffer with fractions collected throughout the elution step at 1 ml/minute.

2.3.6. Protein concentration

Before SEC purification or long-term storage of protein, it was necessary to concentrate the protein. For proteins derived from *Chaetomium thermophilum*, Sartorius Vivaspin 20, PES membrane centrifugal filter units (Generon) were used for concentration with an appropriate molecular weight cut-off (MWCO) for the sample, generally 100 000, 30 000 or 10 000 Daltons. For mammalian proteins, Amicon Ultra-15 centrifugal filter units, regenerated cellulose (Merck Millipore), were used for concentration with an appropriate MWCO: 50 000 or 10 000 Daltons.

2.4. Protein modification

2.4.1. Refolding *in vitro*

For protein constructs that were purified in a misfolded state, denaturation and refolding of the protein was necessary. To 1 part protein at 10 mg/ml, 4 parts of 8 M GuHCl, 7 mM EDTA, 7 mM DTT, 133 mM Tris-HCl was added and incubated for 10 minutes at 65°C to fully unfold the protein. To this, 50x volume of 150 mM Tris-HCl, 10 mM KCl, 250 mM NaCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 800 mM L-arginine, 2 mM reduced L-glutathione and 2 mM oxidised L-glutathione, pH 8.3 was added and the mix left at 13°C.

After 3 days, the sample was dialysed in 10 kDa MWCO Snakeskin Dialysis Tubing (ThermoFisher) against 10x the volume of 50 mM Tris, pH 8.0, 250 mM NaCl, 10 mM $CaCl_2$, with three buffer changes over the course of two days at 4°C, to get a 1000-fold

dilution. This sample was run on a HisTrap™ and then a Superdex 200 10/300 SEC column in order to purify the refolded protein.

2.4.2. Methylation

Methylation was carried out broadly as per the protocol in the reductive alkylation kit sold by Hampton Research. All steps were carried out on ice or at 4°C.

In brief, protein was dialysed into 20 mM HEPES, 150 mM NaCl, 5% w/v glycerol, pH 7.5 and subsequently concentrated to 1-5 mg/ml. To this, 1/50th v/v of 1 M dimethylamine borane (DMAB) and 2/50th v/v 1 M formaldehyde were added and left gently rolling for 2 hours. The same additions were made once more and left for a further 2 hours. 1/100th v/v 1M DMAB was added and left overnight.

The following morning, the reaction was quenched by the addition of 1/8th v/v 1 M glycine and 1/1000th v/v 1 M DTT and incubated for 2 hours. Methylated protein was purified by SEC. The methylation state of the protein was assessed by Dr Abhinav Kumar by tryptic-digest-MS/MS to identify sites of modification. Both lysine di-methylation and methionine oxidation were detected.

2.4.3. Proteolysis

3C protease cleavage

For a 1 ml volume digestion, 20 µl of human rhinovirus (HRV) 3C protease (Novagen) at 2000 U / ml was mixed with 100 µl of 10x 3C protease buffer and 880 µl protein sample. This was incubated at 4°C for 40 hours before affinity purification and SEC.

Other proteases

Proteases trypsin, chymotrypsin, GluC endopeptidase (all from ThermoFisher), proteinase K (New England BioLabs) and *Bacillus Sp.* protease (Sigma-Aldrich) were made up in water; protein was dialysed into 100 mM Tris pH 8.2, 100 mM NaCl, 2 mM $CaCl_2$. Proteases were added at ratios of protein: protease between 1:10 and 1:10000. These were left overnight at 37°C before analysis of the extent of proteolysis.

2.5. Gel electrophoresis

2.5.1. SDS-PAGE

SDS-PAGE was carried out using the Thermo Fisher Scientific Novex® systems following the supplied protocols. For SDS-PAGE, pre-cast 1 mm NuPAGE™ 4-12% Bis-Tris Protein gels were used with 10, 12 or 17 wells. For proteins smaller than 100 kDa, gels were run in NuPAGE® MES buffer for 35 minutes at 200V. For larger proteins, gels were run in NuPAGE® MOPS buffer for 50 minutes at 200V. Gels were stained using InstantBlue protein stain (Expedeon).

Samples were prepared by adding 4x LDS sample buffer (Thermo Fisher Scientific), plus 50 mM DTT if required, heated to 95°C for 1 minute, and then loaded onto the gel.

2.5.2. Western blotting

The iBlot 2 Dry Blotting System (Thermo Fisher Scientific) was used for the transfer of SDS-PAGE gels to polyvinylidene difluoride (PVDF) membranes (iBlot 2 PVDF stack) following the manufacturer's standard protocol for high molecular weight proteins. The membrane was put in 5% w/v milk powder in 1x TBS + 0.1% Tween 20 (TBS-Tween) for 1 hour on a plate rocker. The membrane was then washed with TBS-Tween thrice for 5

minutes each to remove the milk. Primary antibody was diluted into TBS-Tween and added to the membrane, and left on the plate rocker for 1 hour before being washed in the same manner as before. Secondary HRP-Conjugated antibody was diluted into TBS-Tween and added to the membrane for a further hour before being washed. Amersham ECL Prime Western Blotting Reagent (GE Life Sciences) was used to develop the membrane according to manufacturer's protocols and imaged on an LAS-1000 Luminescence Image Analyser (Fuji Medical Systems).

For detection of hexahistidine-tagged proteins, the 6x-His tag monoclonal antibody (3D5), HRP-conjugated antibody (Thermo Fisher Scientific #R93125) did not require a secondary antibody and so the western blot was developed after the second wash step. *Hs* ENTPD5 was detected with rabbit anti-ENTPD5 (abcam #ab108603), *Hs* UGGT1 with mouse anti-UGGT1 (Santa Cruz Biotechnology #sc-374565), *Hs* Sep15 with rabbit anti-Sep15 (Sigma Aldrich #AV48247) and *Hs* glu I with rabbit anti-glu I (abcam #ab155231).

2.5.3. DNA gels

DNA agarose gels were prepared using agarose (a4-0700, geneflow) in TAE buffer. GelRed nucleic acid gel stain (Biotium) was added to molten agar, before pouring into a mould, to allow for visualisation of DNA bands under UV excitation. Once set, gels were run at 150V for 50 minutes in TAE buffer. Samples were prepared by the addition of 6x purple loading dye (New England BioLabs). For DNA fragments greater than 1000bp in size, 0.8% agarose w/v agarose gels were prepared, for those between 500 and 1000 base pairs, a 1.5% w/v agarose gel was used and for fragments predicted to be shorter than 500bp, a 2% w/v agarose gel was used.

2.6. Protein characterisation

2.6.1. Dynamic scanning fluorimetry

Differential Scanning Fluorimetry (DSF) experiments were prepared to discover buffer conditions in which a protein was most thermostable. In MicroAmp™ Fast Optical 96-well reaction plates (Thermo Fisher Scientific), 3 µl of protein between 0.5 – 0.2 mg/ml with 25x SYPRO™ Orange dye (Thermo Fisher Scientific) was added to 27 µl of screening solutions, in duplicate. Conditions were designed to screen a range of pH values, salt concentrations and diverse additives. Plates were sealed with MicroAmp™ clear adhesive film (Thermo Fisher Scientific) and spun at 1500 xg for 5 minutes.

DSF experiments were measured on an Applied Biosystems® 7500 Real-Time PCR System set to measure the ROX reporter. The temperature was equilibrated for 2 minutes at 25°C before being ramped up to 95°C at a rate of 1°C/minute; values were read at every degree increase. Melting temperature (T_M) data was obtained by fitting the data to a Boltzmann sigmoidal curve.

2.6.2. Circular dichroism

Proteins were dialysed into 10 mM $NaPO_4$, 20 mM NaCl, pH 7.4 using Slide-A-Lyzer™ MINI Dialysis Devices, 10K MWCO, 0.1 ml (ThermoFisher) using a 1/1000 dilution thrice over the course of two days. Samples after dialysis were diluted to 0.1 mg/ml and were 0.22 µm filtered. A Jasco J-815 Circular Dichroism spectropolarimeter was coupled to a Peltier thermostat set to 16.87°C. Buffer blanks and sample were collected in a 1 mm Quartz SUPRASIL® high precision cell (Hellma Analytics) using the following settings:

wavelength range of 260-185 nm, measurements every 0.5 nm, a spectral width of 1 nm, a scan rate of 20 nm/minute, a 1 second digital integration time and 5 repeats per wavelength.

The buffer blank was subtracted from the sample and smoothened using the spectrometer's software. Data was converted into Δε and submitted to the DiChroWeb server using the CDSSTR programme and reference data sets 4/7 (optimised for wavelengths 240-190 nm) to derive secondary structure information[262, 263].

2.6.3. SEC-MALS

Size exclusion chromatography multi-angle laser light scattering (SEC-MALS) experiments were carried out by Dr David Staunton of the Biophysics Suite, Department of Biochemistry, University of Oxford. Samples were run on a Superdex 200 10/300 SEC column on a Shimadzu Prominence HPLC system, at a concentration of 1 mg/ml and in a buffer identical to that of the sample. This was connected to a Wyatt Dawn HELEOS-II 8-angle light scattering detector and a Wyatt Optilab rEX refractive index monitor.

2.6.4. Small angle X-ray scattering

A Superdex 200 10/300 SEC column was equilibrated in 20 mM HEPES, 120 mM NaCl, pH 7.4. Three samples were prepared and run:

A. *Chaetomium thermophilum* (*Ct*) UGGT chimera with *Homo sapiens* Sep15 binding loop (UGGT chimera)
B. UGGT chimera + 100 fold molar excess of Sep15-6xHis
C. UGGT chimera + 50 fold molar excess of MBP-Sep15

Fractions containing the protein/complex were concentrated to > 10 mg/ml and a minimum volume of 100 µl. Buffer samples were taken from fractions in the void volume of the SEC run for preparation of dilution series and blanks. Samples were transferred to the BM29 beamline at the European Synchrotron Radiation Facility (ESRF) in Grenoble on dry ice.

At the synchrotron, each sample was 0.22 µm filtered, the OD_{280} taken on a NanoDrop 1000 (Thermo Fisher Scientific) and concentrations in mg/ml recalculated. A dilution series was prepared in ½ dilution steps to produce 6 samples. Samples were loaded into the sample-changer, equilibrated to 20°C. For each dilution of each sample, 40 µl were flowed through a quartz capillary, at 100% transmission, and scattering images collected at 0.5 seconds per image and 10 images per sample. Immediately preceding and following each protein sample, buffer blanks were collected under the same conditions. Beam energy was 12.5 keV (λ=0.99187 Å) and the Pilatus 1M detector was set 2.867 m from the sample.

Data analysis was done using the ATSAS 2.8.3 programme suite[264]. Data was averaged and buffer subtracted using the SAS Data Analysis GUI, using Primus to evaluate the Guinier region and GNOM to estimate the P(r) function[265, 266], before the best data set from each dilution series was chosen, as assessed on the basis of data signal to noise and signs of inter-molecular repulsion or aggregation in the low-resolution portions of the SAXS spectra. *Ab initio* model building was carried out using DAMMIF, which generated 20 models using slow-mode analysis. These were aligned and scored using normalised structural differences (NSD) in DAMAVER, those with the largest NSD were rejected and a most probable model was generated in DAMFILT. Model visualisation used PyMOL

and the PyMOL plug-in, SasPy. SREFLEX and SASREF were used for model manipulation[267-271].

2.6.5. The MESG activity assay

For proteins with phosphatase activity, enzymatic activity could be measured with the MESG assay. This is based on the hydrolysis of a phosphate from a nucleotide that subsequently reacts with 7-methyl-6-thioguanosine (MESG) via a secondary enzymatic reaction catalysed by a nucleoside phosphorylase, leading to an increase in the absorbance at 360nm, as outlined in [272].

MESG was obtained from Berry & Associates, purine-nucleoside phosphorylase (PNPase, Thermo Fisher Scientific # 40809) was purified by SEC on a Superdex 200 16/600 column into 20 mM HEPES, 50 mM NaCl, pH 7.4, fractions containing active protein were pooled, dispensed into 50 µl aliquots and stored at -20°C. All nucleotides were obtained from Sigma-Aldrich and made up in 0.1 M HEPES, pH 7.4.

For a standard 100 µl reaction, in a clear-bottomed 96-well plate, 40 µl of 2 mM MESG was mixed with 2 µl PNPase, 10 µl of a 1 M pH buffer, 5 µl of 20x salts (0.1 M $CaCl_2$, 0.1 M $MgCl_2$, 2 M NaCl), 20 µl of nucleotide, 5 µl of enzyme and made up to 100 µl with H_2O. Reactions were usually started with either nucleotide or enzyme as the last addition. Reactions were measured at A_{360nm} on a CLARIOstar® monochromator microplate reader (BMG labtech) every minute for around 90 minutes. All reactions were carried out in duplicate.

Rates of reaction were measured by taking the linear portion of the results and obtaining a rate of reaction (ΔAbsorbance/min). The background measurement was subtracted

from this value and where necessary, the values were converted to µMol per minute per milligram of enzyme (µMol/min/mg). This was done using an inorganic phosphate (P_i) standard curve that used set concentrations of P_i in lieu of enzyme or nucleotide in the reaction mix. Values were converted as long as the absolute values of measurement were between A_{360nm} of 0.1 – 0.9, the linear portion of the standard curve.

Michaelis-Menten curves were set-up by varying the concentration of nucleotide. The data were processed as above, and analysed in GraphPad Prism 7.0 using the Michaelis-Menten non-linear regression or the allosteric sigmoidal regression if it provided a better fit to the data. IC_{50} curves were set up by the addition of compounds to the reaction mix, allowing time for ligands to associate with enzyme, and using the addition of nucleotide to start the reaction. Data was plotted as a percentage of activity of the control value (no compound) and curves were analysed in GraphPad Prism 7.0 using the log (inhibitor) vs. response -- variable slope non-linear regression, constraining the data to fit the bottom to zero.

2.6.6. Isothermal titration calorimetry

Mm glu I was extensively dialysed into 1x PBS, 0.1 mM TCEP, and the final dialysis buffer retained for preparation of iminosugar inhibitors. Protein was diluted to 25 µM, iminosugar to 300 µM. Calorimetry measurements were carried out on an iTC200 MicroCalorimeter (Malvern Instruments). Heat of dilution experiments were initially run to ensure a minimal background. Titration experiments were carried out with 20x injections of 2µl ligand, each spaced by 180 seconds. Data analysis was carried out on Origin 7 software (OriginLab).

2.7. Crystallisation

2.7.1. Set-up of crystallisation screens

Commercial crystallisation screens were obtained from Hampton Research and Molecular Dimensions. Both commercial sparse screens and custom optimisation screens were dispensed into MRC 2-Well crystallization plates (Swissci) UVXPO, 50 µl into each reservoir. Sitting drops were set up with a Mosquito® system (TTP labtech) as 100 nl protein: 100 nl screen condition in the first sub-well and 200 nl protein: 100 nl screen condition in the second sub-well. Crystal seeds were prepared using MicroSeed Beads (Molecular Dimensions) and following the manufacturer's protocols. Sitting drops with seeds were set up as 150 nl protein with 135 nl screen condition and 15 nl of a 1 in 256 dilution of the crystal seeds. All crystallisation plates were stored in a Rigaku Minstrel™ HT crystallisation hotel and imager at 18°C, with images of wells being taken at both visible and ultraviolet wavelengths at regular intervals over the course of 3 months.

2.7.2. Data collection

Crystals were picked using 50-100 µm nylon loops mounted on pins appropriate for the synchrotron to be used. If the crystallisation condition did not contain an appropriate cryoprotectant, loops containing crystals were passed through a droplet containing 80% mother liquor and 20% ethylene glycol. Crystals were cryocooled in liquid nitrogen prior to shipping to the synchrotron. All data collection took place under a gaseous nitrogen stream at 100 K.

Diamond Light Source (DLS) at Harwell, UK, beamlines I-03, I-04 and I-04-1; ESRF at Grenoble, France, beamline ID30-A and SOLEIL at St Aubin, France, beamline PROXIMA-1 were used to collect data from crystals. Exact parameters used at each of the beamlines are shown in tables in the appendix.

2.7.3. Data processing

Diffraction data were indexed, integrated, scaled, merged and an anisotropic cut-off imposed using the autoPROC toolbox (Global Phasing)[1, 273-275]. Molecular replacement was done using MrBUMP[276] utilising the programme Phaser[277]. Model building and manipulation was carried out using Coot[278] whilst refinement was carried out using BUSTER[279]. Validation of models was done using Coot, the MolProbity[280] and the wwPDB validation service. Structures for publication were deposited to the worldwide Protein Data Bank (wwPDB)[281]. Figures were made using PyMOL[282].

2.8. Cell culture

Cell culture was carried out using aseptic technique in LaminAir HBB2448 tissue culture hoods (Heraeus Instruments). Cell lines were stored in a 37°C and 5% CO_2 incubator. All media and reagents were obtained from Thermo Fisher Scientific unless otherwise stated, and warmed to 37°C before use. Volumes of different media and reagent required are

Flask	Media (ml)	Tryspin-EDTA (ml)	1x PBS (ml)
T175	25	5	10
T75	10	2	10
T25	5	1	5

Table 2.1. Volumes of media, PBS and Trypsin-EDTA used for cell culture in different sized flasks

shown in Table 2.1. Excess media, cells and liquid reagents were disposed of in 2% Virkon® (VWR) and disposed down the drain after mixing for at least 30 minutes.

To passage/split cell lines, the confluency of cells was assessed visually under a light microscope and the dilution to maintain cells within 20-95% confluency was determined. Media was removed from the flask and replaced with 1x PBS. The flask was swirled gently, the PBS removed and replaced with an appropriate volume of trypsin-EDTA (Table 2.1) to lift the cells. The cells were then returned to the incubator for 10 minutes. Flasks were gently rocked and tapped to lift cells, and the cells were resuspended thoroughly through pipetting up and down in a 5 ml serological pipette. A quantity of lifted cells was returned to the flask to obtain a suitable confluency before media was added to the flask and the flask returned to the incubator.

Cell lines were frozen by lifting cells from a T175 flask, pelleting them by centrifuging at 1500 xg for 5 minutes and resuspending the pellets in culture medium + 10% DMSO: 2 ml per T175 flask. Cells were dispensed as 1 ml aliquots in sterile, screw-cap 2 ml tubes and frozen in a Nalgene® cryo 1°C freezing container (ThermoFisher) in accordance to the manufacturer's instructions. After 24 hours, aliquots were stored in a cryostorage tank under liquid nitrogen.

To thaw cells, aliquots were rapidly thawed in a 37°C water bath before they were added to a T175 with 25 ml of pre-warmed media.

2.8.1. HEK293T cells

HEK293T wild-type and the knock-out cell lines of ΔUGGT1, ΔUGGT2 and ΔUGGT1/2 were gifts from Maurizio Molinari (IRB, Università della Svizzera italiana, Switzerland).

Cells were passaged in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS and 1x PenStrep.

2.8.2. Vero cells

Vero E6 cells (ATCC) are an immortalised cell line derived from African Green Monkey (*Chlorocebus sabaeus*) kidney epithelial cells. The growth media used is DMEM supplemented with 10% FBS and 1x PenStrep.

2.8.3. LLCMK2 cells

LLC-MK2 cells (ATCC) are an immortalised cell line derived from Rhesus Monkey (*Macaca mulatta*) epithelial kidney tissue. Cells are grown in Media 199 supplemented with 10% FBS and 1x PenStrep.

2.8.4. CRISPR transfection and screening

Experiments based on the CRISPR/Cas9 system were carried out in Vero E6 cells using the previously described media and broadly to the protocol described by Orkin *et al* [261]. The following protocol describes the knock-out of a single gene.

On *day 0*, Vero E6 cells at a confluency of around 40% in a T25 flask were transfected with the pair of CRISPR plasmids (see 2.1.6. and appendix) for the target gene using the Lipofectamine 3000 system (Thermo Fisher Scientific). In one 1.5 ml Eppendorf tube, 250 µl of OptiMEM (Thermo Fisher Scientific) was mixed with 15 µl of Lipofectamine 3000. In a second tube, 250 µl of OptiMEM was mixed with 18 µl P3000 and 25 µl of each of the two CRISPR plasmids at 500 ng/µl. The contents of the two tubes were mixed and

incubated for 10 minutes at room temperature before adding to the flask in a drop-wise manner. The flask was gently swirled for 10 seconds before returning to the incubator.

On *day 3*, the cells had sufficient time to express the antibiotic resistance gene encoded within PX459. The media was removed and replaced with 5 ml of DMEM supplemented with 10% FBS, 1x PenStrep and 33 µg/ml puromycin (InvivoGen) to kill all cells but those transfected with the highest levels of plasmid containing the resistance gene.

On *day 6*, the puromycin-laced media was removed and replaced with media lacking the antibiotic. Cells were incubated for between 3-7 days to return cells to 20-80% confluency.

On *day 10*, the cells in the flask were lifted with Trypsin-EDTA. Once lifted, the cells were resuspended thoroughly in 10 ml of media. Cells were counted using a Countess II cell counter (Thermo Fisher Scientific) following the manufacturer's protocol. Clonal selection was performed by diluting each knock-out line to 10 cells/ml and dispensing 200 µl of this into each well of three 96-well plates so that there were an estimated 2 cells per well. These plates were returned to the incubator for 14 days or until wells with 'colonies' of 50-500 cells could be identified as well as wells with more than one colony to be marked for exclusion.

Some of the bulk cells were retained and the genome purified using a phenol-chloroform-isoamyl alcohol protocol:

1. Cells were lysed with the addition of 250 µl of 1% w/v SDS, 50mM Tris pH 8.5, 5 mM EDTA, 250 mM NaCl, 0.1 mg/ml proteinase K to a pellet of the cells and incubated at 37°C for 2 hours.

2. To 100 µl of lysed cells, 100 µl of 25:24:1 phenol: chloroform: isoamyl alcohol, pH 8.04 (Sigma-Aldrich) was added. This was shook vigorously for 30 seconds and spun a bench-top centrifuge for 10 minutes at 16,100 xg.
3. The top aqueous phase of the separated mixture (making sure to avoid the interface) was retained and aliquoted into a fresh Eppendorf tube.
4. Steps 2 and 3 were repeated. To the aqueous phase, 250 µl ice-cold 100% ethanol and 10 µl of 3 M Na acetate was added. This was mixed and then stored in a -20°C freezer for 30 minutes.
5. The tube was spun in a centrifuge pre-cooled to 4°C for 30 minutes at 16,100 xg. The supernatant was removed and the pellet resuspended in 500 µl of 70% ice-cold ethanol.
6. This was stored in a -20°C freezer and spun once more. The ethanol was removed and the pellet left to air dry.
7. Finally, the pellet was resuspended in 100 µl nuclease-free water.

The genome was screened using PCR and the Non-Deletion (ND) and Deletion (D) primer pairs to determine if the CRISPR was likely to have been successful, see figure 2.1 for an example.

For the wild-type cells, a band would only be expected for the ND primers. For the bulk knock-out cells, bands in both alleles indicate that some cells have wild-type alleles and some have knocked-out alleles. In a monoclonal populations of cells, it would be expected that some clones only exhibit a Deletion primer band, indicating a full knock-out of the gene.

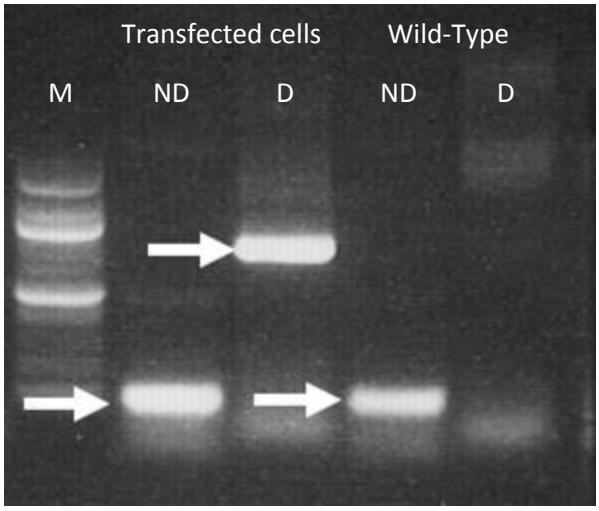


Figure 2.1. DNA gel of PCR screens for CRISPR knock-outs. Markers are 100 bp with the strong bands corresponding to 500 and 1000 bp. 'ND' is non-deletion: bands that only appear if the DNA between the two CRISPR sites has not been excised. 'D' is deletion: bands that appear only if the intervening DNA has been deleted. Arrows highlight the expected band.

On *day 24*, for the wells containing a *single* colony of cells, the media was removed and replaced with 50 µl trypsin-EDTA. After at least 15 minutes, 150 µl of media was added to the cells, immediately mixed and 100 µl aliquoted into each of two identical 96-well plates. Once completed for all of the single-clone wells, 100 µl of media was added to each well of the new plates. One plate was for screening, the other for growth.

Screening took place over the course of a week, cells were left to grow until a likely knock-out candidate was found.

To screen, the media was removed from the wells where there were at least ~100 cells in the "screening plate". 100 µl of lysis buffer described in step 1 of the phenol – chloroform – isoamyl alcohol genome purification protocol was added and the protocol followed for each clone. PCR using Non-Deletion and Deletion primers were used for each clone and the results assessed. If any double allele knock-out cells were identified, these were grown sufficiently to fill 4 T175 flasks at around 80% confluency and frozen for storage. If, after

all screening was completed, the best clone had one band for both Deletion and Non Deletion primers, scarring primers were used to amplify the DNA around each of the two target CRISPR sites and the product was sent for sequencing. Any clones that contained an out-of-frame mutation (via deletion or addition of 1 or 2 bases) were identified as a full knock-out and grown and frozen as described earlier.

2.9. Cellular assays

2.9.1. Toxicity assay

Cytotoxicity of inhibitors was determined using the CellTiter® 96 AQ_{ueous} one solution cell proliferation assay (Promega) according to the manufacturer's protocols.

In short, monolayers of cells, in a 96-well plate, were incubated with inhibitors for a length of time equivalent to the planned antiviral assays. After incubation with the inhibitor, 20 µl of the tetrazolium compound, MTS, was added to the 100 µl of media that covered the cells. Live cells are capable of reducing MTS to a soluble purple compound, allowing colourimetric assessment of the percentage of viable cells. The plate was left in the incubator for 2 hours before reading absorbance at 490 nm on a SpectraMax M5 (Molecular Devices). Absorbance was normalised to untreated cells.

2.9.2. Antiviral assays with Zika virus

Cells were seeded in 24-well plates in triplicate for each cell line or treatment tested, at a cell density of $0.2x10^6$ cells per well, one day prior to infection with Zika virus (strain MR 766). A duplicate plate was prepared so that cell numbers could be assessed in parallel.

On the day of infection with Zika virus, cell counts from the duplicate plate were determined and averaged for each cell line or treatment. Infection of cells was done using a multiplicity of infection (MOI: the ratio of the number of plaque forming units to the number of cells present) of 1. The amount of virus stock required for each well was calculated. From each assay plate, media was removed and cells were washed with ~0.5 ml of HBSS. The virus was put onto cells diluted in 100 µl of MEM media and the plate left on a plate rocker for 90 minutes at room temperature. The virus was removed and replaced with the media appropriate for the cell type, the only difference being that the media were supplemented with 2% FBS instead of 10%. Cells were left in a 37°C, 5% CO_2 incubator for 48 hours.

After 48 hours, the media covering the cells was moved into a deep-well block and spun for 5 minutes at 1,500 xg. Media was then dispensed in 200 µl aliquots into a 96-well tissue culture plate in triplicate; one for qRT-PCR, one for a plaque assay and one as a back-up plate. The plates were sealed in parafilm and stored in a -80°C freezer for further analysis.

2.9.3. Antiviral assays with dengue virus

Dengue antiviral assays were carried out in the same manner as Zika assays except for a few minor changes. All work before addition of virus was carried out in the category 2 designated tissue culture laboratory, all work involving live virus was carried out in a category 3 designated laboratory. Infection of cells by dengue virus was also carried out at a MOI of 1, however, assays were harvested 96 hours after the initial infection.

2.9.4. Plaque assay

Plaque assays were carried out in order to determine the number of plaque forming units (pfu) of virus released from cells. Dengue plaques assays were carried out in broadly the same manner as in [283], with a protocol adapted for Zika by Michelle Hill.

For each individual sample, the following protocols were followed, scaled up for the appropriate number of triplicates and parallel experiments.

For the Zika plaque assay, two days before infection, a 12-well plate was seeded with $0.1x10^6$ Vero cells per well in all wells. For the dengue assay, a 12-well plate was seeded with $0.17x10^6$ LLC-MK2 cells per well, one day before infection.

On the day of infection, plates were checked to ensure an even monolayer of cells had formed. The virus sample was diluted along a 10-fold dilution series, 50 µl of sample in 450 µl MEM in a deep-well block, with dilution factors ranging from 10^{-1} to 10^{-6} (Zika) and from 10^{-1} to 10^{-4} (dengue). The media from the 12-well plate was removed, washed with ~0.5 ml of HBSS and then virus was added. For Zika virus, 100 µl of dilutions 10^{-3} to 10^{-6} were added in triplicate to wells. For dengue virus, 100 µl of dilutions of 10^{-1} to 10^{-4} were added in triplicate to wells. Each well was topped up with a further 100 µl of MEM and the plate was left on a plate rocker for 90 minutes.

For the Zika assay, the virus was removed and replaced with 1 ml of a 1:1 mix of 2% low gelling temperature agarose (ThermoFisher, Cat #A9414) and Zika overlay (see Table 2.2). Plates were left for four days, after which 1 ml of toluidine blue stain (1x PBS, 10% formaldehyde, 0.1% toluidine blue) was added to each well and left for a further 24 hours

at room temperature. Agarose plugs were removed under a stream of water, the plates left to dry, and then plaques counted.

Reagent	Overlay		
	Zika	1st	2nd
Hank's A	50	50	50
Hank's B	50	50	50
MEM Vitamins (100x)	10	10	10
MEM Amino Acids (50x)	10	10	10
FBS	20	50	-
L-glutamine (200mM)	6	6	6
7.5% $NaHCO_3$ pH 8.2	18	20	-
PenStrep (100x)	10	10	10
Adjust to pH:	8.2	8.2	8.3
	To 500 ml with H_2O		

Table 2.2. All volumes are in ml. 10x Hank's A is made up with 160 g of NaCl, 8 g of KCl, 2g of $MgSO_4.7H_2O$, 2.8 g $CaCl_2.2H_2O$ and made up to 1 l with H_2O and autoclaved. Hank's B is made with 1.2 g of Na_2HPO_4, 1.2 g of KH_2PO_4, 20 g glucose and made up to 1 l with H_2O and 0.22 µm filter-sterilised. All overlays were filter sterilised once made up.

For the dengue assay, the virus was removed and replaced with 1ml of a 1:1 mix of 1.8% low gelling temperature agarose and 1st overlay solution (see Table 2.2). Plates were left for five days in the incubator, after which 1 ml of a 1:1 mix of 1.8% agarose and 2nd overlay solution (Table 2.2) supplemented with neutral red was added to each well. This was returned to the incubator for a further 24 hours, after which plaques were counted.

Ideal counts were between 10-50 plaques within a single dilution triplicate. Plaque forming units per ml were back-calculated to the undiluted concentration for comparison. Comparisons between different cell lines and conditions were done using a one-way ANOVA test with Dunnett's correction.

2.9.5. QRT-PCR

Viral RNA was purified from 140 µl of sample supernatant using the QiAMP Viral RNA Minikit (QiaGEN). The Verso qRT-PCR one step, plus ROX, kit (Thermo Fisher Scientific) was used according to the manufacturer's instructions. Reactions were carried out in MicroAmp Fast 96-well reaction plates (Thermo Fisher Scientific) sealed with MicroAmp optical adhesive film (Thermo Fisher Scientific). In each plate, eight standards were run in duplicate and each sample to be tested was also tested in duplicate. Each reaction took place in a total volume of 20 µl and consisted of 6.4 µl nuclease-free H_2O, 0.16 µl forward primer (100 µM), 0.04 µl reverse primer (100 µM), 0.1 µl of 10 µM ROX dye, 0.1 µl of a 20 µM FAM-TAMRA probe, 10 µl 2x 1-step qPCR mix, 1 µl of RT enchancer, 0.2 µl of Verso enzyme mix and 2 µl of purified viral RNA. A master mix of all components save for the viral RNA was prepared and components scaled accordingly.

Sequences of forward and reverse primers, and that of the FAM-TAMRA probes, can be found in the appendix for both Zika and dengue experiments. A standard curve was run for each experiment. For Zika, it consisted of a dilution series from $6x10^7$ to $6x10^1$ followed by a blank, with the sequence of UCGUUGCCCAACACAAGGUGAAGCCUACCUUGACAAGCAAUCAGACACUCA AUAUGUCUGCAAAAGAACAUUAGUGG, from Zika Env, of HPLC purified ssDNA (Integrated DNA technologies). For dengue virus, it consisted of a dilution series from $1x10^7$ to $1x10^1$ plus a control of DNA coding for dengue NS5, prepared by Beatrice Tyrrell by PCR amplification and purification of the gene using the qRT-PCR primers.

An Applied Biosystems 7500 Fast real time PCR system was used to run the experiments. For both viruses the same protocol was used: 48°C for 30 minutes, 95°C for 15 minutes, and then 40 cycles of 95°C for 15 seconds, 55°C for 60 seconds and 60°C for 25 seconds. C_T values were exported from the software and converted to copy number using the standard curve. If C_T values from a sample duplicate had a standard deviation exceeding 0.3 cycles, it was deemed erroneous. Replicates were stacked into columns and a one-way ANOVA test (using Dunnett's correction) was used to ascertain whether a statistically significant difference was obtained.

2.9.6. Immunofluorescence

Experiments were set up in a near identical manner to the Zika and dengue antiviral assays; each of the CRISPR/Cas9 knock-out cell-lines and wild-type Vero cells were plated in triplicate in a 24-well plate at a concentration of $0.07x10^6$ cells/ml. After incubation of cells with virus for 90 minutes, virus was removed and replaced with 1 ml of DMEM + 2% FCS + 1x PenStrep. Plates were incubated for 48 hours, regardless of the virus used for infection.

After 48 hours, media was removed and replaced with approximately 1 ml of a 50% methanol, 50% acetone fixing solution per well. Plates were left at room temperature for 30 minutes before the solution was removed and replaced with approximately 1 ml of wash buffer (1x PBS, 0.1% Tween 20). The plate was left on a plate rocker for 5 minutes before wash buffer was removed. This was replaced with 1 ml of 5% w/v milk powder in wash buffer and the plate left on a plate rocker for 30 minutes. The milk solution was removed and each well washed thrice as outlined earlier.

For the Zika assay, the primary antibody, a mouse anti-Zika Env antibody (BioFront 1176-56) was added at a 1/500 dilution in wash buffer, 100 µl per well. For the dengue assays, the primary antibody was a 3H5.1 mouse anti-dengue Env antibody (Thermo Fisher Scientific) and used at a 1/1000 dilution. Following 1 hour incubation with the primary antibody on a plate rocker, the antibody was removed and the wells washed thrice with wash buffer as before.

A second antibody, an anti-mouse Ig antibody conjugated to Alexafluor488 (Thermo Fisher Scientific A21202) was added to wells at a 1/200 dilution in wash buffer, 100 µl per well. The plate was covered in foil and left on a plate rocker for 1 hour. Following incubation, the secondary buffer was removed and the wells were washed thrice as before, all whilst covered in foil as much as possible. Wash buffer from the final wash was removed and replaced with 4 drops of VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories) diluted 3-fold in 1x PBS. The plates were taken to a dark room and imaged with a Nikon Eclipse TE2000-U (Nikon) inverted-light microscope. For each well, 5 sets of two images were taken: under UV light to image DAPI-stained nuclei and under fluorescent excitation to image the Env protein from either Zika or dengue virus. For each set of images, the number of DAPI-stained cells (total cells in the field) and the number of fluorescently labelled cells (total infected cells in the field-of-view) were counted. A percentage of infected cells was calculated and were arranged by cell-line for analysis in GraphPad Prism 7.0, using a one-way ANOVA and a Dunnett's multiple comparison to compare each cell-line with the wild-type.

3. ER α-glucosidase I

ER α-glucosidase I (glu I) sits at the very entrance to the calnexin cycle and is the first enzyme to act on the N-glycan that has been added co-translationally to a nascent protein. Previously in the Zitzmann laboratory, it has been shown that it is the inhibition of the two ER α-glucosidases that leads to the antiviral effect conferred by iminosugars[153].

Unfortunately, as glucose-mimetics, iminosugars can have off-target effects, including the inhibition of intestinal α-glucosidases and glycolipid glycosyltransferases[51, 284]. One of the aims of this project is to determine the atomic-resolution structure of glu I in complex with an iminosugar inhibitor, with the aim to inform the structure-based drug design needed to make more specific inhibitors.

This chapter continues and builds on from the work completed by Dr Alessandro Caputo in the Zitzmann laboratory[257].

3.1. Cloning of *Mm* glu I

The *Mus musculus* ER α-glucosidase I construct in the pOPINGS vector was originally prepared by Dr Alessandro Caputo with the Oxford Protein Production Facility[257]. The construct consists of residues 59-834 (starting with ^{59}WVLAWL, ending with MAEEY^{834}) of Uniprot entry Q80UM7. The mature protein, after cleavage of the signal peptide encoded by pOPINGS, contains the extra sequences ETG at the N-terminus and ASWSHPQFEKHHHHHH at the C-terminus of the protein construct.

3.2. Expression of *Mm* glu I

For a typical expression, six hundred millilitres of HEK293F culture media, at a cell density of 1x10^6 cells/ml, were transfected with the pOPINGS vector encoding *Mm* glu I at a concentration of 1 µg of vector per ml of culture. The culture was left to express protein for four days before the supernatant was harvested.

3.3. Purification of *Mm* glu I

3.3.1. Purification of *Mm* glu I

Cell culture supernatant was prepared as described in the methods (section 2.3.2) and run on a 5 ml HisTrap excel affinity purification column. Fractions containing *Mm* glu I, as assessed by SDS-PAGE that confirmed a band between 85-95 kDa, were pooled and dialysed thrice into 50x volume of 1x PBS and finally into 50x volume of 1x PBS, 5% glycerol.

The dialysed protein was concentrated to a volume smaller than 5 ml and run on a Superdex 200 16/600 SEC column. Fractions were analysed by SDS-PAGE and α-6xHis-tag western blotting to confirm which fractions *Mm* glu I eluted in.

Samples containing the majority of glu I were concentrated to >4.5 mg/ml using a 50 000 Dalton MWCO centrifugal filter unit. Protein was aliquoted into 100 µl volumes, flash-frozen in liquid nitrogen and stored at -80°C.

3.3.2. Methylation of *Mm* glu I

Methylation of *Mm* glu I was required for its crystallisation. Purification of methylated *Mm* glu I followed that of non-methylated protein until the end of the affinity purification step. All steps preceding the SEC were carried out at 4°C. *Mm* glu I eluted from the HisTrap was pooled and equilibrated twice into 50x volume of 20 mM HEPES, 150 mM NaCl over the course of 16 hours. Protein was subsequently dialysed into the same buffer plus 5% glycerol and the reductive alkylation reactions were carried out as detailed in methods section 2.4.2. Methylated protein was further purified by SEC on a Superdex 200

16/600 column. Fractions containing methylated *Mm* glu I, as analysed by SDS-PAGE, were pooled, concentrated to approximately 5 mg/ml and flash-frozen in the same manner as non-methylated protein.

3.3.3. Mass-spectrometry analysis of methylated-*Mm* glu I

Methylated *Mm* glu I was handed to Dr Abhinav Kumar of the Zitzmann laboratory. *Mm* glu I was digested with trypsin and the peptides were analysed using tandem mass-spectrometry. A sequence coverage of 73% was achieved and covered all but three lysine residues, the amino acid expected to be modified by alkylation. These three lysine

Matched peptides shown in ***bold red***.

```
  1 MARGERRRRA AAAEGARPLE RARAAGRRDG RAGGARGSAS GAALAVVVLA
 51 LAFGLSGRWV LAWLRVRRAL TLHPAPSALP PDSSSPAVAP ELFWGTYRPH
101 VYFGMKTRSP KPLLTGLMWA QQGATPGTPP KLRHTCEQGD GVGPYGWEFH
151 DGRTFGRQHI HDGALRLTTE FVKRPGGQHG GDWSWRVTVE PQASGTPSFP
201 LVSLFFYVVT DGQEVLLPEI GAKGQLKSIS GHTSELGDFR LTLLPPTSPG
251 DTVPKHGSYN VFWSSNPGLP QLTDMVKSRL NSWFQHRPPG ASPDRYLGLP
301 GSLKWEERGP SGQGQFLIQQ VTLKAPFSVE FVFESGSAAT GGNQASGRLV
351 GSQLTQALES HAAAFKERFE KTFQLKEKGL SPEEQALGQV ALSGLLGGIG
401 YFYGQGLVLP DTSMEGSEQK MDPALFPPVP LFSGVPSRSF FPRGFLWDEG
451 FHQLVVQRWD PHLTREALGH WLGLLNADGW IGREQILGDE ARARVPPEFL
501 VQRAAHANPP TLLLPVVHML EGHDPDDLAF LRKAFPRLHA WFSWLHQSQA
551 GPVPLSYRWR GRDLALPTLL NPKTLPSGLD DYPRASHPST AERHLDLRCW
601 VALGARVLSQ LAEQLGETEA AAELGPLAAS LEEPGSLDEL HWAPELGVFA
651 DFGNHTKAVQ LKSRPPQGLV RVVGRPPPRL QYVDALGYVS LFPLLLQLLD
701 PSSPRLGPLL DVLADSRHLW SPFGLRSLSA SSLFYKQRNT EHDPPYWRGA
751 VWLNINYLAL GALHHYGHVE GPHKVQAAKL YHELRANVVR NVRQQYQATG
801 FLWEQYSDQD GRGMGCRPFQ GWTSLVLLIM AEEY
```

Figure 3.1. Sequence coverage of methylated *Mm* glu I by tryptic-digest tandem mass spectrometry. Shown are peptides that were detected, mapped on to the sequence of full-length *Mm* glu I. Underlined are the lysine residues where the methylation state could not be assessed, expression tags are not shown.

residues were located at positions 223, 657 and the C-terminal expression tag (Figure 3.1). Of these, K223 and the expression tag lysine were not observed in the crystal structures, thus having no impact. Lys^{657} is solvent-exposed, however due to the lack of information about it in the MS analysis, it was not labelled as methylated.

Of those residues that MS data was obtained for, lysine residues at positions 111, 131, 227, 304, 366, 371, 376, 378, 533, 573, 662, 736 and 779 were confirmed to be di-methylated. Lysine residues at positions 173, 255, 324 and 420 were found to remain non-methylated.

In addition, the oxidation of certain methionine residues to methionine sulfoxide was discovered on residues 414, 421 and 519. All of these modifications were taken into account when modelling the structure to x-ray diffraction data.

Data Processing Statistics	
	3.26 Å
Resolution Range (Å)	91.717 - 3.261 (3.99 - 3.261)
Cell Dimensions:	
a, b, c (Å)	106.58, 127.99, 133.592
α, β, γ (°)	90.00, 101.163, 90.00
Spacegroup	$P2_1$
Total Reflections	132168 (57206)
Unique Reflections	52553 (23016)
R_{merge}	0.287 (0.861)
R_{meas}	0.406 (1.217)
R_{pim}	0.287 (0.861)
Completeness (%)	95.4 (92.1)
Multiplicity	2.5 (2.5)
CC 1/2	0.907 (0.626)
I/σ	3.7 (1.5)

Table 3.1. Data processing statistics of the first dataset obtained for *Mm* glu I. Processing was carried out using the autoPROC program suite[1]. Values in brackets represent the data for the high-resolution data shell. The resolution cut-off was determined by AutoPROC based on a CC ½ of greater than 0.4.

3.4. *Mm* glu I – the 3.26 Å dataset

3.4.1. Crystallisation and data collection

Methylated *Mm* glu I was screened in the Morpheus HT-96, sitting-drop crystallisation screen. This glu I was prepared at 4.5 mg/ml in 20 mM HEPES, 100 mM NaCl, 0.1 mM TCEP, pH7.5 with the addition of 1 mM M*O*N-DNJ. One hundred nanolitres of this were mixed with 100 nl of the screening condition in each sitting drop of commercial screens by Dr Alessandro Caputo.

After approximately two months in a 17°C crystallisation plate incubator, 2 crystals were observed, growing as thin plates on the edge of well C1 of the Morpheus screen. This condition contained 0.09 M NPS (0.03 M sodium nitrate; 0.03 M sodium phosphate dibasic; 0.03 M ammonium sulfate), 0.1 M Buffer System 1 (imidazole; MES monohydrate (acid)) pH 6.5 and 50% v/v precipitant Mix 1 (40% v/v PEG 500 MME; 20% w/v PEG 20000).

Crystals were mounted in nylon loops, directly cryo-cooled into liquid nitrogen and taken to DLS for data collection (see Appendix for data collection parameters).

3.4.2. Data processing

Diffraction images were processed using the autoPROC suite of integrated programmes. Images were discarded if they showed signs of radiation damage, as assessed by the scaling B factor per frame found in the aimless.log file, leaving 660 frames from the original 900 for scaling and merging.

Statistics from the data processing are shown in table 3.1.

3.4.3. Phase determination

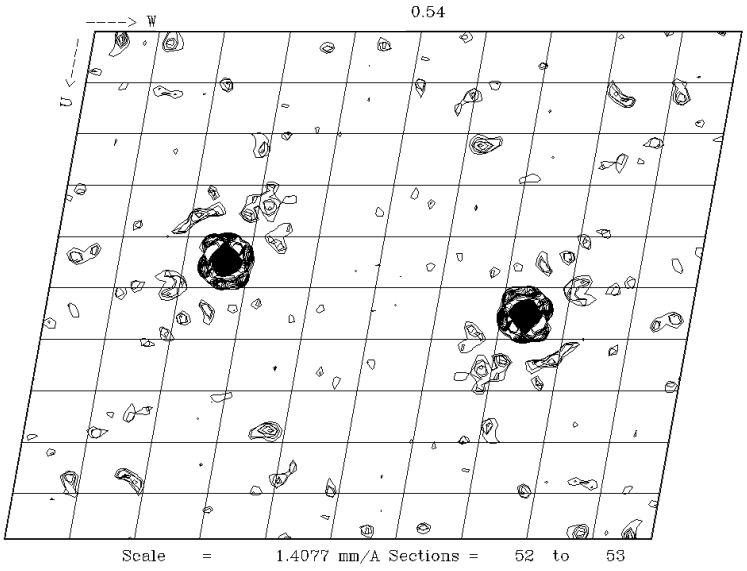
Based on a histogram of solvent content of PDB structures as a function of resolution, the glu I MW as well as the crystal cell volume and symmetry, the most probable number of molecules in the asymmetric unit of the crystal was predicted to be 4, which would indicate the crystal is 53.3% solvent and has a Matthew's coefficient of 2.63 $Å^3/Da$[285].

The CCP4i programme FFT was used to generate a native Patterson to find if there was translational non-crystallographic symmetry present within the asymmetric unit. Due to the packing of the unit cell, the strongest peak observed was found at the origin. A second peak (Figure 3.2.A.) was observed at half the intensity of the origin peak, at fractional coordinates of U=0.44, V=0.5 and W=0.2624, indicating that we have pairs of molecules related by this vector within the asymmetric unit.

The programme molrep (CCP4i) was used to generate a self-rotation function to determine the rotational non-crystallographic symmetry present within the asymmetric unit. In the section corresponding to $\kappa = 180°$ (Figure 3.2.B.), the strong peak observed at the edge of the flattened globe view confirms the 180° rotation necessary for the monoclinic twofold axis of the crystal. A second, less intense peak indicates the direction of another twofold axis within the crystal, with angles of theta = 29°, phi = -61.2° and kappa = 180°: this is likely to be the direction of a twofold axis that is present within the asymmetric unit of the crystal.

Automated molecular replacement was used to find initial phases for the experiment structure factor amplitudes. MrBUMP (CCP4i)[286] used the *Saccharomyces*

A.



B.

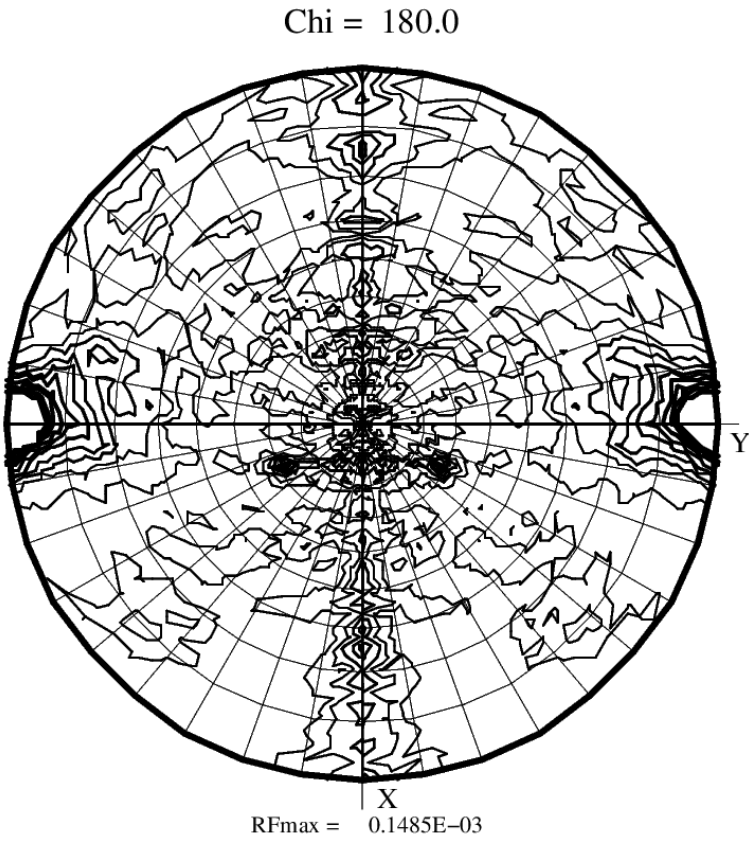


Figure 3.2 A. The V = 0.5 section of the native Patterson of the processed diffraction data. B. The self-rotation function of the processed data in the κ = 180° section.

cerevisiae (*Sc*) structure, 4J5T[42], as the search model, and attempted to place 4 molecules within the asymmetric unit. The best set of phased data was derived from 'Chainsaw' treatment of the search model (side chain truncation based on multiple sequence alignments from MAAFT)[287] followed by molecular replacement carried out by Phaser. Visual inspection of the output model and electron density indicated that the molecules formed a 3D lattice without major clashes nor large areas of unexplained density.

3.4.4. Model building and refinement

Cycles of model building and refinement were carried out in Coot and with autoBUSTER, respectively. Of the four molecules in the asymmetric unit, chains A and B were most similar to one another, as were the pair of chains C and D. Autobuilding programmes were tested on the dataset: none could successfully build a model. After multiple rounds of manual model building, the R_{work}/R_{free} was reduced to 0.2654/0.2723, with the majority of chains in the correct place in the sequence. However, due to the data resolution, side chain conformations were difficult to resolve, nor was any clear electron density observed in the active site. Model building was halted with 11.5% of residues still registering as Ramachandran outliers. The root mean squared deviation (RMSD) of each chain, when superimposed on the original *Sc* glu I structure, varied between 1.88 and 2.27 Å. The focus on generating new, higher resolution crystals was renewed.

3.5. The 2.1 Å resolution structure of *Mm* glu I

3.5.1. Obtaining higher resolution crystals

Morpheus-HT 96 condition C1 was dispensed across an entire MRC 2-Well crystallization plate (Swissci), with 50 µl in each well's reservoir. The drop containing the remnants of the crystals used for the initial 3.26 Å structure was used to prepare a seed stock. The custom screen was finalised by varying the dilution of seed stock used down the plate, from 1/256 in row A in a four-fold dilution series ending at row H.

After 8 days, crystals were observed growing in wells A6 and G3 of this custom screen. Due to the lack of crystals in any other wells containing the same conditions, it is likely that there was an issue with under-nucleation and that the crystals formed independently of the seed-stock.

Data Processing Statistics	
	2.10 Å
Resolution Range (Å)	133.50 - 2.10 (2.21 - 2.10)
Cell Dimensions:	
a, b, c (Å)	108.01, 130.92, 135.37
α, β, γ (°)	90, 99.53, 90
Spacegroup	$P2_1$
Total Reflections	752738 (97509)
Unique Reflections	215063 (31498)
R_{merge}	0.114 (0.884)
R_{meas}	0.157 (1.213)
R_{pim}	0.107 (0.827)
Completeness (%)	99.6 (99.9)
Multiplicity	3.4 (3.1)
CC 1/2	0.996 (0.593)
I/σ	6.5 (1.2)

Table 3.2. Data processing statistics of the dataset obtained for the *Mm* glu I crystal in well A6 of the custom screen. Values in brackets represent the data for the high-resolution data shell.

3.5.2. Data collection and processing

The crystal in well A6 of the custom screen was a singular, trapezoidal prism. Approximately 4 weeks after the screen was set-up, the crystal was fished using a nylon loop and directly cryo-cooled into liquid nitrogen and taken to DLS (see Appendix for data collection parameters).

Two datasets were collected from this crystal: the first led to an incomplete dataset as the crystal was not centred correctly. After re-centring, a second dataset was collected from the same crystal. Fortunately, there were no signs of radiation damage. All of the images were processed using AutoPROC, the statistics of which are presented in table 3.2.

3.5.3. Model building, refinement and validation.

Free R flags were carried over from the original 3.26 Å data using Cad (CCP4i), and a rigid body refinement of the 3.26 Å model was performed against the newly collected 2.1 Å data to produce the starting model and maps for model building.

Cycles of refinement and model building were carried out using autoBUSTER and Coot. Once the ordered protein structure was built, waters were added automatically in Coot. All four chains were eventually modelled separately as the extent of disordered regions differed. MON-DNJ (Ligand ID 6A9) was built into the active site of each chain. In addition, three sulfate ions were modelled into each chain. A glycan could be modelled on Asn^{654} to different extents depending on the chain: 1x $GlcNAc_2$ on chains A and B; 1x $GlcNAc_2Man_1$ (Ligand IDs NAG and BMA) onto each of chains C and D. Finally, PEG was modelled (Ligand ID 7PG) in two positions in chains C and D but just in one position in chains A and B. Methionine sulfoxide and dimethyl lysine were modelled where mass-

spectrometry data suggested they were present. In order to place MON-DNJ into the active site in an unbiased fashion, the iminosugar was removed from the model and the coordinates altered slightly in a random fashion (maximum shift 0.2 Å) with CCP4i-pdbset to overcome any bias from the initial placement of the iminosugar. The model and map were refined once more in autoBUSTER before re-modelling the iminosugar into the unbiased F_O-F_C map.

Model building and refinement statistics

		2.10 Å
Resolution Range (Å)		133.50 - 2.10
Observations		214780
in free R set		10477
R_{work} (%)		18.8
R_{free} (%)		21.6
RMSD bond lengths (°)		0.010
RMSD bond angles (Å)		1.08
Ramachandran outliers		2
Ramachandran favoured (%)		97.06
MolProbity Clashscore		4.85 (98th percentile)
MolProbity Score		1.42 (99th percentile)
Good Rotamers (%)		93.66
B-factor (wilson plot)		33.23
Residues modelled		
	A:	61-193, 198-220, 225-308, 311-339, 348-412, 421-838
	B:	61-193, 198-220, 225-308, 311-339, 347-412, 421-838
	C:	56-220, 225-308, 311-340, 346-412, 421-837
	D:	56-220, 225-308, 311-340, 346-412, 421-837
Waters modelled		
	A	321
	B	359
	C	358
	D	349

Table 3.3. Percentiles are calculated by MolProbity: clashscore based on comparison to a selected set of structures in 2004, overall score based on set from 2006.

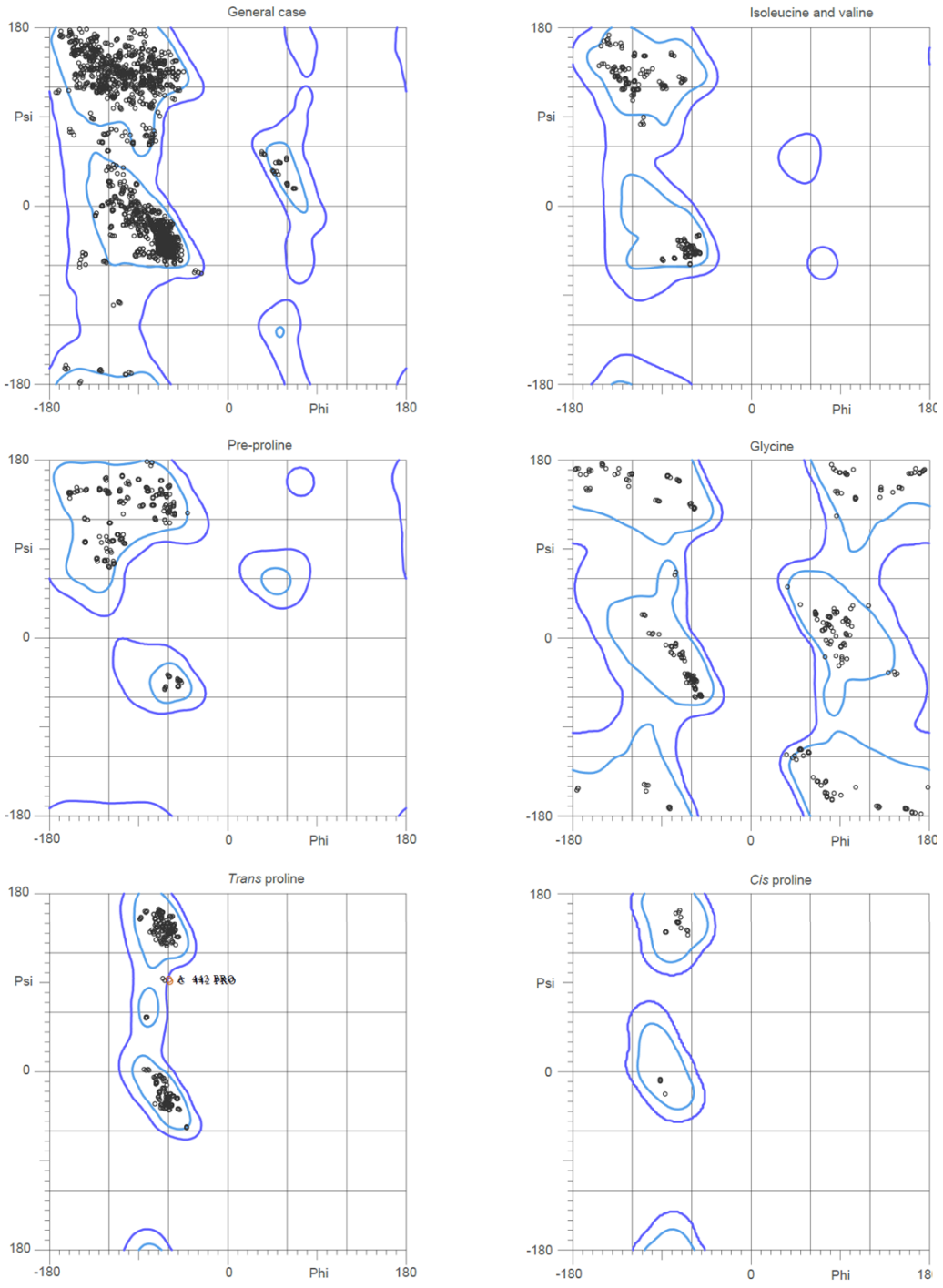


Figure 3.3. Ramachandran plots, generated by MolProbity, of the 2.10 Å structure of *Mm* glu I

Model building and refinement was considered complete once the R_{work} and R_{free} had converged and could not be improved upon and bond angles and lengths were considered acceptable. For full statistics on model building and validation, see figure 3.3 and table 3.3. The final model was deposited to the wwPDB and given the code 5MHF.

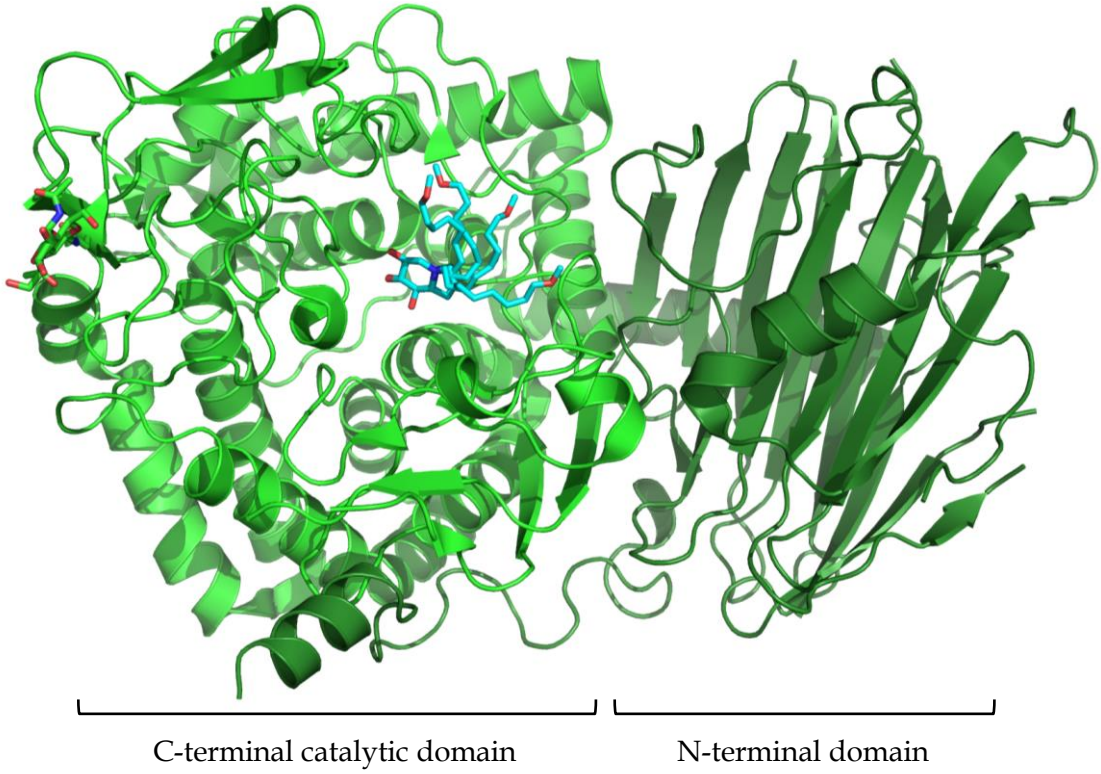


Figure 3.4. Overall architecture of *Mm* glu I. Shown here is chain A as a cartoon model, with stick representations of the glycan attached to N^{654} and MON-DNJ in cyan in the active site.

3.5.4. Analysis of the structure

The overall architecture of the protein (Figure 3.4) is as expected for a homologue of the existing *Sc* processing α-glucosidase I structure, with a C_α RMSD of just 1.86 Å between chain A of the *Mm* glu I model and the yeast structure (PDB ID: 4J5T): 601 aligned C_αs,

35% sequence identity. The major differences are restricted to the loops of the N-terminal domain, the positioning of the N-terminal helix and a number of loops distal to the active site within the catalytic domain.

The protein can be split into two domains: the N-terminal domain is dominated by a super-β-sandwich domain (residues 94 – 339), the role of which is currently unknown, and belongs to Pfam family PF16923 "glycosyl hydrolase family 63 N-terminal domain". The C-terminal domain (residues 381 – 838) contains the active site and is formed as an $(\alpha/\alpha)_6$ helical barrel belonging to Pfam family PF03200 "glycosyl hydrolase family 63 C-terminal domain". On one face of this domain, the active site is nestled in the centre of the helices. The helix 350-374 connects the two domains.

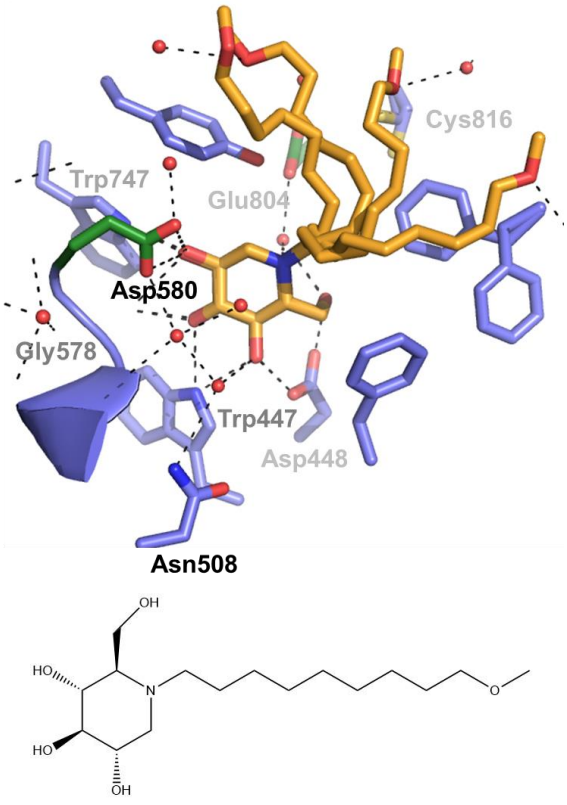


Figure 3.5. Active site of *Mm* glu I. The main chain is coloured in purple, catalytic residues in green and the iminosugar in orange.

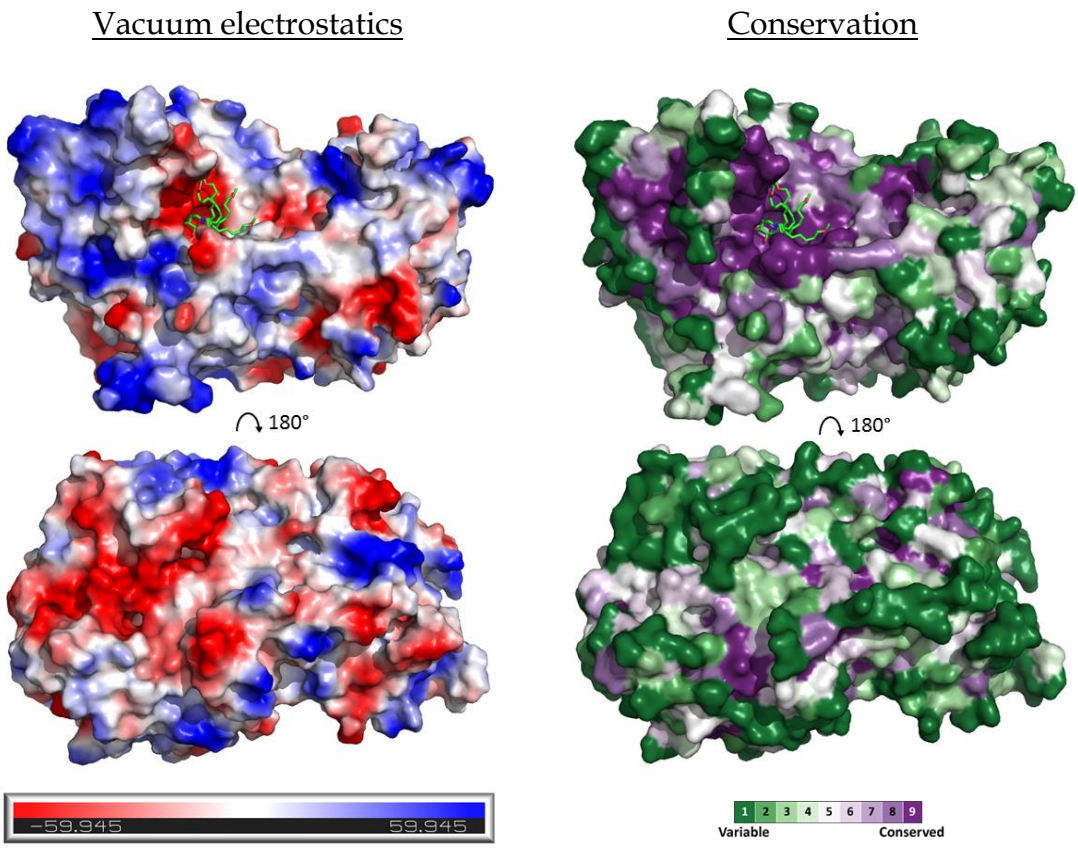


Figure 3.6. On the left, vacuum electrostatics of *Mm* glu I as calculated by PyMol: "averaging charges over a small region of space using a quasi-Coulombic-shaped convolution function" (https://pymolwiki.org/index.php/Protein_contact_potential). On the right, conserved residues as calculated by the ConSurf 2016 server using 150 ER α-glucosidase I homologues.

There is a single predicted N-glycosylation site on the molecule at arginine^{654}. $GlcNAc_2$ can be built to chains A and B whereas $GlcNAc_2Man_1$ can be built onto chains C & D. From the crystallisation condition, a total of 6 chains of poly-ethylene glycol were modelled into the asymmetric unit, all of which occurred at interfaces between different chains as well as 10 molecules of SO_4^-.

The active site contains MON-DNJ, an alkylated glucose-mimic (Figure 3.5). It is predicted to bind in a similar manner to the terminal glucose of the natural substrate, the

tri-glucosylated glycan. The map provided strong electron density to define the position of the deoxynojirimycin ring and the waters immediately surrounding it. However, the orientation of the alkyl tail was not well-defined and was modelled in four conformations to highlight this. The four conformations correspond to the most likely fit in each of the four chains of *Mm* glu I within the asymmetric unit, overlaid onto one another. The occupancies of the four members of the ensemble were then refined within each of the chains.

A number of crystal structures of GH63 family-members have been determined. The structure of mannosylglycerate hydrolase from *Thermus thermophilus HB8* confirms the position of the iminosugar ring as near-identical to that of glucose[288]. It can be postulated that, as glu I is an inverting α-glucosidase, Glu^{804} acts as the base and Asp^{580} acts as the acid in the catalytic mechanism that has been proposed based on structures of a catalytically inert mutant of the *Escherichia coli* homologue, YgjK[289].

None of the methyl-lysine residues are closer than approximately ~12 Å from the iminosugar in the active site, and so are unlikely to have had an effect on iminosugar binding.

Figure 3.6 shows both the vacuum electrostatics and the conservation score of residues plotted on a surface representation of *Mm* glu I. Around the active site, there is a high degree of conservation, as expected between orthologues that share the same substrate and catalytic activity. The electrostatic potential in this pocket is strongly negative. Immediately to the left of the active site (in the view shown), there is a spur (residues of loops 562 – 572, 651 – 684 and 728 – 749), around which there is another patch of high-

sequence conservation, this time positively charged. There are no other large, well-conserved areas that appear to be of immediate interest.

As discussed in the introduction, mutations in the ER α-glucosidase I gene can lead to CDG-IIb. Whilst a number of the observed mutations were point-nonsense and led to early termination of the proteins, there are three described mutations which are missense: $Arg^{486}Thr$, $Phe^{652}Leu$ and $Arg^{110}His$[44, 45]. Figure 3.7 maps these mutations onto the glu I crystal structure. Both R486T and F652L are located in the catalytic domain and could affect the correct folding of the domain. The R110H mutation is located in the N-terminal domain and forms a salt bridge with Glu^{466} in the catalytic domain. Altering this is likely to affect its fold and could explain why it is quickly recognised for ERAD.

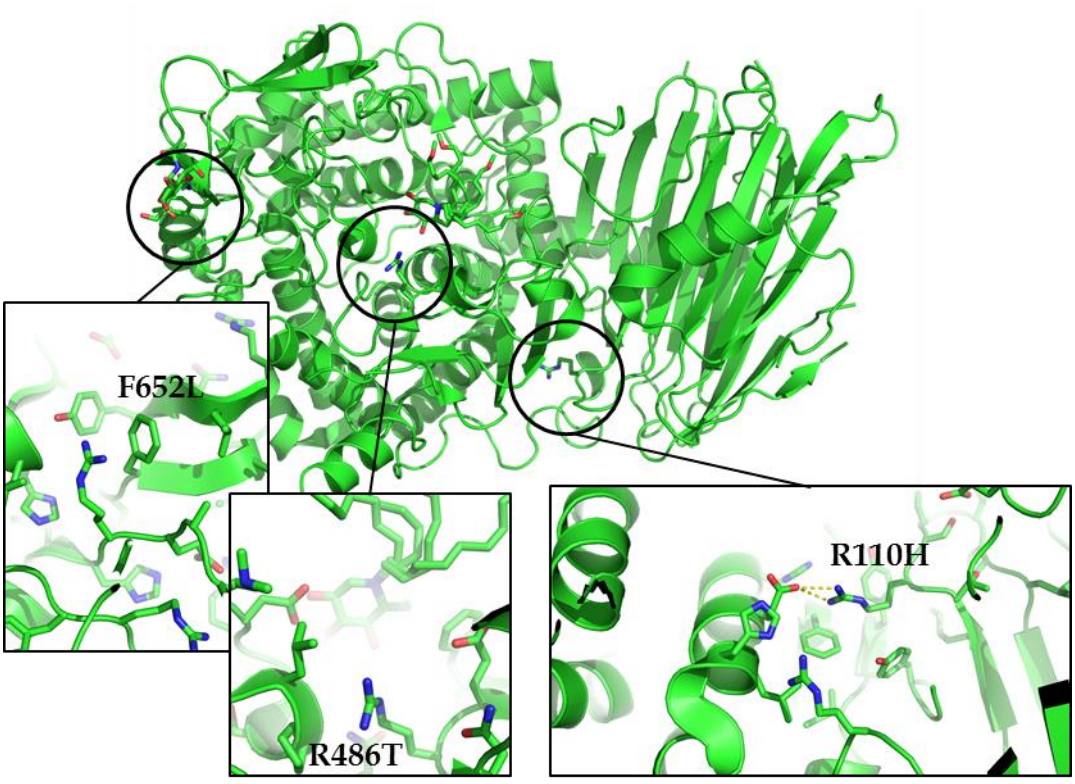


Figure 3.7. The published missense mutations that lead to CDGIIb. F652L and R486T are both within the catalytic domain. R110H disrupts an inter-domain salt-bridge.

3.6. The binding of other iminosugars

3.6.1. Crystallisation

A crystal seed stock was prepared from the crystal plate from which the 2.1 Å dataset of *Mm* glu I was derived. A new crystallisation plate was set up using the same conditions, except that the methylated *Mm* glu I was at a final concentration of 4.7 mg/ml and mixed with a final concentration of 0.25 mM of one of SR1, SR2, SR3, OP1 or AR1 iminosugars (figure 3.8): all derived from 5 mM stocks in DMSO. These were in place of M*O*N-DNJ. After 1-2 months, a number of slow-growing crystals began to form. The crystal morphology was similar to *Mm* glu I: M*O*N-DNJ crystals; however, many looked singular at one end but appeared as bundles of needles at the other end.

3.6.2. Data collection and processing

For each iminosugar, a number of crystals were taken to beamlines and diffraction data were collected. The data collection statistics for the best datasets of each iminosugar co-crystallisation can be found in the Appendix.

These datasets were processed with AutoPROC and all were found to be the same space-group and similar cell dimensions to the original structures (see Table 3.4).

3.6.3. Model building, refinement and validation

As the overall architecture was unlikely to have changed from the previous structures, the 2.1 Å resolution model solved with M*O*N-DNJ was used as a starting model, without

Table 3.4. Data processing statistics

	SR1 - C2_1	SR2 - D3_1	SR3 - E6_1_2	OP1 - F10_1	AR1 - G3_1
Resolution Range (Å)	132.17 - 2.10 (2.22 - 2.10)	129.45 - 2.85 (3.00 - 2.85)	132.13 - 2.72 (2.87 - 2.72)	127.94 - 2.24 (2.36 - 2.24)	92.56 - 2.18 (2.30 - 2.18)
Cell Dimensions:					
a, b, c (Å)	106.999, 130.256, 134.162	106.797, 129.451, 134.323	106.785, 127.904, 134.112	106.655, 127.938, 133.959	107.296, 129.430, 134.405
α, β, γ (°)	90.00, 99.88, 90.00	90.00, 99.92, 90.00	90.00, 99.88, 90.00	90.00, 99.89, 90.00	90.00, 99.84, 90.00
Spacegroup	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Total Reflections	641898 (82457)	278300 (40543)	322870 (47319)	544268 (54118)	585555 (44109)
Unique Reflections	208048 (29840)	83207 (12227)	94666 (13727)	167587 (232600	169592 (14543)
R_{merge}	0.109 (1.468)	0.406 (2.665)	0.321 (2.092)	0.182 (1.304)	0.125 (1.548)
R_{meas}	0.154 (2.076)	0.555 (3.619)	0.444 (2.875)	0.251 (1.822)	0.172 (2.088)
R_{pim}	0.109 (1.467)	0.377 (2.433)	0.306 (1.963)	0.172 (1.271)	0.117 (1.392)
Completeness (%)	99.0 (97.7)	98.7 (99.5)	99.5 (99.5)	99.0 (94.6)	90.7 (53.5)
Multiplicity	3.1 (2.8)	3.3 (3.3)	3.4 (3.4)	3.2 (2.3)	3.5 (3.0)
CC 1/2	0.995 (0.383)	0.958 (0.446)	0.973 (0.405)	0.988 (0.324)	0.996 (0.398)
I/σ	5.5 (0.6)	2.6 (0.5)	2.9 (0.5)	4.1 (0.5)	5.6 (0.7)

Values in brackets represent the figures for the high-resolution shell

Table 3.5. Model building and refinement statistics

		SR1 - C2_1	SR2 - D3_1	SR3 - E6_1_2	OP1 - F10_1	AR1 - G3_1
Resolution Range (Å)		132.17 - 2.10 (2.22 - 2.10)	129.45 - 2.85 (3.00 - 2.85)	132.13 - 2.72 (2.87 - 2.72)	127.94 - 2.24 (2.36 - 2.24)	92.56 - 2.18 (2.30 - 2.18)
Observations		148182	44096	56801	108056	124509
in free R set		7185	2147	2773	5341	6100
R_{work} (%)		0.1877	0.2409	0.1999	0.181	0.1989
R_{free} (%)		0.2202	0.266	0.2409	0.2192	0.2286
RMSD bond lengths (°)		0.010	0.008	0.010	0.010	0.010
RMSD bond angles (Å)		1.140	1.040	1.140	1.110	1.100
Ramachandran outliers		14 (0.49%)	3 (0.11%)	16 (0.56%)	6 (0.21%)	1 (0.04%)
Ramachandran favoured (%)		96.14	96.99	96.01	96.78	97.02
Molprobity Clashscore		11.37 (71st)	2.9 (100th)	12.06 (91st)	12.85 (75th)	11.22 (82nd)
Molprobity Score		2.18 (54th)	1.26 (100th)	2.12 (92nd)	2.00 (81st)	2.21 (67th)
Good Rotamers (%)		89.48	93.26	90.33	92.07	88.17
B-factor (wilson plot)		28.48	38.08	29.13	25.48	36.7
Residues modelled						
	A:	61-339, 348-412, 421-837	61-307, 310-339, 348-412, 421-838	61-308, 311-339, 348-412, 421-838	61-308, 311-339, 348-412, 421-838	61-307, 311-339, 348-412, 421-837
	B:	61-339, 347-412, 421-837	61-307, 311-339, 347-412, 421-838	61-339, 347-412, 421-838	61-308, 311-339, 347-412,421-838	61-307, 311-339, 347-412, 421-837
	C:	56-340, 346-412, 421-837	56-307, 311-340, 346-412, 421-837	56-308, 311-412, 421-837	56-307, 311-340, 346-412, 421-837	56-307, 311-412, 421-837
	D:	56-412, 421-837	56-307, 311-412, 421-837	56-412, 421-837	56-307, 311-412, 421-837	56-307, 311-412, 421-837

ligands present, for a rigid-body refinement, with the FreeR flags carried over from the original 3.26 Å mtz file to these new files.

Little model building was required to finish building of the backbone, as few regions differed from the earlier models. Loops that were disordered and remained unbuilt in the earlier structures were modelled into the electron density maps obtained from these latest crystals, to different extents (see Table 3.5). However, none of the electron density maps from these latest glu I: iminosugar datasets permitted the building of the loop comprised of residues 412 to 421.

Glycan, PEG and water molecules were built into the density. The stereochemical dictionaries of the five new iminosugars were built from their SMILES strings using the GRADE server (Global Phasing). Only the portion of each iminosugar molecule that gave ordered electron density was modelled: typically between 2 and 3 alkyl carbons past the nitrogen in the DNJ ring.

In the structure with SR1, the C-terminal residues of *Mm* glu I were modelled in an alternative conformation from that of the 2.1 Å structure: supporting a second molecule of SR1 in an accessory site.

3.6.4. Iminosugars in the active site

In all of the new crystal structures obtained, an iminosugar could be built into the electron density of the active site of *Mm* glu I. All of the crystallisation drops did contain marginal quantities of the M*O*N-DNJ crystallisation condition. However, the seed stock was made with methylated protein that lacked exposure to an iminosugar, thus only 0.1 µl of liquid containing M*O*N-DNJ was added to the 50 µl of seed-stock, which was then diluted 1 in

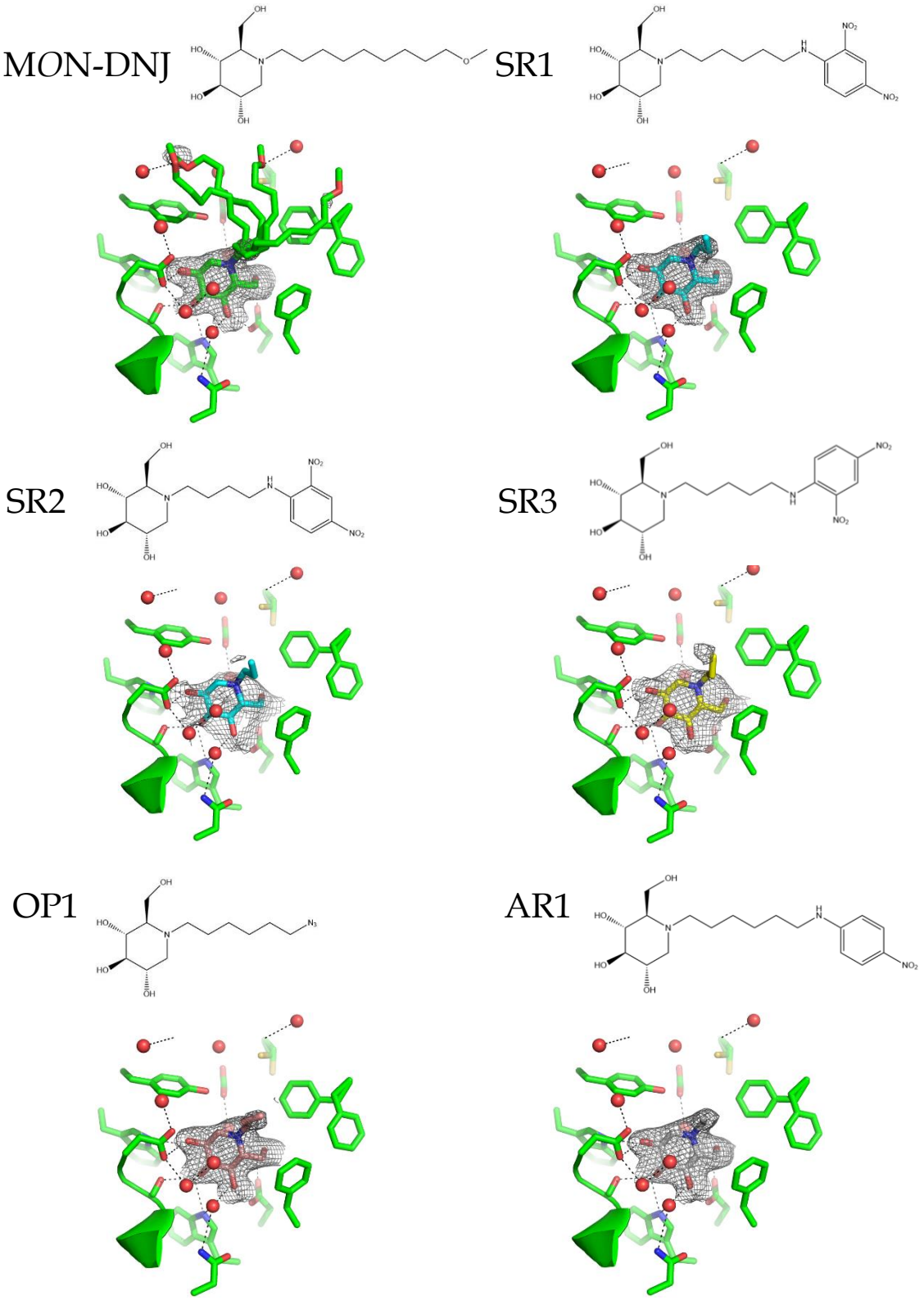


Figure 3.8. Electron density in the active site of *Mm* glu I, when soaked with a panel of iminosugars. Electron density is shown at an I/σ of 1.0 and a carve of 1.8-2.0 around the modelled iminosugar.

64, of which 15 µl were added to these new crystallisation drops. It is therefore unlikely that the iminosugar found in the active site is simply residual MON-DNJ: rather, they are likely to be the new iminosugars that co-crystallisations were trialled with.

Figure 3.8 shows the electron density around each of the iminosugars tested. Similar to the data for the structure with MON-DNJ, none of these other iminosugars had electron density to build the molecule into once past the deoxynojirimycin ring. This strongly suggests that, apart from the deoxynojirimycin head-group, alkylated iminosugar inhibitors are likely flexible in structure. This is in contrast to the study of iminosugar binding to ER α-glucosidase II carried out by Dr Alessandro Caputo *et al.* within the Zitzmann laboratory[46]. Here, the tails of many of the iminosugars had defined

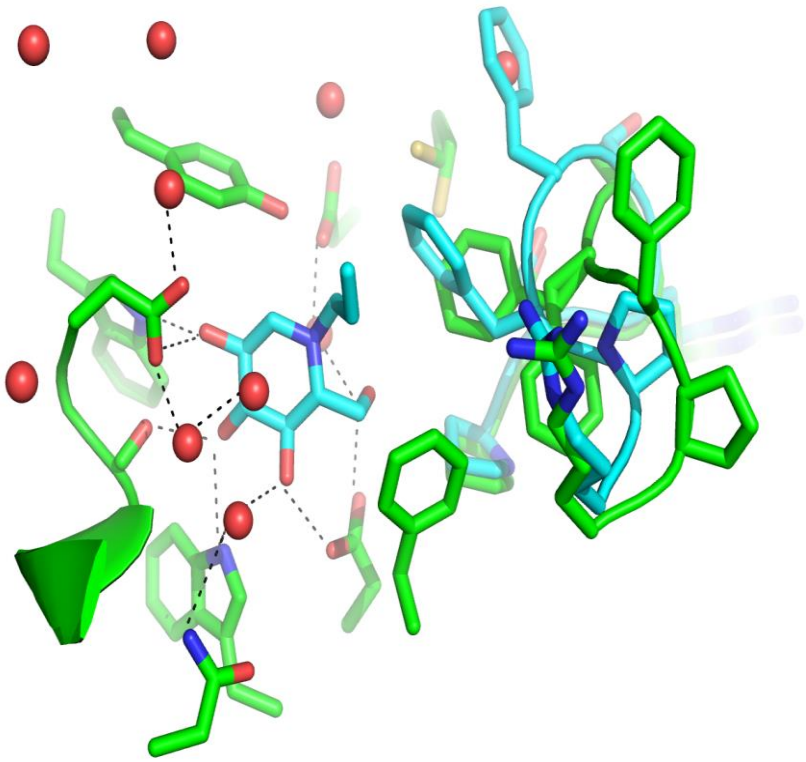


Figure 3.9. Active site of *Mm* glu I with the truncated SR1 iminosugar modelled. The residues in green represent the model from the original MON-DNJ structure, the residues in light blue, from the SR1 structure.

conformations, leading to the conclusion that interaction with a stretch of residues known as the 'exclusion loop' could potentially help design more specific inhibitors of this glucosidase, if this loop could be displaced and forced into an alternate conformation.

Interestingly, a flexible loop is also observed next to the active site in the *Mm* glu I crystal structure, Phe-Phe-Pro, that is also conserved in humans. Figure 3.9 shows this loop in green on the right-hand side, with the initial phenylalanine modelled in two conformations. However, in the SR1 structure, the loop has shifted and the positions of the residues are clearly defined in an alternative, singular, position. Interestingly, no electron density for the 'tail' of the iminosugar can be seen that would lead to this effect. Nonetheless, it is a potential site for interaction with the iminosugar that could lead to a more specific interaction.

3.6.5. An additional iminosugar binding site?

Unexpectedly, the crystal structure of *Mm* glu I in complex with SR1 yielded a C-terminus that could not be satisfactorily explained by the original model in complex with M*O*N-DNJ. Upon re-building of the C-terminus up to Trp^{837}, a large area of density remained unexplained.

A molecule of SR1 was fit into this density: the deoxynojirimycin ring sat nestled further into the α-helical bundle than the dinitrophenyl ring of SR1, which extended towards the C-terminus and appeared to π-stack with the side-chain of tryptophan^{837}.

The position of this second iminosugar binding site, as seen in figure 3.10, is on the opposite face of the $(\alpha/\alpha)_6$-helical barrel from the active site, centred within the helices similar to the iminosugar in the active site.

Whether or not this is a physiological interaction with a sugar moiety is debatable. The tryptophan that the tail of this iminosugar interacts with is part of the expression tag of the protein and thus this iminosugar could not interact in the same way *in vivo*. Yet, the deoxynojirimycin ring is a mimic of glucose, and this ring only interacts with residues that are not part of the expression tag. In the native molecule, this site would be more accessible to sugar binding than in the crystal structure, as the expression tag would not be present to sterically hinder any binding. Although there is less sequence conservation at this accessory site than the in the active site, it is relatively well conserved compared to

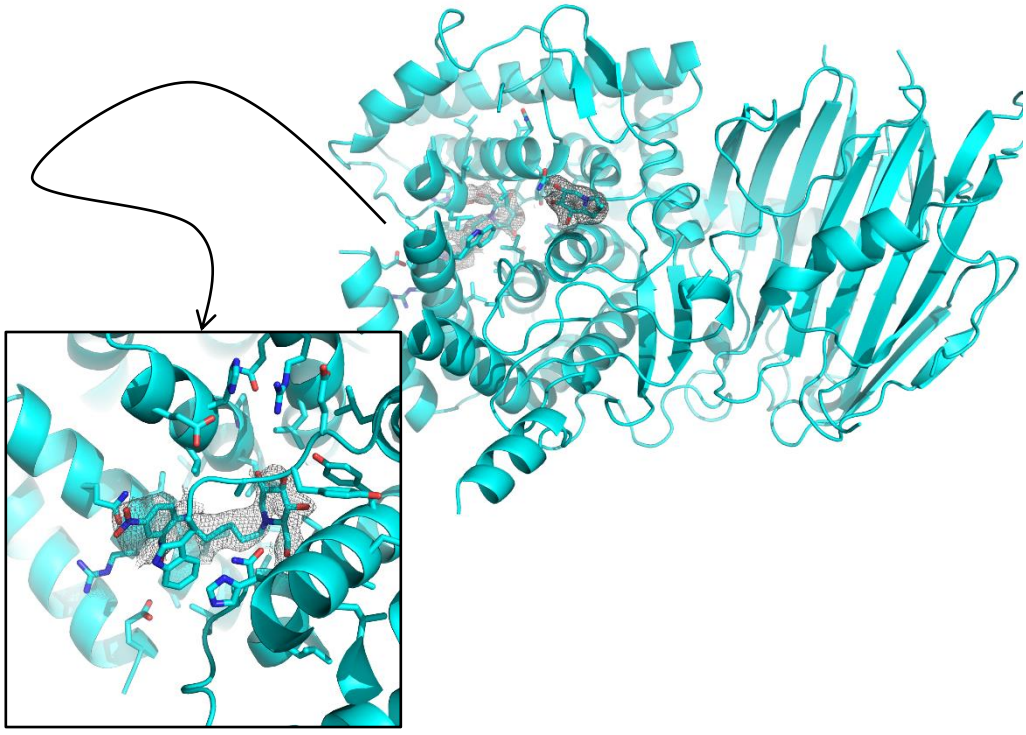


Figure 3.10. Overall structure of *Mm* glu I in complex with SR1. Electron density, at an I/σ of 1.0 and carve = 1.8, shown for both molecules of SR1 found within the structure. In the box, a view of the second iminosugar binding site.

other surface residues on this side of the protein. The three leucine residues (454, 514 and 758) that line this accessory site are all fully conserved, however Gln^{697} and Tyr^{757} both appear to have some slight variation present between orthologues. Whether iminosugar

binding to this site could potentially alter the activity of the protein in some allosteric manner is unknown. The only observed conformational change near the active site is the altered positioning of the nearby loop described in 3.6.4, which could be due to a number of factors.

3.6.6. Likely binding conformation of glucose in the +1 subsite

A crystal structure of *Mm* glu I was not obtained with any of its natural substrates nor products, despite crystallisation screens involving glucose, kojibiose, Glc_3Man_1 and 1-deoxy-5-fluoroglucose. No data of this kind has been produced for any of the eukaryotic members of the GH63 family; however, a number of structures have been solved with potential substrates for the *E.coli* YgjK as well as a GH63 family member from *Thermus thermophilus*, including the structure of an active site mutant in complex with glucose and lactose[40, 288, 289]. When these latter ligands are overlaid into the active site of *Mm* glu I, they mimic the potential conformations of the terminal glucose moieties arranged as they would be if joined with an α(1-2) linkage (figure 3.11).

If thought about in terms of cleavage of the glycan, the deoxynojirimycin ring of iminosugars sits in the -1 site. When looking at the position of the +1 site, it can be seen that in the *Mm* structure, it looks as though the glucose is coordinated in part by the conserved residues, Arg^{483} and Glu^{484}. The arginine is the same residue as the R486T mutation that leads to catalytically inactive protein and CDG-IIb in humans. Here, we hypothesise that this arginine plays a crucial role in coordinating the substrate: alteration of this residues abrogates substrate binding and therefore activity.

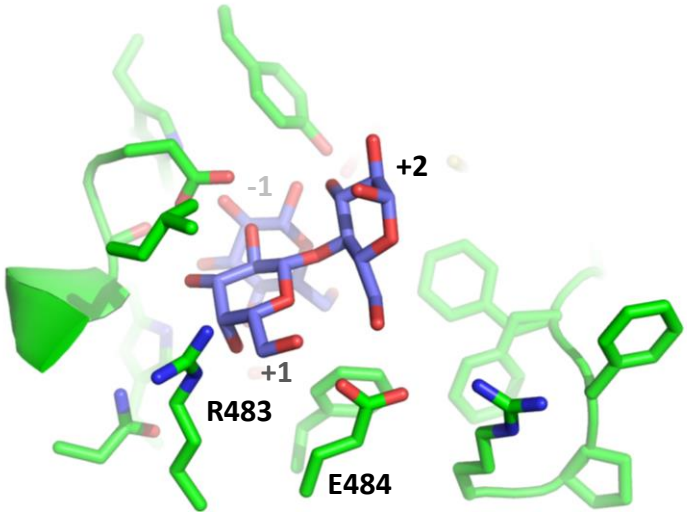


Figure 3.11. Structure of *Mm* glu I (5MHF) with the glucose and α-lactose molecules from an *E.coli* homologue (PDB ID: 5GW7) overlaid in the active site.

As Arg[483] and Glu[484] may play an important role in substrate binding, they may be good targets for specific inhibitors to interact with. A non-hydrolysable α-(1-2)-di-glucose analogue could potentially be a potent, specific inhibitor of glu I, as no other enzymes in the human body target this specific linkage between two glucose residues.

3.7. Ligand binding by ITC

Isothermal titration calorimetry (ITC) is a technique which allows the characterisation of the thermodynamics and stoichiometry of binding of ligands to proteins. Here, ITC was used to determine whether the flexible mode of binding observed in the crystal structures of M*O*N-DNJ to *Mm* glu I was physiological or an artefact of the crystallisation conditions.

A small panel of iminosugars was tested: DNJ, *N*B-DNJ and M*O*N-DNJ. These vary in the length of their alkyl tail: the first simply mimics the sugar ring and the second has an additional butyl chain attached to the nitrogen of the DNJ ring. Despite not crystallising

either ligand with *Mm* glu I, we can infer their position within the active site based on the structures of the longer side chain iminosugars that have been co-crystallised. As ordered side chains were observed up to between 2 and 3 alkyl carbons past the nitrogen in the deoxynojirimycin ring, M*O*N-DNJ is the only member of this small panel that would be predicted to have a significantly flexible tail, due to the lack of ordered electron density observed in the crystals for the distal regions.

The titrations for each of the three ligands are shown in figures 3.12 to 3.14. An approximation for the K_I, the binding affinity, can be determined from the results of each graph. DNJ has an affinity of 3.6 μM (95% CI: 2.4-6.7 μM), *N*B-DNJ an affinity of 463 nM (95% CI: 352-698 nM) and M*O*N-DNJ an affinity of 212 nM (95% CI: 157-325 nM). DNJ therefore has the weakest affinity for glu I, and the addition of the butyl chain in *N*B-DNJ is sufficient to strengthen the affinity almost 10-fold. The much longer chain of M*O*N-DNJ only leads to a 2-fold increase in the strength of the binding over *N*B-DNJ.

ITC can also inform on whether these increases in affinity are due to an increase in the enthalpy change (ΔH) or entropy change (ΔS) of the binding. For DNJ and *N*B-DNJ, the ΔS are -14.1 calories/mole/degree and -15.2 calories/mole/degree respectively, this makes the entropy change of *N*B-DNJ binding more unfavourable than that for DNJ binding. However, we see that a more favourable (for binding) ΔH for *N*B-DNJ outweighs the more unfavourable ΔS and leads to a stronger affinity for glu I than by DNJ.

In the case of M*O*N-DNJ, the ΔH is less favourable than for both *N*B-DNJ and DNJ, although still within the 95% CI of the latter. Yet, the ΔS is just -5.92 calories/mole/degree, a much less unfavourable value which allows for a stronger binding affinity than for *N*B-DNJ.

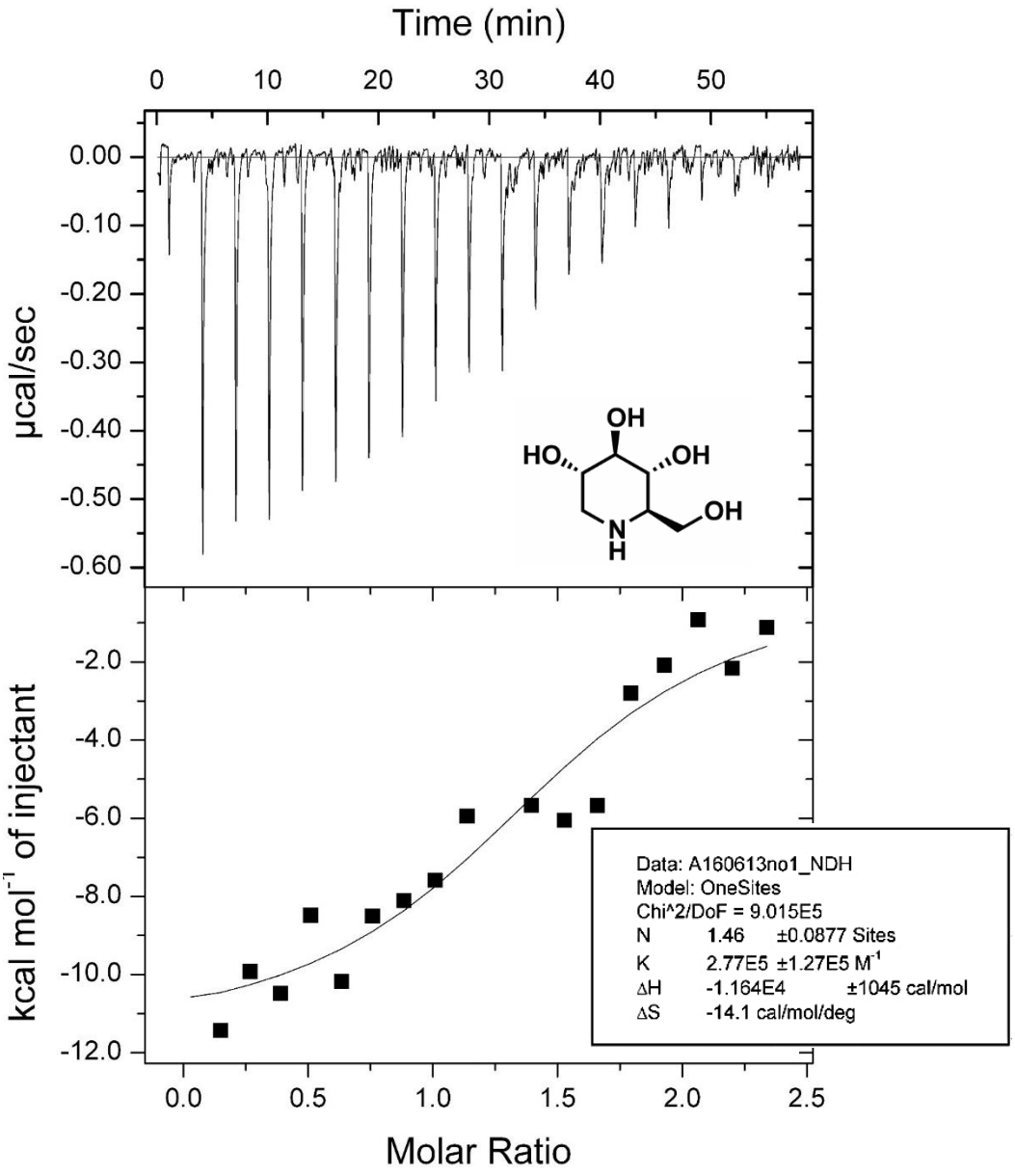


Figure 3.12. Isothermal titration calorimetry data of *Mm* glu I (25 µM) titrated against deoxynojirimycin (300 µM stock).

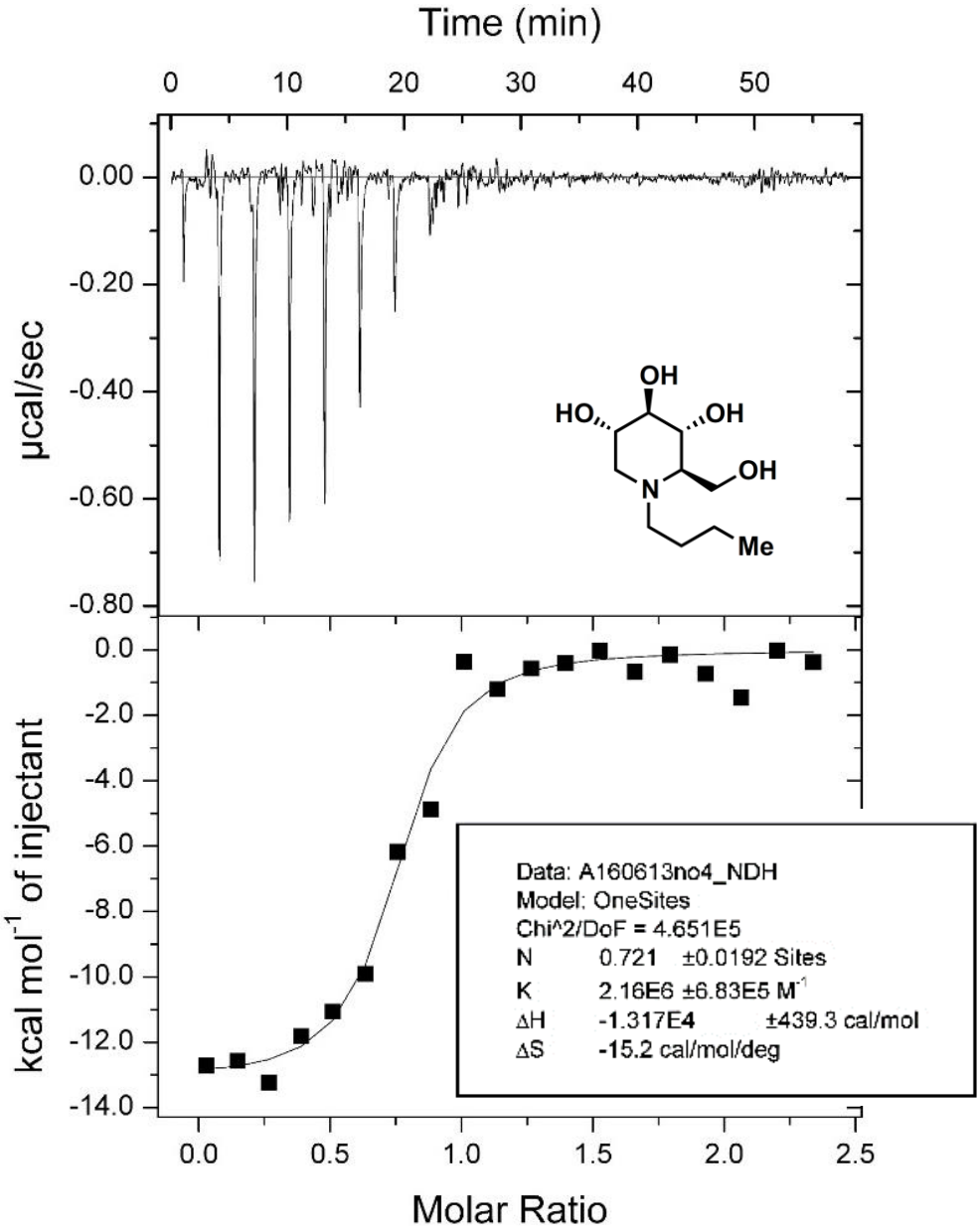


Figure 3.13. Isothermal titration calorimetry data of *Mm* glu I (25 µM) titrated against N-butyl deoxynojirimycin (300 µM stock).

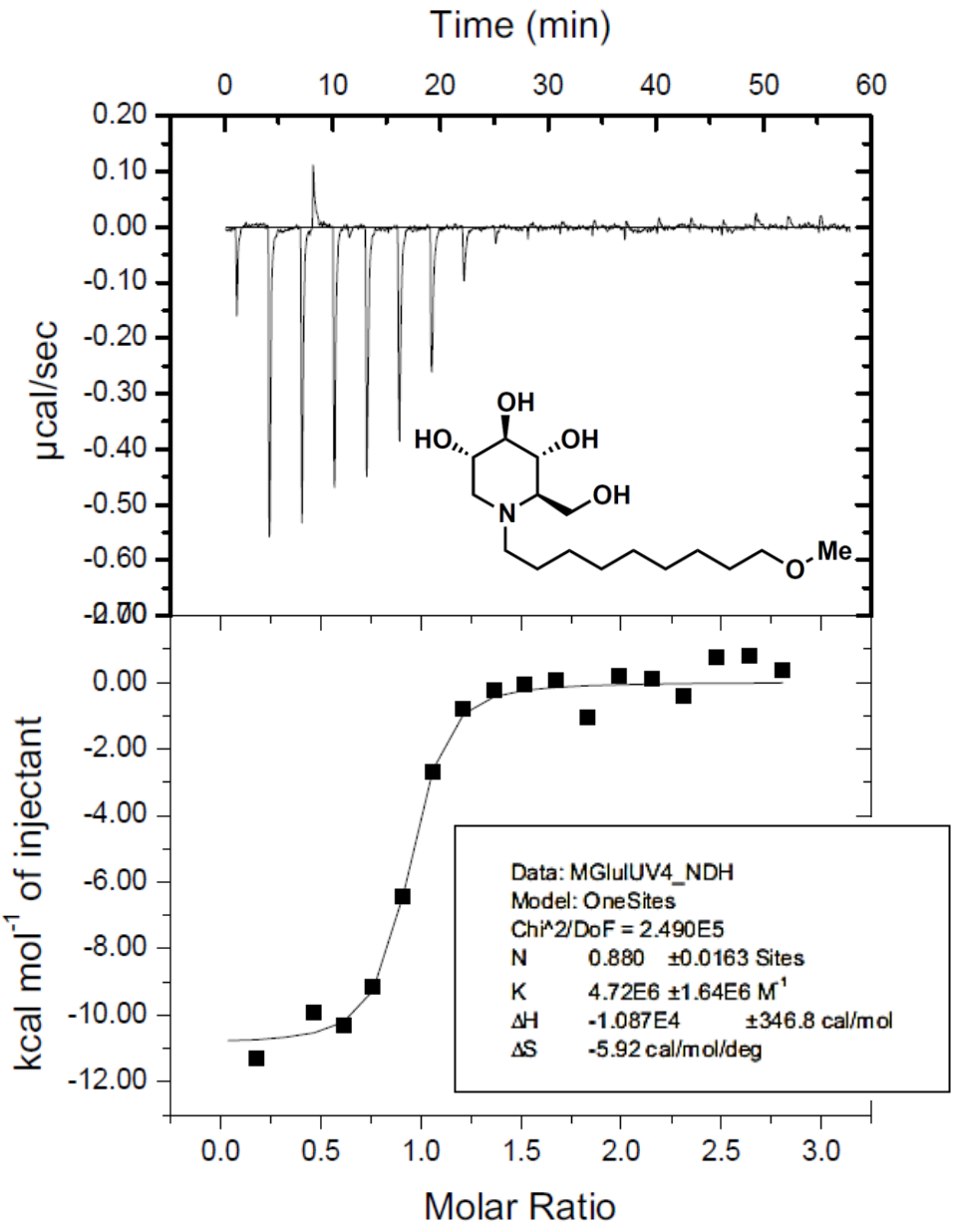


Figure 3.14. Isothermal titration calorimetry data of *Mm* glu I (25 µM) titrated against methoxy-nonyl deoxynojirimycin (300 µM stock).

The ΔS is the measure of the change in entropy in the system, and as such it is a measure of disorder/ order upon binding. Upon binding of a ligand to a protein, the changes in the ligand and protein typically dominate the process and the system inherently becomes more ordered (only the ligand-induced displacement of ordered water into the bulk solvent contributes to disorder), leading to a less favourable, and more negative ΔS. For the binding of MON-DNJ to *Mm* glu I, the more favourable ΔS indicates that the system becomes less ordered when compared to the binding of NB-DNJ. This could be due to a number of reasons, including that the tail of MON-DNJ remains disordered, in agreement with the flexibility observed in the crystal structures. An alternate possibility is that the binding of the iminosugar displaces a number of ordered waters from within the active site of *Mm* glu I, adding some disorder to the system. Further, it is possible that by binding to glu I, the 'water cage' surrounding the alkyl chain is disrupted, imparting a huge entropic cost on the binding reaction. Regardless, these data do not go against the theory that the long-tailed iminosugars bind to *Mm* glu I with a flexible tail.

3.8. Discussion

Extensive structural characterisation of other members of the GH63 family from lower organisms has been performed: including YgjK from *Escherichia coli* and the mannosylglycerate hydrolase from *Thermus thermophilus*[40, 288, 289]. These studies determined the likely catalytic mechanism for this family of enzymes, as well as the manner in which sugar molecules are coordinated in the -1, +1 and +2 subsites.

Presented here is the first structure of an ER α-glucosidase I (also known as processing α-glucosidase or mannosyl-oligosaccharide glucosidase) from a higher vertebrate, the only

known GH63 family member from these species. Previous work on the ER α-glucosidase I has led to the structure of the *Sc* glu I, yet its usefulness in the design of more specific inhibitors was hindered by the occlusion of the active site by the expression tag of a neighbouring molecule of glu I in the crystal lattice[42].

Glu I is a type II single-pass transmembrane protein: formed of two domains that are present in the lumen of the endoplasmic reticulum, a single transmembrane helix and a cytoplasmic portion that contains a number of arginine moieties which are likely responsible for the ER localisation of the protein[38, 290]. Only the lumenal portions of the protein have been expressed and crystallised in this study. The N-terminus of the construct is dominated by a super-β-sandwich domain, the exact function of which is unknown. However, the super β-sandwich is a common motif in carbohydrate binding modules (CBMs) and a molecule of glucose was situated near the centre of this domain in YgjK[40, 291]. This potential CBM could serve a role in localising glu I to areas in the cell where there are high concentrations of nascent glycans; although the lack of high sequence conservation on the surface of this domain suggests that any interaction may not be as crucial as it may have once been, on an evolutionary timescale. It could also be that it only serves a role in stabilising the protein and has long since lost its role as a CBM. The naturally occurring mutation R110H which alters a salt bridge between the N-terminal and C-terminal domains de-stabilises the protein sufficiently to send it prematurely for to the ERAD machinery[45].

The C-terminal domain contains the site of catalysis and is formed of an $(\alpha/\alpha)_6$ helical barrel. The only area of strong surface sequence conservation surrounds the active site, which indicates that this, and this alone, is the primary site of substrate binding.

Supporting this is the fact that the deoxynojirimycin rings of the iminosugar ligands co-crystallised with *Mm* glu I are found in the same orientation and coordination as glucose in existing structures of the GH63 family, implying that iminosugars inhibit glu I by blocking substrate binding to the active site.

One of the primary aims of solving the structure of a mammalian glu I was to help ascertain whether there are any features that could be targeted to make a more specific inhibitor. Iminosugars, as sugar-mimics, target a number of enzymes within the body: Miglitol is used to target intestinal glucosidases to treat diabetes mellitus whereas Zavesca (Miglustat) is used to target the glucosylceramide synthase for substrate reduction therapy in lysosomal folding diseases[209, 210, 220, 292]. Yet, the major side effect of iminosugar treatment is osmotic diarrhoea, resulting from the inhibition of these intestinal glucosidases[51, 293]. Whilst this side effect may be both tolerable and manageable, it would be prudent to minimise the risk of this, and any other potential, side effects. A likely candidate for cross inhibition is glu II, which plays a downstream role in the same ER-resident process as glu I. Discovery of tocopherol-DNJ, an iminosugar specific for glu II over glu I, which can also be specifically targeted to the target both the liver and monocytes[294], shows that it is feasible to generate specific inhibitors of a desired glucosidase.

The DNJ head groups of the iminosugars co-crystallised with *Mm* glu I were all well-defined in the electron density, yet the various alkyl tails were only defined up to a few carbons from the heterocyclic nitrogen. Both the crystallisation data and isothermal titration calorimetry studies do not discount that the alkyl tails are flexible in solution. Unfortunately, the DNJ ring is unlikely to provide any specificity whilst the variable tails

could, yet they have not shown specific interactions. Any observed interactions would have been useful for further study as glu I is the only member of the GH63 family in mammals and so is unlikely to share many substrate binding characteristics with other enzymes within the host, past the coordination of the DNJ ring. Further crystallisation studies with a more diverse panel of iminosugars may discover a conformation of a 'tail' that can be exploited to improve the specificity towards glu I.

However, some features of glu I have been elucidated which may be of interest to the work towards specific inhibitors and could be exploited. Both tyrosine[746] and phenylalanines[440, 441 & 445] are located close to the DNJ ring binding site. π-stacking interactions could be exploited by an aromatic ring on a correctly sized linker to the DNJ head-group. Similarly, cysteine[816] is a polar side-chain located very close to the modelled alkyl tails: if a group were designed to interact with this residue, it may lead to a specific inhibitor with high affinity.

An alternative area of study would be that of the +1 subsite of glu I, as suggested by the alignment of ligands found in the structure a GH63 family member from *Escherichia coli*. Interactions with glutamate[484] and arginine[483] are predicted to coordinate the sugar. The latter residue is so crucial to the binding of substrate that the naturally occurring variant R486T in humans leads to an inactive form of the enzyme and the same severity of CDG-IIb as a nonsense mutation[44, 45]. The 'tails' of all the iminosugars tested thus far have not extended towards the +1 subsite; therefore, it would be of interest to test a non-hydrolysable α-(1-2)-di-glucose analogue of the natural substrate. This disaccharide is not targeted by any other enzymes within the body, although some optimisation may be required to distinguish it from enzymes that recognise sucrose, the α-D-Glc-(1→2)-β-D-

Fru disaccharide. Nonetheless, it is possible that such an inhibitor would have a higher affinity for glu I and could be used at sufficiently low concentrations not to modulate the activity of other proteins.

Of a side-interest is the conformation of the 'flexible loop' ending up in a rigid conformation in the structure of *Mm* glu I in complex with the iminosugar SR1. This is coincidentally the only iminosugar of those tested to bind to a second site in the crystal. Whilst part of the binding interaction is of the 'tail' of the iminosugar to the protein expression tag, it is this interaction which shifts the expression tag sufficiently to allow the DNJ (glucose-mimic) ring to bind to the glu I-proper. It is tempting to speculate that this leads to an allosteric movement and fixes the flexible loop in place, thus altering the recognition of substrate, for either better or for worse. If this were the case, it could provide an alternative site for the design of specific ER α-glucosidase I inhibitors.

4. UGGT and Sep15

UDP-glucose glycoprotein: glucosyltransferase (UGGT) acts as the checkpoint enzyme of the calnexin cycle, selectively re-glucosylating each misfolded glycoprotein and causing its retention in the endoplasmic reticulum by ER lectins, which in turn mediate further attempts at attaining the glycoprotein's correct fold.

Selenoprotein-15 (Sep15 or SELENOF) is a 15 kDa protein that binds with nanomolar affinity to UGGT, an interaction that enhances the re-glucosylation activity of UGGT and which mediates Sep15's retention in the ER[77, 78]. A solution structure of a fragment of *Drosophila melanogaster* Sep15 lacking its 40 N-terminal residues revealed a thioredoxin-like (TRXL) domain[71]. Its unusually large number of cysteine residues, in addition to the presence of selenocysteine in Sep15 from most species, have led to the exploration of a possible role of Sep15 as a redox protein, helping to ensure a glycoprotein's di-sulphide bonds are not mis-paired[71, 89].

Sep15 markedly increases the re-glucosylation activity of both higher eukaryotic UGGT isoforms, UGGT1 and UGGT2, the latter of which only recently had its misfolded glycoprotein re-glucosylation activity confirmed[64]. The UGGT-Sep15 interaction has proven crucial for the quality-control and ER retention of cysteine-rich proteins such as retinol dehydrogenase 11 and IgG and IgM immunoglobulins[79, 295].

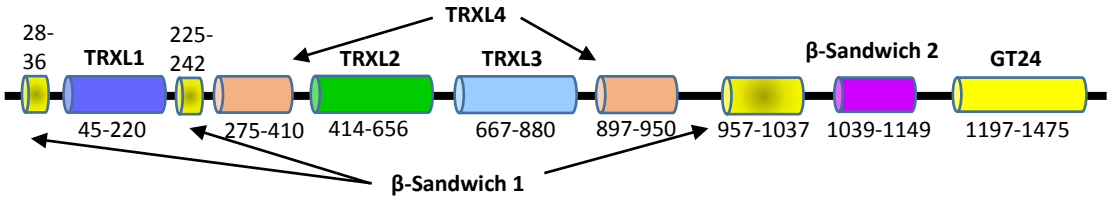
One of the aims of this thesis is to test UGGT as a potential target for novel broadly-acting antivirals. Presented in this chapter is the work which resulted in the highest resolution structure of a full-length UGGT to date - which in turn could form a basis for structure-based lead discovery; the successful cloning and expression of human UGGT1; and a structural investigation of the UGGT-Sep15 interaction, a complex with enhanced re-glucosylation activity that allows UGGT to access a broader range of substrate glycoproteins.

4.1. Cloning

4.1.1. Cloning of a double cysteine *Ct* UGGT mutant

Following on from the work of Dr. Pietro Roversi, where three conformations of *Chaetomium thermophilum* UGGT were determined by X-ray crystallography[68], double cysteine mutations were designed in the hope that they could lock the flexible *Ct* UGGT TRXL2 domain in a closed conformation and provide a better crystallisation target. In particular, based on the "closed" structure (PDB ID 5N2J), cysteines at Asp^{611} and Gly^{1050} were predicted to come close enough in space so as to allow formation of a disulphide bond.

Cloning of the *Ct* UGGT double cysteine mutant was carried out with Mr Andrea Lia, a visiting Erasmus student (Università del Salento, Lecce)*. The *Ct* UGGT construct, after



Gene name	aa_N	aa_C	Vector	Construct:
Hs UGGT1	D45	P236	pOPINTTGneo	TRXL1
Hs UGGT1	D45	K253	pOPINTTGneo	TRXL1 + β-strand
Hs UGGT1	D45	R696	pOPINTTGneo	TRXL1, β-strand, half of TRXL4 and TRXL2
Hs UGGT1	D45	A954	pOPINTTGneo	TRXL1 to TRXL3
Hs UGGT1	D45	L1035	pOPINTTGneo	TRXL1 to TRXL4 (whole domain)
Hs UGGT1	D45	Y1113	pOPINTTGneo	TRXL domains + first β-sandwich (whole domain)
Hs UGGT1	D45	L1222	pOPINTTGneo	TRXL domains + both β-sandwich domains
Hs UGGT1	D45	K1553	pOPINTTGneo	Full length construct
Hs UGGT2	T30	P224	pOPINTTGneo	TRXL1
Hs UGGT2	T30	K241	pOPINTTGneo	TRXL1 + β-strand
Hs UGGT2	T30	I684	pOPINTTGneo	TRXL1, β-strand, half of TRXL4 and TRXL2
Hs UGGT2	T30	A932	pOPINTTGneo	TRXL1 to TRXL3
Hs UGGT2	T30	L1013	pOPINTTGneo	TRXL1 to TRXL4 (whole domain)
Hs UGGT2	T30	Y1091	pOPINTTGneo	TRXL domains + first β-sandwich (whole domain)
Hs UGGT2	T30	I1200	pOPINTTGneo	TRXL domains + both β-sandwich domains
Hs UGGT2	T30	T1514	pOPINTTGneo	Full length construct

Figure 4.1. Design of constructs of *Hs* UGGT1 and *Hs* UGGT2. pOPINTTGneo encodes an N-terminal signal sequence for secreted expression and a C-terminal KHHHHHH- affinity tag. Top panel adapted from Roversi *et al.* (Roversi et al. 2017)

* Under the supervision of Dr Lucia Marti, Institute of Sciences of Food Production, Lecce, Italy

removal of the signal peptide encoded by the pHLsec vector, encodes residues 24-1505 (Uniprot G0SB58) with 'ETG' at the N-terminus and 'GTKHHHHHH' at the C-terminus. Gibson assembly was used to move *Ct* UGGT from pHLsec to the Litmus28i cloning vector (for primers, see Appendix) before site-directed mutagenesis was used to achieve D611C and G1050C mutations, in successive steps. Finally, Gibson assembly was used to return the mutated gene back into the pHLsec mammalian expression vector.

4.1.2. Cloning of *Hs* UGGT1 and Hs UGGT2 with OPPF

Cloning of *Hs* UGGT constructs was carried out at OPPF in collaboration with Dr Jo Nettleship. Constructs were derived from *Hs* UGGT1 and *Hs* UGGT2 - both in the vector pUC57 - bought from GenScript and flanked with an N-terminal AgeI restriction site and a C-terminal KpnI site after the stop codon. Eight constructs were designed for each of the proteins, based on the domain boundaries identified by the crystal structure of *Ct* UGGT (figure 4.1, PDB ID: 5MU1), before the structure had been fully refined and accepted for deposition[68]. Based on the two potential N-glycosylation sites in the sequence, the three disulphide bonds likely present, the structure and by considering the complex topology of the protein, cloning was restricted to mammalian expression vectors. Primers used for the cloning of the constructs can be found in the appendix.

4.1.3. Cloning of the *Hs* Sep15-interacting loop from *Hs* UGGT1 into *Ct* UGGT

A study of the interaction of *Drosophila melanogaster* UGGT with Sep15 using hydrogen-deuterium exchange MS recently highlighted the peptide loop of UGGT that interacts with Sep15. Alignment of *Dm* UGGT and *Ct* UGGT sequences found that the Sep15-

binding loop of *Dm* UGGT was absent from the *Ct* UGGT sequence, consistent with the observation that *Chaetomium thermophilum* does not appear to encode a Sep15 homolgue[296]. Towards a structure of the UGGT-Sep15 complex, we decided to exploit the propensity for *Ct* UGGT to crystallise, by grafting the *Hs* Sep15-binding loop sequence from *Hs* UGGT1 into the corresponding region of *Ct* UGGT.

The loop to be removed from *Ct* UGGT was identified as R^{249} to P^{284} inclusive, by alignment with both fruit fly and human sequences. The loop to be inserted from the *Hs* UGGT1 sequence was T^{261} to P^{317} inclusive, the sequence being ^{261}TQVKGTEVNTTVIGENDPIDEVQGFLFGKLRDLHPDLEGQLKELRKHLVESTNEM AP^{317}. A codon-optimised version of the DNA encoding this portion of *Hs* UGGT1, with 20 base pair complementary overlaps with the corresponding *Ct* UGGT sequence, was ordered as a GeneString™ (Thermo Fisher Scientific) for Gibson assembly (see Appendix for exact sequence of this oligonucleotide).

Ct UGGT DNA was transferred from pHLsec to the Litmus28i cloning vector using Gibson assembly (see Appendix for primers); Litmus28i was digested with AgeI-HF and KpnI-HF. *Ct* UGGT DNA in Litmus28i was amplified with primers 320/321 to exclude the aforementioned residues, the resulting DNA was purified using a PCR clean-up kit. Gibson assembly was used to ligate the Sep15-binding loop from *Hs* UGGT1 into the PCR fragment of linear *Ct* UGGT DNA in Litmus28i, and successful insertion was confirmed by DNA sequencing (M13F/R primers). A second round of Gibson assembly was used to return the construct into pHLsec (see appendix for primers, pHLsec was digested with KpnI-HF and AgeI-HF). Correct insertion of the new construct was confirmed by DNA sequencing, with primers 14/15.

4.1.4. Cloning of *Hs* Sep15

The protein coding sequence for *Hs* Sep15 was obtained from Uniprot (ID O60613) and the signal peptide identified and removed using the SignalP4.1 server. The selenocysteine codon was changed to a cysteine codon and then the DNA sequence was codon optimised for *E.coli* expression using the GeneArt® GeneOptimiser (Thermo Fisher Scientific). Two vectors were chosen for expression; pET303, encoding a hexahistidine purifcation tag at the C-terminus; and pET/M-43, encoding a C-terminal hexahistidine tag and an N-terminal maltose-binding protein (MBP) domain. For cloning into pET303, XbaI and NsiI restriction sites were added to the N- and C- termini respectively; for expression in pET/M-43, NcoI and NotI sites were added. Overlaps of 15bp with the target vector were added to each terminus and the sequences ordered as GeneStrings (Thermo Fisher Scientific). For the full sequences of the DNA ordered, see the appendix.

For Gibson assembly, the pET303 vector was digested with XbaI and NsiI, pET/M-43 with NcoI and NotI, and purified; the linearised vectors were separated on a gel and gel-extracted. Gibson assembly was then performed on the linearised DNA with the respective GeneString fragment. Correct insertion of the DNA was confirmed by DNA sequencing using T7 forward and reverse primers.

4.2. Purification and expression

4.2.1. Expression and purification of the UGGT chimera

Ct UGGT with the *Hs* Sep15-binding loop, "the UGGT chimera" was expressed in HEK293F cells. A 600 ml culture was transfected, supernatant processed as described in

the methods, and run on a 5 ml HisTrap HP column. The eluate was pooled and concentrated in a 100 000 MWCO centrifugal concentrator before further purification by SEC on a Superdex 200 16/600 column.

Protein eluted from the column in 2 high MW peaks, likely corresponding to a monomer and a dimer. Taking the smaller MW peak, and passing the sample through the SEC column once more leads to the appearance of solely the smaller MW peak, indicating that the two species are not in equilibrium. This smaller molecular weight peak was pooled and concentrated to 8.3 mg/ml, flash-frozen in liquid nitrogen in 100 µl aliquots and stored at -80°C.

4.2.2. Expression and purification of *Hs* Sep15 from pET303

Mature *Hs* Sep15 is predicted to be 15 kDa, whereas the expression system with the expression tag of pET303 brings the expected molecular weight of the protein to 16.5 kDa: thus this particular construct will be referred to as Sep16.5 for the rest of this chapter.

BL21 (DE3) *E.coli* cells were transfected with the prepared Sep16.5 vector and expressed as described in the methods. Protein was first purified on a 5 ml HisTrap HP with a gradient elution. Eluted protein was pooled and concentrated using a 10 000 MWCO centrifugal concentrator before passing onto a Superdex 75 16/600 SEC column. Fractions containing Sep16.5, as identified by SDS-PAGE, were pooled and concentrated to 10.64 mg/ml, flash-frozen in 100 µl aliquots in liquid nitrogen and stored at -80°C.

4.2.3. Expression and purification of *Hs* Sep15 in pET/M-43

Expression of *Hs* Sep15 in pET/M-43 leads to a 58 kDa protein with an N-terminal MBP tag, followed by a 3C-cleavage site, the protein of interest and finally a C-terminal hexahistidine tag: a protein that will be called MBP-Sep15.

Protein expression was performed as for Sep16.5. Protein was affinity-purified on a 5 ml HisTrap HP. After elution, the flow-through was repeatedly passed through the column and eluted again until no more protein bound to the column. Elution fractions were pooled and concentrated using a 30 000 MWCO centrifugal concentrator and loaded onto a Superdex 75 16/600 SEC column. Fractions containing the desired protein were pooled and concentrated to 15.6 mg/ml, before being flash-frozen in 100 µl aliquots in liquid nitrogen and stored at -80°C.

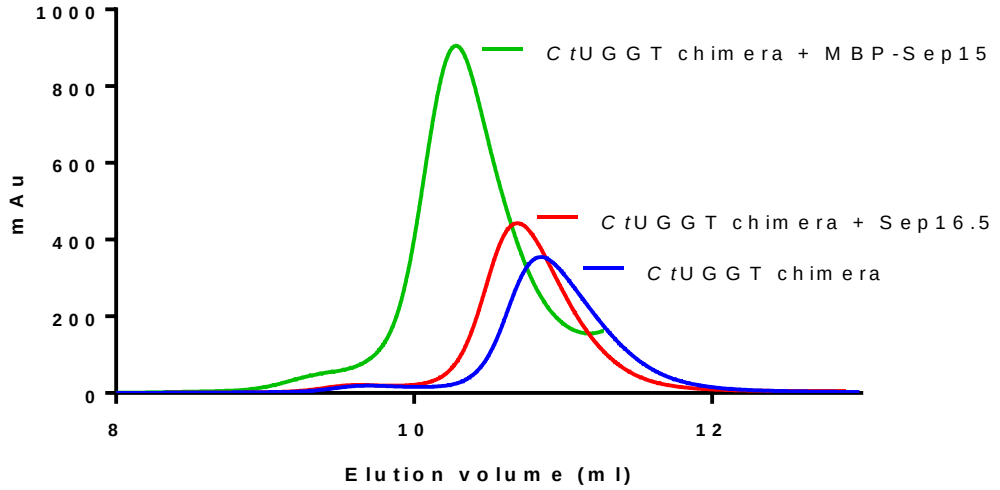


Figure 4.2. The UV absorbance traces of the elution profiles of the UGGT chimera and its complexes with Sep16.5 and MBP-Sep15, run on a Superdex 200 10/300 increase SEC column. Traces are shortened to exclude the Sep16.5 or MBP-Sep15 peaks for simplicity.

4.2.4. Purification of the UGGT chimera – Sep15 complexes

Samples were prepared as outlined in chapter 2.6.4. on an analytical Superdex 200 10/300 SEC column. Formation of the complex was confirmed by the shift in elution volume of the UGGT chimera to that of a higher molecular weight protein, upon exposure to a 10-fold molar excess of a Sep15 construct (see figure 4.2) and confirmed by the presence of Sep16.5 or MBP-Sep15 in the higher molecular weight peak with UGGT, as assessed by SDS-PAGE. The difference in heights of the peak are due to the larger extinction coefficient for the complexes; in the run forming the complex with MBP-Sep15, a larger quantity of protein was run as well. For crystallisation attempts, the UGGT chimera was concentrated to 5.5 mg/ml, the complex with Sep16.5 to 6.25 mg/ml and the complex with MBP-Sep15 to 8.6 mg/ml. For SAXS, the UGGT chimera was concentrated to 12.2 mg/ml, the complex with Sep16.5 to 15.44mg/ml and the complex with MBP-Sep15 to 21.4 mg/ml. All samples were flash-frozen in liquid nitrogen and stored at -80°C. Samples for SAXS were stored on dry ice for the journey to ESRF.

4.2.5. Purification of the double cysteine *Ct* UGGT mutant

Expression and purification of the double mutant of *Ct* UGGT (D611C:G1050C) was carried out in the same manner as for UGGT chimera. Only one peak was observed from the SEC column. The existence of the disulphide bond was confirmed by mass spectrometry by Dr Abhinav Kumar (see [68]).

4.3. Crystallisation trials

4.3.1. The UGGT chimera and its complexes

Of the three crystal structures of *Ct* UGGT, PDB ID 5MU1 ($P6_1$ crystal form) shows a large 'channel' within the crystal lattice, leaving a large solvent space where the predicted Sep15-binding loop should be located[68]. The crystallisation condition that gave rise to the $P6_1$ crystal form was chosen to attempt crystallising the UGGT chimera and its Sep15 complexes.

Well E2 of the Morpheus HT-96 sparse crystallisation screen contains 0.1 M buffer system 1 (imidazole: MES monohydrate), 0.12 M ethylene glycols mix and 50% w/v precipitant mix 2 (all obtained from Molecular Dimensions). An optimisation screen around the condition was prepared, varying the pH of buffer system 1 from 6.2 to 6.9 in 8 steps and varying precipitant mix 2 from 47.5% to 50.5% w/v in 12 steps. For each of the three

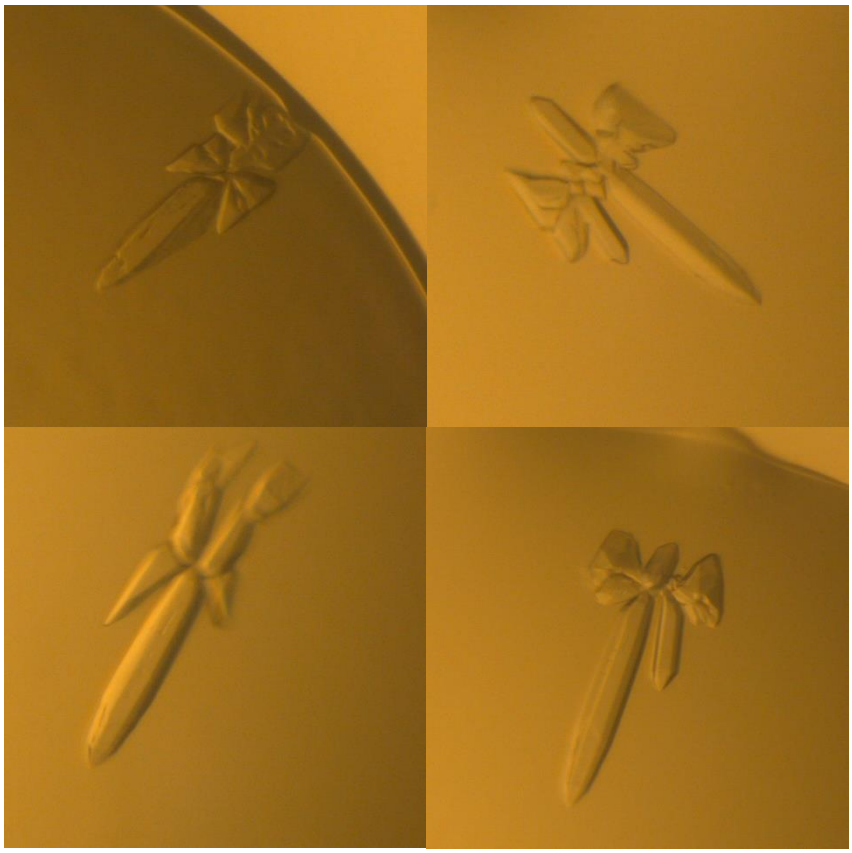


Figure 4.3. Morphology of the *Ct* UGGT crystals nicknamed "the transformers".

prepared samples (the chimera and the two complexes), the crystallisation screens were set up with the first well of this screen containing 100 nl of protein and 100 nl of crystallisation condition, and in the second well with 133 nl of protein and 67 nl of crystallisation conditions. Further sparse screens were trialled for each of the three protein samples in the Morpheus, Proplex and JCSG+ commerical screens, each with the same ratios of protein to crystallisation conditions.

Crystals of the UGGT chimera were the only ones to form in the optimisation screen: at higher concentrations of precipitant mix 2; between 49.95% and 50.5% w/v, irrespective of the pH. All had the same morphology as the crystal form 5MU1, nicknamed "the transformer" because of its morphology (Figure 4.3). Crystals were moved to a 0.5 µl drop containing 2.9 mg/ml of Sep16.5 in the crystallisation condition (prepared by diluting 2 µl of a 29.6 mg/ml stock of Sep16.5 with 18 µl of crystallisation condition from the reservoir). These were left overnight before harvesting them with a nylon loop and cryo-cooling them in liquid nitrogen and taken for DLS for data collection (see appendix for parameters).

4.3.2. Crystallisation of *Ct* UGGT D611C:G1050C

Mutant *Ct* UGGT D611C:G1050C protein (see 4.1.1) was concentrated to 3.8 mg/ml and screened in the Morpheus HT-96 sparse crystallisation screen. A crystal was found in well G2: 0.1 M buffer system 1 pH 6.5, 0.1 M carboxylic acids and 50% v/v precipitant mix 2. The crystal was fished with a nylon loop and cryo-cooled in liquid nitrogen before being taken to the ID30A-1 beamline at ESRF (see appendix for data collection parameters).

4.4. Data processing and refinement

4.4.1. Data processing

Data for the UGGT chimera were collected at DLS (see appendix for data collection parameters) and processed using AutoPROC (Table 4.1). The spacegroup of this crystal, as well as all the other crystals screened at this session, was $P6_122$, whereas the screen was designed to optimise around a condition that gave rise to the $P6_1$ form.

Table 4.1. Data Processing Statistics

	UGGT Chimera	*Ct* UGGT D611C:G1050C
Resolution Range (Å)	143.03 - 3.81 (4.02 - 3.81)	77.85 - 2.78 (2.93 - 2.78)
Cell Dimensions:		
a, b, c (Å)	165.146, 165.146, 253.578	78.23, 142.16, 186.09
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 90.00
Spacegroup	$P6_122$	$P2_12_12_1$
Total Reflections	394429 (59205)	266501 (39593)
Unique Reflections	20696 (2948)	53071 (7648)
R_{merge}	0.364 (3.831)	0.133 (3.955)
R_{meas}	0.382 (4.023)	0.164 (4.816)
R_{pim}	0.118 (1.225)	0.093 (2.708)
Completeness (%)	100.0 (100.0)	99.7 (99.8)
Multiplicity	19.1 (20.1)	5.0 (5.2)
CC 1/2	0.998 (0.501)	0.996 (0.412)
I/σ	9.3 (1.2)	7.8 (0.6)

Table 4.1. Data processing statistics, values in brackets refer to the high-resolution shell.

This result is not entirely unexpected, owing to the fact that screens around the original condition (well E2) mostly produced crystals in the $P6_122$ spacegroup, and only one screen in the $P6_1$ spacegroup. This latter crystal form has never been successfully reproduced since. The resolution of 3.8 Å is typical of a crystal of full-length *Ct* UGGT, and should be sufficient to see density corresponding to the Sep16.5, if present in the lattice.

Data for the *Ct* UGGT D611C:G1050C mutant was collected at ESRF and processed using the AutoPROC programme suite, which put the resolution at 2.78 Å. Staraniso[273] was used to correct for diffraction anisotropy; after correction the data extend to 2.67 Å in the best direction. Anisotropically corrected data were used for model building but not deposition of the structure. Data processing statistics can be found in table 4.1.

Table 4.2. Model building and refinement statistics

	UGGT Chimera	*Ct* UGGT D611C:G1050C
Resolution Range (Å)	143.03 - 3.81	77.85 - 2.78
Observations	20639	51917
in free R set	1012	2510
R_{work} (%)	27.7	23.91
R_{free} (%)	28.54	24.5
RMSD bond angles (°)	0.007	0.008
RMSD bond lengths (Å)	1.112	1.18
Ramachandran outliers	8	4
Ramachandran favoured (%)	95.01	97.08
Molprobity Clashscore	10.34 (97th percentile)	17.76 (91st)
Molprobity Score	2.16 (100th percentile)	2.24 (97th)
Good Rotamers (%)	89.73	92.36%
B-factor (wilson plot)	117.4	111.42
Residues modelled*	27-245, 279-1152, 1191-1438, 1443-1476	27 - 245, 279 - 1152, 1191 - 1474. 5x NAG, 1x BMA

Table 4.2. Model building and refinement statistics. * Residue numbering is based on the original *Ct* UGGT structure (Uniprot G0SB58). Whilst residues 279-284 will actually be part of the new Sep15 binding loop, they were left as they would be for the original sequence – the resolution was not powerful enough to see detail to distinguish side chains.

4.4.2. Model building and refinement

As the crystal form of the *Ct* UGGT chimera was the same as the *Ct* UGGT structure "5MZO", and no indexing ambiguities are possible in the point group 622, no explicit molecular replacement step was taken. A rigid body refinement was executed in

Autobuster using 5MZO as the starting model and the initial model relaxed in its correct placement in the cell.

The R_{work}/R_{free} was lowered to 27.7/28.5%, not far from the published values of 25.1/27.8% for the 3.48 Å structure represented by 5MZO. The quality of this model was deemed sufficient for F_O-F_C difference density analysis to locate the *Hs* UGGT Sep15-binding loop and/or the Sep16.5 that was soaked into the crystals at the time of cryoprotection.

For the D611C:G1050C mutant of UGGT, Matthews coefficient analysis[285] indicated the crystal likely contained one molecule per asymmetric unit. Molrep[297] was used for molecular replacement. The structure that would eventually be deposited as PDB ID 5N2J was used as the initial search model, with the TRXL2 domain missing. After a successful placement of this part of the structure, Molrep was used once more to place the TRXL2 domain from the 5N2J model. Rounds of refinement and model building were done using autoBUSTER and Coot until no more improvements could be made. The model was deposited to the PDB under code 5NV4 and was subsequently used to improve the models of the original three models (PDB IDs 5MZO, 5MU1 and 5N2J) before they were deposited in the PDB. Model building and refinement statistics can be seen in table 4.2.

4.5. The structure of the UGGT chimera

The structures of the refined UGGT chimera structure and *Ct* UGGT 5MZO have an RMSD of just 0.4864 Å over 1373 C_αs, which is not surprising given that the chimera was refined with one rigid body per domain starting from PDB ID 5MZO itself. The only noticeable difference is that the chimeric protein is in a slightly more extended, or 'open',

conformation than 5MZO. Only 6 of the residues modelled in the chimera's structure are of the *Hs* Sep15 binding loop, the electron density in this stretch is indistinguishable from the original *Ct* UGGT structure. An attempt was made to soak Sep16.5 into the crystals during cryoprotection, but despite this, no other residues are visible; indicating that in the chimera, the loop may be still flexible and it is not observably complexed to Sep16.5. If the *Hs* Sep15 binding loop were ordered as part of full-length *Hs* UGGT1, it may not be seen here due to the sequence divergence betweeen *Hs* and *Ct* UGGT.

A look at the crystal packing reveals why it is unlikely a complex formed by soaking Sep16.5 into the ready-made *Ct* UGGT chimera crystal. Figure 4.4. highlights the solvent

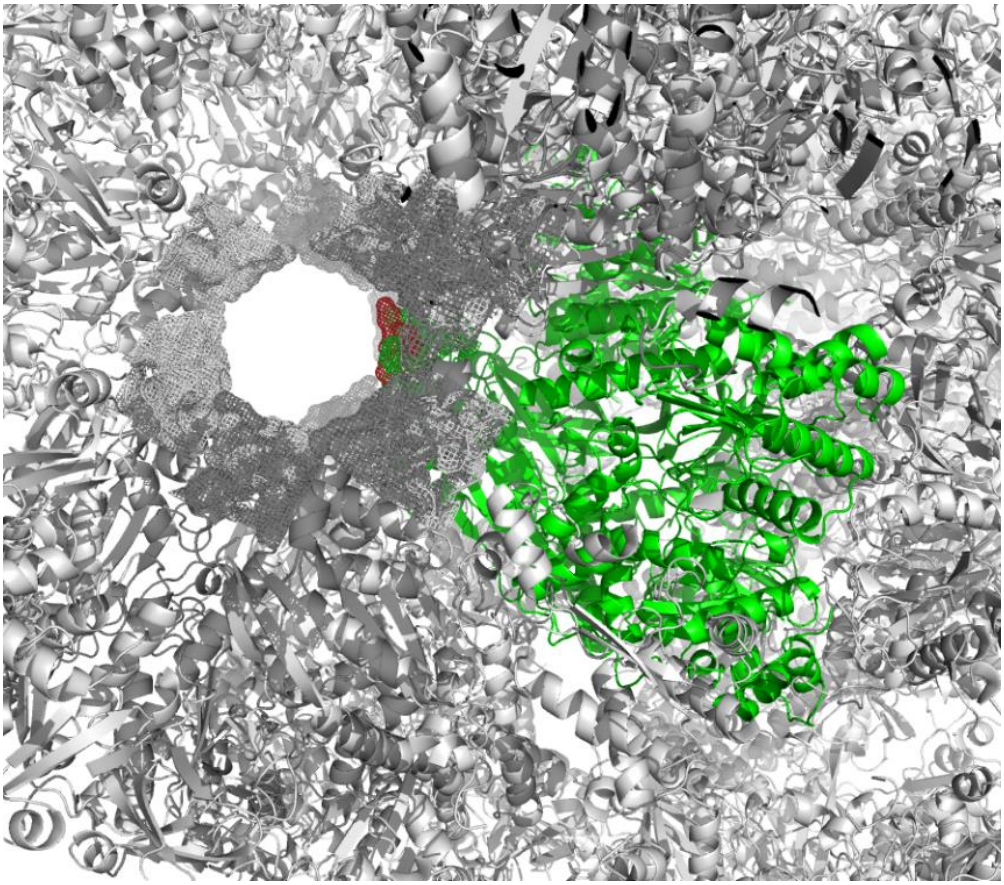


Figure 4.4. Crystal packing of the UGGT chimera crystal. In green is a single molecule of UGGT, in grey are symmetry related molecules. A surface representation is shown for the defined residues of the Sep15 binding loop. The orientation was chosen to highlight the solvent channel to which the loop is exposed to.

channel that the flexible Sep15 binding loop is exposed to: each faces the same loop from an equivalent molecule in the staggered crystal packing. The majority of the 57-residues long Sep15-binding loop is not ordered in the crystal structure but nevertheless, the channel will be occupied by several copies of the disordered loop, which greatly limits the space into which Sep16.5 could bind.

4.6. The structure of *Ct* UGGT D611C:G1050C

The final model can be seen in figure 4.5. The structure clearly shows that the disulphide bond between the two mutated residues, D611C and G1050C, has been properly formed. The disulphide bond links the TRXL2 domain, which has been previously identified as carrying the greatest degree of interdomain conformational mobility, with the second β-sandwich. Whilst the RMSD between this structure and the closest existing structure, the 'closed' form (5N2J), is only 1.17 Å over 1248 C_αs, it is clear that this structure represents an even further 'closed' structure. It is likely that this particular conformation can be reached by the wild-type protein, as the two cysteine residues came close enough to form the disulphide bond. The structure therefore supports high TRXL2 mobility. Should TRXL2 contain part of the misfold recognition epitope(s), this in turn could explain how UGGT can recognise a range of glycoprotein substrates, each with glycans at different sites from the misfold: TRXL2 could just adjust itself so that both glycan and misfold can be accommodated. Importantly, the higher resolution data obtained from this crystal helped to correct a few minor tracing errors in the earlier, lower-resolution structures of *Ct* UGGT. A number of loops remain un-defined, indicating that they are likely naturally flexible.

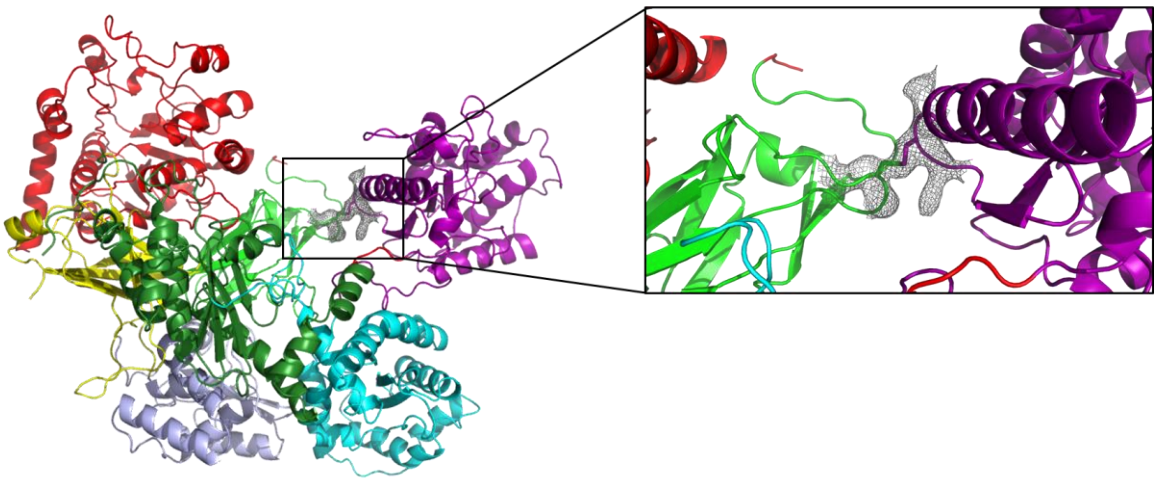


Figure 4.5. Structure of the *Ct* UGGT D611C:G1050C mutant to 2.78 Å resolution. Electron density around the newly engineered disulphide bond is shown at an I/σ of 1 and a carve of 1.7 Å.

4.7. SAXS studies of the *Ct* UGGT chimera

In the absence of ordered electron density in the Sep15 binding region of the UGGT chimera crystals, small angle X-ray scattering (SAXS) was used to obtain low-resolution structural information on the proteins and their complexes in solution.

4.7.1. Collection of data and initial data processing

SAXS data were collected and analysed as described in the methods, section 2.6.4. The averaged and buffer-subtracted data can be seen in figure 4.6. The SAXS data from the UGGT chimera (blue) and its complex with sep16.5 (red) are very similar overall; however, the data do diverge at certain angles, including in the Guinier region, i.e. the data at very low scattering angles, which defines the radius of gyration R_g, as well as differing in the forward scattering intensity $I_{(0)}$. The chimera in complex with MBP-Sep15

differs from the other two in the Guinier region but also has less-pronounced troughs, indicating that the complex is less globular/ spherical than the other two samples.

Although data was collected as a dilution series, only the data from the highest concentration of each sample was used, as it contained the least noise yet did not exhibit any signs of aggregation or inter-molecular repulsion, which can be problematic for some concentrated samples.

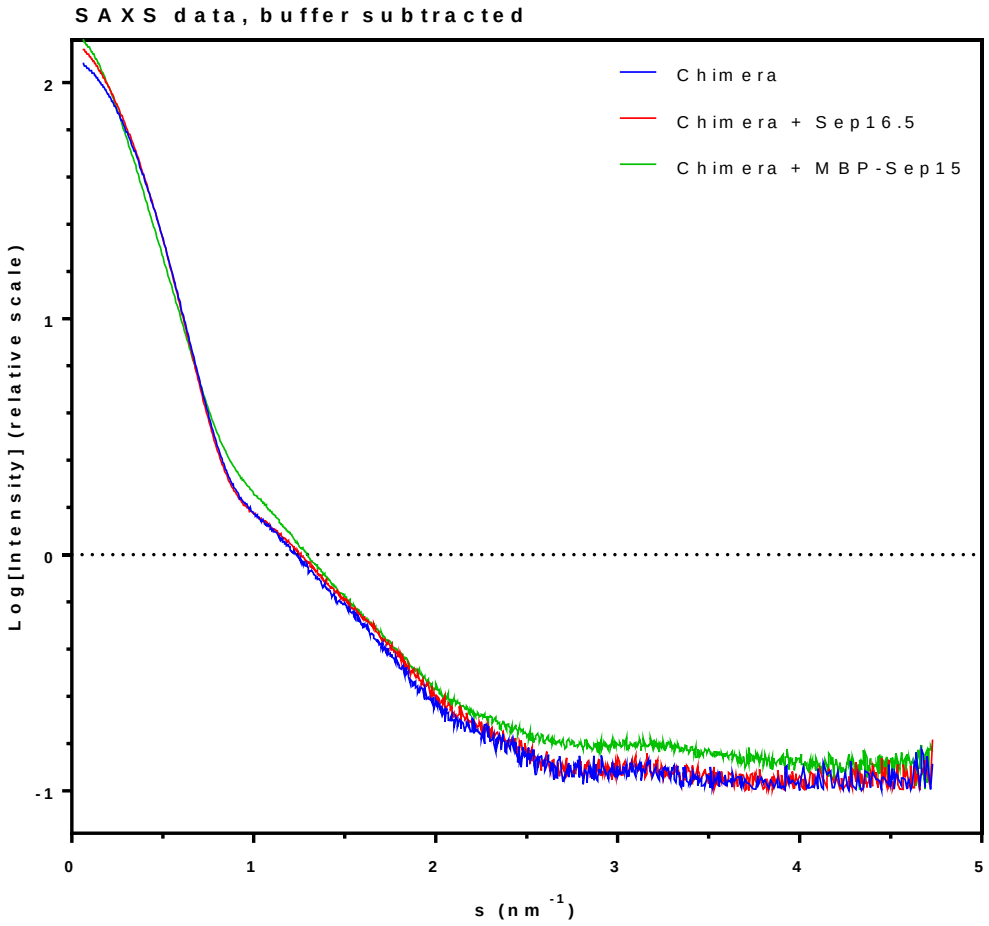


Figure 4.6. Data points from the SAXS measurements of 3 samples taken for UGGT chimera and its complexes with Sep16.5 or MBP-Sep15. The 10 frames for sample were averaged, as were the 20 frames for the buffer blank, the latter was then subtracted from the former to give the curves shown here. Frames with outlier measurements were removed from the analysis.

Based on the analysis of the UGGT chimera alone, the D_{max} was set to 14.1 nm^{-1}, with the R_g/I(0) estimated as 4.61 Å / 117.40 for both the Guinier and P(r) estimations. Usually, a starting value for the D_{max} should be set around 3.5x the value of R_g[298]; here we set it slightly lower as it gave better fit to the data and crystal structures also hold an estimate for the D_{max} – the largest dimensions being approximately 110 Å.

For the UGGT chimera in complex with Sep16.5, the D_{max} was set to 16.32 nm^{-1}, with the R_g/I(0) estimated as 5.30 Å / 136.80 for both the Guinier and P(r) estimations. This was slightly larger than for the chimera alone, yet it was a better fit for the data and is expected as the sample is ~16 kDa larger.

For the UGGT chimera in complex with MBP-Sep15, the D_{max} was set to 18.30 nm^{-1}, with an R_g and I(0) estimated as 5.87 Å and 152.10 for both estimations. Once again, this larger D_{max} would be expected as adding more to a complex is likely to increase the largest intramolecular distances.

4.7.2. *Ab initio modelling*

The integrated *ab initio* model building programme, DAMMIF, as part of the ATSAS suite, was used to produce models that were a good fit to the SAXS data. For each sample, the data processing resulted in the 20 models, shown in figures 4.7 to 4.9.

DAMAVER averages the aligned models, producing a statistically likely average volume. DAMFILT removes low probability and loosely connected features, leaving a most probable, compact model, as can be seen at the bottom of each of the figures. Whilst this model is the most probable envelope, it was not based on the experimental data, thus rarely fits as well as the individual models, and so should be viewed with caution.

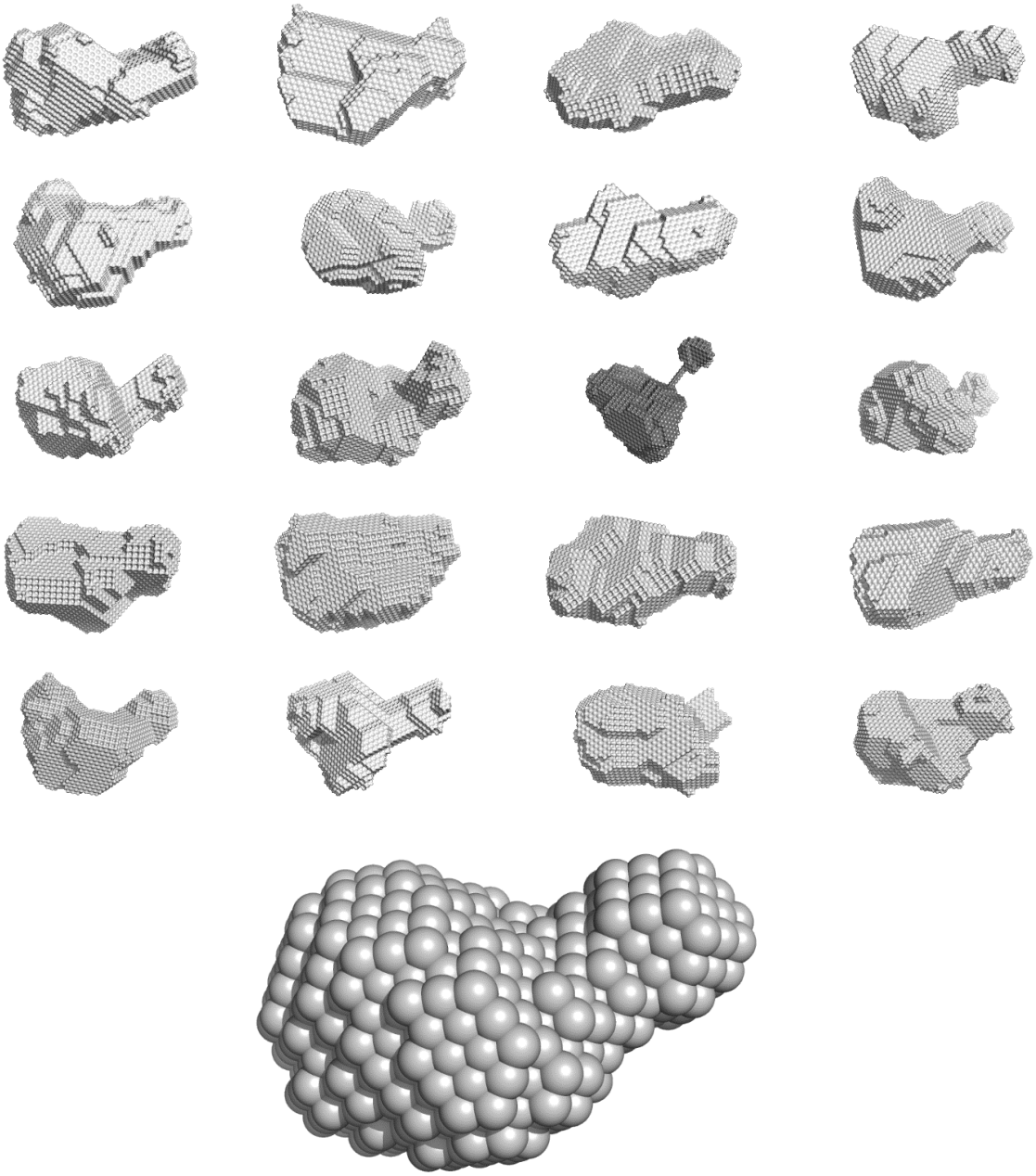


Figure 4.7. *Ab initio* SAXS models of the UGGT chimera. In dark grey is the model excluded from the averaging and model filtering to produce the DAMFILT model (bottom). The models are derived from the highest concentration sample of 12.2 mg/ml with the D_{max} set to 14.1 nm^{-1}. Exclusion criteria: discard models with a normalised special discrepancy (NSD) greater than the mean, plus 2x the standard deviation of the models.

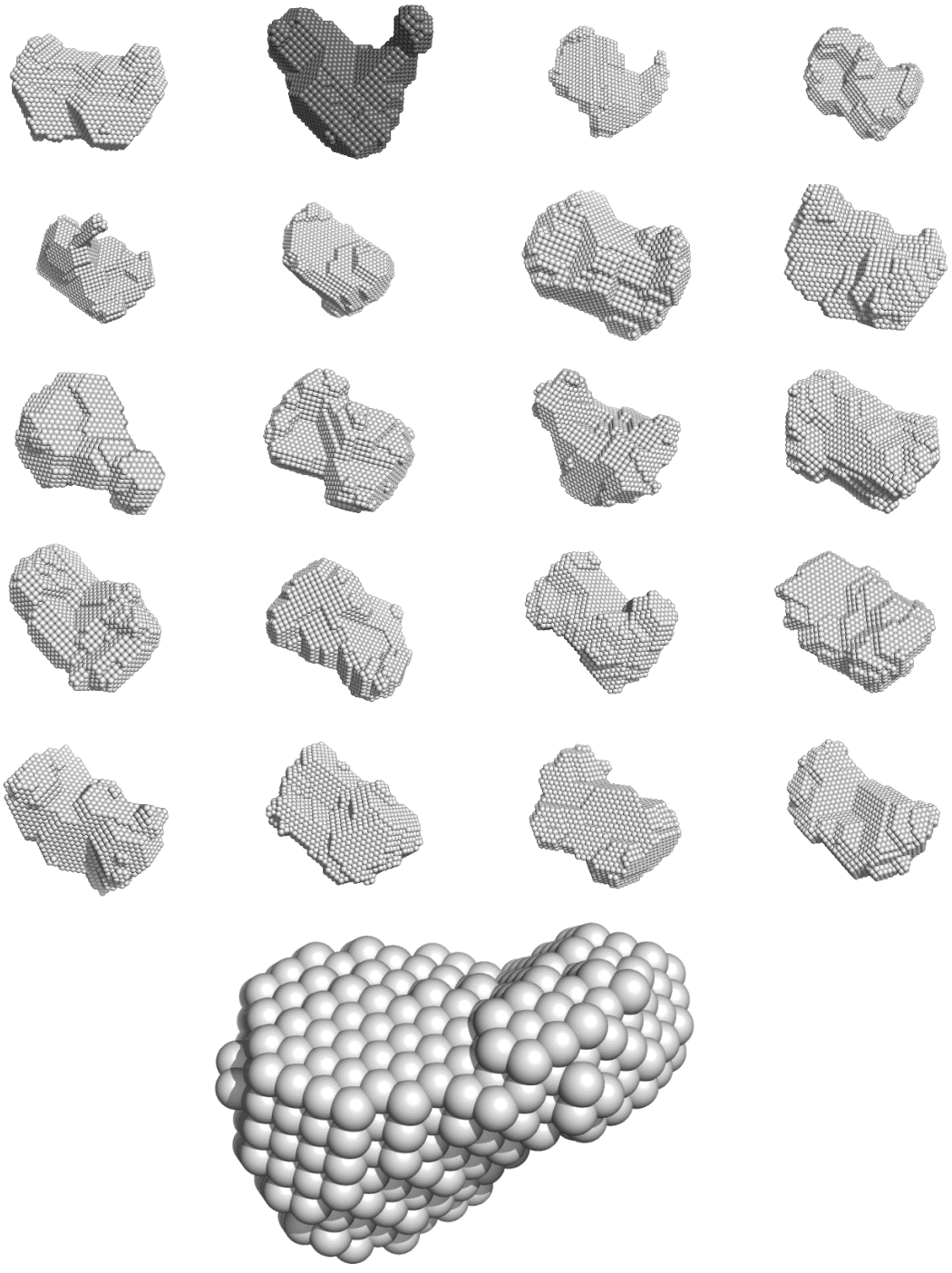


Figure 4.8. *Ab initio* SAXS models of the UGGT chimera in complex with Sep16.5. In dark grey is the model excluded from the averaging and model filtering to produce the DAMFILT model (bottom). The models are derived from the highest concentration sample of 15.44 mg/ml with the D_{max} set to 16.32 nm^{-1}. Exclusion criteria: discard models with an NSD greater than the mean, plus 2x the standard deviation of the models.

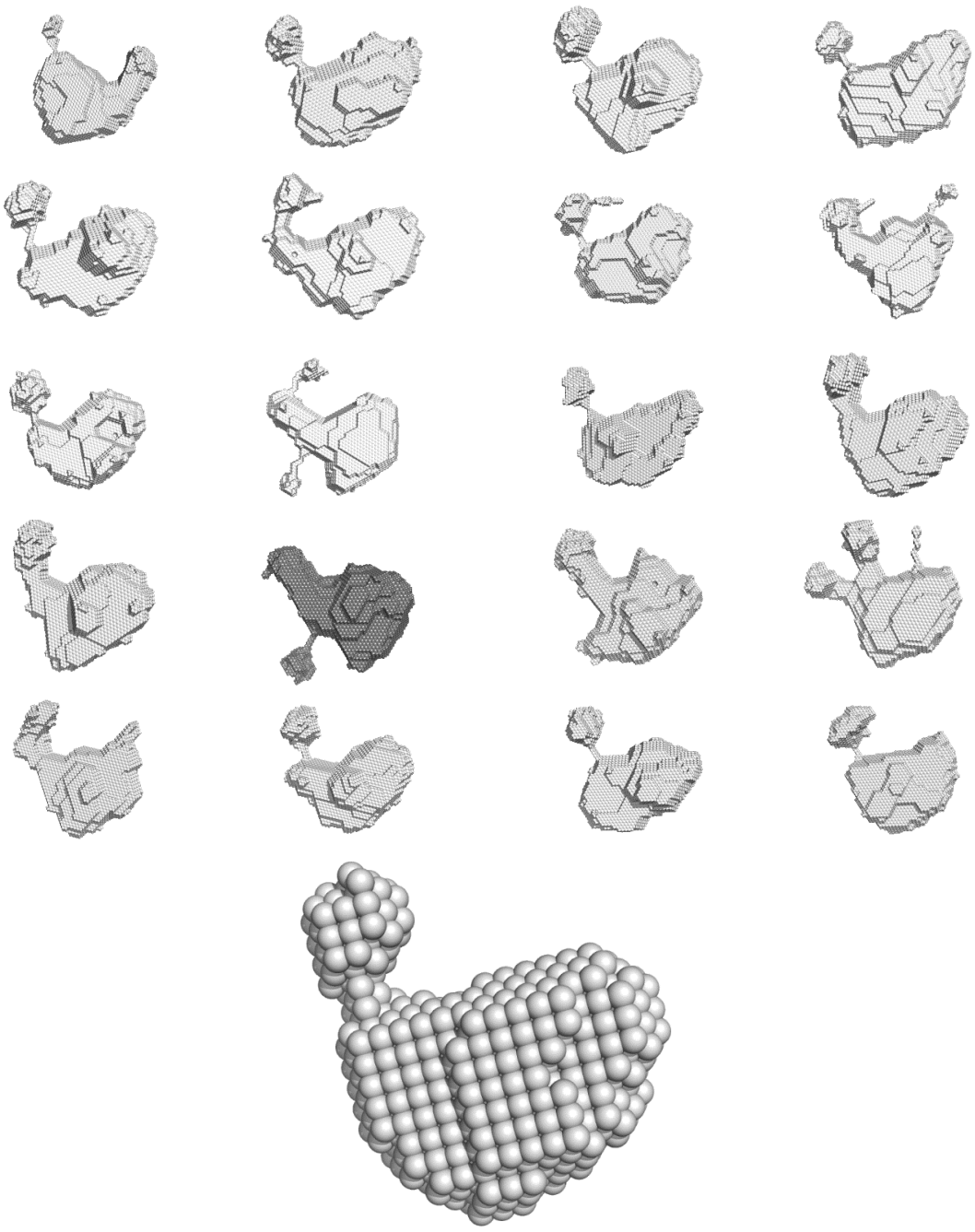


Figure 4.9. *Ab initio* SAXS models of the UGGT chimera in complex with MBP-Sep15. In dark grey is the model excluded from the averaging and model filtering to produce the DAMFILT model (bottom). The models are derived from the highest concentration sample of 21.4 mg/ml with the D_{max} set to 18.3 nm^{-1}. Exclusion criteria: discard models with an NSD greater than the mean, plus 2x the standard deviation of the models.

The models of the UGGT chimera show that, in the large part, the protein is likely to be globular with a small protrusion. Alignment of the crystal structure (PDB ID 5MZO) and the bead model obtained by DAMFILT, with the SasPy plugin SUPALM, confirms good agreement between the models and leads to a normalised spacial discrepancy (NSD) of 3.2598, at a resolution of 48 ± 4 Å. The small protrusion is likely to be the location of the TRXL2 domain highlighted as flexible by the crystal structure[68]. There are pockets within the SAXS envelope into which loops that were disordered in the crystal structure, could fit.

As the SAXS curves of the chimera and the complex with Sep16.5 are so similar, it is no surprise that the output models and the DAMFILT model of the complex appear similar to that of the chimera. SUPALM alignment of the DAMFILT model with 5MZO leads to a NSD of 4.1036, highlighting that the extra volume found in this sample leads to a worse fit. Interestingly, when the *Ct* UGGT crystal is fit to the SAXS model, like that in Figure 4.10, a large unexplained volume is present on the face where the Sep15 binding loop is located: nestled between the GT domain, β-sandwich 1 and TRXL4.

The 20 models of the complex of the UGGT chimera with MBP-Sep15 are all noticeably different to those of the previous two samples. On each of the models, there is a protrusion from the main, globular part of the SAXS envelope. The DAMFILT model would usually exclude groups of atoms attached by a thin linker, however, in this case it appears that the signal is strong enough to warrant its retention in this model. As MBP is the only part of this complex that is not present in the two earlier samples, it follows that this loosely-connected density is likely the MBP part of the MBP-Sep15 fusion protein.

4.7.3. Model of the UGGT chimera in complex with MBP-Sep15

Of the 20 models that fit the experimental data well, one was chosen to which to fit a model of the UGGT chimera in complex with MBP-Sep15. Despite the resolution estimated at 42 ± 3 Å, the SUPALM alignment of *Ct* UGGT (PDB ID: 5N2J) to model 3 (Figure 4.10) fits well and even describes the bumps that are observed in the SAXS envelope. There is, however, a lot of volume leftover. A model of $Sep15_{Cterm}$, obtained from an NMR ensemble (PDB 2A4H) can be fit close to the where the Sep15 binding loop is predicted to reside. There was not enough space to fit MBP within the main volume within the envelope. The protrusion from the main envelope, on the other hand, fits the MBP well, indicating that it is likely to be loosely associated with the UGGT chimera: Sep15 complex.

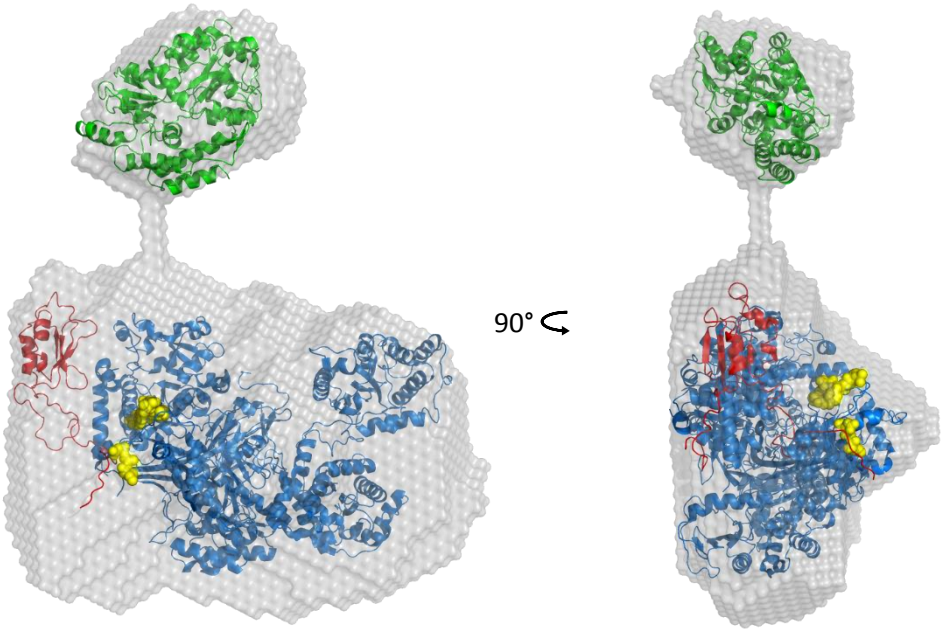


Figure 4.10. Fit of *Ct* UGGT in blue (PDB ID 5N2J), Sep15 in red (PDB ID 2A4H) and MBP in green (PDB ID 1ANF) in a SAXS *ab* initio model of the UGGT chimera: MBP-Sep15 complex. In yellow spheres, the defined ends of the Sep15 binding loop.

4.7.4. Analysis of the conformation of UGGT in solution

The previous crystal structures and cryo-electron microscopy work on *Ct* UGGT have been unable to give information on the conformation in solution: the former because of the need to pack in an ordered fashion and the latter due to the flexible TRXL2 domain

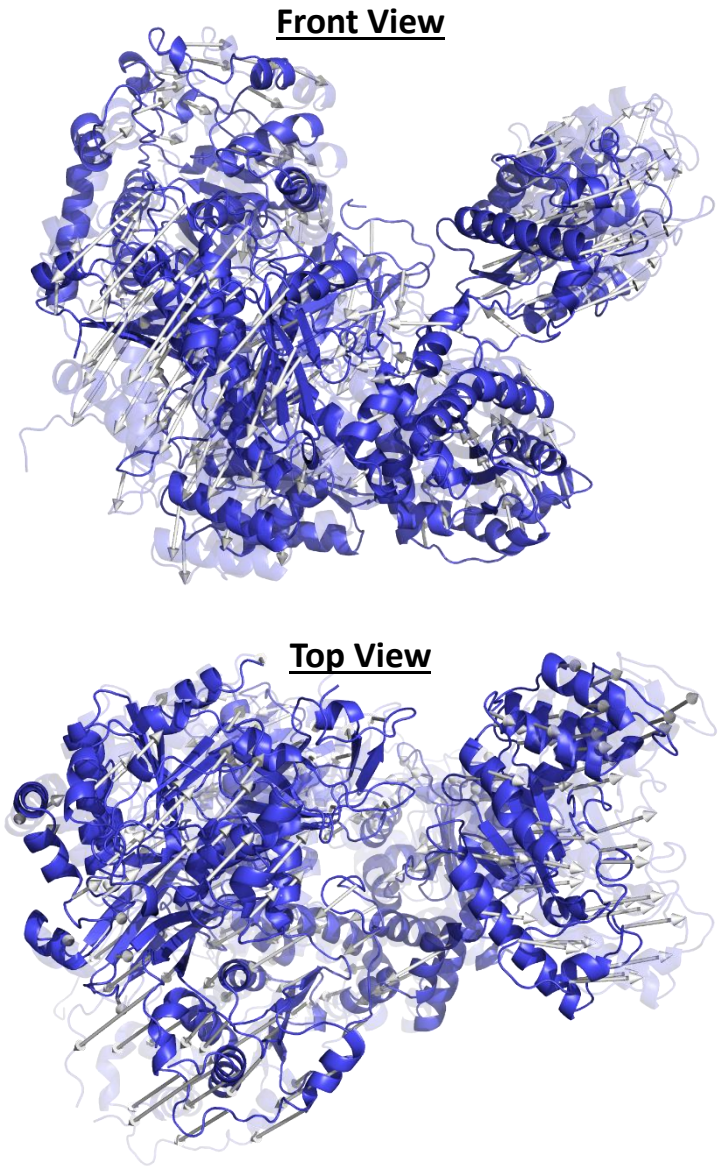


Figure 4.11. SREFLEX analysis of the UGGT chimera SAXS data. In blue, the conformation of UGGT (PDB ID: 5N2J) that fit the data best. In transparent blue, domain shifts to best fit the SAXS data.

not being defined. The structural data collected here was collected at 21°C in a buffer at physiological pH and close to physiological salt concentrations. The programme SREFLEX allows us to fit a protein model (provided by PDB ID 5N2J), to the experimental data, with none of the *ab initio* models required.

Shown in figure 4.11 are the expected domain movements required to best fit the experimental data from the UGGT chimera alone. One of the major movements described by SREFLEX is a large, outwards motion of TRXL4. However, this is also the expected position of the Sep15 binding loop, an undefined volume of around 50 amino acids. It is therefore possible that this domain movement was compensating for the missing residues in the input model. The other domain movement of note is the predicted "swinging out" of TRXL2 to a more open conformation than observed in any of the previous crystal structures. TRXL2 is the most mobile of the UGGT domains, and possibly home to the misfold sensor of the protein; thus it is no surprise that this is where the SAXS data suggest the greatest motion[68].

4.8. Discussion

The overall aims of working on UGGT from a structural perspective are to provide a molecular explanation for its activity and identify functionally relevant sites. An understanding of how this enzyme works at a mechanistic level is desired before achieving the long term goals of discovering UGGT inhibitors and assessing whether they could, like iminosugar inhibition of glu I/II, be viable against viral infections or could ameliorate protein-folding disorders. Previous work within the laboratory yielded the

first full-length structure of UGGT in a number of conformations, highlighting its apparent flexibility.

The 2.78 Å structure of the double cysteine mutant of *Ct* UGGT described here considerably improves the resolution of the full-length UGGT structures previously available. However, even at this resolution, the crystals would be considered generally unsuitable for fragment-based drug discovery, as highly reproducible and better diffracting crystals would be desired[299]. Nonetheless, this is a good start.

An alternative approach has been exemplified by Koichi Kato *et al*[73], who have managed to get 1.35 – 1.4 Å resolution structures of the catalytic GT domain alone from *Thermomyces dupontii*. Such crystals may be a better starting point for fragment based drug discovery although many limitations remain. The relevance of ligands of a fungal UGGT to that of the human homologues must be taken into account, as must the fact that inhibitors of the GT domain could also target one of the many other glycosyltransferases encoded within the genome. As such, targeting other areas unique to UGGT, such as the misfold recognition regions in the N-terminal portion of UGGT, may be a better inhibitory strategy. The work presented here, with the first successful production of recombinant human UGGT1, provides a good basis for further study.

Along similar lines, the site at which the protein Sep15 interacts with UGGT was recently revealed[296]. Sep15 binds to both the UGGT1 and UGGT2 isoforms, the former with nanomolar affinity, and enhances their activity[64, 77, 78]. A point that has often been brought up is that the full inhibition of UGGT may open up an "ER Pandora's box", unleashing glycoproteins that needed to reside in the ER for a greater length of time in order to attain their proper fold: the abrogation of the only known N-glycosylated protein folding

checkpoint may induce abnormal secretion of imperfectly folded complexes and/or slowly folding glycoproteins. For example, it has been shown that correct loading of antigens onto MHC class I depends on UGGT, as does the proper assembling of both MHC and the T-cell receptor[300-302]. However, viral glycoproteins have been shown to be much more sensitive to the alteration of host-cell ER machinery than MHC Class I, indicating that there would be a therapeutic window for potential drugs without affecting crucial host processes[303].

Targeting the interface between UGGT and Sep15 could therefore focus on a unique site among glycosyltransferases as well as avoid interference with the folding of some of UGGT's clients. Modulation of UGGT's activity against a subset of substrates only represents a form of partial inhibition which may circumvent the potential side effects observed from full inhibition of the glucosyltransferase activity. In this respect, a crystal structure of the complex of UGGT and Sep15 would help identification of molecules that could disrupt the interface as well as further elucidate UGGT's role within the cell. *Hs* UGGT1 expressed sufficiently well to attempt crystallisation although the structure is likely too flexible without engineered disulphide bonds that would help restrict its movement. Alignment of the *Ct* UGGT and *Hs* UGGT1 sequences showed that the Sep15 binding loop is not conserved in lower eukaryotes, nor could we get the complex between *Ct* UGGT and *Hs* Sep15 to form *in vitro*. The production of the chimeric UGGT allowed us to obtain samples of complexes of purity and quantity suitable for structural studies.

The chimera only yielded crystals in the non-complexed form. Protein crystallisation is a form of purification, excluding solution molecules from the crystal lattice. The lack of crystal formation of the UGGT chimera: Sep15 samples is additional indirect evidence of

how strongly the two proteins interact, as the complex alone could not disassociate to nucleate the UGGT chimera crystals that formed in the same conditions in the absence of Sep15 constructs.

As well as sparse crystallisation screens, we tried an optimisation screen around the condition that yielded a *Ct* UGGT crystal form presenting a large solvent space surrounding the Sep15 binding loop. However, the original crystal form could not be reproduced (even with *Ct* UGGT). The UGGT chimera crystals that did form were of a different polymorph whose crystal packing only leaves a small solvent volume around the Sep15 binding loop. Not surprisingly, no ordered density for Sep15 was found in the crystals even after soaking.

The three UGGT chimera containing samples were instead devoted to small angle x-ray scattering, and good data were obtained for all. The resolution was limited to around 40 Å, at which the chimera alone appeared globular. Flexing a crystal structure to best fit the SAXS data suggested that the TRXL2 domain may be more flexible than all of the crystal structures had indicated. It is likely that this intrinsic flexibility contributes to UGGT's ability to recognise a diverse array of misfolded glycoproteins of different shapes and sizes.

The SAXS envelopes of the UGGT chimera with the two Sep15 constructs inform about the likely Sep15 position in the complex. Analysis of the UGGT: Sep16.5 SAXS data produces *ab initio* models that are very similar in size and shape to the chimera alone. It is likely, therefore, that Sep15 packs in tightly to UGGT, as we see no evidence for a large, loosely connected domain. The UGGT: MBP-Sep15 SAXS derived models confirm this observation. Superposition of a UGGT crystal structure into the UGGT: MBP-Sep15 SAXS

leaves unexplained volume close to the Sep15 binding loop and the glucosyltransferase domain: it is here that Sep15 is likely to bind. In the light of these data, crystallisation of the complex in one of the existing *Ct* UGGT crystal forms is unlikely, as this region is engaged in crystal contacts in all of them. Clear density in most of the *ab initio* models shows a loosely connected volume that is well fitted in size and shape by MBP connected to Sep15 by a long (cleavable) linker.

This work provides a strong basis for further studies of the UGGT: Sep15 complex and for the eventual discovery of interface disruptors. Based on this work, aiming for existing crystal forms of *Ct* UGGT is likely to be fruitless, so a new emphasis must be on co-crystallising the ready-formed complexes. For higher resolution crystal structures to be used for fragment-based drug discovery, I believe the best strategy is to focus on crystallising isolated domains of interest in the hope that higher resolution and more reproducible crystals can be obtained.

5. ENTPD5

The ER-UDPase, ectonucleoside triphosphate diphosphohydrolase 5 (ENTPD5), resides in the ER as a soluble enzyme that converts UDP, the product of the reglucosylation reaction that UGGT catalyses, to UMP[92]. UMP can escape the ER in exchange for a molecule of UDP-glucose through an ER resident UDP-glucose transporter, thereby allowing for the replenishment of the glucose donor for UGGT's activity and completing the ER UDP-glucose cycle[117]. How ENTPD5 remains in the ER is not yet understood, as

the only targeting sequence that could be identified so far for this enzyme is its signal peptide, which directs towards the ER yet does not mean it is retained there.

The role of ENTPD5 in N-glycosylation was confirmed in 2010, where it was shown that knockdown of ENTPD5 lowered both glycosylation levels and cell surface expression of highly-glycosylated growth factor receptors: expression of wild-type ENTPD5 in these knockdown cell-lines restored their original phenotypes[92, 304]. ENTPD5 also plays a crucial role in the Warburg effect, a process of aerobic glycolysis which accompanies the metabolic changes in many malignant cancers[92, 100, 305]. High-throughput screening found a number of candidate inhibitors of the enzyme which also induced ER stress, as would be expected for modulators of activity of an enzyme involved in protein folding.

However, the hypothesis that inhibition of ENTPD5 may lead to cellular resistance to viral infections, through altering the protein folding environment within the ER, has never been investigated. Described in this chapter is the work towards characterisation of the activity of ENTPD5 and its homologues and the work towards both its structural characterisation and towards its assessment as an antiviral target using known ENTPD5 inhibitors.

5.1. Design of constructs

Expression of ENTPD5 from a variety of organisms has been successfully carried out on a number of occasions from both mammalian and bacterial expression systems[94, 95, 99]. Whilst the human ENTPD5 would be the most relevant for studies to investigate the antiviral therapeutic potential of inhibitors, orthologues from another organism may be more amenable to expression or crystallisation.

The crystal structure of a related phosphatase, *Rattus norvegicus* (*Rn*) ENTPD1, has been determined[306] and was used to help design the boundaries for ENTPD5 constructs. Human and murine forms of the protein were chosen as they were sufficiently distinct to make a difference for crystallisation attempts, whilst still annotated as having the same function *in vivo* (UniProt codes O75356 and Q9WUZ9 respectively). The sequence alignment, carried out using Clustal Omega[307] and eSPript 3.0[308] with secondary structure depictions obtained from PDB 3ZX3, can be found in the appendix. For both *Hs* and *Mm* constructs, mature sequences (lacking the signal peptide as predicted by SignalP 4.1[309]) were ordered as a GeneString (Thermo Fisher Scientific), with codons optimised for expression in mammalian cell lines or *E.coli*. Based on the sequence alignment to 3ZX3, Gibson assembly primers were designed to truncate the protein so that the N-terminus of the sequence starts from ^{44}SASTLY, with the expectation that this will cover the full fold of the protein and will provide a further crystallisation target. Ordered sequences and primers can be found in the appendix.

A homology model of *Hs* ENTPD5 was generated using the FFAS servers[310] and MODELLER[311] based on the coordinates of *Rn*ENTPD2 (PDB ID 4BQZ), another ENTPD family member of known structure[312]. Alignment of the *Hs* ENTPD5 homology model against 4BQZ can be seen in figure 5.1. Whilst the active site residues of 4BQZ accommodate the nucleoside triphosphate well (Figure 5.1.A), the homology model indicates that the same nucleotide would not fit in the active site of ENTPD5 in the same pose due to steric hindrance with the *Hs* ENTPD5 tyrosine^{331} (Figure 5.1.B). A nucleotide with fewer phosphates is a more probable substrate (Figure 5.1.C). The residues of ENTPD5 most likely to interact directly with the substrate consist of ^{54}DAGSTGTR^{61},

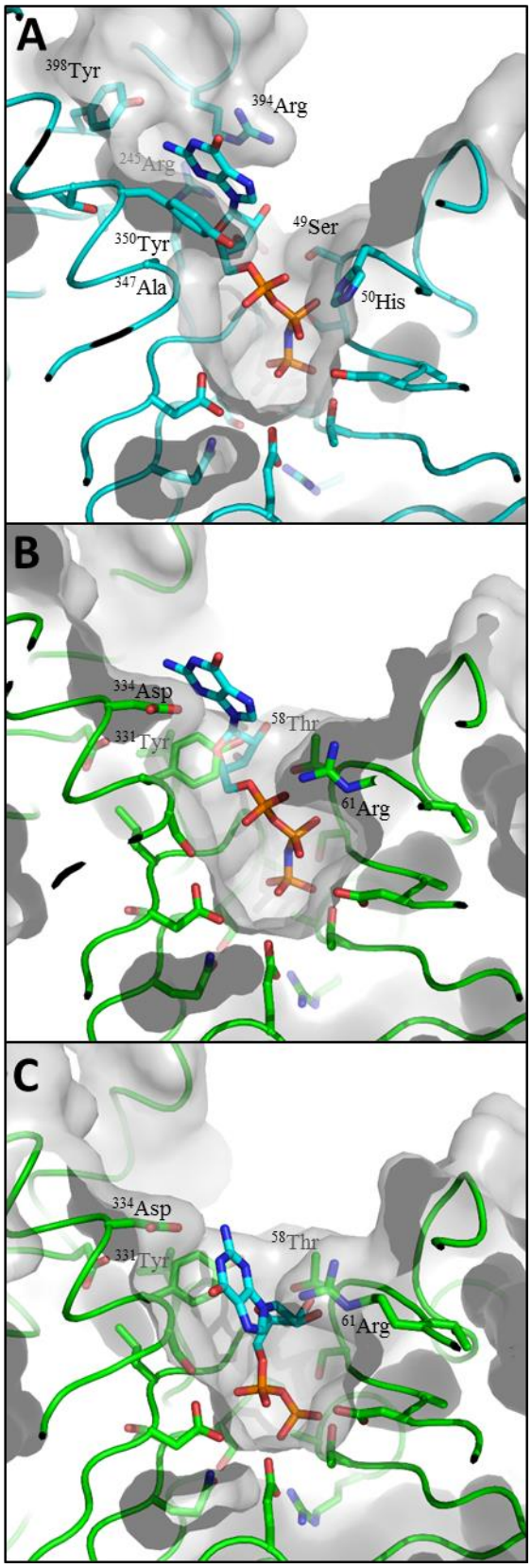


Figure 5.1. Panel A shows the active site of *Rn*ENTPD2, PDB 4BQZ in complex with the nucleotide analogue GMPPNP. Panel B shows the homology model of *Hs*ENTPD5 with the GMPPNP ligand of 4BQZ still present in the same orientation. Panel C shows how a diphosphate nucleotide could fit in the active site of ENTPD5. The conformation of the ligand was manually altered to produce the fit.

^{130}TAG^{132}, R^{134}, E^{172} (catalytic residue), ^{199}DLGGASTQ^{206}, L^{247}, ^{330}SY^{331}, D^{334} and D^{379}. ConSurf[313] was used to determine the most conserved residues of ENTPD5 genes originating from a range of species and found that all of these residues were conserved.

Within many labs including ours, the use of protein homologues from eukaryotic thermophilic fungi has proven useful in the procurement of crystal structures[68, 314, 315]. To this end, we searched the *Chaetomium thermophilum* genome for an ENTPD homologue using NCBI blast. The closest match is the membrane-bound, Golgi-located *Ct* ENTPD6, a fairly under-characterised member of the ENTPD family, which shares a similar substrate specificity towards nucleoside diphosphates[114, 316]. Alignment of *Hs* ENTPD5 and *Ct* ENTPD6 reveals the predicted active site residues are fully conserved, indicating that the substrate specificities are likely very similar.

Due to its potential applicability towards the characterisation of *Hs* ENTPD5, *Ct* ENTPD6 constructs were designed (see appendices). The predicted trans-membrane-helix and cytosolic portion of the protein were excluded, leaving just the soluble portion of the protein, residues 104 to 564.

5.2. Cloning of DNA constructs

The pET/M-33 and pET/M-43 vectors were linearised with NcoI-HF and XhoI, pET303 with XbaI and XhoI, and pHLsec with KpnI-HF and AgeI-HF. Linear inserts were either ordered directly as GeneStrings (Thermo Fisher Scientific) or produced by PCR using the primers for either full-length or truncated constructs (see Appendix). Gibson assembly was used to build all of the combinations of DNA vectors and constructs shown in table 5.1.

Successful cloning was confirmed by sequencing to ensure insertion occurred where expected and the inserts were still in frame.

Organism	Vector	Full-length?	Truncated
Mus musculus	pET303	✓	✓
	pET/M-43	✓	✓
	pET/M-33	✓	✓
	pHLsec	✓	✓
Homo sapiens	pET303	✓	✓
	pET/M-43	✓	✓
	pET/M-33	✓	✓
	pHLsec	✓	✓
Chaetomium thermophilum	pET303	X	X
	pET/M-43	X	✓
	pET/M-33	X	X
	pHLsec	X	✓

Table 5.1. DNA constructs of *Mm* and *Hs* ENTPD5 and *Ct* ENTPD6

ENTPD6 amplified from *Ct* cDNA (a gift from Ed Hurt, Heidelberg) was initially cloned into pET/M-43. A non-coding sequence of DNA, likely an un-spliced intron, was still present in the amplified gene, which would lead to early termination of the protein. Site-directed mutagenesis was carried out using the primers 296_E6_insert_fwd and 297_E6_insert_rev (see Appendix) to remove the intron. Gibson assembly was then used to move the repaired gene into pHLsec.

5.3. Expression Trials

Of the mammalian constructs of ENTPD5 that were tested, only the fusions with either glutathione S-transferase or maltose binding protein (MBP) led to soluble protein. Truncated constructs (based on the structures of closely related family members) led to

How the expression went

Organism	Vector	Full-length or truncated	Soluble Expression Level
Human	pET/M-33	Full-length	++
	pET/M-43	Full-length	+
	pET303	Full-length	-
	pHLsec	Full-length	-
	pET/M-33	Truncated	++
	pET/M-43	Truncated	+++++
	pET303	Truncated	-
	pHLsec	Truncated	-
Mouse	pET/M-33	Full-length	+
	pET/M-43	Full-length	+
	pET303	Full-length	-
	pHLsec	Full-length	-
	pET/M-33	Truncated	+++
	pET/M-43	Truncated	++++
	pET303	Truncated	-
	pHLsec	Truncated	-
Ct	pET/M-43	Truncated	-
	pHLsec	Truncated	+++++

Table 5.2. Expression levels of the different constructs. pET vectors were expressed in *E.coli*, pHLsec in HEK293F cells.

higher expression levels than the full-length protein. Truncated MBP - *Hs* ENTPD5*, originating from the pET/M-43 vector, led to the highest expression levels and was selected for further expression and characterisation. The lowest tested concentration of IPTG (0.05 mM) for induction of protein expression yielded the highest levels of protein. Dropping the temperature after induction of protein to 16°C and letting expression continue for a further 20 hours before harvesting further improved the protein expression yield.

The UDPase homologue from *Chaetomium thermophilum* achieved high yields in the mammalian expression system, yet did not express in *E.coli* as an MBP fusion.

* MBP-*Hs* ENTPD5 will be referred to simply as MBP-ENTPD5

5.4. Purification, refolding and methylation

5.4.1. MBP- *Hs* ENTPD5

The MBP fusion of *Hs* ENTPD5 was purified from an *E.coli* cell lysate by purification on a 5 ml MBPTrap column and eluted with a buffer containing 10 mM maltose. Flow-through from the first run was collected and subsequently run through a 5 ml MBPTrap column re-equilibrated into binding buffer, as the column was overloaded with protein on the first run. This process was repeated until no more elution of protein was observed. As the protein is predicted to contain two disulphide bonds that are unlikely to form in an *E.coli* expression system, disulphide shuffling and refolding was performed, as described in chapter 2.4.1. Protein was concentrated to approximately 10 mg/ml prior to the refolding procedure. Refolded protein was purified on a 5 ml HisTrap column followed by SEC on a Superdex 200 16/60 column. Protein was concentrated to between 3 and 10 mg/ml, flash-frozen in liquid nitrogen as 100 µl aliquots, and stored at -80°C.

Methylation of MBP-ENTPD5 was carried out as in 2.4.2. Methylated protein was purified on a Superdex 200 16/600 SEC column, fractions containing protein were pooled and concentrated to 4.86 mg/ml and flash-frozen for storage at -80°C.

Removal of the MBP-tag was achieved by 3C-cleavage (2.4.3). Protein was run on a 1 ml HisTrap™ affinity column so that cleaved *Hs* ENTPD5 would bind whereas the MBP would flow through. The elution fractions were pooled and run on a Superdex 200 10/300 SEC column; fractions containing cleaved ENTPD5 were pooled, concentrated to between 0.2 to 3.1 mg/ml, flash-frozen and stored at -80°C.

5.4.2. *Ct* ENTPD6

Mammalian cells were used for this expression and therefore no disulphide shuffling nor refolding step was required. Cell supernatant was prepared as described in 2.3.2. and purified on a 5 ml HisTrap column followed by a SEC run on a Superdex 200 16/60 column. Protein was concentrated to between 15 and 70 mg/ml, flash frozen in 100 µl aliquots and stored at -80°C.

Methylation of *Ct* ENTPD6 was carried out as described in 2.4.2. The methylated protein was purified on a Superdex 200 16/600 SEC column and eluted at the same volume as non-methylated protein. Protein was concentrated to 8.4 mg/ml and flash-frozen for storage at -80°C.

Limited proteolysis of *Ct*ENTPD6 was attempted using trypsin, chymotrypsin, GluC endoproteinase (GluC), proteinase K and a protease from *Bacillus sp.* (see 2.4.3). Only proteolysis with GluC led to a stable single fragment, approximately 2 kDa smaller than the non-proteolysed fragment, when using a ratio of 1 part GluC to 20 parts *Ct* ENTPD6. As GluC self-inactivates, fresh GluC was added to a new batch of protein every 12 hours for 48 hours to ensure that *Ct* ENTPD6 was fully cleaved. Cleaved protein was run through a Superdex 75 16/600 SEC column followed by a MonoQ column. Protein eluted from the latter column in 2 peaks, as can be seen in figure 5.2. The second peak contains *Ct* ENTPD6 that does not appear to be proteolysed and behaves as usual, as a multimer; as seen on the non-reducing gel (figure 5.2) a disulphide bonded dimeric species is seen in the second peak, as well as the monomer (fractions B9 – C5). The earlier peak, represented by fractions B.3. to B.8., contains bands at a smaller molecular weight

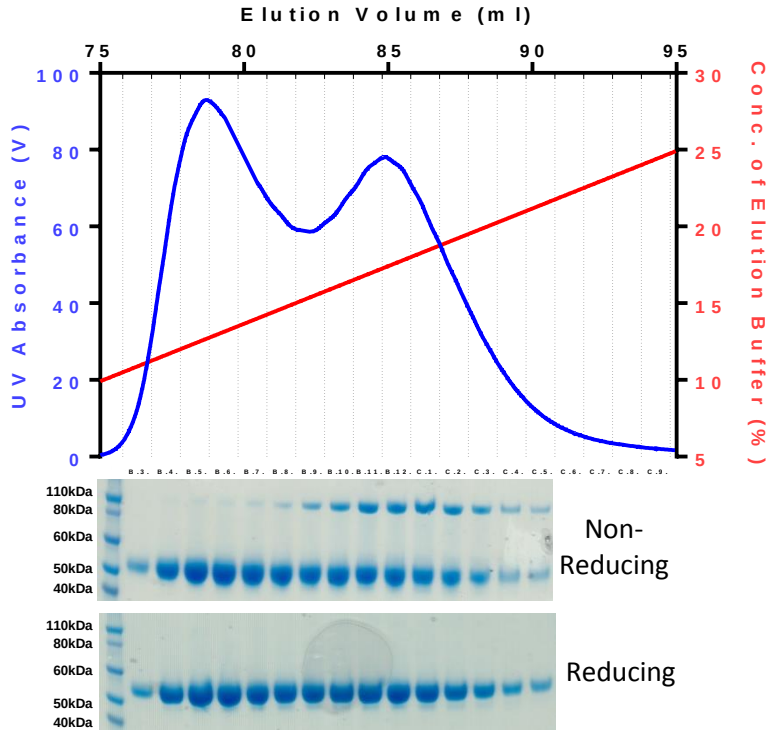


Figure 5.2. Upper panel, blue curve: elution profile of proteolysed *Ct* ENTPD6 from a MonoQ column of GluC-digested *Ct*E6 from a MonoQ 5/150 anion exchange column, with corresponding SDS-PAGE gels; red line: molarity of salt elution buffer. Bottom panel: SDS-PAGE analysis of the Mono Q elution fractions.

contains monomer only, suggesting that the proteolysis gave rise to a monomeric form of *Ct* ENTPD6. Fractions represented by lanes 1-5 were pooled, concentrated to 15.1 mg/ml and flash-frozen for storage at -80°C.

5.5. Characterisation and activity assays

5.5.1. SEC-MALS

Before refolding of the MBP fusion of *Hs* ENTPD5, the purified protein was sent for SEC-MALS analysis. The results showed that most of the protein eluted in, or near, the void volume of the Superdex 200 10/300 SEC column used for separation, indicating that most

of this soluble protein was probably eluting as multimers. In the same run, a very broad peak was observed next to the void, of molecular weights between 1 and 10 MDa, reflecting even larger aggregates. Finally, a small peak was observed that corresponded to around 100 kDa. The predicted molecular weight of the fusion was 85.8 kDa, which indicates that this peak likely corresponds to the monomer. Following the refolding procedure, only the monomer peak remained, indicating that the multimers and aggregates were removed or refolded during the process.

Pure *Ct* ENTPD6 elutes from the SEC column as a single peak, which was sent for SEC-MALS analysis. The results estimate that the molecular weight of this molecule is approximately 147 kDa. *Ct* ENTPD6 from this expression system is predicted to be 51.6 kDa, suggesting that this construct of *Ct* ENTPD6 behaves as a trimer in solution. When taking into account the results of figure 5.2., where an SDS-resistant band corresponding to a dimer can be observed, it could be postulated that it exists as an asymmetric trimer where a disulphide bond linked dimer may be associating with a further monomer.

5.5.2. Dynamic scanning fluorimetry

Dynamic scanning fluorimetry (DSF) was carried out to determine in what conditions the *Hs* ENTPD5 and *Ct* ENTPD6 were most stable. A titration of pH against melting temperature (T_M) of the protein was carried out. A buffer system of a 2:3:3 ratio of malonic acid: imidazole: boric acid[317] was used to span the entire pH range with the same buffer chemistry throughout. Figure 5.3 shows the titration of a pH 4.0 buffer against a pH 10.0 buffer, where the pH follows the ideal profile closely but not exactly. For the purposes of the DSF experiment, this is inconsequential as we can use the measured pH. Figure 5.4.A

shows how the pH affects the stability of both *Hs* ENTPD5 and *Ct* ENTPD6. For both enzymes, the protein is most stable at a pH of between 6.5 and 7.2, a physiological pH likely to be found in the secretory system between the ER and cis-Golgi apparatus[318].

Part B of the same figure shows how salt concentration affects the stability of the protein. For *Ct* ENTPD6, the protein appears most stable with between 50 and 100 mM NaCl, whereas *Hs* ENTPD5 shows greatest stability at the highest concentrations of salt tested (500 mM). Finally, part C of the figure shows how different additives affect protein stability. The *Hs* ENTPD5 has already been described as requiring divalent cations for its enzymatic activity. Addition of excess (100 mM) cations has a destabilising effect on both enzymes tested whilst 10 mM EDTA helps to stabilise the protein. Whilst cations are important for activity, conditions in which the protein is most stable increases the likelihood of crystallisation, typically because of a lower entropic cost upon moving from

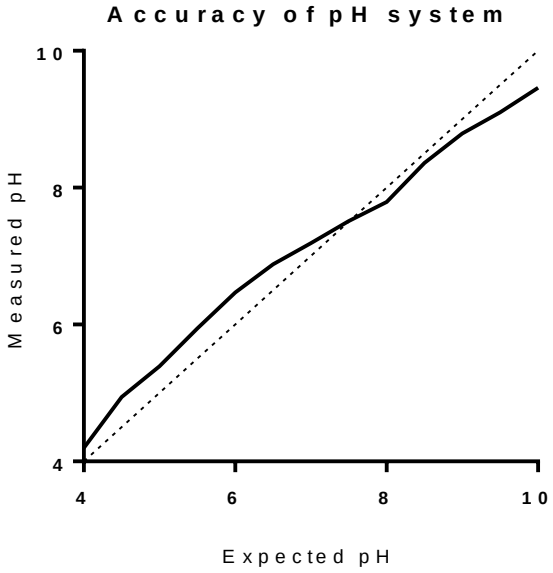


Figure 5.3. Expected pH (dotted line) vs measured pH (solid line) of a 2:3:3 molar ratio of malonic acid: imidazole: boric acid. The expected pH is if mixing a pH 4 and 10 solution in different quantities led to a linear relationship in the resultant pH.

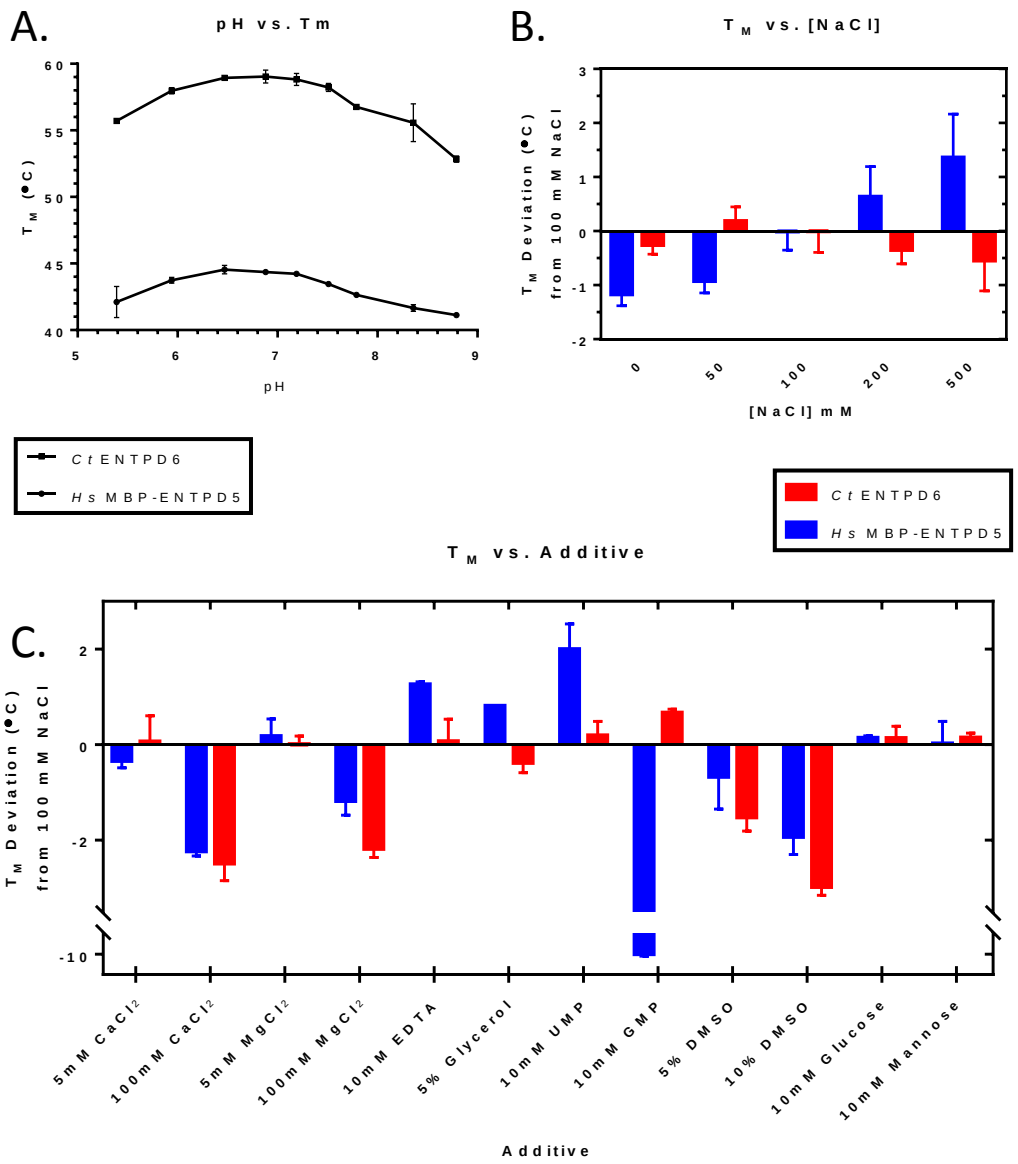


Figure 5.4. DSF experiments to determine conditions in which protein constructs are most stable. A. T_M of *Ct* ENTPD6 was measured in a 50 mM 2:3:3 malonic acid: imidazole: boric acid buffer system, with 100 mM NaCl at 0.1 mg/ml. T_M of MBP–ENTPD5 was measured in a 20 mM 2:3:3 malonic acid: imidazole: boric acid buffer system, with 150 mM NaCl at 0.025 mg/ml of protein. B. T_M of *Ct* ENTPD6 was measured in a 50 mM HEPES buffer system, pH 7.2 at 0.1 mg/ml and graph centred on 58.08°C. T_M of MBP-ENTPD5 was measured in a 20 mM HEPES buffer system, pH 7.5, at 0.025 mg/ml of protein and centred on 43.5°C. C. T_M of *Ct* ENTPD6 was measured in a 50 mM HEPES buffer system, 100 mM NaCl, pH 7.2 at 0.1 mg/ml and graph centred on 58.08°C. T_M of MBP–ENTPD5 was measured in a 20 mM HEPES buffer system, 150 mM NaCl, pH 7.5, at 0.025 mg/ml of protein and centred on 43.5°C. Buffers were made up with the concentration of additives shown in the figure.

the solution to the crystal lattice. Uridine 5'-monophosphate is the likely product of the reaction of both enzymes; its addition to *Hs* ENTPD5 has a strong stabilising effect, whilst it has little effect on the *Ct* homologue. Conversely, guanidine 5'-monophosphate stabilises the *Ct* ENTPD6 but strongly destabilises the *Hs* ENTPD5. DMSO is not tolerated well by either enzyme whilst glucose or mannose have little or no effect on them. These data can help inform what conditions may be best for enzymatic studies as well as crystallisation.

5.5.3. Activity assay optimisation

The MESG (7-methyl 6-thioguanosine) assay was used to measure activity of *Hs* ENTPD5 and *Ct* ENTPD6 as it allows for the continuous collection of data and therefore measurement of rates of reaction in the linear portion of the reaction curves; this gives the

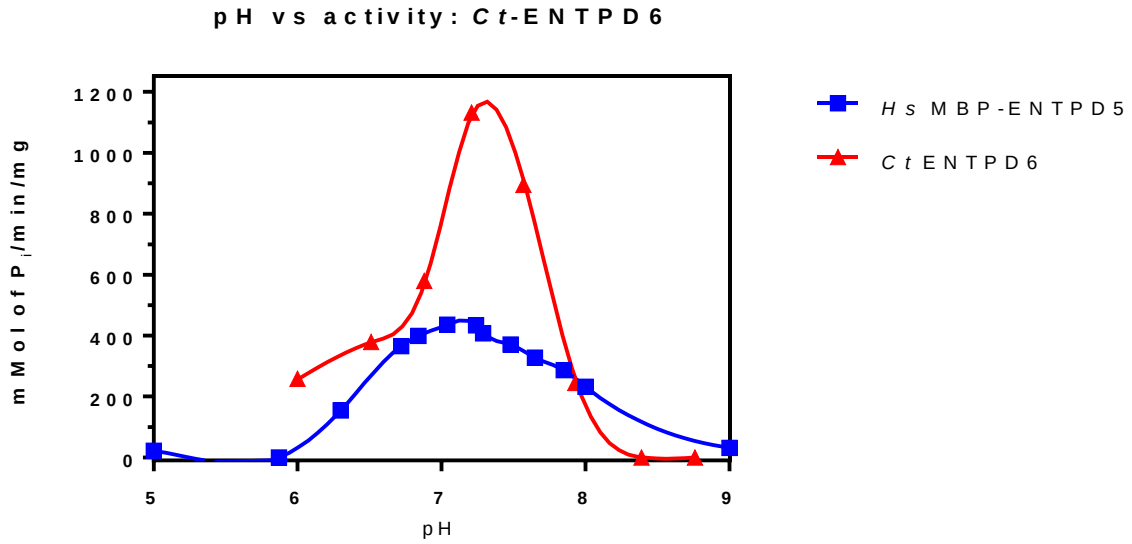


Figure 5.5. Activity in the MESG assay of the two different enzymes shown, as the pH (of 0.1M HEPES buffer) is varied, using 10 mM GDP as the substrate. Assays were run at 37°C.

MESG assay some immediate advantages over other well-known endpoint assays that measure free inorganic phosphate only once, such as the malachite green assay[319].

In preparation for the activity assays, the effect of pH on the activity of the two enzymes was determined; points were joined using a LOWESS regression (Figure 5.5). For both enzymes, the pH optimum falls between 7.0 and 7.2, identical to that of the ER. The maximal rate of reaction of the *Ct* homologue is approximately 2.5x greater than that of the human construct, however, as the latter is an MBP fusion protein, the molar concentration of MBP-ENTPD5 is approximately half that of the *Ct* construct per mg of protein.

A published value for the maximal activity of *Hs* ENTPD5 against GDP was 50000 µMol of P_i/hour/mg of enzyme[99]. When compared to the equivalent value 26580 µMol of P_i/hour/mg for the construct tested here, the two values are very similar.

5.5.4. Michaelis-Menten curves

Michaelis-Menten curves can be used to characterise the relationship between an enzyme and its substrates during the quasi steady-state of the reaction. V_{max} represents the maximal rate of reaction of the system that an enzyme could achieve, whereas the Michaelis constant (K_M) is the substrate concentration at which the rate is half V_{max}, which means it can be used as an approximate binding constant.

V_{max}/K_M, the specificity constant, allows comparison of the efficiency of the enzymatic reaction of an enzyme with different substrates. Table 5.3 highlights the different constants obtained from both MBP-ENTPD5 and *Ct* ENTPD6 with a range of different potential nucleotide substrates. By determining Michaelis-Menten curves, we can

compare the activity of the two enzymes to one another and observe whether they have similar substrate profiles.

Nucleotide	*Hs*ENTPD5			*Ct* ENTPD6		
	K_M (mM)	V_{max} (µMol P_i/min/mg)	V_{max}/K_M	K_M (mM)	V_{max} (µMol P_i/min/mg)	V_{max}/K_M
UTP	21.51	205.6	9.6	4.73	0.1	0.0
UDP	0.86	185.3	215.5	0.87	104.3	120.2
UMP	-	-	-	-	-	-
GTP	22.17	75.8	3.4	1.95	19.5	10.0
GDP	4.12	258.9	62.8	1.04	124.3	119.5
GMP	-	-	-	-	-	-
CTP	-	-	-	1.00	0.4	0.4
CDP	1.76	9.7	5.5	0.79	14.1	17.9
CMP	-	-	-	-	-	-
ATP	28.88	32.4	1.1	2.88	1.0	0.3
ADP	-	-	-	0.71	2.1	3.0
AMP	-	-	-	-	-	-
ITP	24.28	44.5	1.8	1.40	19.8	14.2
IDP	3.83	294.7	76.9	0.15	42.1	279.7
IMP	-	-	-	-	-	-

Table 5.3.This table shows the catalytic constants of MBP-ENTPD5 and *Ct* ENTPD6 with a range of potential substrates. Michaelis-Menten curves were set up to a maximal nucleotide concentration of 20 mM; at higher concentrations, nucleotides acidified the reactions. All curves were fit to a standard Michaelis-Menten curve except for that of GDP against *Ct* ENTPD6, which was fit to substrate inhibition model.

None of the nucleotide-monophosphates were potential substrates for either of the two enzymes, as was expected. For both of the enzymes, the nucleotide diphosphates were the best substrates, with the smallest K_M values and the largest V_{max} values. Interestingly, some nucleotide triphosphates were hydrolysed by both enzymes, albeit at a much lower rate than the nucleotide diphosphates. This is in contrast to what was predicted in the homology models at the start of the chapter, indicating that there is some flexibility in the enzymes to accommodate these.

Figure 5.6 shows the V_{max}/K_M values plotted so that substrate efficiency profiles of the two enzymes can be compared. Clearly, the enzymes are only relatively efficient processing UDP, GDP and IDP, with a much lower efficiency for CDP and very small values for the nucleotide triphosphates.

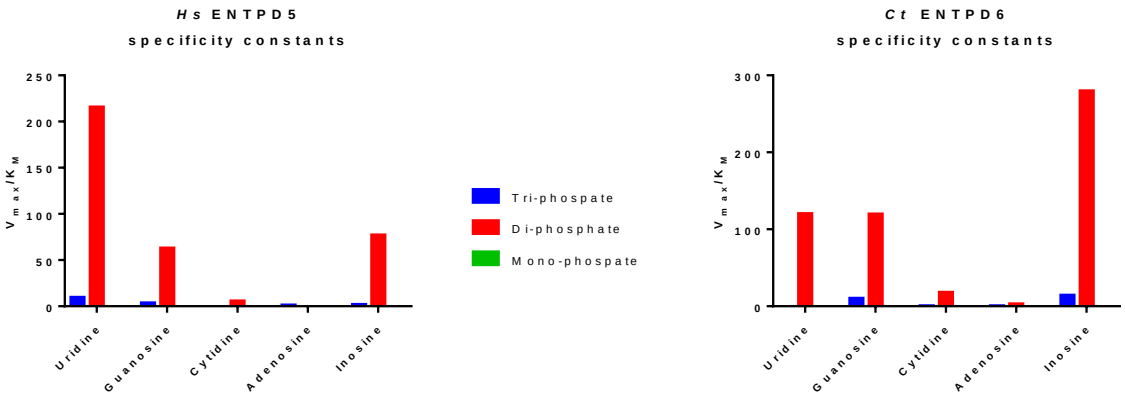


Figure 5.6. Graphs of the specificity constants of *Hs* ENTPD5 and *Ct* ENTPD6 against the range of potential substrates.

Overall, *Ct* ENTPD6 appears to have a very similar substrate profile to *Hs* ENTPD5 and is likely to carry out the same hydrolysis reactions *in vivo*. Within the ER and *cis*-Golgi, both enzymes are likely to encounter UDP and GDP, as by-products of glycosyl-transferase catalysed hydrolysis of nucleotide sugars[320]. In order for these diphosphate nucleotides to leave the ER, they must be hydrolysed to a nucleotide monophosphate. These can then be exchanged for a nucleotide-sugar by one of a large number of nucleotide-sugar antiporters located in the ER and Golgi membranes[116]. Adenosine triphosphate is the other nucleotide that these enzymes are likely to come into contact with in the environment of the ER. However, ATP is used as an energy source by a number of proteins to catalyse their reactions, including by Grp78, a folding chaperone

within the ER[321]. Hydrolysis by either *Hs* ENTPD5 or *Hs* ENTPD6 would be a waste of energy and detrimental to the cell, consistent with both enzymes having evolved no specificity towards it.

5.5.5. *Hs* ENTPD5 E172Q: the catalytic residue mutation

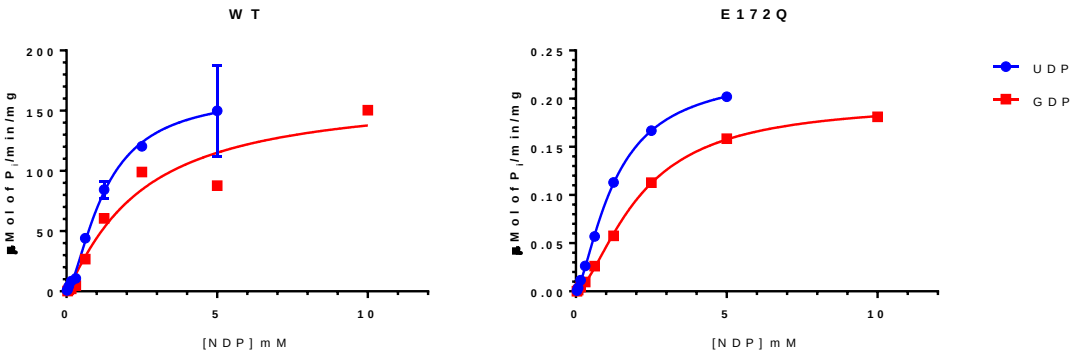


Figure 5.7. Michaelis-Menten curves for the wild-type and E172Q mutants of *Hs* ENTPD5.

Nucleotide	Wild-type *Hs*ENTPD5		*E172Q Hs* ENTPD5	
	K_M (mM)	V_{max} (µMol P_i/min/mg)	K_M (mM)	V_{max} (µMol P_i/min/mg)
UDP	1.26	165.7	1.30	0.230
GDP	2.39	164.3	2.10	0.197

Table 5.4 K_M and V_{max} values for wild-type and the E172Q mutant of *Hs* ENTPD5. 350-fold more protein was used in the reactions for the mutant in order to get a large enough signal to detect.

An inactive mutant of the *Hs* ENTPD5 was cloned and expressed with the aim to stabilise the complex of the protein with its substrate and obtain crystals. Site-directed mutagenesis was successfully carried out to change the codon from GAG to CAG. *Hs* ENTPD5 E172Q was expressed as an MBP-fusion of the truncated *Hs* construct and refolded as before. Michaelis-Menten curves were determined for UDP and GDP, as the main two substrates for the wild-type enzyme (Figure 5.7). Although much less active (note the scale of the y-axis) this 'inactive' mutant clearly retains some catalytic activity.

Table 5.4 lists the catalytic constants derived from these measurements. Whilst the V_{max} of the E172Q mutant is approximately 1000-fold lower than that of the wild-type, the K_M is almost the same. This suggests that E172 does not play a role in binding the substrate, hence the unaffected K_M (affinity for the substrate). Residual hydrolysis of the substrate is possible, perhaps as the substrate is forced into a conformation where the phosphate-phosphate bond is more susceptible to hydrolysis.

5.5.6. Inhibition of the ER-UDPase

High-throughput screening of ENTPD5 discovered novel inhibitors, in an attempt to develop anti-cancer drugs[304]. A number of inhibitors from that study were chosen that induced ER stress, as assessed by increased levels of GRP78. This choice increases the likelihood that these molecules were ENTPD5 inhibitors, as affecting ENTPD5 levels would likely cause a knock-on effect on the calnexin cycle and therefore induce ER stress.

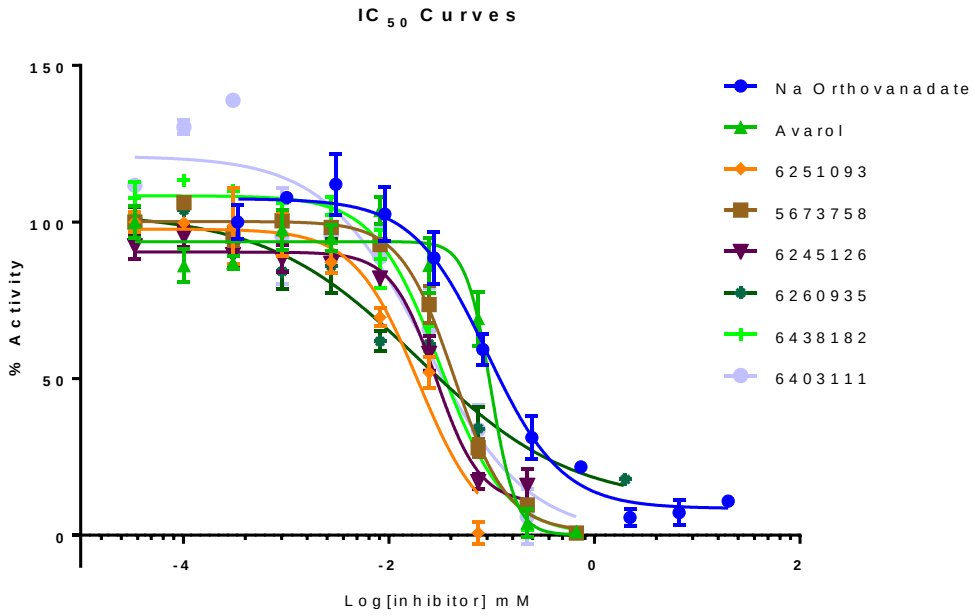
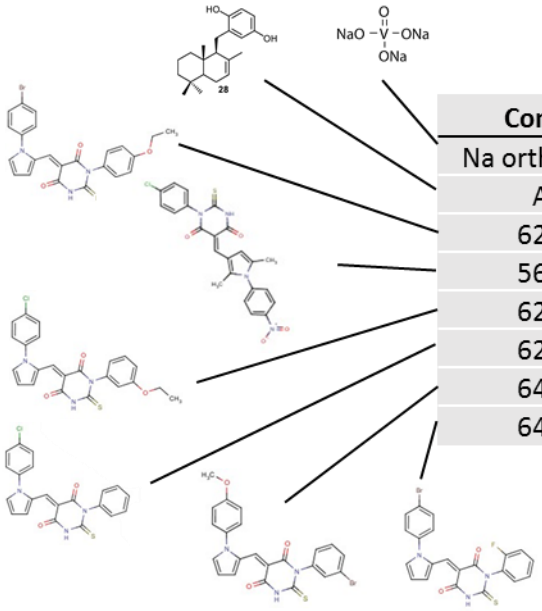


Figure 5.8. IC_{50}s of a range of potential *Hs* ENTPD5 inhibitors. Assays were run with a UDP concentration of 1 mM and 0.00001 mg of enzyme per reaction.

Sodium orthovanadate is a well-known structural mimic of phosphate and was therefore likely also an inhibitor of the enzyme.

Compounds were soluble in DMSO at 10 mM; however, when diluted into assay buffer, many crashed out of solution above concentrations of around 200 µM. The IC_{50} curves for the compounds tested are shown in figure 5.8 and their respective IC_{50} values and structures in table 5.5.

Sodium orthovanadate inhibits ENTPD5 with an IC_{50} value of 91 µM, similar to that of avarol, a sesquiterpenoid hydroquinone originally isolated from a marine sponge and also reported to have antiviral activity, albeit by a different mechanism of action than through ENTPD5 inhibition[322, 323]. The six compounds that were chosen from the high-throughput screen all contain a uracil-like ring and were more potent inhibitors than avarol. In my hands, compounds 5673758 and 6403111 were the most soluble; the latter was the more potent of the two. The next step for the ENTPD5 inhibitors was to see if they could lead to an antiviral effect *in cellula* (5.7.2)



Compound	IC_{50} (µM)
Na orthovanadate	91.0
Avarol	98.2
6251093	19.4
5673758	43.0
6245126	28.6
6260935	19.3
6438182	31.5
6403111	22.7

Table 5.5. Structures of the potential *Hs* ENTPD5 inhibitors and their IC_{50} values against the enzyme.

5.6. Attempts at crystallisation

5.6.1 Crystallisation of MBP-ENTPD5

Sparse-screen crystallisation plates were set up of the refolded MBP-ENTPD5 at protein to mother liquid volume ratios of 1:1 and 2:1, unless otherwise stated. Initially, a concentration of 4.74 mg/ml passed the pre-crystallisation test (PCT, from Hampton Research) and screened in Morpheus (Molecular Dimensions), Index (Hampton Research) and JCSG+ (Molecular Dimensions) plates.

Guanosine 5'-[α, β-methylene] diphosphate (GPCP) is a non-hydrolysable analogue of the ENTPD5 substrate, GDP. After incubation with a 30 fold molar excess of GPCP, MBP-ENTPD5 was plated at 4.94 mg/ml in the Morpheus, Index, JCSG+, Pact Premier (Molecular Dimensions), Proplex (Molecular Dimensions) and PEG/Ion (Hampton Research) crystallisation screens.

Once the DSF assays were completed, the protein was dialysed into 20 mM HEPES, 50 mM NaCl, 10 mM EDTA and 10 mM UMP. A Proplex screen was set-up at 7.66 mg/ml in this buffer but the protein precipitated out of solution almost immediately. A dilution to 5.0 mg/ml was made and Morpheus, Index, Pact Premier, JCSG+ and PEG/Ion plates were set up, but with protein to mother liquor volume ratios of 1:1 and 3:1. A low salt concentration of 50 mM was used despite the protein showing greater stability at 500 mM NaCl, in order to allow for vapour diffusion from the sitting drop to the mother liquor reservoir; high concentrations of salt in the protein buffer will slow down drop evaporations when equilibrating against a low-salt screen condition.

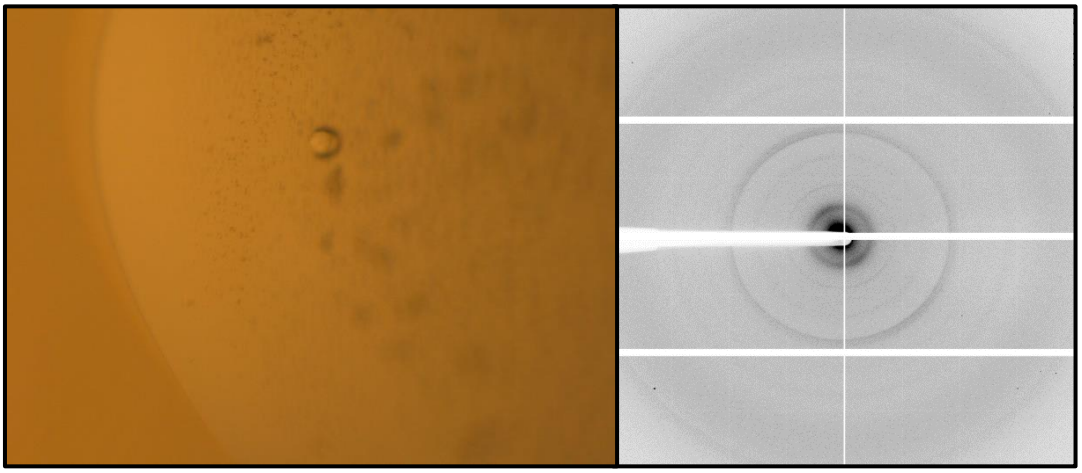


Figure 5.9. MBP-ENTPD5. On the left is a visible light image of the object grown in well B8.1 of the PEG/Ion screen, exposed to X rays at ESRF. On the right is the resultant X-ray diffraction image. Protein powder rings are visible.

A number of crystalline-like objects were identified in these droplets. These were screened for X-ray diffraction at ESRF. All but one crystal gave diffraction patterns reminiscent of salt crystal diffraction. The sample from well B8.1 of the PEG/Ion screen prepared with the GPCP substrate analogue (figure 5.9), gave a protein powder diffraction image.

It was observed that at 5.0 mg/ml of MBP-ENTPD5, 10 mM UMP, the protein was much more susceptible to precipitating out of solution. The solution was diluted to approximately 2.5 mg/ml protein using H_2O the sample re-screened in the Morpheus sparse-screen. A lot of UV light positive precipitation was observed in these screens as well. As nucleotides are also visible under UV light, it was likely that the UMP was precipitating. From this point onwards, UMP was omitted from protein crystallisation buffers.

The MBP part of the fusion was removed using 3C protease cleavage and the resultant *Hs* ENTPD5 re-purified and concentrated to 3.1 mg/ml and plated in a Morpheus screen.

A further batch of protein was methylated, concentrated to 4.86 mg/ml and plated in Stura Footprint Combination (Molecular Dimensions), Pact Premier, Index, Morpheus and JCSG+ crystallisation screens. None of these further approaches were successful at producing single crystals.

5.6.2. Crystallisation of *Ct* ENTPD6

Ct ENTPD6 was purified, concentrated to 7.5 mg/ml and initially used to set up crystallisation screens at ratios of protein to screening condition ratios of 1:1 and 2:1, in Morpheus, Morpheus 2 (Molecular Dimensions), Index, JCSG+, PEG/Ion and MIDAS (Molecular Dimensions) screens.

Only a few of the drops showed sample precipitation, so the protein was concentrated to 28.6 mg/ml and used in Morpheus, Proplex, JCSG+, PEG/Ion, Index and MIDAS screens. Once again, very few conditions produced any precipitated protein, although many looked like a protein 'gel' had formed, similar to phase separation. Protein was concentrated once more to 58.8 mg/ml and screened in Morpheus and Proplex plates at protein to mother liquor volume ratios of 1:1 and 3:1. Once again, few conditions led to precipitated protein but many appeared with phase-separation or a gel-like appearance.

Ct ENTPD6 was methylated, which did not affect the trimeric state of the protein. Protein, both methylated and non-methylated, was plated with and without 2.5 mM GMP in the Stura Footprint combination screen and the Morpheus screen. Finally, monomeric *Ct* ENTPD6 was obtained by partially digesting the protein with the GluC endoproteinase. This construct was concentrated to 14.35 mg/ml and plated in Morpheus, JSCG+, Index, Pact Premier and Proplex crystallisation screens.

None of the approaches for any of the protein constructs, treatments or additives, produced any protein crystals. The protein may not be amenable to crystallisation, due to its flexibility being too great, or monomer/ multimer heterogeneity. Other approaches that could be tried would be to develop antibodies for the folded protein and attempt co-crystallisation with a specific Fc fragment derived from the antibody.

5.7. Antiviral properties of ENTPD5 inhibitors

In order to initially test whether abrogation of ENTPD5 activity could hold promise as an antiviral strategy, ENTPD5 inhibition was tested to see if it could affect Zika virus release or infectivity from infected monolayers of kidney epithelial cells.

5.7.1. Toxicity assay

Toxicity assays were run to see what concentrations of ENTPD5 inhibitors it is possible to expose cells to without this leading to a cytotoxic effect. The assay is based on the reduction of a tetrazolium compound, MTS, which only occurs in viable cells, thus allowing it to be used to assess the cytotoxicity of inhibitors by following the absorbance at 490 nm.

Based on their IC_{50} results, their solubility and the variety in their molecular structures, the inhibitors 5673758, 6251093 and 6403111 were chosen for screening in the toxicity assay. Results are shown in figure 5.10.

All three compounds show cytotoxic effects that occur at concentrations where DMSO alone does not have an effect. The toxic concentration at which 50% cells are non-viable (CC_{50}) is 33.45 µM, 36.41 µM and 24.37 µM for 6403111, 6251093 and 5673758, respectively.

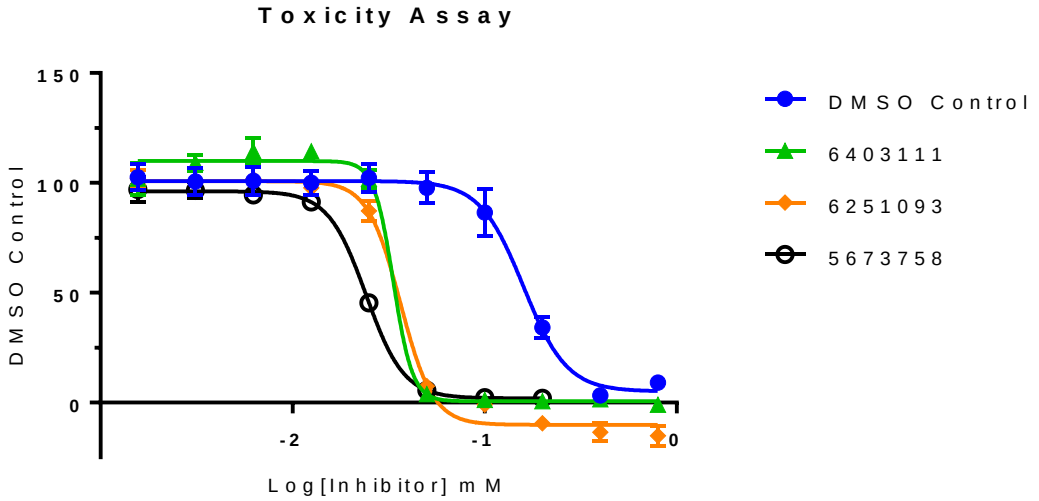


Figure 5.10. Effect of ENTPD5 inhibitors on cells as measured by an MTS cytotoxicity assay. Background levels of absorbance by the compounds were corrected for by running an identical set of measurements with inhibitor but without cells.

The latter has a less potent IC_{50} than the other two yet has a lower CC_{50}, indicating that the toxic effect does not occur through just ENTPD5 alone. Cell viability at 25 µM of either 6403111 or 6251093 is actually greater than 90%, suggesting that it may be possible to use this concentration in the Zika assay. Whilst the toxicity assay exposed cells to compound for four days, in the Zika assay, this would only occur for 2 days, and so cytotoxic effects by ENTPD5 inhibitor alone may not be as pronounced by this point.

Based on the most favourable combination of its IC_{50}, solubility, and toxicity, the compound 6403111 was chosen for use in a Zika antiviral assay.

5.7.2. 6403111 in the Zika antiviral assay

Cells were exposed to 6403111 at various concentrations diluted into culture medium immediately after the initial infection of Vero wild-type cells with Zika virus (MOI=1)

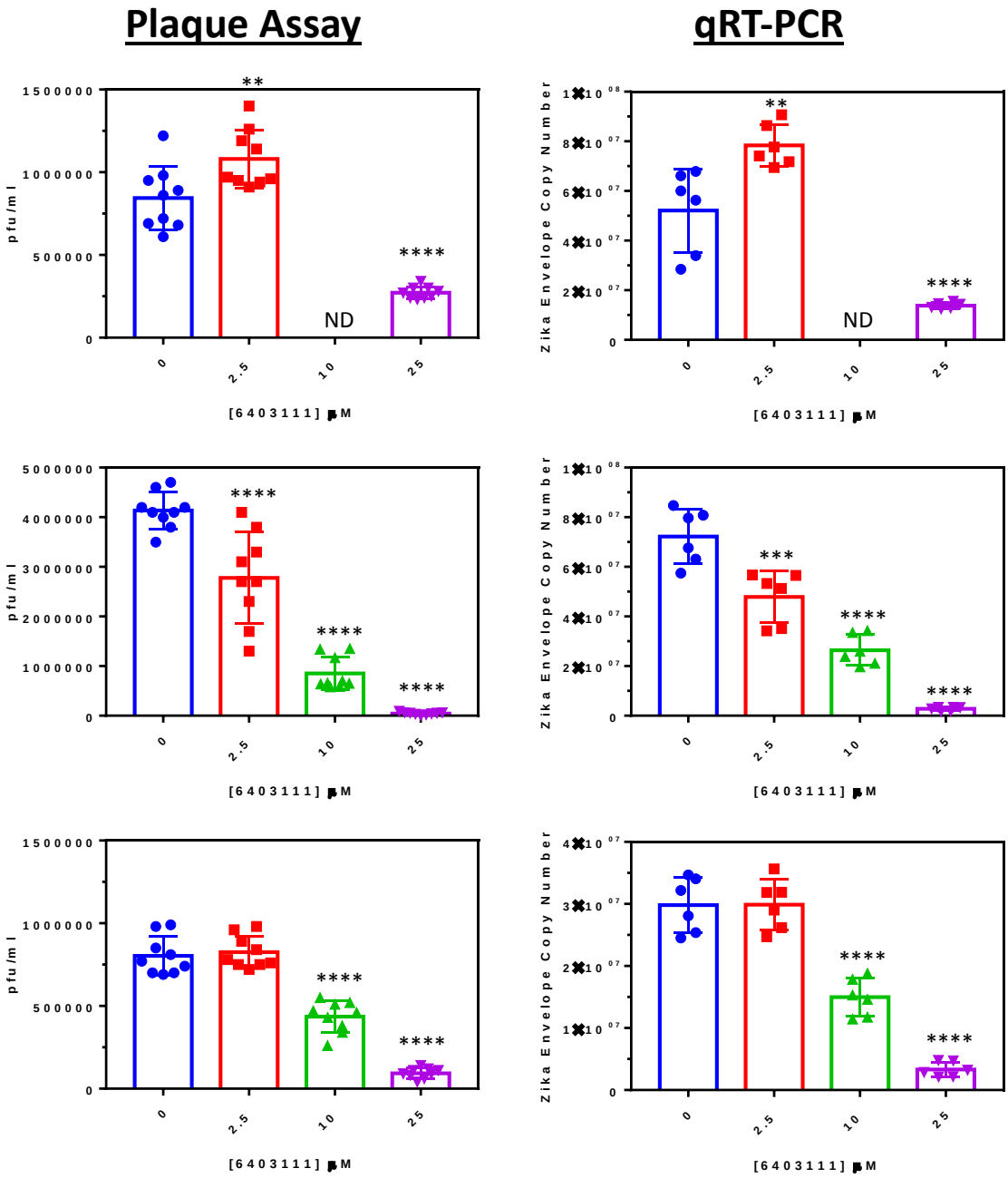


Figure 5.11. The left hand-side graphs represent 3 repeats of the antiviral assay with release of infectious virus measured by the plaque assay. The right-hand side graphs show the corresponding results of the qRT-PCR measurements to detect total virus release. ND = No Data. * = p value of ≤0.0322, ** = p value of of ≤0.00211, *** = p value of of ≤0.0002, ****, of ≤0.0001.

and removal of the viral inoculum by washing the cells. Cells were left for 48 hours before the supernatants were harvested for measurements of total virus and infectious virus release, in the qRT-PCR and plaque assay, respectively.

Figure 5.11 shows the results from these assays, with three repeats of the assay. The results and trends for the plaque assay closely correspond to those of the qRT-PCR for each of the repeats. This indicates that the virus released from cells, as measured by the qRT-PCR, is not any less infectious than that released from untreated cells.

At 25 μM of 6403111, there is a consistently significant drop of Zika virus release, when compared to untreated cells of approximately 10-20 fold. At this concentration of compound in the toxicity assay, only 5-10% of cells were non-viable. By visual inspection, the large majority of cells were viable at the time of harvesting supernatants for the assay, which indicates that this compound has an antiviral effect that is not just due to the fact that dead cells are unable to produce virus. Potential toxicity to the cells resulting from a combination of the virus and ENTPD5 inhibitor was not studied by any other method.

At 10 μM, the inhibitor also consistently produced an antiviral effect but to a lesser extent compared to the higher concentration. No toxic effects to the cells were observed at this concentration of compound. At just 2.5 μM 6403111, there was no consistent antiviral effect observed. In one repeat, significantly more virus was released, in the next, significantly less and in the last, there was no difference to untreated cells at this concentration. This highlights why relying on p values from a single experiment can lead to false conclusions; more repeats are necessary before reliable conclusions can be drawn from the data.

5.8. Discussion

Production of small amounts of recombinant and tissue-purified ENTPD5 has been reported in the literature. Here, I aimed to express and purify *Hs* ENTPD5 in milligram quantities for use in crystallisation studies, and for use by structure-based lead discovery.

Different expression vectors and purification tags were tested, with the MBP-fusion constructs producing the largest quantities of soluble protein. Initially this protein, however, acted like a large multimer or aggregate and had much lower activity than has been reported in the literature. The denaturing and subsequent refolding of the protein proved successful in leading to a monomeric and fully active form of the protein.

In addition to the full-length mammalian constructs, a second set of constructs was designed with a truncated N-terminus. The original rationale behind this idea was derived from observations of other ENTPD family members: that for crystallisation to occur, a potential membrane interaction loop (MIL) had to be removed, inclusion of this loop prevented crystallisation of rat ENTPD1[306]. However, rat ENTPD2 crystallised with this loop intact[324]. Sequence alignment of ENTPD family-members indicated that *Hs* ENTPD5 does not contain this loop in the same position, yet a very similar sequence can be found directly following the signal peptide of *Hs* ENTPD5. It was this sequence that was removed to produce the truncated construct of *Hs* ENTPD5, which led to the best expression of all mammalian constructs.

An alternative target was an ENTPD5 homologue from *Chaetomium thermophilum*, however, the ENTPD5 itself could not be found in that genome. The closest homologue was *Ct* ENTPD6, whose homology model displays a very similar active site to *Hs*

ENTPD5. This is the first reported expression of *Ct* ENTPD6, which shares similar substrate specificities to *Hs* ENTPD5. It may be that the Golgi-located ENTPD6 makes up for the lack of an ER-located ENTPD5 in *Chaetomium thermophilum*.

A mutant in which the *Hs* ENTPD5 predicted catalytic residue, E172, was mutated to a glutamine was produced in the hope that a stable complex between substrate and protein could be formed. Residual catalytic activity still remained, so that crystallisation trials with this mutant would likely not lead to a complex with the substrate. In the proposed *Hs* ENTPD5 catalytic mechanism, E172 would nucleophilically attack a water which subsequently attacks the nucleotide's terminal phosphate[325]. It is possible that water is primed spontaneously and can attack the nucleotide, which can still bind the mutant enzyme.

A wide range of crystallisation screens were prepared with many different concentrations and preparations of *Hs* ENTPD5 and *Ct* ENTPD6. Next to no crystal occurred in any of the conditions tested. Methylation, proteolysis and co-crystallisation with the reaction products were all unsuccessful. However, these constructs do show that large quantities of protein can be produced for crystallisation and other approaches may yet be successful.

Finally, the first evidence that inhibition of ENTPD5 may be useful as an antiviral strategy has been obtained, indicating that affecting the passage of nucleotides and nucleotide sugars within the ER may help to stop viral proteins from utilising the host glycoprotein folding machinery. Development of less cytotoxic and more specific inhibitors will be needed if the current results are to be built upon. Inhibitors that are broadly specific to the ENTPD family members that hydrolyse UDP and GDP may be actually more desirable than those specific for just one. Targeting just one protein may not be as effective

if ENTPD5 and ENTPD6 show redundancy within the cell. Also ENTPD5/6 are not the only proteins that have been reported to hydrolyse UDP within the secretory pathway; soluble calcium-activated nucleotidase 1 (SCAN1), an apyrase family member, is also ER/Golgi localised and has a similar substrate specificity to the two ENTPD family members[326]. Future work should focus on the dissection of the mechanisms and protein targets behind the antiviral activity of the ENTPD5 inhibitors and their cytotoxicity.

6. CRISPR/Cas9 knock-out cell lines

Within the space of the last 10 years, a technique commonly referred to as "CRISPR/Cas9" has rapidly become a staple of hundreds of research laboratories around the world. It is an easily modifiable system that can be used to target unique sites within a target cell's genome to generate total gene knock-outs or alter the sequence of the genetic material by using template DNA. Other advances allow for inducible

CRISPR/Cas9 operons, single-stranded DNA cutting or the use of inactive nuclease systems to recruit other proteins to the unique sites within the genome for regulation or imaging[327-330].

One of the original aims of this D.Phil project was to explore whether other proteins involved in the calnexin cycle could be, in principle, valid antiviral targets. Rather than use RNA-based silencing techniques, it was decided that the CRISPR/Cas9 system would be used to generate full knock-outs of the genes encoding these proteins and to assess whether the susceptibility of the resulting cell lines to viral infection and whether viral production in these cells were altered.

The proteins chosen for knock-out were: both ER α-glucosidases, these would serve as a control, as inhibition by iminosugars is already established to have anti-dengue activity; UGGT1 and UGGT2, which act as the quality control checkpoints of the calnexin cycle; malectin, an under-characterised ER-lectin which is known to bind di-glucosylated glycans; SLC35B1, an antiporter hypothesised to transport UGGT's substrate, UDP-glucose, into the ER and export UMP; and finally ENTPD5, the ER-UDPase that has been suggested in an earlier chapter to be a target for antiviral drugs[64, 92, 124, 331]. All of the knock-outs were produced in Vero E6 cells (ATCC code CRL-1586), a cell line well established within the laboratory for antiviral assays, derived from a female African Green Monkey (*Chlorocebus sabaeus*) and deficient in type-I interferons, which makes Veros more susceptible to many viral infections. The work presented here represents the fruits of this labour.

6.1. Bioinformatics and construct design

The genome sequence of *Chlorocebus sabaeus* was only released in 2014[332]. The work presented here is among some of the first CRISPR experiments attempted in Vero cells, with the first papers describing CRISPR of Vero cells published in 2016[333, 334].

Two sets of primers were designed for each gene to be knocked out – in neighbouring exons if possible, with the aim that a large excision would take place between the two sites and remove a large slice of DNA from the genomic loci. The primers used to generate the single guide RNA (sgRNA) that targeted unique sites within the genome are recorded in the appendix. Screening primers, designed to look for signs of deletions in the genome by PCR, and scarring (sequencing) primers, to sequence the genome around each targeted loci, were designed as in Orkin *et al*[261] (see Appendix). The extent of gene alteration required to produce a 'knock-out' is under debate, however, here I will define it as a mutation that would lead to an inactive protein – although all of the mutations described here would likely lead to much more severe effects on the protein targeted.

For the knock-out of ER α-glucosidase I, the CRISPR sgRNAs target exon 1 (around Arg^{100}) and exon 4 (around Asp^{746}), meaning that if the full excision does not occur successfully, the first target could lead to a frameshift that knocks-out both lumenal domains of the protein and the latter would affect part of the catalytic domain, including one of the catalytic residues.

For the knock-out of ER α-glucosidase II, the CRISPR sgRNAs target the *GANAB* gene, exon 3 (around residue Thr^{99}) and exon 6 (around residue Ala^{194}). Once again, if the full excision does not occur successfully, frameshift mutations at either site would disrupt the vast majority of the protein.

For the knock-out of UGGT1, the sgRNAs target around 5000bp before the start codon in the genomic sequence (although much closer in the mature mRNA) as well as exon 2 (around Val^{40}). In this case, either a full-excision or a frameshift mutation at the second CRISPR site would be sufficient to knock-out the protein: the former would remove both of the possible start codons and the latter would lead to early termination of the protein.

For the UGGT2 knock-out, the CRISPR sgRNAs target exon 3 (around Gln^{123}) and exon 5 (near Glu^{189}); excision of the intervening sequence or frameshift mutations at either site would be sufficient to produce a knock-out.

Malectin is the smallest protein to be knocked-out. The sgRNAs target exon 2 (around Ala^{128}) and a site 700 bp after the C-terminus of the protein as no sites specific to only these loci could be found elsewhere within the protein. A knock-out would be achieved with the excision of the intervening DNA or a frameshift mutation in exon 2.

The proposed UDP-glucose: UMP antiporter, SLC35B1, two sgRNAs target the gene at exon 1 (around Asp^{10}) and the 5' end of exon 9 – positioned at the exon boundary. An excision of the intervening DNA would be ideal, although frameshift mutations at exon 1 would also suffice.

Finally, for the knock-out of ENTPD5, sgRNAs target exon 2 (Val^{21}) and exon 5 (around Pro^{160}), both of which are N-terminal to the catalytic residues and would lead to a knock-out whether through an excision or a frameshift mutation.

One thing to note, however, is that many splice isoforms of each of the proteins have been recorded, and for some of these, the CRISPR/Cas9 mutations induced by our choice of

sgRNA would not be as damaging. It is therefore imperative to pick clones where the resultant protein has been altered to the largest extent.

6.2. Cloning, transfection and clonal selection

Cloning of the target DNA sequences was carried out as outlined in the methods (2.1.6), with a minor change from the established, published protocols to allow for a much more efficient annealing step: the buffer supplied by NEB with the polynucleotide kinase lacked ATP, and substitution of this buffer with the ATP-containing T4 DNA ligase buffer led to correct insertion of the intended sequence in nearly 100% of colonies.

Transfection of Vero E6 cells took place as described in the materials and methods, using a Lipofectamine 3000 based, transient transfection. Subsequent clonal selection was carried out as described, simply by diluting the cells so that a single clonal population grew in wells of a 96-well tissue culture plate.

6.3. Screening of clones

6.3.1. PCR screening

For each knockout, once a suitable number of clones had reached a cell number between approximately 50 and 500, clones were split into two identical populations: one for screening by PCR and one for further growth.

PCR screening was carried out as outlined in the methods (2.8.4). In brief, for each clone, two PCR reactions were carried out, one with a pair of 'non-deletion' (ND) primers and a pair of 'deletion' (D) primers. The former primer pair will produce a band of the

expected size if there are alleles present in the sample that have not had DNA excised between the two CRISPR sites in the gene; the latter will only produce a band if there are alleles present in which the intervening DNA has been excised. Of course, the presence of a ND band does not mean that a successful knock-out has not occurred, as frameshift mutations involving the insertion or deletion of a few bases could still have occurred.

All clonal populations from the ER α-glucosidase I knock-out attempts gave rise to a PCR 'ND' band and in a large number of clones, a 'D' band was also present. These clones appeared to be heterozygous for the excision of the intervening DNA between the two CRISPR sites. These clones were retained for further screening to ascertain whether knock-out had been successful: as the allele surviving the excision and giving rise to the ND band could have undergone frameshift mutation.

Knock-out of the ER α-glucosidase II was much more successful, as two clones, B1 and H1, both produced a 'D' band but no 'ND' band, as can be seen in figure 6.1. Therefore, both alleles had the intervening DNA between the two CRISPR sites fully removed and two knock-outs were clearly achieved.

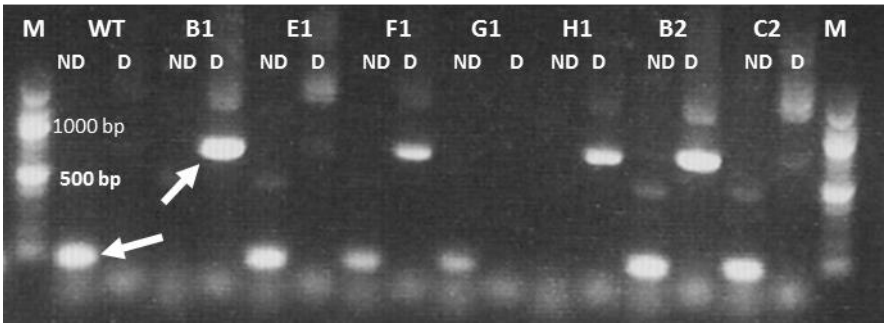


Figure 6.1. PCR screening of ER α-glucosidase II clonal populations. WT refers to the mother 'Vero E6' cell line, in which only a ND band is expected (lower arrow). A band is expected in the D lanes (higher arrow) where DNA between the two CRISPR sites has been removed from an allele.

UGGT1 screening resembled the glu II one, in that clone F7 produced a 'D' band but not a 'ND' band. This indicates that the CRISPR removed the DNA between the two sites targeted for CRISPR and has fully removed the start of the protein, including the start codons.

PCR screening of the UGGT2 knock-out attempts, similarly to those of glu I, did not yield any clones that solely had a 'D' band, but did produce a number of clones which were heterozygous for the large deletion of DNA. These clones were retained for further screening to find out whether any knock-outs were produced. This was also the case for the knock-outs of malectin, the knock-outs of SLC35B1 and the knock-outs of ENTPD5.

6.3.2. Sequencing to confirm knock-outs

For those cell lines in which only clonal populations heterozygous for the intended DNA excisions were isolated, it was necessary to sequence the CRISPR sites to see if the 'surviving' allele contained an indel that could lead to a frameshift mutation and therefore lack a functional gene product. In order to do this, the 'scarring' primers (see Appendix) were used to PCR amplify a 500 base region around each of the target CRISPR

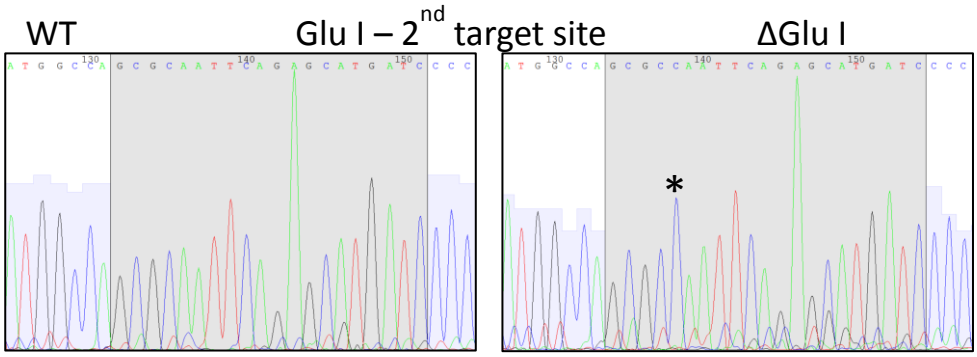


Figure 6.2. DNA sequencing of the second targeted CRISPR site in the gene encoding glu I. Shaded areas highlight the DNA sequence targeted by the sgRNA in the non-mutated clone. The asterisk indicates the insertion found in the knock-out clone.

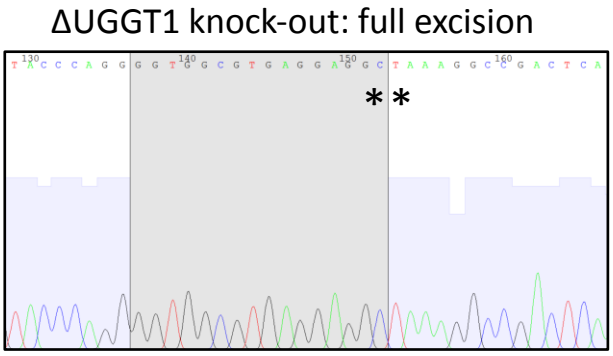


Figure 6.3. DNA sequencing of the UGGT1 knock-out. Asterisks represent the site of ligation where the DNA between the two target sites were removed. Shaded areas highlight the sgRNA target sequence in the wild-type gene.

sites that was sent for sequencing. This was not necessary for the homozygous knock-outs of glu II and UGGT1, although DNA sequencing of the latter was carried out for related work.

Figure 6.2 shows the second site targeted for CRISPR cleavage for the Δglu I clone used in further experiments. The left panel highlights the DNA sequence targeted by the sgRNA in a non-mutated clone. The right panel shows that in one of the mutated clones there is an extra cytosine base that would lead to a frameshift mutation from that point onwards. This in turn leads to an early termination and the disruption of three of the α-helices that form part of the catalytic domain and includes the catalytic residue Glu^{804}. *In cellulo*, of the two alleles present in the clonal population, one has had the DNA between the two CRISPR sites excised, with of ~80% of the protein removed, while the other allele can only produce catalytically inactive protein if it manages to fold at all and is not eliminated by ERAD.

The full knock-outs of glu II and UGGT1 (confirmed by removal of the DNA between the two CRISPR sites in both alleles) did not need to be sequenced, although ΔUGGT1 was

sequenced using the 'Deletion' screening primers (Figure 6.3). Cleavage by the CRISPR/Cas9 system is predicted to occur three bases 5' of the PAM sequence[335]. This is exactly what has occurred at both of the target sites: the remaining DNA has ligated, removing the intervening ~5000 base pairs.

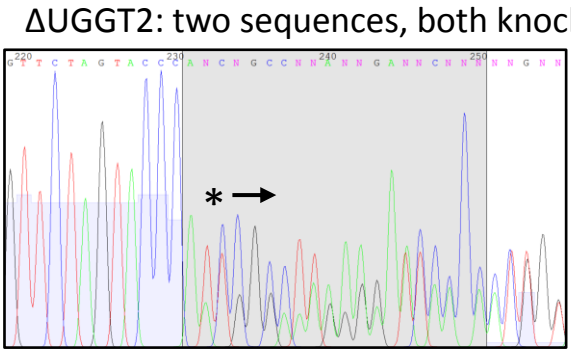


Figure 6.4. DNA sequencing of the UGGT2 knock-out. Asterisk represents the site of mutation. Two distinct DNA species were observed, neither of which resembled the wild-type sequence. The shaded area represents the region targeted by sgRNA in a non-mutated clone.

The knock-out of UGGT2 was mutated at the second CRISPR target site, still in the part of the gene encoding the N-terminal portion of the protein. The mutation can be seen in figure 6.4: two distinct species were observed in the clonal population, representing two different alleles. Neither of these sequences were equivalent to the wild-type sequence and due to the N-terminal positioning of the mutation in the sequence of the protein, the large majority of the protein has been disrupted.

Figure 6.5 shows both of the targeted CRISPR sites of malectin. The left panels highlight the wild-type sequence and the targeted DNA, the right hand panel highlights the position of a large insertion (1st target site) or large deletion (2nd target site) at the respective loci, causing frameshift mutations and knock-out of the gene.

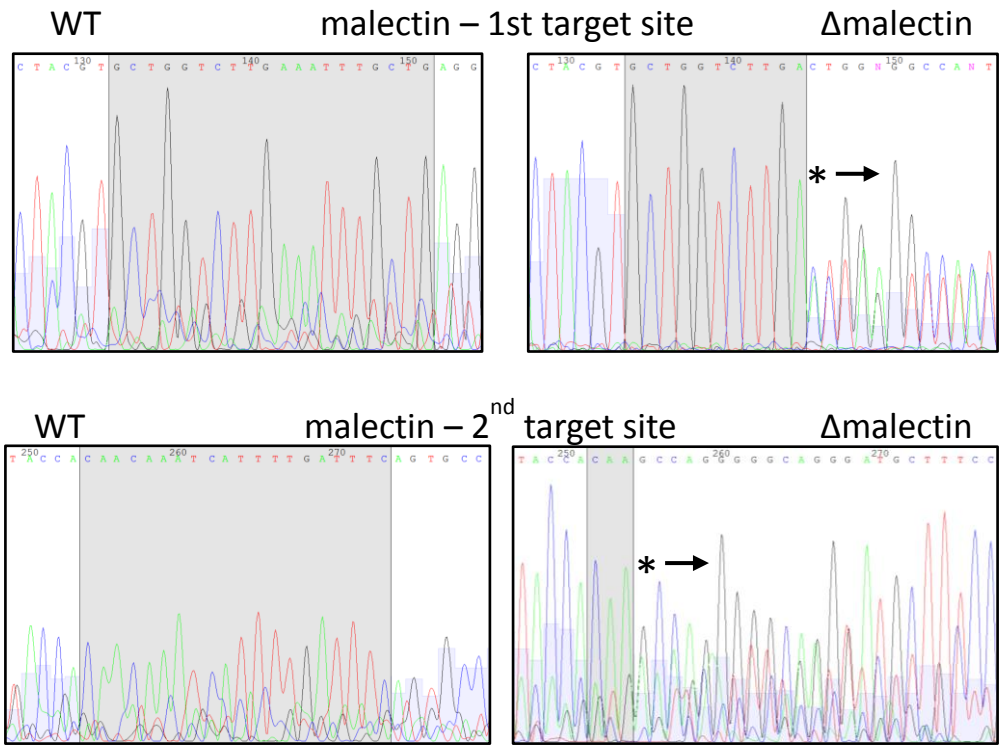


Figure 6.5. DNA sequencing of the malectin knock-out, both target CRISPR sites are shown. Asterisks highlight at what point the mutation occurs. Shaded regions are the sgRNA target sites.

The PCR screening of the SLC35B1 knock-out clones only found wild-type-like PCR bands at the two CRISPR sites for those clones for which a PCR band was produced. However, for one of the clones, the 'scarring' primers could not amplify the intended PCR fragment for sequencing from the second site, indicating that it was altered to the extent that the primers could not bind and that a knock-out has actually been achieved.

Finally, the knock-out of ENTPD5 produced many clones that all appeared heterozygous in the PCR screening with both 'D' and 'ND' PCR bands observed. However, sequencing of the CRISPR sites using the 'scarring' primers found that many of them contained frameshift mutations at both sites. Figure 6.6 shows the clone that was used for further work. The left-hand side panels highlight the targeted sequences in a wild-type cell line,

the right-hand side panels highlight a deletion and an insertion of a single base for the first and second CRISPR target sites, respectively, inducing frameshift mutations.

Overall, each set of DNA sequencing experiments confirmed that at least one clone of each knock-out achieved a desired mutation to knock-out the protein.

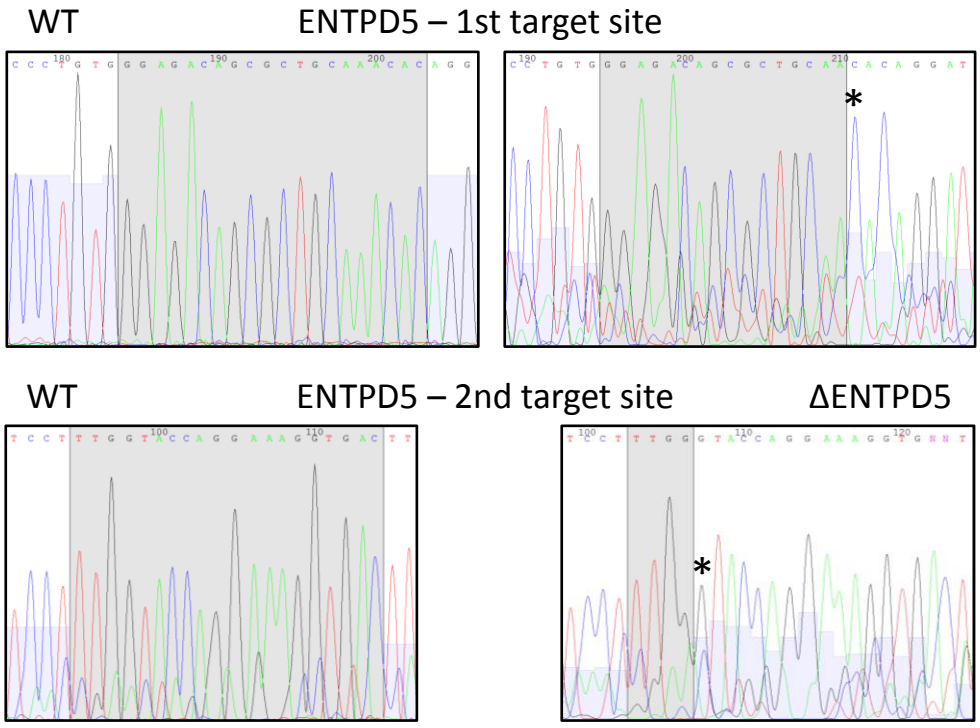


Figure 6.6. DNA sequencing of the ENTPD5 knock-out cell line, both target CRISPR sites are shown. Asterisks highlight the point mutations that lead to a frameshift mutation. Shaded regions are the sgRNA target sites.

6.3.3. Free oligosaccharide analysis

Free oligosaccharide (FOS) analysis is a technique which measures the levels glycans derived from degraded glycoproteins in the cell. Alonzi *et al*, optimised the analysis by fluorescently labelling FOS samples with 2-amino anthranilic acid, allowing for the

detection of FOS species that resulted from ER-derived proteins that have been sent for ERAD and destruction via the proteosome[336].

As such, FOS analysis could be used to confirm successful removal of the catalytically active protein from the two glucosidase knock-outs by a means other than DNA based methods. The knock-out of ER α-glucosidase I should lead to a build up of tri-glucosylated FOS species, whereas that of ER α-glucosidase II would lead to an increase of di-glucosylated glycans. Figure 6.7 shows the FOS analysis of the Vero wild-type and glu I / II knock-out cell lines, carried out by Dr Dominic Alonzi. The upper trace of FOS derived from the wild-type cell line shows a much less pronounced profile than the other two samples. This is because the analysis can detect FOS derived from glycoproteins sent for degradation – in a healthy, wild-type cell line, there is no reason to expect high-levels of proteins being diverted to ERAD. The middle panel shows the results derived from the glu I knock-out line. Here, we can see peaks appear that correspond to tri-glucosylated species, namely $Glc_3Man_9GlcNAc_1$ and $Glc_3Man_5GlcNAc_1$. This indicates that in this cell line, firstly, more proteins are being sent for destruction, and secondly, that there is a lack of ER α-glucosidase I activity, confirming the knock-out. The last panel shows the results derived from the glu II knock-out cell line. Here, the marked peaks are those from $Glc_2Man_9GlcNAc_1$ and $Glc_2Man_5GlcNAc_1$, that indicate once again, more proteins are being sent for ERAD in this cell line, and lack glu II activity. These traces confirm that the activity of glu I and II have been removed in the respective knock-out cell lines.

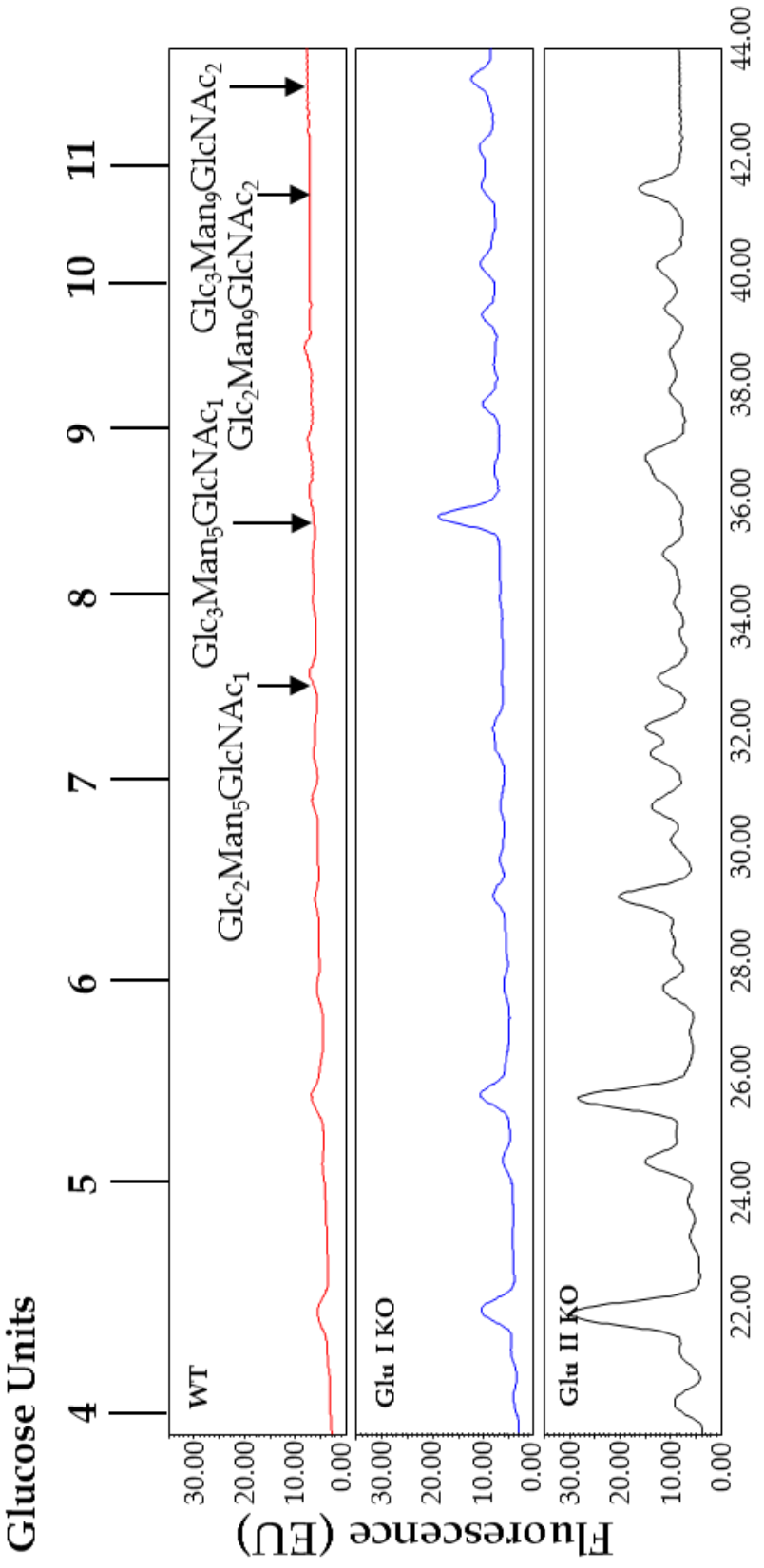


Figure 6.7. Elution traces of FOS samples from three cell lines, Vero wild-type (WT), Vero glu I knock-out and Vero glu II knock-out. Labelled are the peaks of interest corresponding to inhibition of the two glucosidases. Data and figure produced by Dr. Dominic Alonzi.

However, this technique cannot tell us more about any of the other knock-outs – as knock-out of UGGT1/2 or any of the other downstream enzymes would only lead to a lack of mono-glucosylated glycans at best – a short-lived glycosylation species, and thus would be hard to detect.

6.4. Immunofluorescence experiments

One of the caveats of using a knock-out cell line to assess whether the altered gene would be a valid targets for antiviral drugs, is that the cell line has already been exposed to the effect of the knock-outs for many passages before exposure to virus. *In vivo*, cells would not generally be exposed to antiviral drugs until after viral infection.

For example, if the receptor that the virus hijacks in order to enter the cell, were to fold under the control of the knocked out protein. In this case, the results from this knock-out cell line would confer great antiviral efficacy, despite it solely being due to the lack of viral entry and the consequent lack of viral replication. For both of the viruses used within this study, Zika and dengue, the exact nature of the receptors or molecules which mediate viral entry into the host cell is unknown. Zika virus has been shown to be able to use Axl[337], which encodes six N-glycosylation sites, to aid entry into the cell. Dengue virus, on the other hand, has many potential receptors in mammalian cells, including DC-SIGN[338], CD14[339] and the mannose receptor[340], all of which contain a number of N-glycosylation sites. The fact that the viral receptors are glycoproteins themselves highlights the importance of determining whether this could affect the results presented in this chapter.

Immunofluorescence was used to visualise viral envelope proteins two days after infection, which should be sufficient to show which cells have been infected.

Figure 6.8 shows the data obtained from infection of knock-out cell lines with Zika virus. Infection of the virus assays use a MOI of 1, meaning that on average, every cell will be infected with a single, replicating, virus. This is reflected in the percentage of infected cells for each of the cell lines, all of which are almost 100% infected. All of the cell lines, except for the knock-out of ENTPD5, do not exhibit significant differences from the wild, indicating that virus can enter all cells equally. However, for the ENTPD5 knock-out, there is a significant reduction, from around 98% of cells infected in the wild-type line, to 89% in this knock-out. This suggests fewer viruses are entering these cells, presumably

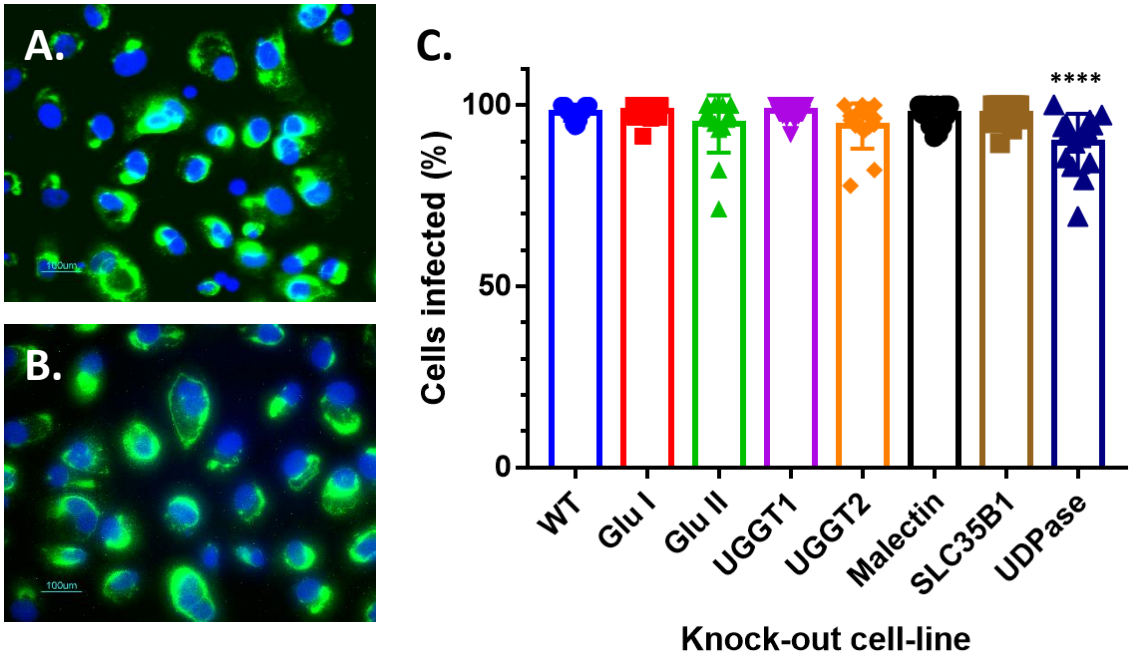


Figure 6.8. Immunofluorescence of Zika-infected cells. Panels A and B show representative images obtained under the inverted microscope of cells 48 hours post infection. In blue are the DAPI-stained nuclei, in green, the FITC staining from the conjugated secondary antibody. Panel C shows the percentage of cells infected for each of the cell lines. **** = p-value of <0.0001. There were 50-150 cells per image, 5 images per well, 3 wells per cell line.

because of a downstream effect of the knock-out of the ENTPD5. However, this reduction is likely to be independent of the calnexin cycle as if it were, we would expect to see an effect in the knock-out cell lines of other components such as UGGT1. For the experiments that follow, the slight resistance of the ENTPD5 knockout to Zika virus infection must be taken into account for any interpretation.

The immunofluorescence results for dengue virus infection, seen in figure 6.9, are much more striking but the reasons behind the results must be carefully considered. For both glucosidase knock-out cell lines, a reduction in the percentage of infected cells to 30-40% was observed, whereas all other lines have no statistically signifcant difference to the wild-type line (95% of cells infected).

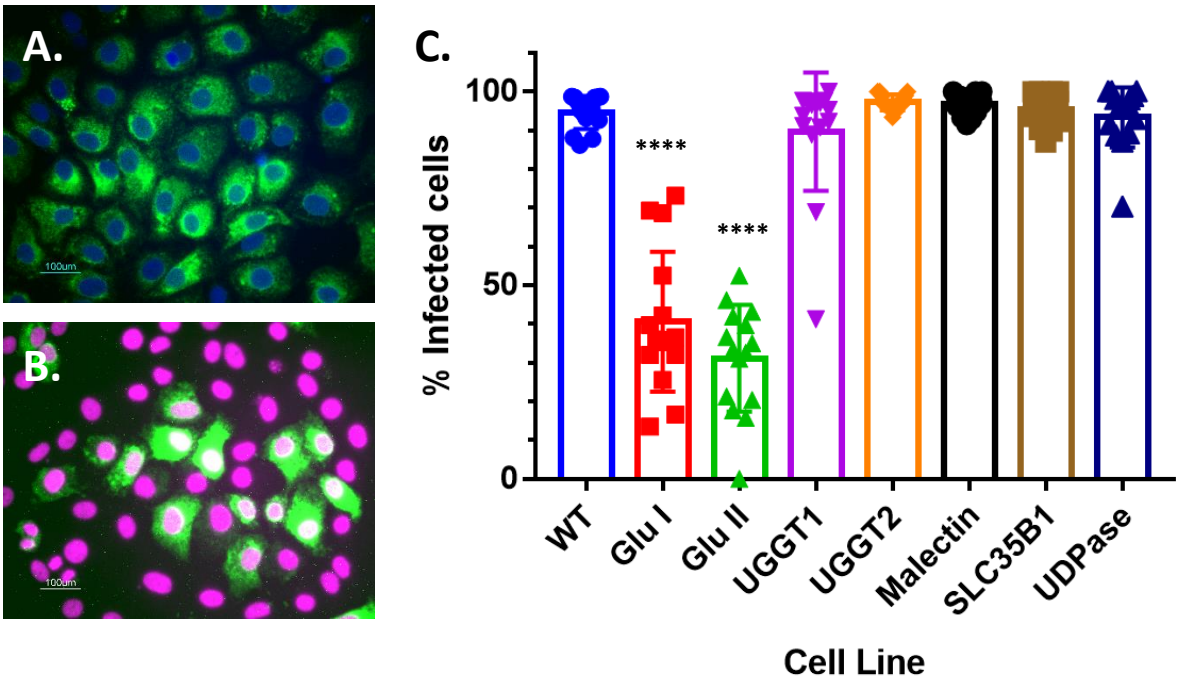


Figure 6.9. Immunofluorescence of dengue virus infected cells. Panels A and B show representative images obtained under the inverted microscope of cells 48 hours post infection. In blue are the DAPI-stained nuclei, in green, the FITC staining from the conjugated secondary antibody. Panel C shows the percentage of cells infected for each of the cell lines. **** = p-value of <0.0001. There were 50-150 cells per image, 5 images per well, 3 wells per cell line.

Whilst knock-out of either glucosidase is certain to affect the glycosylation of any of the existing glycosylated entry rececptors, it still may not affect function or fold of the receptor. From existing work with iminosugar inhibition of ER α-glucosidase II to combat dengue virus infection, we would expect a very strong antiviral effect within the cell, post viral entry, as iminosugar inhibitors that act through glu II are only added post-infection (MOI = 1) yet still lead to a near total loss of secreted virus[153]. Therefore, we expect fairly low levels of virus within these cell lines due to reduced production or increased degradation of the virus. In order to visualise any Env protein for the glucosidase knock-outs, the brightness and contrast had to be altered to nearly maximum (Figure 6.9, panel B), which was not the case for other cell lines. This is indicative of low levels of Env protein within the cell and may mean that virus is entering the cell but dengue Env protein is expressed/maintained at such low levels that it is barely detectable and thus responsible for the apparent low percentage of infected cells. Nevertheless, for future experiments, this approximate 3-fold drop in percentage of dengue-infected cells infected will be taken into account.

6.5. Zika antiviral assays

In order to measure whether the generated CRISPR knockout cell lines were more or less susceptible to viral infection, two different methods were used to measure virus released from cells. The first, qRT-PCR, measures the absolute quantity of viral RNA in the supernatant, the probes for which, in this case, target the sequence for Zika Env protein. The second, known as the plaque assay, uses dilution of the supernatant to assess how many plaque forming units are present in the original supernatant. Therefore, the first

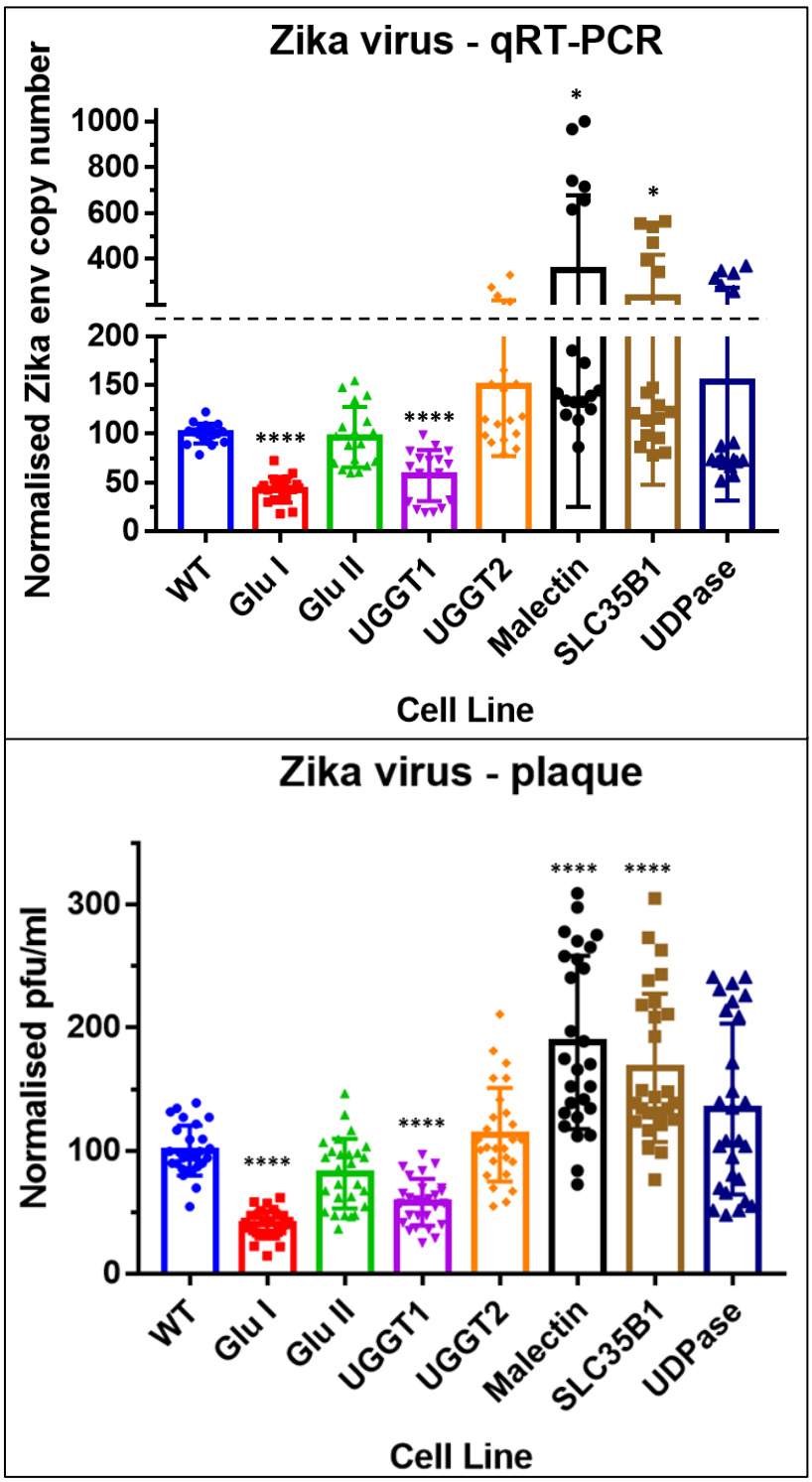


Figure 6.10. Top panel: qRT-PCR results showing Zika virus copy number as represented by Env protein RNA, normalised to the wild-type values. Bottom panel: Plaque assays results showing the normalised plaque forming units of infectious virus released from the cell lines shown. P-values: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001

method measures total virus released and the second measures total infectious virus released from cells.

In figure 6.10, the results from both measures of secreted virus are presented. The results are normalised from three separate biological replicates, each of which were in technical replicate: n = 6 for qRT-PCR, n = 9 for the plaque assay. A one-way ANOVA analysis with a Dunnett's multiple comparison test was used to compare each cell line's results with that of the wild-type line, to see if there were any significant differences.

The ER α-glucosidase I knock-out cell lines releases less than half the quantity of Zika virus the wild-type cells release, as measured by both qRT-PCR and the plaque assay. This indicates that Zika virus release is somewhat dependent on the actions of glu I within the cells, be it due to a direct effect on the folding of Zika virus glycoproteins or indirectly by affecting another protein that Zika virus requires. However, knock-out of glu II, the second enzyme required for glycoproteins to associate with the ER lectins, calnexin and calreticulin, has no effect on the release of virus from infected cells.

Knock-out of UGGT1 from Vero cells also reduces the susceptibility of the cells to Zika virus infection, as measured by both assays. UGGT1 is an enzyme downstream of both of the ER α-glucosidases, and so any effect observed in the knock-out of UGGT1 should also be observed in the glu II knock-out. This is not the case, and therefore may hint at a possible unknown mechanism by which glu II is bypassed, or a calnexin-independent role that UGGT1 and glu I are both involved in.

Knock-out of UGGT2, like glu II, has no effect on the release of Zika virus from infected cells. This highlights the different roles that the two UGGT isoforms play within mammalian cells, as they do not recognise the same client glycoproteins.

Knock-outs of both malectin and the nucleotide-sugar antiporter, SLC35B1, appear to be proviral: more virus is released from these knock-out lines when compared to wild-type cells. This would indicate that in wild-type cells, both proteins actually play a role in reducing the release of Zika virus from infected cells. For SLC35B1, it casts doubt on its role as the sole UDP-glucose: UMP antiporter, as if this were the case, we would expect an effect to mimic that of UGGT1. It is therefore likely that at the very least, the antiporter is involved in other processes rather than simply the supply of substrate to the quality control checkpoint of the calnexin cycle.

Finally, the knock-out of ENTPD5, the ER UDPase, has no significant effect. Once again, if its role were simply to hydrolyse the inhibitory product of UGGT, we would expect the knock-out to mimic the effect of the UGGT1 knock-out. The fact that deletion of ENTPD5 has no effect is observed indicates that there may be redundant systems within the ER to carry out UDP hydrolysis, or that ENTPD5 plays other important roles within the ER.

Of note, immunofluorescence data suggested that fewer cells would actually be infected with Zika virus in ENTPD5 knock-out cells (Figure 6.8). However, this does not seem to lead to significantly less virus being released from the cell, possibly due to greater variation within these two assays used to measure virus release.

Also of note is the contrast of these results involving the knock-out of ENTPD5 with the results obtained using ENTPD5 inhibitors, the latter of which appeared to be strongly

antiviral against Zika virus infection. Other nucleotide diphosphate hydrolases within the ER, as mentioned in section 5.8, could explain those results if the inhibitors would target them all, unlike the effect of the ENTPD5 knock-out.

Overall, the results from this work cast some doubt that the Zika virus glycoproteins from this strain (MR 766) rely on calnexin cycle mediated folding. However, inhibitors of either glu I or UGGT1, two major platers commonly known only in the context of the calnexin cycle, have potential as antivirals against this virus.

6.6. Dengue virus - antiviral assays

The idea behind these sets of experiments is much the same as for the Zika virus work: observe whether the knock-outs of the different proteins affect how much virus is released from infected cells. In this set of work, the two glucosidases have already been shown to possibly affect the entry receptors for dengue virus, leading to approximately just 30% of cells being infected if this is the case.

For these dengue virus experiments, the glucosidase knock-outs can be used as controls as we know that inhibition of their activity with iminosugars leads to a strong antiviral effect which the results shown in figure 6.11 strongly agree with. The plaque assay measures infectious virus and the levels of infectious dengue virus were barely detectable. qRT-PCR is a much more sensitive technique, but even in this case the levels of detectable virus were reduced around 100-fold. Even if fewer cells are actually infected by dengue virus in these cell lines, it is clear that they also have a large reduction in the secretion of virus, validating them once more as targets for antiviral drugs.

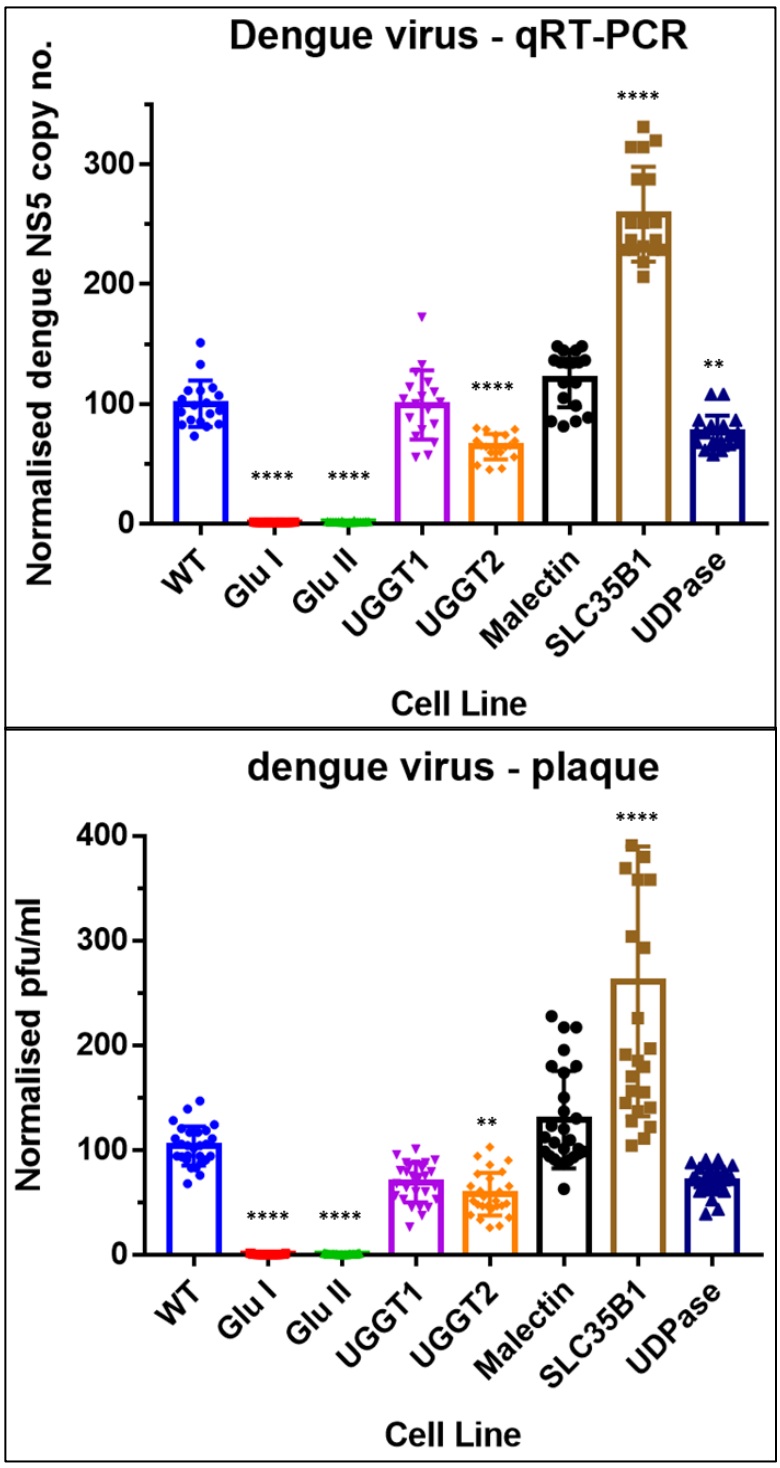


Figure 6.11. Top panel: qRT-PCR results showing dengue virus copy number as represented by NS5 RNA, normalised to the wild-type values. Bottom panel: Plaque assays results showing the normalised plaque forming units of infectious virus released from the cell lines shown. P-values: * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001

Contrary to the observations for Zika virus-infected cells, the UGGT1 knock-out cell does not appear to significantly affect viral secretion, whereas the UGGT2 knock-out cells appear to have a reduced level of viral secretion. Once again, these results highlight how client glycoproteins appear to prefer a single one of the two UGGT isoforms present in mammalian cells. In this case, inhibition of viral secretion looks like it may be via the calnexin cycle, evidenced by the effects of both glucosidase knock-out lines, and it is possible that UGGT2 is the enzyme that dengue virus relies upon for re-glucosylation of its misfolded glycoproteins. The magnitude of the reduction of viral secretion from infected cells is much less in the UGGT2 knock-out, but this is unsurprising; any viral particles that folded correctly do not require UGGT2's activity, only the subset that has not achieved its correct fold is required.

Knock-out of malectin, which appeared pro-viral in the case of Zika virus infection, has no effect on dengue virus secretion, indicating that the two viruses utilise different pathways. However, for the cell line lacking SLC35B1, there is once again an increase in virus secreted from infected cells, as measured by both qRT-PCR and the plaque assay. Due to the effect being the opposite of that observed for UGGT2, it is unlikely that it is having any effect due to alteration of the concentration of UGGT substrate within the ER, but some other, unknown mechanism.

Finally, knock-out of ENTPD5, the ER-UDPase, appears to show a reduction in virus released from cells to around 75% of wild-type levels but only significant if measured by qRT-PCR. In the plaque assay, the mean is reduced to 70% of wild-type levels but the p-value is non-significant (0.07). Nevertheless, the trend is that knock-out of ENTPD5 could prove to be slightly antiviral towards dengue virus.

Overall, the results show that the existing antiviral drug targets, ER α-glucosidase I and II, are unsurprisingly validated as antiviral targets. The inhibition of UGGT2 may also prove to be a valid target for antivirals, although the level of inhibition is of a different magnitude to that of the glucosidases, which may limit its usefulness.

6.7. Discussion

Within the last decade, the appearance of the CRISPR/Cas9 system and its derivatives have changed how scientific problems have been approached within the laboratory and beyond[341]. Presented in this chapter is the successful application of the CRISPR/Cas9 system to knock out target genes within Vero cells, a commonly used cell line for antiviral assays. The work has thoroughly characterised the mutations introduced and the knock-out cell lines have been tested in antiviral assays with Zika and dengue viruses.

Proteins were chosen for knock out for a number of reasons. The two ER α-glucosidases were picked as they have already been characterised as antivirals against dengue virus, thus these could act as controls[153]. The two isoforms of UGGT were chosen due to their role as folding sensors in the calnexin cycle and as enzymes they have an easily identifiable active site to be targeted with inhibitors[64, 342]. The putative UDP-glucose: UMP antiporter, SLC35B1, and the ER-UDPase, ENTPD5, are both implicated in a calnexin cycle associated cycle that supplies substrate to UGGT and degrades the inhibitory product, as seen in figure 1.5. If a knock-out of UGGT has an effect on viral replication, it is conceivable that knock-outs of the UDP-glucose: UMP antiporter or an ER UDPase knock-out may also have an effect. Finally, malectin, an orphaned ER-resident lectin may also be able to recognise viral glycoproteins at an early stage of folding, and thus a knock-

out could help investigating the role of di-glucosylated glycan intermediates in glycoprotein folding[124].

Sequencing of clonal populations for each of the genes to be targeted found a knock-out in each case, despite it not always being the intended full excision of DNA that was designed. FOS analysis of the glucosidase knock-outs also confirmed that their activity had been abrogated. The analysis further indicated an altered folding environment in each of the knock-outs, as evidenced by an increase in detectable FOS, which would suggest more glycoproteins being directed out of the ER for degradation.

The immunofluorescence experiments were carried out to check if the percentage of cells infected for each of the cell lines was altered. This is important, as it is unrealistic to suggest that cells would be exposed to an antiviral drug for many days prior to exposure to virus, mirroring the effect that a knock-out of a drug target may have. In the large majority of cases, no differences were observed for the ability of viruses to infect the different cell lines; the knock-out of ENTPD5 did appear to lead to a slight reduction in Zika virus infection, although this did not lead to a subsequent reduction in virus secretion. The reasons behind this reduced infectivity are unknown but the primary putative entry receptor, Axl, is a glycoprotein: if ENTPD5 acted through the calnexin cycle to bring about this effect, it would also be seen in the other knock-outs, especially the glucosidases.

In the immunofluorescence experiments with dengue virus, both glucosidase knock-outs appeared to have significantly fewer infected cells. Yet, existing data regarding the inhibition of the glucosidases with iminosugar inhibitors show that at high concentrations, a near total reduction of viral secretion is achieved[153]. It would follow that

a low level of virus is expected in these cells and could explain why so few appeared infected: the levels may have simply been undetectable. Nonetheless, strong, anti-dengue effects were observed in the plaque assay and qRT-PCR experiments for both glucosidase knock-outs, which were far stronger than the low percentage of infected cells could explain.

The results presented in this chaper helped to validate this approach of screening candidates of putative antiviral targets. The known reliance of dengue virus secretion on the ER glucosidases was clearly evident from the results of these assays.

The results of the glucosidase knock-outs on Zika virus secretion was not as clear-cut as for dengue virus. Only the knock-out of ER α-glucosidase I led to a significant reduction in secreted Zika virus and to only a level of approximately 40% that of wild-type. These data predict the efficacy of iminosugars against Zika virus would be limited, due to two reasons:

1. Viral replication will be inhibited to a lesser extent than that for dengue virus
2. Iminosugars have a greater affinity *in vivo* for glu II, the knock-out of which does not affect Zika virus secretion from cells

It is of interest that the glu I knock-out had an effect against Zika virus but the glu II knock-out did not. Both glucosidases are required for the association of a nascent glycoprotein with the ER lectins calnexin and calreticulin. This suggests that the effect seen for glu I is independent of calnexin-mediated folding. A secondary explanation would be that glycoproteins bearing the glu I product, a glycoprotein with two glucose moieties, interacts with malectin, the ER-lectin that can recognise such a species. Yet, the

knock-out of malectin leads to a higher secretion of Zika virus, inconsistent with this idea. It is therefore likely that the antiviral effect of the glu I knock-out is via a mechanism which is currently unknown.

Against Zika virus, the UGGT1 knock-out cell line is the only cell line other than that of the glu I knock out that led to a reduction in the secretion of Zika virus. As a mechanism via the calnexin cycle has already been ruled out, it is likely that this comes about via an alternate mechanism. It has recently been suggested that UGGT1 can be exported from the ER and to act as a scaffold for a virus' RNA polymerase machinery[75]. Our results do not discount this mechanism.

Unexpectedly, UGGT1 knockout does not have an effect against dengue virus whereas that of UGGT2 does. Conventionally, UGGT1 has been referred to as the active isoform whereas no role for UGGT2 was postulated[61]. More recently, a role of UGGT2 in reglucosylation of misfolded proteins has been all but confirmed[64, 343]. The data presented here corroborate these recent results, and may suggest that some dengue virus glycoproteins fold under the control of UGGT2 but not UGGT1, whilst the reverse may be true for ZIKV glycoproteins. Whilst the data suggest that an inhibitor of UGGT2 may reduce dengue virus replication, it would be a large jump to propose it as an alternative to iminosugars, as the reduced levels of secreted virus are still not comparable to that of cells in which the glucosidases are inhibited or knocked out.

Of the two proteins suggested to play a role in the UDP-glucose cycle (Figure 1.5), ENTPD5 does not impact either virus whereas the proposed UDP-Glc: UMP antiporter appears to cause a proviral effect when removed from cells. The latter is exactly opposite to the anticipated effect, but neither agree with the effects seen with either UGGT isoform

knock-out, suggesting the model proposed earlier in this thesis is, at best, incomplete. It could be that SLC35B1 is not the UDP-glucose: UMP antiporter, or that there are redundant systems in place in ER, as has been suggested for the ER-UDPase, ENTPD5. Indeed in yeast, disruption of all nucleotide sugar transporter genes whose productions localise to the ER or have an unknown location only partially abrogated UDP-Glc entrance into the ER[344]. Further study would be required to characterise levels of UDP-Glc, UDP and UMP in the ER, involving the knock-out of multiple putative proteins that could be providing the redundancy to either ENTPD5 or SLC35B1.

Of interest is the observation that the results from both qRT-PCR and plaque assay always mirror one-another: indicating there is no effect on viral specific infectivity, as has been suggested as the antiviral mechanism of some iminosugars against certain viral infections. Regardless, no clear new alternative targets for broad-spectrum antivirals have been identified, supporting continued optimisation of the inhibition of the glucosidases.

7. Conclusions

Presented in this thesis is work carried out over the past three and a half years to characterise proteins involved in the calnexin cycle and to assess, in principle, whether they could represent novel targets for broad-spectrum antivirals.

Here, the first atomic-resolution structure of a mammalian ER α-glucosidase I was determined. Previous studies on the *S.cerevisiae* orthologue found the active site occluded by an expression tag, forming crystal contacts that could not be displaced without

disruption of the crystal packing. The *Mus musculus* structures presented here are in complex with iminosugar inhibitors and reveal how the deoxynojirimycin ring is coordinated within the active site of the protein. Whilst none of the ligands tested had an alkyl 'tail' in a defined position, the structures may allow for the design of new, specific inhibitors that interact with side chains unique to glu I.

UDP-glucose: glycoprotein glucosyltransferase acts as the quality control checkpoint for the calnexin cycle. The modulation of UGGT activity has been hypothesised to act as both a broadly-acting antiviral strategy and also to ameliorate protein-folding diseases. High resolution structures would facilitate the discovery of ligands through structure-based lead discovery. Our 2.78 Å resolution structure of a double cysteine mutant of *Chaetomium thermophilum* UGGT is a step towards this and is the highest resolution full-length structure of a UGGT to date. SAXS studies of the chimeric complex between UGGT and Sep15, a co-chaperone that enhances the activity of UGGT, have yielded the first low resolution models of the two proteins together. The interface is a potential site for small molecule modulators of UGGT.

The ER-UDPase, otherwise known as ENTPD5, was also a focus of study in this D.Phil. ENTPD5's role in the Warburg effect, an important process involved in many cancers, has been previously confirmed and led to high-throughput screening and isolation of a number of lead compounds. Structural characterisation of ENTPD5 could aid in the further development of inhibitors and lead to a deeper understanding on a mechanistic level of how this protein works. Quantities of protein suitable for crystallisation were obtained for a human ENTPD5 construct and that of the *Chaetomium thermophilum* paralog, ENTPD6. Analysis of their substrate specificities found that they are both

broadly active as nucleotide diphosphatases with a similar activity towards UDP and GDP. An ENTPD5 inhibitor appeared to greatly reduce the levels of both total and infectious virus released from cells infected with Zika. However, a knock-out cell line of ENTPD5 did not demonstrate the same antiviral resistance. The discrepancy between the inhibitor study and the knock-out cell line may be due to redundant nucleotide diphosphatases within the ER that are inhibited by the molecule tested.

CRISPR/Cas9 was used to test whether selected calnexin cycle components could be valid targets for antiviral inhibitors. Two members of the *Flaviviridae*, Zika virus and dengue virus, were used based on their known reliance on N-glycosylation for proper folding or function. The results unexpectedly demonstrated that Zika virus does not rely on the calnexin cycle to the same extent that dengue virus does. For cells infected with dengue virus, no other knock-out cell lines were able to match the level of reduced susceptibility to virus secretion seen in the knock-outs of the ER glucosidases. These data support continued optimisation of ER glucosidase inhibition as the best calnexin cycle target for development of antiviral drugs.

Of note was that Zika and dengue viruses relied on UGGT1 and UGGT2 to different extents, confirming the likely role of the latter in the folding of glycoproteins. This is especially surprising given how similar Zika and dengue viruses are from both a genomic and structural perspective. Further studies are required to assess whether the knock-outs of these proteins carry the same phenotypes in other cell-lines. For example, the pathogenicity of Zika virus is in part due to its effect in neuronal cells, and so assays in this system would be of interest to test. Also, a wide variety of enveloped viruses outside of the *Flaviviridae* family encode glycoproteins; it is therefore of interest to study the

reliance of further viruses on the calnexin cycle: perhaps some may rely on the activities of the UGGTs to a much greater extent than that seen here.

Appendix A

Primers used in this thesis

Chapter Reference	Full Primer Name.	Sequence (5' to 3')	Tm (°C)	GC content (%)	Comment
2.9.5	Zika MR766 1086	TCGTTGCCCAACACAAG	63	53	Primers for qRT-PCR against Zika Env
2.9.5	Zika MR766 1162r	CCACTAATGTTCTTTTGCAGACAT	69	60	Primers for qRT-PCR against Zika Env
2.9.5	Zika MR766 1107 probe	(6FAM))AGCCTACCTTGACAAGCAA TCAGACACTCAA(TAM)			FAM-TAMRA probe against Zika Env
2.9.5	F NS5 5′ primer	ACAAGTCGAACAACCTGGTCC	67	52	Primers for qRT-PCR against Denv NS5, forward primer
2.9.5	R NS5 3′ primer	GCCGCACCATTGGTCTTCTC	69	60	Primers for qRT-PCR against Denv NS5, reverse primer
2.9.5	NS5_qRT_probe	(6FAM))CCAGTGGAATCATGGGAGG AAATCCCA(TAM)			FAM-TAMRA probe against Denv NS5
4.1.1	CtUGGT_pHL2Lit_Fwd	GTGCTGCTGATGGGTTGCGTAGCTG AAACCGGTGTCGCAGCCTCT	72	60	Gibson assembly primers to move CtUGGT from its pHLsec vector to the Litmus28i cloning vector
4.1.1	CtUGGT_pHL2Lit_Rev	TCATTAGTGATGGTGATGGTGGTGC TTGGTACCCTCCCGAACCGTCTTG	72	53	Gibson assembly primers to move CtUGGT from its pHLsec vector to the Litmus28i cloning vector
4.1.1	D611C_F	TCCACGGGAGTGCAATTGGCTGC	60	61	Site-directed-mutagenesis primers to mutate Asp 611 to Cys
4.1.1	D611C_R	ATTGGGAAACCATCAAAGAAG	60	38	Site-directed-mutagenesis primers to mutate Asp 611 to Cys
4.1.1	G1050C_F	CGAGATTCCCTGCGCTCATGCCC	65	65	Site-directed-mutagenesis primers to mutate Gly 1050 to Cys
4.1.1	G1050C_R	CGCGAATGGCCCTCAATA	65	56	Site-directed-mutagenesis primers to mutate Gly 1050 to Cys
4.1.1	CtUGGT_Lit2pHL_Fwd	CGACTCACTATAGGCCTTGACTAG AGGGTACCCTCCCGAACCGTCTTG	72	56	Gibson assembly primers to move CtUGGT from the litmus28i vector back in the pHLsec
4.1.1	CtUGGT_Lit2pHL_Rev	TCCACGAAGCTTCCCATGGTGACGT CACCGGTGTCGCAGCCTCT	72	61	Gibson assembly primers to move CtUGGT from the litmus28i vector back in the pHLsec
4.1.2.	OPPF_UGGT1_Fwd	gcgtagctgaaaccggcGACTCAAAAGCCA TTACAACCTCTCT	60	51	OPPF In-Phusion Forward Primer for the PCR of Hs UGGT1 constructs
4.1.2.	OPPF_UGGT2_Fwd	gcgtagctgaaaccggcACGGTCGCCGCGT CCAAGTC	60	68	OPPF In-Phusion Forward Primer for the PCR of Hs UGGT2 constructs
4.1.2.	OPPF_UGGT1_Rev1	gtgatggtgatgtttGGGATTAAATATATA ATGTCTGAATACATA	60	29	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 1
4.1.2.	OPPF_UGGT1_Rev2	gtgatggtgatgtttCTTAATGGCCAATTCC ACGCCATAG	60	45	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 2
4.1.2.	OPPF_UGGT1_Rev3	gtgatggtgatgtttTCGTGGAACAACATT TGGCTGATTC	60	43	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 3
4.1.2.	OPPF_UGGT1_Rev4	gtgatggtgatgtttAGCATCCACCTTCATT ACCAAGTC	60	44	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 4
4.1.2.	OPPF_UGGT1_Rev5	gtgatggtgatgtttTAAAGGCATGTCAGA AAGTTTGGATT	60	37	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 5
4.1.2.	OPPF_UGGT1_Rev6	gtgatggtgatgtttGTATTCCAGCTCATAC TCAGCAGCC	60	48	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 6
4.1.2.	OPPF_UGGT1_Rev7	gtgatggtgatgtttCAAGTCTTCGTTCACC ATATCTGCC	60	45	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 7
4.1.2.	OPPF_UGGT1_Rev8	gtgatggtgatgtttTTTCTGAGGACCTTCT CGGCTTGG	60	49	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT1, Construct 8
4.1.2.	OPPF_UGGT2_Rev1	gtgatggtgatgtttTGGTTTCTGAATATAA TGGCGAAGA	60	38	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 1
4.1.2.	OPPF_UGGT2_Rev2	gtgatggtgatgtttCTTAATTGCTAGCTCC ACACCATAC	60	43	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 2
4.1.2.	OPPF_UGGT2_Rev3	gtgatggtgatgtttTATACGGGGTACAAC ATTATTCCTAT	60	37	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 3
4.1.2.	OPPF_UGGT2_Rev4	gtgatggtgatgtttGGCATCAACTTTCAT AATAAAGTCACT	60	36	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 4
4.1.2.	OPPF_UGGT2_Rev5	gtgatggtgatgtttTAAAGGGGCTTCTGA AAGCCTGCC	60	49	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 5
4.1.2.	OPPF_UGGT2_Rev6	gtgatggtgatgtttGTATTCTAGTTCATAT TCTGCTGTAAC	60	36	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 6
4.1.2.	OPPF_UGGT2_Rev7	gtgatggtgatgtttGATATCTTCCTTAATT TTGTCTGTTTC	60	33	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 7
4.1.2.	OPPF_UGGT2_Rev8	gtgatggtgatgtttTGTCAAAATTGTATCT TGCTTCTTGTTT	60	33	OPPF In-Phusion Reverse Primer for the PCR of Hs UGGT2, Construct 8

4.1.3.	LitUGGT_Fwd	TCCACGAAGCTTCCCATGGTGACGT CACCGGTGTCGCAGCCTCT	72	61	Primers to Gibson Assembly Ct UGGT from pHLsec to litmus28i
4.1.3.	LitUGGT_Rev	CGACTCACTATAGGCCTTGACTAG AGGGTACCCTCCCGAACCGTCTTG	72	56	Primers to Gibson Assembly Ct UGGT from pHLsec to litmus28i
4.1.3.	320_Ctloop_fwd	CTGGAGAAATCCGAACTTG	59	47	Primers to amplify Ct_UGGT in litmus28i without the 249-284 loop.
4.1.3.	321_Ctloop_rev	GTCATCAATGACGATGTAATC	59	38	Primers to amplify Ct_UGGT in litmus28i without the 249-284 loop.
4.1.3.	CtUGGT_pHLsec_F	GGTTGCGTAGCTGAAACCGGTCAA GTCGCAGCCTCTCCA	72	59	Primers to Gibson Assembly Ct UGGT from Litmus28i to pHLsec
4.1.3.	CtUGGT_pHLsec_R	GATGGTGGTGCTTGGTACCCTCCCG AACCGTCTTGAC	72	59	Primers to Gibson Assembly Ct UGGT from Litmus28i to pHLsec
4.1.3.	M13F	CAG GAA ACA GCT ATG ACC	48	50	Sequencing primer for the M13F site (used for Litmus28i)
4.1.3.	M13R	TGT AAA ACG ACG GCC AGT	48	50	Sequencing primer for the M13F site (used for Litmus28i)
4.1.3.	T7_Promoter (F)	TAA TAC GAC TCA CTA TAG GG	48	53	Sequencing primer for the T7F site (pET-vectors)
4.1.3.	T7_Terminator (REV)	GCT AGT TAT TGC TCA GCG G	51	50	Sequencing primer for the T7F site (pET-vectors)
5.1.	3_HsENTPD5_303_fwd	CTTATCTCTAGAATGGTCTCCCACC GGAAC	69	44	Primer to clone *Hs* ENTPD5 from pET/M vectors to pET303 by addition of 5' noise, XbaI site and start codon
5.1.	4_pET303_His_rev	GCTTTGTTAGCAGCCGGATCTCAGT G	70	54	Primer to clone C-term to N-term, starting after the HisTag and stop codon, of pET/M and pET vectors
5.1.	5_MmENTPD5_303_fwd	CTTATCTCTAGAATGGTCTTTTACC GCGAGCA	71	50	Primer to clone *Mm* ENTPD5 from pET/M vectors to pET303 by addition of 5' noise, XbaI site and start codon
5.1.	6_HsE5_Tru_pETM_f	TTATCCCATGGTCATCAACGTGTCC GCCAG	73	53	Primer to truncate *Hs* ENTPD5 in pET/M vectors to start from N' SASTLY, adds a restriction site for easy cloning back into pET/M
5.1.	7_pETM_Rev	GGCTTTGTTAGCAGCCGGATCTCAG TG	72	56	Primer to clone C-term to N-term, starting after the HisTag and stop codon, of pET/M and pET vectors
5.1.	8_MmE5_Tru_pETM_f	TTATCCCATGGTCATCAACGTGTCC GCCG	75	55	Primer to truncate *Mm* ENTPD5 in pET/M vectors to start from N' SASTLY, adds a restriction site for easy cloning back into pET/M
5.1.	9_HsE5_Tru_303_f	CTTATCTCTAGAATGATCAACGTGT CCGCCAG	70	47	Primer to truncate *Hs* ENTPD5 in pET/M vectors to start from N' SASTLY, adds a restriction site for easy cloning back into pET303
5.1.	10_MmE5_Tru_303_f	CTTATCTCTAGAATGATCAACGTGT CCGCCGG	70	50	Primer to truncate *Mm* ENTPD5 in pET/M vectors to start from N' SASTLY, adds a restriction site for easy cloning back into pET303
5.1.	11_HsE5_Tru_pHL_f	CTTATCACCGGTATCAACGTGTCCG CCAG	73	55	Primer to truncate *Hs* ENTPD5 in pHLsec vector to start from N' SASTLY, adds a restriction site for easy cloning back into pHLsec
5.1.	12_pHLsec_rev	CAGCCTGCACCTGAGGAGTGAATT	70	54	Primer to clone C-term to N-term, starting after the HisTag and stop codon, of pHLsec
5.1.	13_MmE5_Tru_pHL_f	CTTATCACCGGTATCAACGTGTCCG CCG	73	57	Primer to truncate *Mm* ENTPD5 in pHLsec vector to start from N' SASTLY, adds a restriction site for easy cloning back into pHLsec
5.1.	14_pHLsec_fwd	TTCTCTCCGTGCTGCTGATG	65	55	Forward primer for sequencing pHLsec
5.1.	15_pHLsec_rev	GCACCTGAGGAGTGAATTATCGC	66	52	Reverse primer for sequencing pHLsec
5.1.	16_pETM_fwd	CTGGAAGTTCTGTTCCAGGGG	66	57	Forward primer for sequencing pETM vectors from the cleavage site onwards
5.1.	17_pETM_rev	CTTTGTTAGCAGCCGGATCTCA	65	50	Reverse primer for sequencing pETM vectors
5.1.	292_CtE6_pHL_fwd	GCTGATGGGTTGCGTAGCTGAAAC CGGTACGCGCGCCTACTCCAAG	71	67	Gibson assembly primers to amplify Ct ENTPD6 cDNA and then clone into pHLsec
5.1.	293_CtE6_pHL_rev	ATGGTGATGGTGGTGCTTGGTACCA TGAATCTCCTTCACCTTGCACTGC	69	48	Gibson assembly primers to amplify Ct ENTPD6 cDNA and then clone into pHLsec
5.1.	294_CtE6_pETM_fwd	CTGGAAGTTCTGTTCCAGGGGCCC ATGGACACGCGCGCCTACTCCAAG	71	67	Gibson assembly primers to amplify Ct ENTPD6 cDNA and then clone into pETM vectors
5.1.	295_CtE6_pETM_rev	TGGTGGTGCTCGAGTGCGGCCGCA TGAATCTCCTTCACCTTGCACTGC	69	48	Gibson assembly primers to amplify Ct ENTPD6 cDNA and then clone into pETM vectors
5.2.	296_E6_insert_fwd	GCTTGCCGCTTCTGTCTAAGG	69	57	SDM primers to remove the insert in the CtENTPD6 DNA that was cloned into pETM vector. Annealing temp = 70°C
5.2.	297_E6_insert_rev	TTGCCCCCAAGCACCACC	72	67	SDM primers to remove the insert in the CtENTPD6 DNA that was cloned into pETM vector. Annealing temp = 70°C
6.2	68_CRSPR_CsG1_1	CACCGGCGGTAGGTTCCCCAGAAG			CRISPR/Cas9 oligo to target ER α-glucosidase I, for insertion into PX459 plasmid, primer pair 1
6.2	69_CRSPR_CsG1_1	AAACCTTCTGGGGAACCTACCGCC			CRISPR/Cas9 oligo to target ER α-glucosidase I, for insertion into PX459 plasmid, primer pair 1
6.2	70_CRSPR_CsG1_2	CACCGATCATGCTCTGAATTGCGC			CRISPR/Cas9 oligo to target ER α-glucosidase I, for insertion into PX459 plasmid, primer pair 2
6.2	71_CRSPR_CsG1_2	AAACGCGCAATTCAGAGCATGATC			CRISPR/Cas9 oligo to target ER α-glucosidase I, for insertion into PX459 plasmid, primer pair 2
6.2	76_CRSPR_CsG2_1	CACCGGCTTCAGAAGAACATGACT			CRISPR/Cas9 oligo to target ER α-glucosidase II, for insertion into PX459 plasmid, primer pair 1
6.2	77_CRSPR_CsG2_1	AAACAGTCATGTTCTTCTGAAGCC			CRISPR/Cas9 oligo to target ER α-glucosidase II, for insertion into PX459 plasmid, primer pair 1
6.2	78_CRSPR_CsG2_2	CACCGGATCAAAAGACCCAGCTGA			CRISPR/Cas9 oligo to target ER α-glucosidase II, for insertion into PX459 plasmid, primer pair 2
6.2	79_CRSPR_CsG2_2	AAACTCAGCTGGGTCTTTTGATCC			CRISPR/Cas9 oligo to target ER α-glucosidase II, for insertion into PX459 plasmid, primer pair 2

6.2	116_CRSPR_CsU1_1	CACCGGTGGCGTGAGGAGGCCTC			CRISPR/Cas9 oligo to target UGGT1, for insertion into PX459 plasmid, primer pair 1
6.2	117_CRSPR_CsU1_1	AAACTGAGGCCTCCTCACGCCACC			CRISPR/Cas9 oligo to target UGGT1, for insertion into PX459 plasmid, primer pair 1
6.2	118_CRSPR_CsU1_2	CACCGTGGCTGTTCTCCTCAGTAA			CRISPR/Cas9 oligo to target UGGT1, for insertion into PX459 plasmid, primer pair 2
6.2	119_CRSPR_CsU1_2	AAACTTACTGAGGAGAACAGCCAC			CRISPR/Cas9 oligo to target UGGT1, for insertion into PX459 plasmid, primer pair 2
6.2	124_CRSPR_CsU2_1	CACCGCTGAAACATCTGAATAGC			CRISPR/Cas9 oligo to target UGGT2, for insertion into PX459 plasmid, primer pair 1
6.2	125_CRSPR_CsU2_1	AAACAGCTATTCAGATGTTTCAGC			CRISPR/Cas9 oligo to target UGGT2, for insertion into PX459 plasmid, primer pair 1
6.2	126_CRSPR_CsU2_2	CACCGTGATTCTCTATGCCGAATT			CRISPR/Cas9 oligo to target UGGT2, for insertion into PX459 plasmid, primer pair 2
6.2	127_CRSPR_CsU2_2	AAACAATTCGGCATAGAGAATCAC			CRISPR/Cas9 oligo to target UGGT2, for insertion into PX459 plasmid, primer pair 2
6.2	132_CRSPR_CsMl_1	CACCGCTGGTCTTGAAATTTGCTG			CRISPR/Cas9 oligo to target malectin, for insertion into PX459 plasmid, primer pair 1
6.2	133_CRSPR_CsMl_1	AAACCAGCAAATTTCAAGACCAGC			CRISPR/Cas9 oligo to target malectin, for insertion into PX459 plasmid, primer pair 1
6.2	134_CRSPR_CsMl_2	CACCGAAATCAAAATGATTTGTTG			CRISPR/Cas9 oligo to target malectin, for insertion into PX459 plasmid, primer pair 2
6.2	135_CRSPR_CsMl_2	AAACCAACAAATCATTTTGATTTC			CRISPR/Cas9 oligo to target malectin, for insertion into PX459 plasmid, primer pair 2
6.2	140_CRSPR_CsSLC_1	CACCGCAGCTCCCTGGTGCCCGAC			CRISPR/Cas9 oligo to target SLC35B1, for insertion into PX459 plasmid, primer pair 1
6.2	141_CRSPR_CsSLC_1	AAACGTCGGGCACCAGGGAGCTGC			CRISPR/Cas9 oligo to target SLC35B1, for insertion into PX459 plasmid, primer pair 1
6.2	142_CRSPR_CsSLC_2	CACCGGTAGTCTCTCTCTTCCTAG			CRISPR/Cas9 oligo to target SLC35B1, for insertion into PX459 plasmid, primer pair 2
6.2	143_CRSPR_CsSLC_2	AAACCTAGGAAGAGAGAGACTAC C			CRISPR/Cas9 oligo to target SLC35B1, for insertion into PX459 plasmid, primer pair 2
6.2	220_CsE5_3_fwd	CACCGGAGACAGCGCTGCAAACAC			CRISPR/Cas9 oligo to target ENTPD5, for insertion into PX459 plasmid, primer pair 1
6.2	221_CsE5_3_rev	AAACGTGTTTGCAGCGCTGTCTCC			CRISPR/Cas9 oligo to target ENTPD5, for insertion into PX459 plasmid, primer pair 1
6.2	222_CsE5_4_fwd	CACCGTCACCTTTCCTGGTACCAA			CRISPR/Cas9 oligo to target ENTPD5, for insertion into PX459 plasmid, primer pair 2
6.2	223_CsE5_4_rev	AAACTTGGTACCAGGAAAGGTGAC			CRISPR/Cas9 oligo to target ENTPD5, for insertion into PX459 plasmid, primer pair 2
6.2	72_Int_CsG1	TGGCAGGATCTCTGAAGTGG	59	55	Non-Deletion Screening Primers for ER α-glucosidase I knock-out
6.2	73_Int_CsG1	CCAGGGCATGGGTCAGTAGA	61	60	Non-Deletion Screening Primers for ER α-glucosidase I knock-out
6.2	74_Ext_CsG1	GACCTCTGCCAGGATGGCT	61	63	Deletion Screening Primers for ER α-glucosidase I knock-out
6.2	75_Ext_CsG1	CCAGGGCTATGTGGGATGAC	60	60	Deletion Screening Primers for ER α-glucosidase I knock-out
6.2	80_Int_CsG2	GGCTGAGGGACCCTACAAGA	61	60	Non-Deletion Screening Primers for ER α-glucosidase II knock-out
6.2	81_Int_CsG2	TAAGCAGAAGGCTTCGGTCC	60	55	Non-Deletion Screening Primers for ER α-glucosidase II knock-out
6.2	82_Ext_CsG2	AGCGACAGAGAAGCATACGG	60	55	Deletion Screening Primers for ER α-glucosidase II knock-out
6.2	83_Ext_CsG2	TGAATGTCTCCTCCCAGGCT	60	55	Deletion Screening Primers for ER α-glucosidase II knock-out
6.2	120_Int_CsU1	CGTCACATTGAGCTTTCGCC	60	55	Non-Deletion Screening Primers for UGGT1 knock-out
6.2	121_Int_CsU1	CGCGAATAGGGGAAGGGATG	60	60	Non-Deletion Screening Primers for UGGT1 knock-out
6.2	122_Ext_CsU1	GCGTGTCGGTTTCTCACTGG	60	60	Deletion Screening Primers for UGGT1 knock-out
6.2	123_Ext_CsU1	GAGAGGTTGTAATGGCTTTTGAGT	59	42	Deletion Screening Primers for UGGT1 knock-out
6.2	128_Int_CsU2	CTGATGAGCCACCACCAGATG	61	57	Non-Deletion Screening Primers for UGGT2 knock-out
6.2	129_Int_CsU2	GCAGCTTTCTTCAGCAGCTTTT	61	45	Non-Deletion Screening Primers for UGGT2 knock-out
6.2	130_Ext_CsU2	TCCTGAAGAAAGCTGGACAGT	59	48	Deletion Screening Primers for UGGT2 knock-out
6.2	131_Ext_CsU2	TGGCGAAGAACATACAGAATTTCC	60	42	Deletion Screening Primers for UGGT2 knock-out
6.2	136_Int_CsMl	CTATGCCTCGGACAACAGCA	60	55	Non-Deletion Screening Primers for malectin knock-out
6.2	137_Int_CsMl	CTGCTCTCTACTCGCTCTGC	60	60	Non-Deletion Screening Primers for malectin knock-out
6.2	138_Ext_CsMl	AAACTGCCAATCCTGCGTTC	59	50	Deletion Screening Primers for malectin knock-out
6.2	139_Ext_CsMl	TAGCTTCAGAGCTGCCAAGG	60	55	Deletion Screening Primers for malectin knock-out

6.2	144_Int_CsSLC	TATTTTGGTCCCCTGACCTGC	60	52	Non-Deletion Screening Primers for SLC35B1 knock-out
6.2	145_Int_CsSLC	GGGATTGGCGAAGAGGATCA	60	55	Non-Deletion Screening Primers for SLC35B1 knock-out
6.2	146_Ext_CsSLC	GATGTCCGGCTGGATCTGTC	60	60	Deletion Screening Primers for SLC35B1 knock-out
6.2	147_Ext_CsSLC	GGAACTCAGTCCCATATTCCTATT	58	42	Deletion Screening Primers for SLC35B1 knock-out
6.2	224_CsE5_int2_fd	TGTGTTCTGGCAGTAAGCGT	60	50	Non-Deletion Screening Primers for ENTPD5 knock-out
6.2	225_CsE5_int2_rv	GAGAGGTTGGAGCCCTATGC	60	60	Non-Deletion Screening Primers for ENTPD5 knock-out
6.2	226_CsE5_ext2_fd	GGCCCCAGTTCCTTTCAGTT	60	55	Deletion Screening Primers for ENTPD5 knock-out
6.2	227_CsE5_ext2_rv	TGGAACCAGATACCCTTCGG	59	55	Deletion Screening Primers for ENTPD5 knock-out
6.2	240_CsG1_screen5	CCGTCGTGGTCCTGTCTTTG	61	60	'Scarring' primer pair to amplify a 647 base pair sequence around the first CRISPR site of the ER α-glucosidase I
6.2	241_CsG1_screen5	CTGGGCTGAGAAAAGGGTGT	60	55	'Scarring' primer pair to amplify a 647 base pair sequence around the first CRISPR site of the ER α-glucosidase I
6.2	242_CsG1_screen6	GCCCTTGGCTATGTCAGTCT	59	55	'Scarring' primer pair to amplify a 607 base pair sequence around the second CRISPR site of the ER α-glucosidase I
6.2	243_CsG1_screen6	GGGCTATGTGGGATGACAGT	59	55	'Scarring' primer pair to amplify a 607 base pair sequence around the second CRISPR site of the ER α-glucosidase I
6.2	244_CsG2_screen3	GACAGAGAAGCATACGGCCA	60	55	'Scarring' primer pair to amplify a 608 base pair sequence around the first CRISPR site of the ER α-glucosidase II
6.2	245_CsG2_screen3	TGGCAAGGTGCTATGCTAGT	59	50	'Scarring' primer pair to amplify a 608 base pair sequence around the first CRISPR site of the ER α-glucosidase II
6.2	246_CsG2_screen4	TCCTCCAATGCCTTCGTTCC	60	50	'Scarring' primer pair to amplify a 649 base pair sequence around the second CRISPR site of the ER α-glucosidase II
6.2	247_CsG2_screen4	ATGTCACCACAGAACTCACTCA	59	45	'Scarring' primer pair to amplify a 649 base pair sequence around the second CRISPR site of the ER α-glucosidase II
6.2	248_CsUGGT1_screen3	CTGGCAGGTCCCACACTAAA	60	55	'Scarring' primer pair to amplify a 663 base pair sequence around the first CRISPR site of the UGGT1
6.2	249_CsUGGT1_screen3	GCGTGTCGGTTTCTCACTGG	62	60	'Scarring' primer pair to amplify a 663 base pair sequence around the first CRISPR site of the UGGT1
6.2	250_CsUGGT1_screen4	GTGGAGTGGAAAACCATTTTGTTG	59	42	'Scarring' primer pair to amplify a 754 base pair sequence around the second CRISPR site of the UGGT1
6.2	251_CsUGGT1_screen4	TATGGTTGTGTGCAAGCTGGT	60	48	'Scarring' primer pair to amplify a 754 base pair sequence around the second CRISPR site of the UGGT1
6.2	252_CsUGGT2_screen3	GCTAGCATGCCTCTTTGTCC	59	55	'Scarring' primer pair to amplify a 604 base pair sequence around the first CRISPR site of the UGGT2
6.2	253_CsUGGT2_screen3	TGGCTATCGGCACTAAATCTCA	59	45	'Scarring' primer pair to amplify a 604 base pair sequence around the first CRISPR site of the UGGT2
6.2	254_CsUGGT2_screen4	CCCCAGAAGAGACAGAATCCC	60	57	'Scarring' primer pair to amplify a 821 base pair sequence around the second CRISPR site of the UGGT2
6.2	255_CsUGGT2_screen4	GCTAGACCGTAAGAGTCCTGG	59	57	'Scarring' primer pair to amplify a 821 base pair sequence around the second CRISPR site of the UGGT2
6.2	256_CsMl_screen3	TGTGTCTTGCCTCTGATGTGC	61	52	'Scarring' primer pair to amplify a 411 base pair sequence around the first CRISPR site of malectin
6.2	257_CsMl_screen3	TCTGGCTTTGCTTTCTGTCACT	60	45	'Scarring' primer pair to amplify a 411 base pair sequence around the first CRISPR site of malectin
6.2	258_CsMl_screen4	TGACCAGCACTGGGAGTTAG	59	55	'Scarring' primer pair to amplify a 629 base pair sequence around the second CRISPR site of malectin
6.2	259_CsMl_screen4	GGTCAAGTGTTCTGTTGCTGC	60	52	'Scarring' primer pair to amplify a 629 base pair sequence around the second CRISPR site of malectin
6.2	260_CsSLC_screen3	AGACGGAGAGGAAACATGCAAA	60	45	'Scarring' primer pair to amplify a 615 base pair sequence around the first CRISPR site of SLC35B1
6.2	261_CsSLC_screen3	ATGTCCGGCTGGATCTGTCG	62	60	'Scarring' primer pair to amplify a 615 base pair sequence around the first CRISPR site of SLC35B1
6.2	262_CsSLC_screen4	GGAACTCAGTCCCATATTCCTATTG	59	44	'Scarring' primer pair to amplify a 967 base pair sequence around the second CRISPR site of SLC35B1
6.2	263_CsSLC_screen4	GGTGGTTTATGCCTGTAATTCTT	57	39	'Scarring' primer pair to amplify a 967 base pair sequence around the second CRISPR site of SLC35B1
6.2	310_CR33_fwd	AGCGTATTTGTTTCTGGGAGGA	60	45	'Scarring' primer pair to amplify a ~500 base pair sequence around the first CRISPR site of ENTPD5
6.2	311_CR33_rev	CCCACATTACCCTGGTTTAGC	59	52	'Scarring' primer pair to amplify a ~500 base pair sequence around the first CRISPR site of ENTPD5
6.2	312_CR34_fwd	CTTTCTCTTCAGGACCCCGT	59	55	'Scarring' primer pair to amplify a ~500 base pair sequence around the second CRISPR site of ENTPD5
6.2	313_CR34_rev	TGGGAAGGCTCCTCCTAGTT	60	55	'Scarring' primer pair to amplify a ~500 base pair sequence around the second CRISPR site of ENTPD5

Appendix B

X-ray diffraction data collection parameters

	3.26 Å *Mm* glu I crystal	2.1 Å *Mm* glu I crystal	SR1 - C2_1
Beamline	DLS I02	DLS I04-1	ESRF ID30A-1
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS3 2M
Wavelength (Å)	0.97949	0.92819	0.966
Oscillation per image (°)	0.2	0.05	0.05
Number of images	900	3600	3300
Exposure per image (s)	0.2	0.05	0.1
Beam size (µm)	100x20	70	50
Transmission (%)	100	100	100
Detector Resolution Limit (Å)	2.7	2.2	2.43

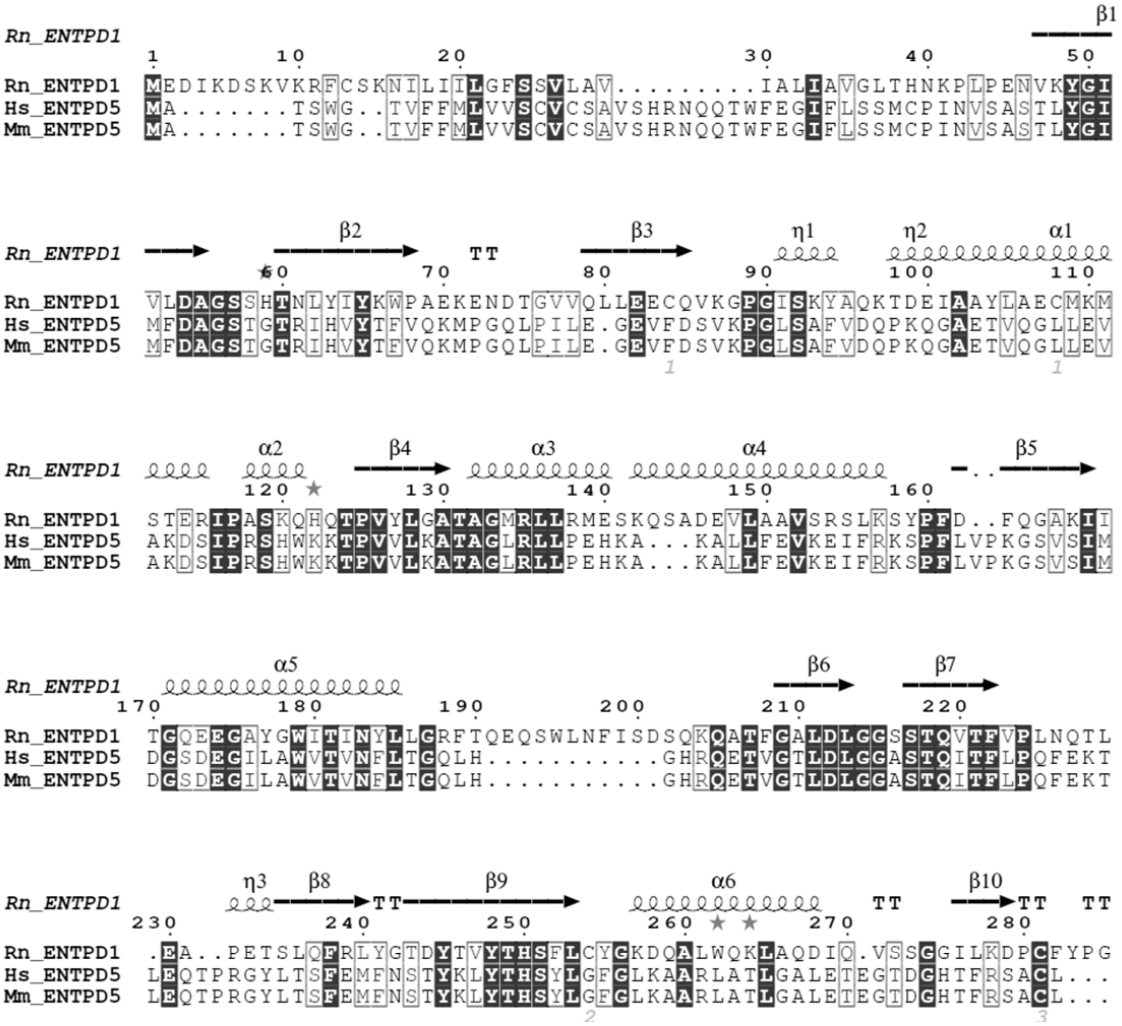
	SR2 - D3_1	SR3 - E6_1_2	OP1 - F10_1
Beamline	DLS I04	DLS I04	DLS I04
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Wavelength (Å)	1.28281	0.9795	0.9795
Oscillation per image (°)	0.1	0.1	0.1
Number of images	1800	1800	1800
Exposure per image (s)	0.22	0.25	0.25
Beam size (µm)	63x50	63x50	63x51
Transmission (%)	100	100	100
Detector Resolution Limit (Å)	2.5	2.5	2.5

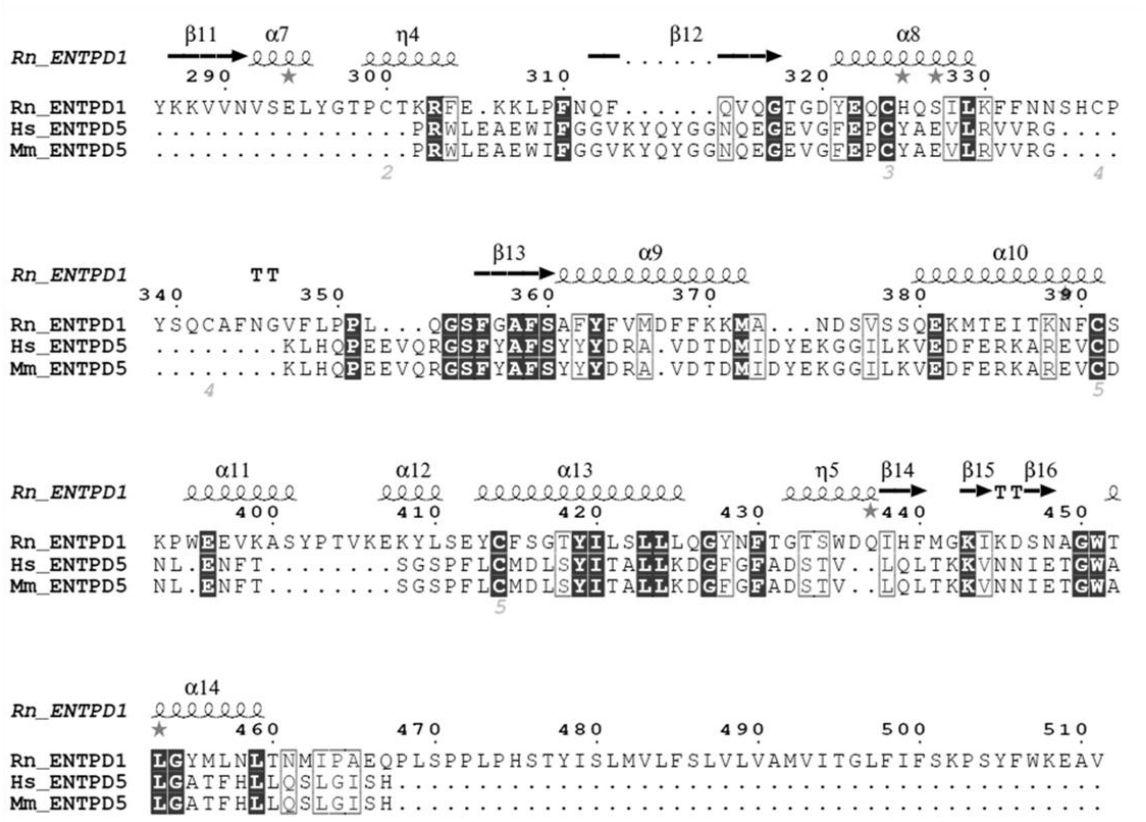
	AR1 - G3_1	*Ct* UGGT Chimera	*Ct* UGGT D611C:G1050C
Beamline	EMBL P14	DLS I03	ESRF ID30A-1
Detector	EIGER 16M	PILATUS3 6M	PILATUS3 2M
Wavelength (Å)	0.97659	0.9763	0.966
Oscillation per image (°)	0.1	0.2	0.05
Number of images	1800	900	2720
Exposure per image (s)	0.5	0.1	0.088
Beam size (µm)		20x20	100
Transmission (%)	100	100	100
Detector Resolution Limit (Å)	1.76	3.5	2.96

Appendix C

Sequences used in the cloning of constructs throughout this thesis

Alignment of *Rn* ENTPD1 (PDB 3ZX3) with *Hs* ENTPD5 and *Ms* ENTPD5





Codon optimised sequences for *Hs* ENTPD5, lacking the signal peptide and including overlap for Gibson assembly cloning (highlighted in grey).

For cloning into pET/M vectors

*GTTCTGTTCCAGGGGC**C***CATGGTCATCAACGTGTCCGCCAGCACCCTGTACGGCA
TCATGTTCGATGCCGGCAGCACCGGCACCCGGATCCACGTGTACACCTTCGTGC
AGAAGATGCCTGGCCAGCTGCCCATCCTGGAAGGCGAGGTGTTCGACAGCGTG
AAGCCTGGCCTGAGCGCCTTCGTGGACCAGCCTAAACAGGGCGCCGAGACAGT
GCAGGGCCTGCTGGAAGTGGCCAAGGACAGCATCCCCAGAAGCCACTGGAAG
AAAACCCCCGTGGTGCTGAAGGCCACAGCCGGCCTGAGACTGCTGCCTGAGCA
CAAGGCCAAGGCCCTGCTGTTCGAAGTGAAAGAGATCTTCCGCAAGAGCCCAT
TCCTGGTGCCCAAGGGCAGCGTGTCCATCATGGATGGCAGCGACGAGGGCATC
CTGGCCTGGGTCACCGTGAACTTTCTGACCGGACAGCTGCACGGCCACCGGCAG
GAAACAGTGGGCACACTGGATCTGGGCGGAGCCTCTACCCAGATCACCTTCCTG
CCCCAGTTCGAGAAAACACTGGAACAGACCCCCAGAGGCTACCTGACCAGCTT

CGAGATGTTCAACAGCACCTACAAGCTGTACACCCACAGCTACCTGGGCTTCGG
CCTGAAGGCCGCCAGACTGGCTACACTGGGAGCCCTGGAAACCGAGGGCACCG
ACGGCCACACCTTCAGATCTGCCTGCCTGCCCAGATGGCTGGAAGCCGAGTGGA
TCTTCGGCGGCGTGAAGTACCAGTACGGCGGCAACCAGGAAGGCGAAGTGGGC
TTCGAGCCCTGCTATGCCGAGGTGCTGAGAGTCGTGCGGGGCAAGCTGCACCAG
CCTGAGGAAGTGCAGCGGGGCAGCTTCTACGCCTTCAGCTACTACTACGACCGG
GCCGTGGACACCGACATGATCGACTATGAGAAGGGCGGCATCCTGAAGGTGGA
AGATTTCGAGCGGAAGGCCCGGGAAGTGTGCGACAACCTGGAAAACTTTACCA
GCGGCAGCCCCTTCCTGTGCATGGACCTGAGCTACATCACCGCCCTGCTGAAGG
ACGGCTTCGGCTTCGCCGATAGCACCGTGCTGCAGCTGACCAAGAAAGTGAAC
AACATCGAGACAGGCTGGGCCCTGGGCGCCACCTTTCATCTGCTGCAGAGCCTG
GGCATCAGCCAC**CTCGAG***CACCACCACCACCAC*

For cloning into pHLsec

*GGTTGCGTAGCTGAAA****ACCGGT***GTCTCCCACCGGAACCAGCAGACTTGGTTTGAG
GGCATCTTCCTGAGCAGCATGTGCCCCATCAACGTGTCCGCCAGCACCCTGTAC
GGCATCATGTTCGATGCCGGCAGCACCGGCACCCGGATCCACGTGTACACCTTC
GTGCAGAAGATGCCTGGCCAGCTGCCCATCCTGGAAGGCGAGGTGTTCGACAG
CGTGAAGCCTGGCCTGAGCGCCTTCGTGGACCAGCCTAAACAGGGCGCCGAGA
CAGTGCAGGGCCTGCTGGAAGTGGCCAAGGACAGCATCCCCAGAAGCCACTGG
AAGAAAACCCCCGTGGTGCTGAAGGCCACAGCCGGCCTGAGACTGCTGCCTGA
GCACAAGGCCAAGGCCCTGCTGTTCGAAGTGAAAGAGATCTTCCGCAAGAGCC
CATTCCTGGTGCCCAAGGGCAGCGTGTCCATCATGGATGGCAGCGACGAGGGC
ATCCTGGCCTGGGTCACCGTGAACTTTCTGACCGGACAGCTGCACGGCCACCGG
CAGGAAACAGTGGGCACACTGGATCTGGGCGGAGCCTCTACCCAGATCACCTT
CCTGCCCCAGTTCGAGAAAACCCTGGAACAGACCCCCAGAGGCTACCTGACCA
GCTTCGAGATGTTCAACAGCACCTACAAGCTGTACACCCACAGCTACCTGGGCT
TCGGCCTGAAGGCCGCCAGACTGGCTACACTGGGAGCCCTGGAAACCGAGGGC
ACCGACGGCCACACCTTCAGATCTGCCTGCCTGCCCAGATGGCTGGAAGCCGA
GTGGATCTTCGGCGGCGTGAAGTACCAGTACGGCGGCAACCAGGAAGGCGAAG
TGGGCTTCGAGCCCTGCTATGCCGAGGTGCTGAGAGTCGTGCGGGGCAAGCTGC
ACCAGCCTGAGGAAGTGCAGCGGGGCAGCTTCTACGCCTTCAGCTACTACTACG

ACCGGGCCGTGGACACCGACATGATCGACTATGAGAAGGGCGGCATCCTGAAG
GTGGAAGATTTCGAGCGGAAGGCCCGGGAAGTGTGCGACAACCTGGAAAACTT
TACCAGCGGCAGCCCCTTCCTGTGCATGGACCTGAGCTACATCACCGCCCTGCT
GAAGGACGGCTTCGGCTTCGCCGATAGCACCGTGCTGCAGCTGACCAAGAAAG
TGAACAACATCGAGACAGGCTGGGCCCTGGGCGCCACCTTTCATCTGCTGCAGA
GCCTGGGCATCAGCCAC**GGTACC***AAGCACCACCATCAC*

Both encode the following peptide sequence.

VSHRNQQTWFEGIFLSSMCPINVSASTLYGIMFDAGSTGTRIHVYTFVQKMPGQLPILE
GEVFDSVKPGLSAFVDQPKQGAETVQGLLEVAKDSIPRSHWKKTPVVLKATAGLRLL
PEHKAKALLFEVKEIFRKSPFLVPKGSVSIMDGSDEGILAWVTVNFLTGQLHGHRQET
VGTLDLGGASTQITFLPQFEKTLEQTPRGYLTSFEMFNSTYKLYTHSYLGFGLKAARLA
TLGALETEGTDGHTFRSACLPRWLEAEWIFGGVKYQYGGNQEGEVGFEPCYAEVLR
VVRGKLHQPEEVQRGSFYAFSYYYDRAVDTDMIDYEKGGILKVEDFERKAREVCDNL
ENFTSGSPFLCMDLSYITALLKDGFGFADSTVLQLTKKVNNIETGWALGATFHLLQSL
GISH

Codon optimised sequences for *Mm* ENTPD5, lacking the signal peptide and including overlap for Gibson assembly cloning (highlighted in grey).

For cloning into pET/M vectors

*GTTCTGTTCCAGGGGC**C***CATGG**TCTTTTACCGCGAGCAGCAGACTTGGTTTGAGG
GCGTGTTCCTGAGCAGCATGTGCCCCATCAACGTGTCCGCCGGCACCTTCTACG
GCATCATGTTCGACGCCGGCAGCACCGGCACCAGAATCCACGTGTACACCTTCG
TGCAGAAAACAGCCGGCCAGCTGCCCTTCCTGGAAGGCGAGATCTTCGACAGC
GTGAAGCCTGGCCTGAGCGCCTTCGTGGACCAGCCTAAACAGGGCGCCGAGAC
AGTGCAGGAACTGCTGGAAGTGGCCAAGGACAGCATCCCCAGAAGCCACTGGG
AGAGAACCCCCGTGGTGCTGAAGGCTACCGCCGGACTGAGACTGCTGCCTGAG
CAGAAGGCTCAGGCCCTGCTGCTGGAAGTGGAAGAGATCTTCAAGAACAGCCC
TTTCCTGGTGCCCGACGGCAGCGTGTCCATCATGGACGGCTCTTACGAGGGCAT
CCTGGCCTGGGTCACCGTGAACTTTCTGACCGGACAGCTGCACGGCAGAGGCCA

GGAAACAGTGGGCACACTGGATCTGGGCGGAGCCTCTACCCAGATCACCTTCCT
GCCCCAGTTCGAAAAGACACTGGAACAGACCCCCAGAGGCTACCTGACCAGCT
TCGAGATGTTCAACAGCACCTTCAAGCTGTACACCCACAGCTACCTGGGCTTCG
GCCTGAAGGCCGCCAGACTGGCTACACTGGGAGCCCTGGAAGCCAAGGGCACA
GACGGCCACACATTCAGAAGCGCCTGCCTGCCCAGATGGCTGGAAGCCGAGTG
GATCTTTGGCGGCGTGAAGTACCAGTACGGCGGCAACCAGGAAGGCGAAATGG
GCTTCGAGCCCTGCTACGCCGAGGTGCTGAGAGTGGTGCAGGGCAAGCTGCAC
CAGCCTGAGGAAGTGCGGGGCTCTGCCTTCTACGCCTTCAGCTACTACTACGAC
AGAGCCGCCGACACCCACCTGATCGACTACGAGAAAGGCGGCGTGCTGAAGGT
GGAAGATTTCGAGAGAAAGGCCAGAGAAGTGTGCGACAACCTGGGCAGCTTCA
GCAGCGGCAGCCCCTTCCTGTGTATGGACCTGACCTACATCACCGCCCTGCTGA
AGGACGGCTTCGGCTTCGCTGATGGCACCCTGCTGCAGCTGACCAAGAAAGTG
AACAACATCGAGACAGGCTGGGCCCTGGGCGCCACCTTTCATCTGCTGCAGAGC
CTGGGCATCACCAGC**CTCGAG***CACCACCACCACCAC*

For cloning into pHLsec

*GGTTGCGTAGCTGAAA***ACCGGT**GTCTTTTACCGCGAGCAGCAGACTTGGTTTGAGG
GCGTGTTCCTGAGCAGCATGTGCCCCATCAACGTGTCCGCCGGCACCTTCTACG
GCATCATGTTCGACGCCGGCAGCACCGGCACCAGAATCCACGTGTACACCTTCG
TGCAGAAAACAGCCGGCCAGCTGCCCTTCCTGGAAGGCGAGATCTTCGACAGC
GTGAAGCCTGGCCTGAGCGCCTTCGTGGACCAGCCTAAACAGGGCGCCGAGAC
AGTGCAGGAACTGCTGGAAGTGGCCAAGGACAGCATCCCCAGAAGCCACTGGG
AGAGAACCCCCGTGGTGCTGAAGGCTACCGCCGGACTGAGACTGCTGCCTGAG
CAGAAGGCTCAGGCCCTGCTGCTGGAAGTGGAAGAGATCTTCAAGAACAGCCC
TTTCCTGGTGCCCGACGGCAGCGTGTCCATCATGGACGGCTCTTACGAGGGCAT
CCTGGCCTGGGTCACCGTGAACTTTCTGACCGGACAGCTGCACGGCAGAGGCCA
GGAAACAGTGGGCACACTGGATCTGGGCGGAGCCTCTACCCAGATCACCTTCCT
GCCCCAGTTCGAAAAGACCCTGGAACAGACCCCCAGAGGCTACCTGACCAGCT
TCGAGATGTTCAACAGCACCTTCAAGCTGTACACCCACAGCTACCTGGGCTTCG
GCCTGAAGGCCGCCAGACTGGCTACACTGGGAGCCCTGGAAGCCAAGGGCACA
GACGGCCACACATTCAGAAGCGCCTGCCTGCCCAGATGGCTGGAAGCCGAGTG
GATCTTTGGCGGCGTGAAGTACCAGTACGGCGGCAACCAGGAAGGCGAAATGG

GCTTCGAGCCCTGCTACGCCGAGGTGCTGAGAGTGGTGCAGGGCAAGCTGCAC
CAGCCTGAGGAAGTGCGGGGCTCTGCCTTCTACGCCTTCAGCTACTACTACGAC
AGAGCCGCCGACACCCACCTGATCGACTACGAGAAAGGCGGCGTGCTGAAGGT
GGAAGATTTCGAGAGAAAGGCCAGAGAAGTGTGCGACAACCTGGGCAGCTTCA
GCAGCGGCAGCCCCTTCCTGTGTATGGACCTGACCTACATCACCGCCCTGCTGA
AGGACGGCTTCGGCTTCGCTGATGGCACCCTGCTGCAGCTGACCAAGAAAGTG
AACAACATCGAGACAGGCTGGGCCCTGGGCGCCACCTTTCATCTGCTGCAGAGC
CTGGGCATCACCAGC**GGTACC***AAGCACCACCATCAC*

Both encode the following peptide sequence

VFYREQQTWFEGVFLSSMCPINVSAGTFYGIMFDAGSTGTRIHVYTFVQKTAGQLPFL
EGEIFDSVKPGLSAFVDQPKQGAETVQELLEVAKDSIPRSHWERTPVVLKATAGLRLL
PEQKAQALLLEVEEIFKNSPFLVPDGSVSIMDGSYEGILAWVTVNFLTGQLHGRGQET
VGTLDLGGASTQITFLPQFEKTLEQTPRGYLTSFEMFNSTFKLYTHSYLGFGLKAARLA
TLGALEAKGTDGHTFRSACLPRWLEAEWIFGGVKYQYGGNQEGEMGFEPCYAEVLR
VVQGKLHQPEEVRGSAFYAFSYYYDRAADTHLIDYEKGGVLKVEDFERKAREVCDN
LGSFSSGSPFLCMDLTYITALLKDGFGFADGTLLQLTKKVNNIETGWALGATFHLLQS
LGITS

cDNA predicted sequence desired for *Ct* ENTPD6

ACGCGCGCCTACTCCAAGGACAAGCCGCTCATCCAATATGTGCTCATGATCGAC
GCCGGCAGCACCGGCTCTCGTATCCACGTCTACAAGTTCAACAATTGCGGTCCC
GCCCCTGAGCTCGAAGCTGAGGTGCTGTTCAAGCAGACTGCCAAGATTGAGGG
CCAGAGCAGCGGTCTCAGCGCCTACAAGGACGATCCTGAAGCCGCTGCCCGTA
GCTTGGACCAGCTCCTGGACGAGGCCGTCAAGGTCGTGCCCGAAAAACTGCAG
GGCTGCAGCCCGATCGCGGTAAAGGCCACGGCCGGCCTGCGGCTCATCGGCCC
CGAGAAGTCCCAGAAGATCCTCGAGGCGGTCCGCAACCGTCTTGAGACGAAGT
ACCCCTTCCCCGTTGTCTCTCGTGAAGAGAACGGCGTAGCGATCATGGACGGCT
CGGACGAGGGTGTGTATGCCTGGATCACGACCAACTACCTGCTCGGAAAGATTG
GCGGCCCCGACCCAGACGAAACAGCCGCAGTGTTCGATCTTGGCGGCGGCTCT

ACTCAAATCGTCTTCCAGCCCAGCTTCGAGGGCCTCAAGTCCGGCGGCATGCCC
GACAAGCTCGCCGACGGTGACCACAAGTACGAGCTCGACTTCGGCGGCCGCAA
GTTCACGCTCTACCAGCACTCCCACCTGGGCTACGGCCTGATGTCCGCCCGCGA
AGCTATCCTGGCCACTCTCCTCGCCGACCTGCACGAAGCGCACAAGGCCTCCGG
CTCGACGGAGTGGATCCACAAGCAGCCCATTGTCAACCCATGCTACAGTGCCG
GCATGTCTAAGGTGGCCAAGGTGCCCATCCCCTCAGACCATCCCCTCGCTGCGC
TCGCGACCAACGGCAAGCTGGAAGTTAACATGACCGGCCCGGAGGTTGGCGCC
CCGGCACAATGCCGTAGCCTTGCCGAGCGTATTCTCAAGAAAGATGCTCCCTGC
ACGCTGACGCCATGCTCTTTCAATGGCGTGCACCAGCCAAGCATCGCGAAAACC
TTCGCCCGCGAGCCAGTGTACTTCCTGTCGTACTTCTACGACAGGACCCAGCCG
CTCGGCATGCCGTCGTCCTTTACCCTGCGGGAGATGGCAGACCTGGCAGCGCGA
GTGTGCGGTGGCAAGAAGGAGTGGGATGTTTTCACTAGCATTCCAGGCGCCATG
GAGGAGCTGAGCGATAGGCCGGAGCATTGTATGGATTTGAACTTCATGGTCGCG
TTGACGCATACTGGGTATGAGATGCCGATTGATAGAGAGGTGAGAATCGCGAA
GCAGATTAAGGGAAATGAGCTTGGGTGGTGCTTGGGGGCAAGCTTGCCGCTTCT
GTCTAAGGGATCGGGTTGGCAGTGCAAGGTGAAGGAGATTCAT

Which encodes for

TRAYSKDKPLIQYVLMIDAGSTGSRIHVYKFNNCGPAPELEAEVLFKQTAKIEGQSSG
LSAYKDDPEAAARSLDQLLDEAVKVVPEKLQGCSPIAVKATAGLRLIGPEKSQKILEA
VRNRLETKYPFPVVSREENGVAIMDGSDEGVYAWITTNYLLGKIGGPDPDETAAVFD
LGGGSTQIVFQPSFEGLKSGGMPDKLADGDHKYELDFGGRKFTLYQHSHLGYGLMS
AREAILATLLADLHEAHKASGSTEWIHKQPIVNPCYSAGMSKVAKVPIPSDHPLAAL
ATNGKLEVNMTGPEVGAPAQCRSLAERILKKDAPCTLTPCSFNGVHQPSIAKTFARE
PVYFLSYFYDRTQPLGMPSSFTLREMADLAARVCGGKKEWDVFTSIPGAMEELSDRPE
HCMDLNFMVALTHTGYEMPIDREVRIAKQIKGNELGWCLGASLPLLSKGSGWQCK
VKEIH

Codon-optimised sequence of the *Hs* Sep15-binding loop from UGGT1, with homology arms for insertion into *Ct* UGGT (highlighted in grey)

ATTACATCGTCATTGATGACACCCAAGTGAAGGGCACCGAAGTGAACACCACC
GTGATCGGCGAGAACGACCCCATCGATGAGGTGCAGGGCTTCCTGTTTGGCAAG
CTGAGAGATCTGCACCCCGACCTGGAAGGCCAGCTGAAAGAGCTGAGAAAGCA
CCTGGTGGAAAGCACCAACGAGATGGCCCCTCTGGAGAAATCCGAACTTGC

(Highlighted are the 20bp of complementary bases for overlap with the target sequence)

This would give a final mature protein sequence coding for:

(Highlighted is the inserted sequence)

ETGQVAASPSINVALKAAFPSPPYLVELLETAASDNTTIYYSLLDRIAKGHFAEATTDK
ALYEKFLEVLRDDGHMDPEALSAFKLALSLRTATPRVEAHYQYYTATVEPSLSGTQE
GCDQWFLIDGEQYCSPTLDTSHGKVKGEDQLRTLPFDRKFGVGSRDVILYADITSKSF
APFHEVAMDLAKKGKASYRVRYRRSPSHSRESLSVNGYGVELVLKRTDYIVIDDTQV
KGTEVNTTVIGENDPIDEVQGFLFGKLRDLHPDLEGQLKELRKHLVESTNEMAPLEK
SELAALGMKAASFVMQSEKPFEALLKLTQDFPKYSNSLGSQNVSAEFEAEHRGNREV
FLPEGSNVLWLNGLHLIDRQIQPFGLVDLLTRERKLIKSVLDLGLTGQQAVDLLGHAE
VAHAKSGDDEPRRFDWRDDIEEGQVIIWLNNLEKDKRYKSFSPSIWVLIHHFGHGLP
QIRRDVFNLVVPVDLTKADDVKIVVEGLLSFVKRLIPVRFGFVPLTPTGQAIDQAKVV
YYLLENYGLAAATAYLEKSYEEQSTGQPNERIFNEVIKDKSLRPDGVELSFKDIFISEKH
EKQIHLSKHWVERLRAGGDVPTVFFDGFPIPREDNWLRVMNHRLMQDLQALQQAG
YFGMLNESMWLPGFFLEKALSRRNTLIFPEDKNELTVLNVNKIYIENHDLMSKVPVIE
ASKESTRDDWAALTVVADLDDIEGQELVYYALRFRKSNDGVRLDIVHNPKDTSRSPS
VLAQRLKSREDKLLDFTRFLDLETALETGEFEPDVAYDASLANFLASSNMKAGDNFVI
LNGRVLGPITSADDFKKEDFEVFLQAERRTRILPVYKALEDLGLDDKVSGPLSAAKLT
SVTALSTISDLPQGIFDNAPTVRTTLFKQWNSTYTSFEVGDASTATIFFVAVINPASEIG
QRWVAVLKVLSELEGVHLRVFLNPTVMIEELPVKRFYRYVLSSSPSFDESGKVKALSA
RFTGVPRETLLVVGMDVPPAWLVTSKVAVDDLDNLRIKDIKAKRGTEHVEAIYELEH
ILIEGHSREIPGAHAPRGVQLVLETENNPHFADTIIMANLGYFQFKANPGVYNIRLKE
GRSSEIFTLESVGAKGWGPIPGDDNTEVVLMDFQGTTLYPRLRRKPGMEEEDVLEPST

KSGEESGSGARNLVSRGIKFAEGLLGRGNKAAEATKSVSKTEHAEINIFSVASGHLYE
RMLNIMMASVMHHTNHTVKFWFIEQFLSPSFKDFIPHMAAEYGFKYEMVTYKWPH
WLRQQKEKQREIWGYKILFLDVLFPLSLDKVIFVDADQIVRTDMYDLVEHPLDGAPY
GFAPMCDSRVEMEGYRFWKTGYWANYLKGKPYHISALYVVDLQRFRELAAGDRLR
QQYHALSADPNSLANLDQDLPNHMQFTIPIATLPQEWLWCETWCSDETLKDARTID
LCNNPMTKEPKLDRARRQVPEWTKYDEEIAELARRVREEKPKKKEEEKVQKNPKSR
RLDGDEEEVKTVREGTKHHHHHH

GeneString Sequences of *Hs* Sep15 ordered for Gibson assembly into the *E.coli* expression vectors pET303 and pET/M-43.

For Gibson assembly into pET303:

CTTTAAGAAGGAGGTCTAGAATGTTTGGTGCAGAATTTTCAAGCGAAGCATGTC
GTGAACTGGGTTTTAGCAGCAATCTGCTGTGTAGCAGCTGTGATCTGCTGGGTCA
GTTTAATCTGCTGCAGCTGGATCCGGATTGTCGTGGTTGTTGTCAAGAAGAAGC
ACAGTTCGAAACCAAAAAACTGTATGCCGGTGCAATTCTGGAAGTTTGTGGTTG
TAAACTGGGTCGTTTTCCGCAGGTTCAGGCATTTGTTCGTAGCGATAAACCGAA
ACTGTTTCGTGGTCTGCAGATCAAATATGTTCGTGGTAGCGATCCGGTTCTGAAA
CTGCTGGATGATAATGGTAATATTGCAGAGGAACTGAGCATCCTGAAATGGAAT
ACCGATAGCGTTGAAGAATTTCTGAGCGAAAAACTGGAACGCATTATGCATCTC
GAGCACC

For Gibson Assembly into pET/M-43:

TCTGTTCCAGGGGCCCATGGACTTTGGTGCAGAATTTTCAAGCGAAGCATGTCG
TGAACTGGGTTTTAGCAGCAATCTGCTGTGTAGCAGCTGTGATCTGCTGGGTCAG
TTTAATCTGCTGCAGCTGGATCCGGATTGTCGTGGTTGTTGTCAAGAAGAAGCAC
AGTTCGAAACCAAAAAACTGTATGCCGGTGCAATTCTGGAAGTTTGTGGTTGTA
AACTGGGTCGTTTTCCGCAGGTTCAGGCATTTGTTCGTAGCGATAAACCGAAAC
TGTTTCGTGGTCTGCAGATCAAATATGTTCGTGGTAGCGATCCGGTTCTGAAACT
GCTGGATGATAATGGTAATATTGCAGAGGAACTGAGCATCCTGAAATGGAATA

CCGATAGCGTTGAAGAATTTCTGAGCGAAAAACTGGAACGTATTGCGGCCGCA
CTCGAGCACCAC

Mature protein coding sequence of Sep15 in pET303:

MFGAEFSSEACRELGFSSNLLCSSCDLLGQFNLLQLDPDCRGCCQEEAQFETKKLYAG
AILEVCGCKLGRFPQVQAFVRSDKPKLFRGLQIKYVRGSDPVLKLLDDNGNIAEELSIL
KWNTDSVEEFLSEKLERIMHLEHHHHHH

Mature protein coding sequence of Sep15 in pET/M-43: (contains an N-terminal His tag)

MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGD
GPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEA
LSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAF
KYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAM
TINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLE
NYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMS
AFWYAVRTAVINAASGRQTVDEALKDAQTPGSLEVLFQGPMDFGAEFSSEACRELGF
SSNLLCSSCDLLGQFNLLQLDPDCRGCCQEEAQFETKKLYAGAILEVCGCKLGRFPQ
VQAFVRSDKPKLFRGLQIKYVRGSDPVLKLLDDNGNIAEELSILKWNTDSVEEFLSEK
LERIAAALEHHHHHH

References

1. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox, *Acta Crystallographica Section D: Biological Crystallography 67*, 293-302.
2. Choi, J. H., and Croyle, M. A. (2013) Emerging Targets and Novel Approaches to Ebola Virus Prophylaxis and Treatment, *BioDrugs 27*, 565-583.
3. Goeijenbier, M., van Kampen, J. J., Reusken, C. B., Koopmans, M. P., and van Gorp, E. C. (2014) Ebola virus disease: a review on epidemiology, symptoms, treatment and pathogenesis, *The Netherlands journal of medicine 72*, 442-448.
4. Ioos, S., Mallet, H. P., Leparc Goffart, I., Gauthier, V., Cardoso, T., and Herida, M. (2014) Current Zika virus epidemiology and recent epidemics, *Médecine et Maladies Infectieuses 44*, 302-307.
5. Koonin, E. V., and Dolja, V. V. (2013) A virocentric perspective on the evolution of life, *Current Opinion in Virology 3*, 546-557.
6. Sayce, A. C., Miller, J. L., and Zitzmann, N. (2010) Targeting a host process as an antiviral approach against dengue virus, *Trends in Microbiology 18*, 323-330.
7. Wolf, M. C., Freiberg, A. N., Zhang, T., Akyol-Ataman, Z., Grock, A., Hong, P. W., Li, J., Watson, N. F., Fang, A. Q., Aguilar, H. C., Porotto, M., Honko, A. N., Damoiseaux, R., Miller, J. P., Woodson, S. E., Chantasirivisal, S., Fontanes, V., Negrete, O. A., Krogstad, P., Dasgupta, A., Moscona, A., Hensley, L. E., Whelan, S. P., Faull, K. F., Holbrook, M. R., Jung, M. E., and Lee, B. (2010) A broad-spectrum antiviral targeting entry of enveloped viruses, *Proceedings of the National Academy of Sciences 107*, 3157.
8. Aman, M. J., Kinch, M. S., Warfield, K., Warren, T., Yunus, A., Enterlein, S., Stavale, E., Wang, P., Chang, S., Tang, Q., Porter, K., Goldblatt, M., and Bavari, S. (2009) Development of a broad-spectrum antiviral with activity against Ebola virus, *Antiviral Research 83*, 245-251.
9. Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S., and Strauss, J. H. (2002) Structure of Dengue Virus: Implications for Flavivirus Organization, Maturation, and Fusion, *Cell 108*, 717-725.
10. Rajcani, J., and Vojvodova, A. (1998) The role of herpes simplex virus glycoproteins in the virus replication cycle, *Acta virologica 42*, 103-118.
11. Sanchez, A., Yang, Z.-Y., Xu, L., Nabel, G. J., Crews, T., and Peters, C. J. (1998) Biochemical Analysis of the Secreted and Virion Glycoproteins of Ebola Virus, *Journal of Virology 72*, 6442-6447.
12. Wyatt, R., and Sodroski, J. (1998) The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens, *Science 280*, 1884.
13. He, Y., Bowman, V. D., Mueller, S., Bator, C. M., Bella, J., Peng, X., Baker, T. S., Wimmer, E., Kuhn, R. J., and Rossmann, M. G. (2000) Interaction of the poliovirus receptor with poliovirus, *Proceedings of the National Academy of Sciences 97*, 79.

14. Bergelson, J. M., Cunningham, J. A., Droguett, G., Kurt-Jones, E. A., Krithivas, A., Hong, J. S., Horwitz, M. S., Crowell, R. L., and Finberg, R. W. (1997) Isolation of a Common Receptor for Coxsackie B Viruses and Adenoviruses 2 and 5, *Science 275*, 1320.
15. Apweiler, R., Hermjakob, H., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochimica et biophysica acta 1473*, 4-8.
16. Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011) Protein translocation across the ER membrane, *Biochimica et biophysica acta 1808*, 912-924.
17. Aebi, M. (2013) N-linked protein glycosylation in the ER, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1833*, 2430-2437.
18. Mellquist, J. L., Kasturi, L., Spitalnik, S. L., and Shakin-Eshleman, S. H. (1998) The Amino Acid Following an Asn-X-Ser/Thr Sequon Is an Important Determinant of N-Linked Core Glycosylation Efficiency, *Biochemistry 37*, 6833-6837.
19. Pfeffer, S., Dudek, J., Gogala, M., Schorr, S., Linxweiler, J., Lang, S., Becker, T., Beckmann, R., Zimmermann, R., and Förster, F. (2014) Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon, *Nature Communications 5*, 3072.
20. Schweden, J., Borgmann, C., Legler, G., and Bause, E. (1986) Characterization of calf liver glucosidase I and its inhibition by basic sugar analogs, *Archives of Biochemistry and Biophysics 248*, 335-340.
21. Parodi, A. J. (2000) Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation, *Biochemical Journal 348*, 1-13.
22. Miyagawa, A., Totani, K., Matsuo, I., and Ito, Y. (2010) Promiscuous activity of ER glucosidase II discovered through donor specificity analysis of UGGT, *Biochemical and Biophysical Research Communications 403*, 322-328.
23. Trombetta, S. E., Bosch, M., and Parodi, A. J. (1989) Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes, *Biochemistry 28*, 8108-8116.
24. Trombetta, S. E., Ganan, S. A., and Parodi, A. J. (1991) The UDP-Glc:glycoprotein glucosyltransferase is a soluble protein of the endoplasmic reticulum, *Glycobiology 1*, 155-161.
25. Sousa, M. C., Ferrero-Garcia, M. A., and Parodi, A. J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase, *Biochemistry 31*, 97-105.
26. Olivari, S., Cali, T., Salo, K. E. H., Paganetti, P., Ruddock, L. W., and Molinari, M. (2006) EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation, *Biochemical and Biophysical Research Communications 349*, 1278-1284.
27. Stanley, P., Schachter, H., and Taniguch, N. (2009) Chapter 8 N-glycans, In *Essentials of glycobiology, 2nd edition* (Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., and Etzler, M. E., Eds.) 2nd ed., Cold Spring Harbor NY.
28. Hall, M. K., Weidner, D. A., Dayal, S., and Schwalbe, R. A. (2014) Cell surface N-glycans influence the level of functional E-cadherin at the cell–cell border, *FEBS Open Bio 4*, 892-897.
29. Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., and Dwek, R. A. (2001) Glycosylation and the Immune System, *Science 291*, 2370.
30. Li, L.-B., Chen, N., Ramamoorthy, S., Chi, L., Cui, X.-N., Wang, L. C., and Reith, M. E. A. (2004) The Role of N-Glycosylation in Function and Surface Trafficking of the Human Dopamine Transporter, *Journal of Biological Chemistry 279*, 21012-21020.

31. Melikian, H. E., Ramamoorthy, S., Tate, C. G., and Blakely, R. D. (1996) Inability to N-glycosylate the human norepinephrine transporter reduces protein stability, surface trafficking, and transport activity but not ligand recognition, *Molecular pharmacology 50*, 266-276.
32. Porath, B., Gainullin, V. G., Cornec-Le Gall, E., Dillinger, E. K., Heyer, C. M., Hopp, K., Edwards, M. E., Madsen, C. D., Mauritz, S. R., Banks, C. J., Baheti, S., Reddy, B., Herrero, J. I., Banales, J. M., Hogan, M. C., Tasic, V., Watnick, T. J., Chapman, A. B., Vigneau, C., Lavainne, F., Audrezet, M. P., Ferec, C., Le Meur, Y., Torres, V. E., and Harris, P. C. (2016) Mutations in GANAB, Encoding the Glucosidase IIalpha Subunit, Cause Autosomal-Dominant Polycystic Kidney and Liver Disease, *Am J Hum Genet 98*, 1193-1207.
33. Chang, J., Block, T. M., and Guo, J.-T. (2015) Viral resistance of MOGS-CDG patients implies a broad-spectrum strategy against acute virus infections, *Antiviral therapy 20*, 257-259.
34. Hubbard, S. C., and Robbins, P. W. (1979) Synthesis and processing of protein-linked oligosaccharides in vivo, *Journal of Biological Chemistry 254*, 4568-4576.
35. Grinna, L. S., and Robbins, P. W. (1979) Glycoprotein biosynthesis. Rat liver microsomal glucosidases which process oligosaccharides, *Journal of Biological Chemistry 254*, 8814-8818.
36. Deprez, P., Gautschi, M., and Helenius, A. (2005) More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle, *Molecular Cell 19*, 183-195.
37. Shailubhai, K., Pukazhenthi, B. S., Saxena, E. S., Varma, G. M., and Vijay, I. K. (1991) Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain, *Journal of Biological Chemistry 266*, 16587-16593.
38. Schutze, M. P., Peterson, P. A., and Jackson, M. R. (1994) An N-terminal double-arginine motif maintains type II membrane proteins in the endoplasmic reticulum, *The EMBO journal 13*, 1696-1705.
39. Hardt, B., Kalz-Fuller, B., Aparicio, R., Volker, C., and Bause, E. (2003) (Arg)3 within the N-terminal domain of glucosidase I contains ER targeting information but is not required absolutely for ER localization, *Glycobiology 13*, 159-168.
40. Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A., and Tonozuka, T. (2008) Structural Insights into the Substrate Specificity and Function of Escherichia coli K12 YgjK, a Glucosidase Belonging to the Glycoside Hydrolase Family 63, *Journal of molecular biology 381*, 116-128.
41. Miyazaki, T., Ichikawa, M., Yokoi, G., Kitaoka, M., Mori, H., Kitano, Y., Nishikawa, A., and Tonozuka, T. (2013) Structure of a bacterial glycoside hydrolase family 63 enzyme in complex with its glycosynthase product, and insights into the substrate specificity, *The FEBS Journal 280*, 4560-4571.
42. Barker, M. K., and Rose, D. R. (2013) Specificity of Processing α-Glucosidase I Is Guided by the Substrate Conformation: CRYSTALLOGRAPHIC AND IN SILICO STUDIES, *Journal of Biological Chemistry 288*, 13563-13574.
43. Volker, C., De Praeter, C. M., Hardt, B., Breuer, W., Kalz-Fuller, B., Van Coster, R. N., and Bause, E. (2002) Processing of N-linked carbohydrate chains in a patient with glucosidase I deficiency (CDG type IIb), *Glycobiology 12*, 473-483.
44. De Praeter, C. M., Gerwig, G. J., Bause, E., Nuytinck, L. K., Vliegenthart, J. F. G., Breuer, W., Kamerling, J. P., Espeel, M. F., Martin, J.-J. R., De Paepe, A. M., Chan, N. W. C., Dacremont, G. A., and Van Coster, R. N. (2000) A Novel Disorder Caused by Defective

Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency, *The American Journal of Human Genetics 66*, 1744-1756.

45. Sadat, M. A., Moir, S., Chun, T.-W., Lusso, P., Kaplan, G., Wolfe, L., Memoli, M. J., He, M., Vega, H., Kim, L. J. Y., Huang, Y., Hussein, N., Nievas, E., Mitchell, R., Garofalo, M., Louie, A., Ireland, D. C., Grunes, C., Cimbro, R., Patel, V., Holzapfel, G., Salahuddin, D., Bristol, T., Adams, D., Marciano, B. E., Hegde, M., Li, Y., Calvo, K. R., Stoddard, J., Justement, J. S., Jacques, J., Long Priel, D. A., Murray, D., Sun, P., Kuhns, D. B., Boerkoel, C. F., Chiorini, J. A., Di Pasquale, G., Verthelyi, D., and Rosenzweig, S. D. (2014) Glycosylation, Hypogammaglobulinemia, and Resistance to Viral Infections, *New England Journal of Medicine 370*, 1615-1625.
46. Caputo, A. T., Alonzi, D. S., Marti, L., Reca, I.-B., Kiappes, J. L., Struwe, W. B., Cross, A., Basu, S., Lowe, E. D., Darlot, B., Santino, A., Roversi, P., and Zitzmann, N. (2016) Structures of mammalian ER α-glucosidase II capture the binding modes of broad-spectrum iminosugar antivirals, *Proceedings of the National Academy of Sciences 113*, E4630-E4638.
47. Trombetta, E. S., Simons, J. F., and Helenius, A. (1996) Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit, *The Journal of biological chemistry 271*, 27509-27516.
48. Olson, L. J., Orsi, R., Peterson, F. C., Parodi, A. J., Kim, J.-J. P., D'Alessio, C., and Dahms, N. M. (2015) Crystal Structure and Functional Analyses of the Lectin Domain of Glucosidase II: Insights into Oligomannose Recognition, *Biochemistry 54*, 4097-4111.
49. Munro, S. (2001) The MRH domain suggests a shared ancestry for the mannose 6-phosphate receptors and other N-glycan-recognising proteins, *Current Biology 11*, R499-R501.
50. Totani, K., Ihara, Y., Matsuo, I., and Ito, Y. (2008) Effects of Macromolecular Crowding on Glycoprotein Processing Enzymes, *Journal of the American Chemical Society 130*, 2101-2107.
51. Fischl, M. A., Resnick, L., Coombs, R., Kremer, A. B., Pottage, J. C., Jr., Fass, R. J., Fife, K. H., Powderly, W. G., Collier, A. C., Aspinall, R. L., and et al. (1994) The safety and efficacy of combination N-butyl-deoxynojirimycin (SC-48334) and zidovudine in patients with HIV-1 infection and 200-500 CD4 cells/mm3, *Journal of acquired immune deficiency syndromes 7*, 139-147.
52. Reddy, B. V., and Chapman, A. B. (2017) A Patient with a Novel Gene Mutation Leading to Autosomal Dominant Polycystic Kidney Disease, *Clinical journal of the American Society of Nephrology : CJASN 12*, 1695-1698.
53. Li, A., Davila, S., Furu, L., Qian, Q., Tian, X., Kamath, P. S., King, B. F., Torres, V. E., and Somlo, S. (2003) Mutations in PRKCSH Cause Isolated Autosomal Dominant Polycystic Liver Disease, *The American Journal of Human Genetics 72*, 691-703.
54. Drenth, J. P. H., te Morsche, R. H. M., Smink, R., Bonifacino, J. S., and Jansen, J. B. M. J. (2003) Germline mutations in PRKCSH are associated with autosomal dominant polycystic liver disease, *Nature Genetics 33*, 345.
55. Davila, S., Furu, L., Gharavi, A. G., Tian, X., Onoe, T., Qian, Q., Li, A., Cai, Y., Kamath, P. S., King, B. F., Azurmendi, P. J., Tahvanainen, P., Kääriäinen, H., Höckerstedt, K., Devuyst, O., Pirson, Y., Martin, R. S., Lifton, R. P., Tahvanainen, E., Torres, V. E., and Somlo, S. (2004) Mutations in SEC63 cause autosomal dominant polycystic liver disease, *Nature Genetics 36*, 575.

56. Fedeles, S. V., Gallagher, A.-R., and Somlo, S. (2014) Polycystin-1: a master regulator of intersecting cystic pathways, *Trends in Molecular Medicine 20*, 251-260.
57. Fujimoto, K., and Kornfeld, R. (1991) alpha-Glucosidase II-deficient cells use endo alpha-mannosidase as a bypass route for N-linked oligosaccharide processing, *The Journal of biological chemistry 266*, 3571-3578.
58. Ganan, S., Cazzulo, J. J., and Parodi, A. J. (1991) A major proportion of N-glycoproteins are transiently glucosylated in the endoplasmic reticulum, *Biochemistry 30*, 3098-3104.
59. Zapun, A., Petrescu, S. M., Rudd, P. M., Dwek, R. A., Thomas, D. Y., and Bergeron, J. J. M. (1997) Conformation-Independent Binding of Monoglucosylated Ribonuclease B to Calnexin, *Cell 88*, 29-38.
60. Tokunaga, F., Brostrom, C., Koide, T., and Arvan, P. (2000) Endoplasmic Reticulum (ER)-associated Degradation of Misfolded N-Linked Glycoproteins Is Suppressed upon Inhibition of ER Mannosidase I, *Journal of Biological Chemistry 275*, 40757-40764.
61. Arnold, S. M., Fessler, L. I., Fessler, J. H., and Kaufman, R. J. (2000) Two homologues encoding human UDP-glucose:glycoprotein glucosyltransferase differ in mRNA expression and enzymatic activity, *Biochemistry 39*, 2149-2163.
62. Arnold, S. M., and Kaufman, R. J. (2003) The Noncatalytic Portion of Human UDP-glucose:Glycoprotein Glucosyltransferase I Confers UDP-glucose Binding and Transferase Function to the Catalytic Domain, *Journal of Biological Chemistry 278*, 43320-43328.
63. Ferris, S. P., Jaber, N. S., Molinari, M., Arvan, P., Kaufman, R. J., and Gilmore, R. (2013) UDP-glucose:glycoprotein glucosyltransferase (UGGT1) promotes substrate solubility in the endoplasmic reticulum, *Molecular Biology of the Cell 24*, 2597-2608.
64. Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., Fujikawa, K., Kajihara, Y., and Ito, Y. (2014) Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active, *Glycobiology 24*, 344-350.
65. Ito, Y., Takeda, Y., Seko, A., Izumi, M., and Kajihara, Y. (2015) Functional analysis of endoplasmic reticulum glucosyltransferase (UGGT): Synthetic chemistry's initiative in glycobiology, *Seminars in Cell & Developmental Biology 41*, 90-98.
66. Izumi, M., Dedola, S., Ito, Y., and Kajihara, Y. (2015) Chemical Synthesis of Homogeneous Glycoproteins for the Study of Glycoprotein Quality Control System, *Israel Journal of Chemistry 55*, 306-314.
67. Zhu, T., Satoh, T., and Kato, K. (2014) Structural insight into substrate recognition by the endoplasmic reticulum folding-sensor enzyme: crystal structure of third thioredoxin-like domain of UDP-glucose:glycoprotein glucosyltransferase, *Sci. Rep. 4*.
68. Roversi, P., Marti, L., Caputo, A. T., Alonzi, D. S., Hill, J. C., Dent, K. C., Kumar, A., Levasseur, M. D., Lia, A., Waksman, T., Basu, S., Soto Albrecht, Y., Qian, K., McIvor, J. P., Lipp, C. B., Siliqi, D., Vasiljević, S., Mohammed, S., Lukacik, P., Walsh, M. A., Santino, A., and Zitzmann, N. (2017) Interdomain conformational flexibility underpins the activity of UGGT, the eukaryotic glycoprotein secretion checkpoint, *Proceedings of the National Academy of Sciences 114*, 8544.
69. Wilkinson, B., and Gilbert, H. F. (2004) Protein disulfide isomerase, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1699*, 35-44.
70. Hagiwara, M., Maegawa, K.-i., Suzuki, M., Ushioda, R., Araki, K., Matsumoto, Y., Hoseki, J., Nagata, K., and Inaba, K. (2011) Structural Basis of an ERAD Pathway Mediated by the ER-Resident Protein Disulfide Reductase ERdj5, *Molecular Cell 41*, 432-444.
71. Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Arac, D., Chelliah, Y., Amezcua, C. A., Rizo, J., Gladyshev, V. N., and Deisenhofer, J. (2006) NMR structures of the selenoproteins

Sep15 and SelM reveal redox activity of a new thioredoxin-like family, *The Journal of biological chemistry 281*, 3536-3543.

72. Serve, O., Kamiya, Y., Maeno, A., Nakano, M., Murakami, C., Sasakawa, H., Yamaguchi, Y., Harada, T., Kurimoto, E., Yagi-Utsumi, M., Iguchi, T., Inaba, K., Kikuchi, J., Asami, O., Kajino, T., Oka, T., Nakasako, M., and Kato, K. (2010) Redox-Dependent Domain Rearrangement of Protein Disulfide Isomerase Coupled with Exposure of Its Substrate-Binding Hydrophobic Surface, *Journal of molecular biology 396*, 361-374.
73. Satoh, T., Song, C., Zhu, T., Toshimori, T., Murata, K., Hayashi, Y., Kamikubo, H., Uchihashi, T., and Kato, K. (2017) Visualisation of a flexible modular structure of the ER folding-sensor enzyme UGGT, *Scientific Reports 7*, 12142.
74. Molinari, M., Galli, C., Vanoni, O., Arnold, S. M., and Kaufman, R. J. (2005) Persistent glycoprotein misfolding activates the glucosidase II/UGT1-driven calnexin cycle to delay aggregation and loss of folding competence, *Mol Cell 20*, 503-512.
75. Huang, P. N., and Jheng, J. R. (2017) UGGT1 enhances enterovirus 71 pathogenicity by promoting viral RNA synthesis and viral replication, *PLOS Pathogens 13*, e1006375.
76. Kalcklosch, M., Kyriakopoulos, A., Hammel, C., and Behne, D. (1995) A New Selenoprotein Found in the Glandular Epithelial Cells of the Rat Prostate, *Biochemical and Biophysical Research Communications 217*, 162-170.
77. Korotkov, K. V., Kumaraswamy, E., Zhou, Y., Hatfield, D. L., and Gladyshev, V. N. (2001) Association between the 15-kDa Selenoprotein and UDP-glucose:Glycoprotein Glucosyltransferase in the Endoplasmic Reticulum of Mammalian Cells, *Journal of Biological Chemistry 276*, 15330-15336.
78. Labunskyy, V. M., Ferguson, A. D., Fomenko, D. E., Chelliah, Y., Hatfield, D. L., and Gladyshev, V. N. (2005) A Novel Cysteine-rich Domain of Sep15 Mediates the Interaction with UDP-glucose:Glycoprotein Glucosyltransferase, *Journal of Biological Chemistry 280*, 37839-37845.
79. Yim, S. H., Everley, R. A., Schildberg, F. A., Lee, S. G., Orsi, A., Barbati, Z. R., Karatepe, K., Fomenko, D. E., Tsuji, P. A., Luo, H. R., Gygi, S. P., Sitia, R., Sharpe, A. H., Hatfield, D. L., and Gladyshev, V. N. (2018) Role of Selenof as a Gatekeeper of Secreted Disulfide-Rich Glycoproteins, *Cell reports 23*, 1387-1398.
80. Labunskyy, V. M., Yoo, M.-H., Hatfield, D. L., and Gladyshev, V. N. (2009) Sep15, a Thioredoxin-like Selenoprotein, is Involved in the Unfolded Protein Response and Differentially Regulated by Adaptive and Acute ER Stresses, *Biochemistry 48*, 8458-8465.
81. Kumaraswamy, E., Malykh, A., Korotkov, K. V., Kozyavkin, S., Hu, Y., Kwon, S. Y., Moustafa, M. E., Carlson, B. A., Berry, M. J., Lee, B. J., Hatfield, D. L., Diamond, A. M., and Gladyshev, V. N. (2000) Structure-Expression Relationships of the 15-kDa Selenoprotein Gene: POSSIBLE ROLE OF THE PROTEIN IN CANCER ETIOLOGY, *Journal of Biological Chemistry 275*, 35540-35547.
82. Jun Hu, Y., Korotkov, K. V., Mehta, R., Hatfield, D. L., Rotimi, C. N., Luke, A., Prewitt, T. E., Cooper, R. S., Stock, W., Vokes, E. E., Dolan, M. E., Gladyshev, V. N., and Diamond, A. M. (2001) Distribution and Functional Consequences of Nucleotide Polymorphisms in the 3'-Untranslated Region of the Human Sep15 Gene, *Cancer Research 61*, 2307-2310.
83. Apostolou, S., Klein, J. O., Mitsuuchi, Y., Shetler, J. N., Poulikakos, P. I., Jhanwar, S. C., Kruger, W. D., and Testa, J. R. (2004) Growth inhibition and induction of apoptosis in mesothelioma cells by selenium and dependence on selenoprotein SEP15 genotype, *Oncogene 23*, 5032-5040.

84. Bockmuhl, U., Schwendel, A., Dietel, M., and Petersen, I. (1996) Distinct patterns of chromosomal alterations in high- and low-grade head and neck squamous cell carcinomas, *Cancer Res 56*, 5325-5329.
85. Kohno, T., and Yokota, J. (1999) How many tumor suppressor genes are involved in human lung carcinogenesis?, *Carcinogenesis 20*, 1403-1410.
86. Nasr, M. A., Hu, Y. J., and Diamond, A. M. (2004) Allelic loss at the Sep15 locus in breast cancer, *Cancer Ther 1*, 293-298.
87. Jablonska, E., Gromadzinska, J., Sobala, W., Reszka, E., and Wasowicz, W. (2008) Lung cancer risk associated with selenium status is modified in smoking individuals by Sep15 polymorphism, *European journal of nutrition 47*, 47-54.
88. Benelli, J. L., de Medeiros, R. M., Matte, M. C. C., de Melo, M. G., Almeida, S. E. d. M., and Fiegenbaum, M. (2016) Role of SEP15 Gene Polymorphisms in the Time of Progression to AIDS, *Genetic Testing and Molecular Biomarkers*.
89. Kasaikina, M. V., Fomenko, D. E., Labunskyy, V. M., Lachke, S. A., Qiu, W., Moncaster, J. A., Zhang, J., Wojnarowicz, M. W., Jr., Natarajan, S. K., Malinouski, M., Schweizer, U., Tsuji, P. A., Carlson, B. A., Maas, R. L., Lou, M. F., Goldstein, L. E., Hatfield, D. L., and Gladyshev, V. N. (2011) Roles of the 15-kDa selenoprotein (Sep15) in redox homeostasis and cataract development revealed by the analysis of Sep 15 knockout mice, *The Journal of biological chemistry 286*, 33203-33212.
90. Trombetta, E. S., and Helenius, A. (1999) Glycoprotein reglucosylation and nucleotide sugar utilization in the secretory pathway: identification of a nucleoside diphosphatase in the endoplasmic reticulum, *The EMBO journal 18*, 3282.
91. Führing, J. I., Cramer, J. T., Schneider, J., Baruch, P., Gerardy-Schahn, R., and Fedorov, R. (2015) A Quaternary Mechanism Enables the Complex Biological Functions of Octameric Human UDP-glucose Pyrophosphorylase, a Key Enzyme in Cell Metabolism, *Scientific Reports 5*, 9618.
92. Fang, M., Shen, Z., Huang, S., Zhao, L., Chen, S., Mak, T. W., and Wang, X. (2010) The ER UDPase ENTPD5 Promotes Protein N-Glycosylation, the Warburg Effect, and Proliferation in the PTEN Pathway, *Cell 143*, 711-724.
93. Chadwick, B. P., and Frischauf, A. M. (1998) The CD39-like gene family: identification of three new human members (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of the gene family from Drosophila melanogaster, *Genomics 50*, 357-367.
94. Mulero, J. J., Yeung, G., Nelken, S. T., and Ford, J. E. (1999) CD39-L4 Is a Secreted Human Apyrase, Specific for the Hydrolysis of Nucleoside Diphosphates, *Journal of Biological Chemistry 274*, 20064-20067.
95. Mulero, J. J., Yeung, G., Nelken, S. T., Bright, J. M., McGowan, D. W., and Ford, J. E. (2000) Biochemical Characterization of CD39L4, *Biochemistry 39*, 12924-12928.
96. Paez, J. G., Recio, J. A., Rouzaut, A., and Notario, V. (2001) Identity between the PCPH proto-oncogene and the CD39L4 (ENTPD5) ectonucleoside triphosphate diphosphohydrolase gene, *International journal of oncology 19*, 1249-1254.
97. Atkinson, B., Dwyer, K., Enjyoji, K., and Robson, S. C. (2006) Ecto-nucleotidases of the CD39/NTPDase family modulate platelet activation and thrombus formation: Potential as therapeutic targets, *Blood Cells, Molecules, and Diseases 36*, 217-222.
98. O'Keeffe, M. G., Thorne, P. R., Housley, G. D., Robson, S. C., and Vlajkovic, S. M. (2010) Distribution of NTPDase5 and NTPDase6 and the regulation of P2Y receptor signalling in the rat cochlea, *Purinergic Signalling 6*, 249-261.
99. Murphy-Piedmonte, D. M., Crawford, P. A., and Kirley, T. L. (2005) Bacterial expression, folding, purification and characterization of soluble NTPDase5 (CD39L4) ecto-

nucleotidase, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1747*, 251-259.

100. Shen, Z., Huang, S., Fang, M., and Wang, X. (2011) ENTPD5, an Endoplasmic Reticulum UDPase, Alleviates ER Stress Induced by Protein Overloading in AKT-Activated Cancer Cells, *Cold Spring Harbor Symposia on Quantitative Biology 76*, 217-223.
101. Recio, J. A., Páez, J. G., Sanders, S., Kawakami, T., and Notario, V. (2002) Partial Depletion of Intracellular ATP Mediates the Stress-Survival Function of the PCPH Oncoprotein, *Cancer Research 62*, 2690-2694.
102. Tirado, O. M., Mateo-Lozano, S., Sanders, S., Dettin, L. E., and Notario, V. (2003) The PCPH Oncoprotein Antagonizes the Proapoptotic Role of the Mammalian Target of Rapamycin in the Response of Normal Fibroblasts to Ionizing Radiation, *Cancer Research 63*, 6290-6298.
103. Villar, J., Arenas, M. I., MacCarthy, C. M., Blánquez, M. J., Tirado, O. M., and Notario, V. (2007) PCPH/ENTPD5 Expression Enhances the Invasiveness of Human Prostate Cancer Cells by a Protein Kinase Cδ–Dependent Mechanism, *Cancer Research 67*, 10859-10868.
104. Read, R., Hansen, G., Kramer, J., Finch, R., Li, L., and Vogel, P. (2009) Ectonucleoside Triphosphate Diphosphohydrolase Type 5 (Entpd5)-Deficient Mice Develop Progressive Hepatopathy, Hepatocellular Tumors, and Spermatogenic Arrest, *Veterinary Pathology Online 46*, 491-504.
105. Villar, J., Quadri, H. S., Song, I., Tomita, Y., Tirado, O. M., and Notario, V. (2009) PCPH/ENTPD5 Expression Confers to Prostate Cancer Cells Resistance against Cisplatin-Induced Apoptosis through Protein Kinase Cα–Mediated Bcl-2 Stabilization, *Cancer Research 69*, 102-110.
106. MacCarthy, C. M., and Notario, V. (2013) The ENTPD5/mt-PCPH oncoprotein is a catalytically inactive member of the ectonucleoside triphosphate diphosphohydrolase family, *International journal of oncology 43*, 1244-1252.
107. Zadran, S., Amighi, A., Otiniano, E., Wong, K., and Zadran, H. (2012) ENTPD5-mediated modulation of ATP results in altered metabolism and decreased survival in gliomablastoma multiforme, *Tumor Biol. 33*, 2411-2421.
108. Curry, N. L., Mino-Kenudson, M., Oliver, T. G., Yilmaz, Ö. H., Yilmaz, V. O., Moon, J. Y., Jacks, T., Sabatini, D. M., and Kalaany, N. Y. (2013) Pten-Null Tumors Cohabiting the Same Lung Display Differential AKT Activation and Sensitivity to Dietary Restriction, *Cancer Discovery 3*, 908-921.
109. Perilli, L., Vicentini, C., Agostini, M., Pizzini, S., Pizzi, M., D'Angelo, E., Bortoluzzi, S., Mandruzzato, S., Mammano, E., Rugge, M., Nitti, D., Scarpa, A., Fassan, M., and Zanovello, P. (2014) Circulating miR-182 is a biomarker of colorectal adenocarcinoma progression, *Oncotarget 5*, 6611-6619.
110. Beckenkamp, A., Santana, D. B., Bruno, A. N., Calil, L. N., Casali, E. A., Paccez, J. D., Zerbini, L. F., Lenz, G., Wink, M. R., and Buffon, A. (2014) Ectonucleotidase expression profile and activity in human cervical cancer cell lines, *Biochemistry and Cell Biology 92*, 95-104.
111. Vogiatzi, F., Brandt, D. T., Schneikert, J., Fuchs, J., Grikscheit, K., Wanzel, M., Pavlakis, E., Charles, J. P., Timofeev, O., Nist, A., Mernberger, M., Kantelhardt, E. J., Siebolts, U., Bartel, F., Jacob, R., Rath, A., Moll, R., Grosse, R., and Stiewe, T. (2016) Mutant p53 promotes tumor progression and metastasis by the endoplasmic reticulum UDPase ENTPD5, *Proceedings of the National Academy of Sciences*.
112. Schneikert, J., and Stiewe, T. (2017) Pro-metastatic p53 mutants control folding of N-glycoproteins, *Cell cycle (Georgetown, Tex.) 16*, 591-592.

113. Ivanenkov, V. V., Murphy-Piedmonte, D. M., and Kirley, T. L. (2003) Bacterial Expression, Characterization, and Disulfide Bond Determination of Soluble Human NTPDase6 (CD39L2) Nucleotidase: Implications for Structure and Function, *Biochemistry 42*, 11726-11735.
114. Hicks-Berger, C. A., Chadwick, B. P., Frischauf, A.-M., and Kirley, T. L. (2000) Expression and Characterization of Soluble and Membrane-bound Human Nucleoside Triphosphate Diphosphohydrolase 6 (CD39L2), *Journal of Biological Chemistry 275*, 34041-34045.
115. Wang, T. F., and Guidotti, G. (1998) Golgi localization and functional expression of human uridine diphosphatase, *The Journal of biological chemistry 273*, 11392-11399.
116. Song, Z. (2013) Roles of the nucleotide sugar transporters (SLC35 family) in health and disease, *Molecular Aspects of Medicine 34*, 590-600.
117. Orellana, A., Moraga, C., Araya, M., and Moreno, A. (2016) Overview of Nucleotide Sugar Transporter Gene Family Functions Across Multiple Species, *Journal of molecular biology*.
118. Norambuena, L., Marchant, L., Berninsone, P., Hirschberg, C. B., Silva, H., and Orellana, A. (2002) Transport of UDP-galactose in Plants: IDENTIFICATION AND FUNCTIONAL CHARACTERIZATION OF AtUTr1, ANARABIDOPSIS THALIANA UDP-GALACTOSE/UDP-GLUCOSE TRANSPORTER, *Journal of Biological Chemistry 277*, 32923-32929.
119. Reyes, F., Marchant, L., Norambuena, L., Nilo, R., Silva, H., and Orellana, A. (2006) AtUTr1, a UDP-glucose/UDP-galactose Transporter from Arabidopsis thaliana, Is Located in the Endoplasmic Reticulum and Up-regulated by the Unfolded Protein Response, *Journal of Biological Chemistry 281*, 9145-9151.
120. Reyes, F., León, G., Donoso, M., Brandizzí, F., Weber, A. P. M., and Orellana, A. (2010) The nucleotide sugar transporters AtUTr1 and AtUTr3 are required for the incorporation of UDP-glucose into the endoplasmic reticulum, are essential for pollen development and are needed for embryo sac progress in Arabidopsis thaliana, *The Plant Journal 61*, 423-435.
121. Moreno, I., Moreno, A., Donoso, M., Moraga, C., Nebel, J. C. and Orellana, A. (2011) Dissecting functional domains on nucleotide sugar transporters In *XXXIV Reunión anual Sociedad de Bioquímica y Biología Molecular de Chile*, Unpublished, Pedro de Valdivia, Chile.
122. Hadley, B., Maggioni, A., Ashikov, A., Day, C. J., Haselhorst, T., and Tiralongo, J. (2014) Structure and function of nucleotide sugar transporters: Current progress, *Computational and Structural Biotechnology Journal 10*, 23-32.
123. Schallus, T., Jaeckh, C., Fehér, K., Palma, A. S., Liu, Y., Simpson, J. C., Mackeen, M., Stier, G., Gibson, T. J., Feizi, T., Pieler, T., and Muhle-Goll, C. (2008) Malectin: A Novel Carbohydrate-binding Protein of the Endoplasmic Reticulum and a Candidate Player in the Early Steps of Protein N-Glycosylation, *Molecular Biology of the Cell 19*, 3404-3414.
124. Schallus, T., Fehér, K., Sternberg, U., Rybin, V., and Muhle-Goll, C. (2010) Analysis of the specific interactions between the lectin domain of malectin and diglucosides, *Glycobiology 20*, 1010-1020.
125. Qin, S.-Y., Hu, D., Matsumoto, K., Takeda, K., Matsumoto, N., Yamaguchi, Y., and Yamamoto, K. (2012) Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins, *The Journal of biological chemistry 287*, 38080-38089.
126. Tannous, A., Pisoni, G. B., Hebert, D. N., and Molinari, M. (2015) N-linked sugar-regulated protein folding and quality control in the ER, *Semin Cell Dev Biol 41*, 79-89.

127. Gelderblom, H. R. (1996) Chapter 41Structure and Classification of Viruses, In *Medical Microbiology. 4th edition.* (Baron, S., Ed.), University of Texas Medical Branch at Galveston, Galveston (TX).
128. Berlanga, M. (2010) *Brock Biology of Microorganisms (11th edn). Michael T. Madigan, John M. Martinko (eds)*.
129. Vella, F. (1994) Molecular biology of the cell (third edition): By B Alberts, D Bray, J Lewis, M Raff, K Roberts and J D Watson. pp 1361. Garland Publishing, New York and London. 1994, *Biochemical Education 22*, 164-164.
130. Mustafa, M. S., Rasotgi, V., Jain, S., and Gupta, V. (2015) Discovery of fifth serotype of dengue virus (DENV-5): A new public health dilemma in dengue control, *Medical Journal Armed Forces India 71*, 67-70.
131. Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E., Battle, K. E., Simmons, C. P., and Hay, S. I. (2014) Global spread of dengue virus types: mapping the 70 year history, *Trends in Microbiology 22*, 138-146.
132. Kliks, S. C., Nisalak, A., Brandt, W. E., Wahl, L., and Burke, D. S. (1989) Antibody-Dependent Enhancement of Dengue Virus Growth in Human Monocytes as a Risk Factor for Dengue Hemorrhagic Fever, *The American Journal of Tropical Medicine and Hygiene 40*, 444-451.
133. Guzman, M. G., Halstead, S. B., Artsob, H., Buchy, P., Farrar, J., Gubler, D. J., Hunsperger, E., Kroeger, A., Margolis, H. S., Martínez, E., Nathan, M. B., Pelegrino, J. L., Simmons, C., Yoksan, S., and Peeling, R. W. (2010) Dengue: a continuing global threat, *Nature Reviews Microbiology 8*, S7.
134. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein, *Proceedings of the National Academy of Sciences 100*, 6986.
135. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2004) Structure of the dengue virus envelope protein after membrane fusion, *Nature 427*, 313.
136. Idris, F., Muharram, S. H., and Diah, S. (2016) Glycosylation of dengue virus glycoproteins and their interactions with carbohydrate receptors: possible targets for antiviral therapy, *Archives of Virology 161*, 1751-1760.
137. Mackenzie, J. M., Jones, M. K., and Young, P. R. (1996) Immunolocalization of the Dengue Virus Nonstructural Glycoprotein NS1 Suggests a Role in Viral RNA Replication, *Virology 220*, 232-240.
138. Fan, J., Liu, Y., and Yuan, Z. (2014) Critical role of Dengue Virus NS1 protein in viral replication, *Virologica Sinica 29*, 162-169.
139. Flamand, M., Megret, F., Mathieu, M., Lepault, J., Rey, F. A., and Deubel, V. (1999) Dengue Virus Type 1 Nonstructural Glycoprotein NS1 Is Secreted from Mammalian Cells as a Soluble Hexamer in a Glycosylation-Dependent Fashion, *Journal of Virology 73*, 6104-6110.
140. Somnuke, P., Hauhart, R. E., Atkinson, J. P., Diamond, M. S., and Avirutnan, P. (2011) N-linked glycosylation of dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement, *Virology 413*, 253-264.
141. Courageot, M. P., Frenkiel, M. P., Dos Santos, C. D., Deubel, V., and Despres, P. (2000) Alpha-glucosidase inhibitors reduce dengue virus production by affecting the initial steps of virion morphogenesis in the endoplasmic reticulum, *J Virol 74*, 564-572.

142. Sanchez, I. J., and Ruiz, B. H. (1996) A single nucleotide change in the E protein gene of dengue virus 2 Mexican strain affects neurovirulence in mice, *The Journal of general virology 77 (Pt 10)*, 2541-2545.
143. Pokidysheva, E., Zhang, Y., Battisti, A. J., Bator-Kelly, C. M., Chipman, P. R., Xiao, C., Gregorio, G. G., Hendrickson, W. A., Kuhn, R. J., and Rossmann, M. G. (2006) Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN, *Cell 124*, 485-493.
144. Bryant, J. E., Calvert, A. E., Mesesan, K., Crabtree, M. B., Volpe, K. E., Silengo, S., Kinney, R. M., Huang, C. Y., Miller, B. R., and Roehrig, J. T. (2007) Glycosylation of the dengue 2 virus E protein at N67 is critical for virus growth in vitro but not for growth in intrathoracically inoculated Aedes aegypti mosquitoes, *Virology 366*, 415-423.
145. Mondotte, J. A., Lozach, P. Y., Amara, A., and Gamarnik, A. V. (2007) Essential role of dengue virus envelope protein N glycosylation at asparagine-67 during viral propagation, *J Virol 81*, 7136-7148.
146. Dejnirattisai, W., Supasa, P., Wongwiwat, W., Rouvinski, A., Barba-Spaeth, G., Duangchinda, T., Sakuntabhai, A., Cao-Lormeau, V.-M., Malasit, P., Rey, F. A., Mongkolsapaya, J., and Screaton, G. R. (2016) Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus, *Nature Immunology 17*, 1102.
147. Arredondo-García, J. L., Hadinegoro, S. R., Reynales, H., Chua, M. N., Rivera Medina, D. M., Chotpitayasunondh, T., Tran, N. H., Deseda, C. C., Wirawan, D. N., Cortés Supelano, M., Frago, C., Langevin, E., Coronel, D., Laot, T., Perroud, A. P., Sanchez, L., Bonaparte, M., Limkittikul, K., Chansinghakul, D., Gailhardou, S., Noriega, F., Wartel, T. A., Bouckenooghe, A., and Zambrano, B. (2018) Four-year safety follow-up of the tetravalent dengue vaccine efficacy randomized controlled trials in Asia and Latin America, *Clinical Microbiology and Infection 24*, 755-763.
148. Hadinegoro, S. R., Arredondo-García, J. L., Capeding, M. R., Deseda, C., Chotpitayasunondh, T., Dietze, R., Hj Muhammad Ismail, H. I., Reynales, H., Limkittikul, K., Rivera-Medina, D. M., Tran, H. N., Bouckenooghe, A., Chansinghakul, D., Cortés, M., Fanouillere, K., Forrat, R., Frago, C., Gailhardou, S., Jackson, N., Noriega, F., Plennevaux, E., Wartel, T. A., Zambrano, B., and Saville, M. (2015) Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease, *New England Journal of Medicine 373*, 1195-1206.
149. Gailhardou, S., Skipetrova, A., Dayan, G. H., Jezorwski, J., Saville, M., Van der Vliet, D., and Wartel, T. A. (2016) Safety Overview of a Recombinant Live-Attenuated Tetravalent Dengue Vaccine: Pooled Analysis of Data from 18 Clinical Trials, *PLOS Neglected Tropical Diseases 10*, e0004821.
150. Plennevaux, E., Moureau, A., Arredondo-García, J. L., Villar, L., Pitisuttithum, P., Tran, N. H., Bonaparte, M., Chansinghakul, D., Coronel, D. L., L'Azou, M., Ochiai, R. L., Toh, M.-L., Noriega, F., and Bouckenooghe, A. (2018) Impact of Dengue Vaccination on Serological Diagnosis: Insights From Phase III Dengue Vaccine Efficacy Trials, *Clinical Infectious Diseases 66*, 1164-1172.
151. Taguwa, S., Maringer, K., Li, X., Bernal-Rubio, D., Rauch, J. N., Gestwicki, J. E., Andino, R., Fernandez-Sesma, A., and Frydman, J. (2015) Defining Hsp70 Subnetworks in Dengue Virus Replication Reveals Key Vulnerability in Flavivirus Infection, *Cell 163*, 1108-1123.
152. Rathore, A. P., Paradkar, P. N., Watanabe, S., Tan, K. H., Sung, C., Connolly, J. E., Low, J., Ooi, E. E., and Vasudevan, S. G. (2011) Celgosivir treatment misfolds dengue virus NS1

protein, induces cellular pro-survival genes and protects against lethal challenge mouse model, *Antiviral Res 92*, 453-460.

153. Sayce, A. C., Alonzi, D. S., Killingbeck, S. S., Tyrrell, B. E., Hill, M. L., Caputo, A. T., Iwaki, R., Kinami, K., Ide, D., Kiappes, J. L., Beatty, P. R., Kato, A., Harris, E., Dwek, R. A., Miller, J. L., and Zitzmann, N. (2016) Iminosugars Inhibit Dengue Virus Production via Inhibition of ER Alpha-Glucosidases—Not Glycolipid Processing Enzymes, *PLOS Neglected Tropical Diseases 10*, e0004524.
154. Cao-Lormeau, V. M., Blake, A., Mons, S., Lastere, S., Roche, C., Vanhomwegen, J., Dub, T., Baudouin, L., Teissier, A., Larre, P., Vial, A. L., Decam, C., Choumet, V., Halstead, S. K., Willison, H. J., Musset, L., Manuguerra, J. C., Despres, P., Fournier, E., Mallet, H. P., Musso, D., Fontanet, A., Neil, J., and Ghawché, F. (2016) Guillain-Barré Syndrome outbreak caused by ZIKA virus infection in French Polynesia, *Lancet (London, England) 387*, 1531-1539.
155. Mlakar, J., Korva, M., Tul, N., Popovic, M., Poljsak-Prijatelj, M., Mraz, J., Kolenc, M., Resman Rus, K., Vesnaver Vipotnik, T., Fabjan Vodusek, V., Vizjak, A., Pizem, J., Petrovec, M., and Avsic Zupanc, T. (2016) Zika Virus Associated with Microcephaly, *The New England journal of medicine 374*, 951-958.
156. Dick, G. W. A., Kitchen, S. F., and Haddow, A. J. (1952) Zika Virus (I). Isolations and serological specificity, *Transactions of The Royal Society of Tropical Medicine and Hygiene 46*, 509-520.
157. Ai, J.-W., Zhang, Y., and Zhang, W. (2016) Zika virus outbreak: 'a perfect storm', *Emerging Microbes & Infections 5*, e21.
158. Rodriguez-Morales, A. J. (2016) Zika and microcephaly in Latin America: An emerging threat for pregnant travelers?, *Travel Medicine and Infectious Disease 14*, 5-6.
159. Lucchese, G., and Kanduc, D. (2016) Zika virus and autoimmunity: From microcephaly to Guillain-Barré syndrome, and beyond, *Autoimmunity Reviews 15*, 801-808.
160. Cao-Lormeau, V.-M., Blake, A., Mons, S., Lastère, S., Roche, C., Vanhomwegen, J., Dub, T., Baudouin, L., Teissier, A., Larre, P., Vial, A.-L., Decam, C., Choumet, V., Halstead, S. K., Willison, H. J., Musset, L., Manuguerra, J.-C., Despres, P., Fournier, E., Mallet, H.-P., Musso, D., Fontanet, A., Neil, J., and Ghawché, F. (2016) Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: a case-control study, *The Lancet 387*, 1531-1539.
161. Colón-González, F. J., Peres, C. A., Steiner São Bernardo, C., Hunter, P. R., and Lake, I. R. (2017) After the epidemic: Zika virus projections for Latin America and the Caribbean, *PLOS Neglected Tropical Diseases 11*, e0006007.
162. Fauci, A. S., and Morens, D. M. (2016) Zika Virus in the Americas — Yet Another Arbovirus Threat, *New England Journal of Medicine 374*, 601-604.
163. Wang, A., Thurmond, S., Islas, L., Hui, K., and Hai, R. (2017) Zika virus genome biology and molecular pathogenesis, *Emerging Microbes & Infections 6*, e13.
164. Kostyuchenko, V. A., Lim, E. X., Zhang, S., Fibriansah, G., Ng, T. S., Ooi, J. S., Shi, J., and Lok, S. M. (2016) Structure of the thermally stable Zika virus, *Nature 533*, 425-428.
165. Sirohi, D., Chen, Z., Sun, L., Klose, T., Pierson, T. C., Rossmann, M. G., and Kuhn, R. J. (2016) The 3.8Å resolution cryo-EM structure of Zika Virus(), *Science (New York, N.Y.) 352*, 467-470.
166. Beasley, D. W., Whiteman, M. C., Zhang, S., Huang, C. Y., Schneider, B. S., Smith, D. R., Gromowski, G. D., Higgs, S., Kinney, R. M., and Barrett, A. D. (2005) Envelope protein glycosylation status influences mouse neuroinvasion phenotype of genetic lineage 1 West Nile virus strains, *J Virol 79*, 8339-8347.

167. Zhu, Z., Chan, J. F.-W., Tee, K.-M., Choi, G. K.-Y., Lau, S. K.-P., Woo, P. C.-Y., Tse, H., and Yuen, K.-Y. (2016) Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for virological factors potentially associated with the rapidly expanding epidemic, *Emerging Microbes &Amp; Infections 5*, e22.
168. Haddow, A. D., Schuh, A. J., Yasuda, C. Y., Kasper, M. R., Heang, V., Huy, R., Guzman, H., Tesh, R. B., and Weaver, S. C. (2012) Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage, *PLoS Neglected Tropical Diseases 6*, e1477.
169. Chambers, T. J., Halevy, M., Nestorowicz, A., Rice, C. M., and Lustig, S. (1998) West Nile virus envelope proteins: nucleotide sequence analysis of strains differing in mouse neuroinvasiveness, *The Journal of general virology 79 (Pt 10)*, 2375-2380.
170. Fontes-Garfias, C. R., Shan, C., Luo, H., Muruato, A. E., Medeiros, D. B. A., Mays, E., Xie, X., Zou, J., Roundy, C. M., Wakamiya, M., Rossi, S. L., Wang, T., Weaver, S. C., and Shi, P. Y. (2017) Functional Analysis of Glycosylation of Zika Virus Envelope Protein, *Cell reports 21*, 1180-1190.
171. Mossenta, M., Marchese, S., Poggianella, M., Slon Campos, J. L., and Burrone, O. R. (2017) Role of N-glycosylation on Zika virus E protein secretion, viral assembly and infectivity, *Biochemical and Biophysical Research Communications 492*, 579-586.
172. Annamalai, A. S., Pattnaik, A., Sahoo, B. R., Muthukrishnan, E., Natarajan, S. K., Steffen, D., Vu, H. L. X., Delhon, G., Osorio, F. A., Petro, T. M., Xiang, S.-h., and Pattnaik, A. K. (2017) Zika Virus Encoding Non-Glycosylated Envelope Protein is Attenuated and Defective in Neuroinvasion, *Journal of Virology*.
173. Wen, D., Li, S., Dong, F., Zhang, Y., Lin, Y., Wang, J., Zou, Z., and Zheng, A. (2018) N-glycosylation of Viral E Protein Is the Determinant for Vector Midgut Invasion by Flaviviruses, *mBio 9*.
174. Lennemann, N. J., Rhein, B. A., Ndungo, E., Chandran, K., Qiu, X., and Maury, W. (2014) Comprehensive Functional Analysis of N-Linked Glycans on Ebola Virus GP1, *mBio 5*, e00862-00813.
175. Wang, B., Wang, Y., Frabutt, D. A., Zhang, X., Yao, X., Hu, D., Zhang, Z., Liu, C., Zheng, S., Xiang, S. H., and Zheng, Y. H. (2017) Mechanistic understanding of N-glycosylation in Ebola virus glycoprotein maturation and function, *292*, 5860-5870.
176. Dowling, W., Thompson, E., Badger, C., Mellquist, J. L., Garrison, A. R., Smith, J. M., Paragas, J., Hogan, R. J., and Schmaljohn, C. (2007) Influences of glycosylation on antigenicity, immunogenicity, and protective efficacy of ebola virus GP DNA vaccines, *J Virol 81*, 1821-1837.
177. Lennemann, N. J., Walkner, M., Berkebile, A. R., Patel, N., and Maury, W. (2015) The Role of Conserved N-Linked Glycans on Ebola Virus Glycoprotein 2, *The Journal of Infectious Diseases 212*, S204-S209.
178. Organisation, W. H. (2018) Seasonal Influenza Fact Sheet (http://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)).
179. Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., Gust, I. D., Hampson, A. W., Hay, A. J., Hurt, A. C., de Jong, J. C., Kelso, A., Klimov, A. I., Kageyama, T., Komadina, N., Lapedes, A. S., Lin, Y. P., Mosterin, A., Obuchi, M., Odagiri, T., Osterhaus, A. D. M. E., Rimmelzwaan, G. F., Shaw, M. W., Skepner, E., Stohr, K., Tashiro, M., Fouchier, R. A. M., and Smith, D. J. (2008) Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses, *Vaccine 26*, D31-D34.
180. Bouvier, N. M., and Palese, P. (2008) The biology of influenza viruses, *Vaccine 26*, D49-D53.

181. Abe, Y., Takashita, E., Sugawara, K., Matsuzaki, Y., Muraki, Y., and Hongo, S. (2004) Effect of the Addition of Oligosaccharides on the Biological Activities and Antigenicity of Influenza A/H3N2 Virus Hemagglutinin, *Journal of Virology 78*, 9605-9611.
182. Daniels, R., Kurowski, B., Johnson, A. E., and Hebert, D. N. (2003) N-Linked Glycans Direct the Cotranslational Folding Pathway of Influenza Hemagglutinin, *Molecular Cell 11*, 79-90.
183. Wagner, R., Wolff, T., Herwig, A., Pleschka, S., and Klenk, H.-D. (2000) Interdependence of Hemagglutinin Glycosylation and Neuraminidase as Regulators of Influenza Virus Growth: a Study by Reverse Genetics, *Journal of Virology 74*, 6316-6323.
184. Pieren, M., Galli, C., Denzel, A., and Molinari, M. (2005) The use of calnexin and calreticulin by cellular and viral glycoproteins, *Journal of Biological Chemistry 280*, 28265-28271.
185. Frankel, A. D., and Young, J. A. T. (1998) HIV-1: Fifteen Proteins and an RNA, *Annual review of biochemistry 67*, 1-25.
186. Cihlar, T., and Fordyce, M. (2016) Current status and prospects of HIV treatment, *Current Opinion in Virology 18*, 50-56.
187. Behrens, A.-J., Vasiljevic, S., Pritchard, Laura K., Harvey, David J., Andev, Rajinder S., Krumm, Stefanie A., Struwe, Weston B., Cupo, A., Kumar, A., Zitzmann, N., Seabright, Gemma E., Kramer, Holger B., Spencer, Daniel I. R., Royle, L., Lee, Jeong H., Klasse, Per J., Burton, Dennis R., Wilson, Ian A., Ward, Andrew B., Sanders, Rogier W., Moore, John P., Doores, Katie J., and Crispin, M. (2016) Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein, *Cell reports 14*, 2695-2706.
188. Fenouillet, E., Gluckman, J. C., and Bahraoui, E. (1990) Role of N-linked glycans of envelope glycoproteins in infectivity of human immunodeficiency virus type 1, *Journal of Virology 64*, 2841-2848.
189. Montefiori, D. C., Robinson Jr, W. E., and Mitchell, W. M. (1988) Role of protein N-glycosylation in pathogenesis of human immunodeficieny virus type 1, *Proceedings of the National Academy of Sciences of the United States of America 85*, 9248-9252.
190. Lee, W. R., Syu, W. J., Du, B., Matsuda, M., Tan, S., Wolf, A., Essex, M., and Lee, T. H. (1992) Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1, *Proceedings of the National Academy of Sciences 89*, 2213.
191. Reitter, J. N., Means, R. E., and Desrosiers, R. C. (1998) A role for carbohydrates in immune evasion in AIDS, *Nature Medicine 4*, 679.
192. Raska, M., Takahashi, K., Czernekova, L., Zachova, K., Hall, S., Moldoveanu, Z., Elliott, M. C., Wilson, L., Brown, R., Jancova, D., Barnes, S., Vrbkova, J., Tomana, M., Smith, P. D., Mestecky, J., Renfrow, M. B., and Novak, J. (2010) Glycosylation patterns of HIV-1 gp120 depend on the type of expressing cells and affect antibody recognition, *The Journal of biological chemistry 285*, 20860-20869.
193. Wang, W., Nie, J., Prochnow, C., Truong, C., Jia, Z., Wang, S., Chen, X. S., and Wang, Y. (2013) A systematic study of the N-glycosylation sites of HIV-1 envelope protein on infectivity and antibody-mediated neutralization, *Retrovirology 10*, 14-14.
194. Inouye, S., Tsuruoka, T., and Nida, T. (1966) The structure of nojirimycin, a piperidinose sugar antibiotic, *The Journal of antibiotics 19*, 288-292.
195. Yagi, M., Kouno, T., Aoyagi, Y., and Murai, H. (1976) The Structure of Moranoline, a Piperidine Alkaloid from Morus Species, *Nippon Nōgeikagaku Kaishi 50*, 571-572.

196. Asano, N., Nash, R. J., Molyneux, R. J., and Fleet, G. W. J. (2000) Sugar-mimic glycosidase inhibitors: natural occurrence, biological activity and prospects for therapeutic application, *Tetrahedron: Asymmetry 11*, 1645-1680.
197. Hettkamp, H., Legler, G., and Bause, E. (1984) Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides, *European journal of biochemistry / FEBS 142*, 85-90.
198. Hohenschutz, L. D., Bell, E. A., Jewess, P. J., Leworthy, D. P., Pryce, R. J., Arnold, E., and Clardy, J. (1981) Castanospermine, A 1,6,7,8-tetrahydroxyoctahydroindolizine alkaloid, from seeds of Castanospermum australe, *Phytochemistry 20*, 811-814.
199. Asano, N. (2000) Alkaloidal Sugar Mimetics: Biological Activities and Therapeutic Applications, *Journal of Enzyme Inhibition 15*, 215-234.
200. Saunier, B., Kilker, R. D., Tkacz, J. S., Quaroni, A., and Herscovics, A. (1982) Inhibition of N-linked complex oligosaccharide formation by 1-deoxynojirimycin, an inhibitor of processing glucosidases, *Journal of Biological Chemistry 257*, 14155-14161.
201. Hettkamp, H., Bause, E., and Legler, G. (1982) Inhibition by nojirimycin and 1-deoxynojirimycin of microsomal glucosidases from calf liver acting on the glycoprotein oligosaccharides Glc_{1–3}Man₉GlcNAc<sub>2</sub&gt, *Bioscience Reports 2*, 899.
202. Scofield, A. M., Fellows, L. E., Nash, R. J., and Fleet, G. W. J. (1986) Inhibition of mammalian digestive disaccharidases by polyhydroxy alkaloids, *Life Sciences 39*, 645-650.
203. Andersson, U., Butters, T. D., Dwek, R. A., and Platt, F. M. (2000) N-butyldeoxygalactonojirimycin: a more selective inhibitor of glycosphingolipid biosynthesis than N-butyldeoxynojirimycin, in vitro and in vivo, *Biochemical Pharmacology 59*, 821-829.
204. Rhinehart, B. L., Robinson, K. M., King, C.-H. R., and Liu, P. S. (1990) Castanospermine-glucosides as selective disaccharidase inhibitors, *Biochemical Pharmacology 39*, 1537-1543.
205. Overkleeft, H. S., Renkema, G. H., Neele, J., Vianello, P., Hung, I. O., Strijland, A., van der Burg, A. M., Koomen, G.-J., Pandit, U. K., and Aerts, J. M. F. G. (1998) Generation of Specific Deoxynojirimycin-type Inhibitors of the Non-lysosomal Glucosylceramidase, *Journal of Biological Chemistry 273*, 26522-26527.
206. Li, Y.-G., Ji, D.-F., Zhong, S., Lv, Z.-Q., Lin, T.-B., Chen, S., and Hu, G.-Y. (2011) Hybrid of 1-deoxynojirimycin and polysaccharide from mulberry leaves treat diabetes mellitus by activating PDX-1/insulin-1 signaling pathway and regulating the expression of glucokinase, phosphoenolpyruvate carboxykinase and glucose-6-phosphatase in alloxan-induced diabetic mice, *Journal of Ethnopharmacology 134*, 961-970.
207. Joubert, P. H., Bam, W. J., and Manyane, N. (1986) Effect of an alpha-glucosidase inhibitor (BAY m 1099) on post-prandial blood glucose and insulin in type II diabetics, *European Journal of Clinical Pharmacology 30*, 253-255.
208. Yoshikuni, Y. (1991) Alpha-Glucosidase Inhibitor, 1-Deoxynojirimycin and Its N-Substituted Derivatives as Anti-Diabetic Agents or Anti-Human Immunodeficiency Virus Agents, *Trends in Glycoscience and Glycotechnology 3*, 184-192.
209. Campbell, L. K., Baker, D. E., and Campbell, R. K. (2000) Miglitol: Assessment of its Role in the Treatment of Patients with Diabetes Mellitus, *Annals of Pharmacotherapy 34*, 1291-1301.

210. Mitrakou, A., Tountas, N., Raptis, A. E., Bauer, R. J., Schulz, H., and Raptis, S. A. (1998) Long-term effectiveness of a new α-glucosidase inhibitor (BAY m1099-miglitol) in insulin-treated Type 2 diabetes mellitus, *Diabetic Medicine 15*, 657-660.
211. Kingma, P. J., Menheere, P. P. C. A., Sels, J. P., and Kruseman, A. C. N. (1992) α-Glucosidase Inhibition by Miglitol in NIDDM Patients, *Diabetes Care 15*, 478.
212. Horne, G., Wilson, F. X., Tinsley, J., Williams, D. H., and Storer, R. (2011) Iminosugars past, present and future: medicines for tomorrow, *Drug Discovery Today 16*, 107-118.
213. Winchester, B. G. (2009) Iminosugars: from botanical curiosities to licensed drugs, *Tetrahedron: Asymmetry 20*, 645-651.
214. Pastores, G. M., Barnett, N. L., and Kolodny, E. H. (2005) An open-label, noncomparative study of miglustat in type I Gaucher disease: Efficacy and tolerability over 24 months of treatment, *Clinical Therapeutics 27*, 1215-1227.
215. Grabowski, G. A. (2008) Phenotype, diagnosis, and treatment of Gaucher's disease, *The Lancet 372*, 1263-1271.
216. Abian, O., Alfonso, P., Velazquez-Campoy, A., Giraldo, P., Pocovi, M., and Sancho, J. (2011) Therapeutic Strategies for Gaucher Disease: Miglustat (NB-DNJ) as a Pharmacological Chaperone for Glucocerebrosidase and the Different Thermostability of Velaglucerase Alfa and Imiglucerase, *Molecular Pharmaceutics 8*, 2390-2397.
217. Lyseng-Williamson, K. A. (2014) Miglustat: A Review of Its Use in Niemann-Pick Disease Type C, *Drugs 74*, 61-74.
218. Wraith, J. E., and Imrie, J. (2009) New therapies in the management of Niemann-Pick type C disease: clinical utility of miglustat, *Therapeutics and Clinical Risk Management 5*, 877-887.
219. Lachmann, R. H., te Vruchte, D., Lloyd-Evans, E., Reinkensmeier, G., Sillence, D. J., Fernandez-Guillen, L., Dwek, R. A., Butters, T. D., Cox, T. M., and Platt, F. M. (2004) Treatment with miglustat reverses the lipid-trafficking defect in Niemann–Pick disease type C, *Neurobiology of Disease 16*, 654-658.
220. Patterson, M. C., Vecchio, D., Prady, H., Abel, L., and Wraith, J. E. (2007) Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study, *The Lancet Neurology 6*, 765-772.
221. Tropak, M. B., Reid, S. P., Guiral, M., Withers, S. G., and Mahuran, D. (2004) Pharmacological Enhancement of β-Hexosaminidase Activity in Fibroblasts from Adult Tay-Sachs and Sandhoff Patients, *Journal of Biological Chemistry 279*, 13478-13487.
222. Platt, F. M., Neises, G. R., Reinkensmeier, G., Townsend, M. J., Perry, V. H., Proia, R. L., Winchester, B., Dwek, R. A., and Butters, T. D. (1997) Prevention of Lysosomal Storage in Tay-Sachs Mice Treated with N-Butyldeoxynojirimycin, *Science 276*, 428.
223. Markham, A. (2016) Migalastat: First Global Approval, *Drugs 76*, 1147-1152.
224. Parenti, G., Zuppaldi, A., Gabriela Pittis, M., Rosaria Tuzzi, M., Annunziata, I., Meroni, G., Porto, C., Donaudy, F., Rossi, B., Rossi, M., Filocamo, M., Donati, A., Bembi, B., Ballabio, A., and Andria, G. (2007) Pharmacological enhancement of mutated alpha-glucosidase activity in fibroblasts from patients with Pompe disease, *Molecular therapy : the journal of the American Society of Gene Therapy 15*, 508-514.
225. Wennekes, T., Meijer, A. J., Groen, A. K., Boot, R. G., Groener, J. E., van Eijk, M., Ottenhoff, R., Bijl, N., Ghauharali, K., Song, H., O'Shea, T. J., Liu, H., Yew, N., Copeland, D., van den Berg, R. J., van der Marel, G. A., Overkleeft, H. S., and Aerts, J. M. (2010) Dual-Action Lipophilic Iminosugar Improves Glycemic Control in Obese Rodents by Reduction of Visceral Glycosphingolipids and Buffering of Carbohydrate Assimilation, *Journal of Medicinal Chemistry 53*, 689-698.

226. Langeveld, M., van den Berg, S. A. A., Bijl, N., Bijland, S., van Roomen, C. P., Houben-Weerts, J. H., Ottenhoff, R., Houten, S. M., van Dijk, K. W., Romijn, J. A., Groen, A. K., Aerts, J. M., and Voshol, P. J. (2012) Treatment of genetically obese mice with the iminosugar N-(5-adamantane-1-yl-methoxy-pentyl)-deoxynojirimycin reduces body weight by decreasing food intake and increasing fat oxidation, *Metabolism - Clinical and Experimental 61*, 99-107.
227. van Eijk, M., Aten, J., Bijl, N., Ottenhoff, R., van Roomen, C. P. A. A., Dubbelhuis, P. F., Seeman, I., Ghauharali-van der Vlugt, K., Overkleeft, H. S., Arbeeny, C., Groen, A. K., and Aerts, J. M. F. G. (2009) Reducing Glycosphingolipid Content in Adipose Tissue of Obese Mice Restores Insulin Sensitivity, Adipogenesis and Reduces Inflammation, *PLOS ONE 4*, e4723.
228. Pan, Y. T., Hori, H., Saul, R., Sanford, B. A., Molyneux, R. J., and Elbein, A. D. (1983) Castanospermine inhibits the processing of the oligosaccharide portion of the influenza viral hemagglutinin, *Biochemistry 22*, 3975-3984.
229. Elbein, A. D., Solf, R., Dorling, P. R., and Vosbeck, K. (1981) Swainsonine: an inhibitor of glycoprotein processing, *Proceedings of the National Academy of Sciences 78*, 7393.
230. Olden, K., Breton, P., Grzegorzewski, K., Yasuda, Y., Gause, B. L., Oredipe, O. A., Newton, S. A., and White, S. L. (1991) The potential importance of swainsonine in therapy for cancers and immunology, *Pharmacology & Therapeutics 50*, 285-290.
231. You, N., Liu, W., Wang, T., Ji, R., Wang, X., Gong, Z., Dou, K., and Tao, K. (2012) Swainsonine inhibits growth and potentiates the cytotoxic effect of paclitaxel in hepatocellular carcinoma in vitro and in vivo, *Oncology reports 28*, 2091-2100.
232. Roberts, J. D., Klein, J. L., Palmantier, R., Dhume, S. T., George, M. D., and Olden, K. (1998) The role of protein glycosylation inhibitors in the prevention of metastasis and therapy of cancer, *Cancer detection and prevention 22*, 455-462.
233. Wojtowicz, K., Januchowski, R., Sosinska, P., Nowicki, M., and Zabel, M. (2016) Effect of brefeldin A and castanospermine on resistant cell lines as supplements in anticancer therapy, *Oncology reports 35*, 2896-2906.
234. Hadwiger, A., Niemann, H., Käbisch, A., Bauer, H., and Tamura, T. (1986) Appropriate glycosylation of the fms gene product is a prerequisite for its transforming potency, *The EMBO journal 5*, 689-694.
235. Allan, G., Ouadid-Ahidouch, H., Sanchez-Fernandez, E. M., Risquez-Cuadro, R., Fernandez, J. M. G., Ortiz-Mellet, C., and Ahidouch, A. (2013) New Castanospermine Glycoside Analogues Inhibit Breast Cancer Cell Proliferation and Induce Apoptosis without Affecting Normal Cells, *PLOS ONE 8*, e76411.
236. Shaheen, P. E., Stadler, W., Elson, P., Knox, J., Winquist, E., and Bukowski, R. M. (2005) Phase II study of the efficacy and safety of oral GD0039 in patients with locally advanced or metastatic renal cell carcinoma, *Investigational new drugs 23*, 577-581.
237. Li, Z., Xu, X., Huang, Y., Ding, L., Wang, Z., Yu, G., Xu, D., Li, W., and Tong, D. (2012) Swainsonine Activates Mitochondria-mediated Apoptotic Pathway in Human Lung Cancer A549 Cells and Retards the Growth of Lung Cancer Xenografts, *International Journal of Biological Sciences 8*, 394-405.
238. Schlesinger, S., Malfer, C., and Schlesinger, M. J. (1984) The formation of vesicular stomatitis virus (San Juan strain) becomes temperature-sensitive when glucose residues are retained on the oligosaccharides of the glycoprotein, *Journal of Biological Chemistry 259*, 7597-7601.

239. Schlesinger, S., Koyama, A. H., Malfer, C., Gee, S. L., and Schlesinger, M. J. (1985) The effects of inhibitors of glucosidase I on the formation of Sindbis virus, *Virus Research 2*, 139-149.
240. Datema, R., Romero, P. A., Legler, G., and Schwarz, R. T. (1982) Inhibition of formation of complex oligosaccharides by the glucosidase inhibitor bromoconduritol, *Proc Natl Acad Sci U S A 79*, 6787-6791.
241. Schwarz, R. T., and Datema, R. (1982) The lipid pathway of protein glycosylation and its inhibitors: the biological significance of protein-bound carbohydrates, *Advances in carbohydrate chemistry and biochemistry 40*, 287-379.
242. Datema, R., Romero, P. A., Rott, R., and Schwarz, R. T. (1984) On the role of oligosaccharide trimming in the maturation of Sindbis and influenza virus, *Arch Virol 81*, 25-39.
243. Repp, R., Tamura, T., Boschek, C. B., Wege, H., Schwarz, R. T., and Niemann, H. (1985) The effects of processing inhibitors of N-linked oligosaccharides on the intracellular migration of glycoprotein E2 of mouse hepatitis virus and the maturation of coronavirus particles, *Journal of Biological Chemistry 260*, 15873-15879.
244. Pinter, A., Honnen, W. J., and Li, J. S. (1984) Studies with inhibitors of oligosaccharide processing indicate a functional role for complex sugars in the transport and proteolysis of Friend mink cell focus-inducing murine leukemia virus envelope proteins, *Virology 136*, 196-210.
245. Taylor, D. L., Fellows, L. E., Farrar, G. H., Nash, R. J., Taylor-Robinson, D., Mobberley, M. A., Ryder, T. A., Jeffries, D. J., and Tyms, A. S. (1988) Loss of cytomegalovirus infectivity after treatment with castanospermine or related plant alkaloids correlates with aberrant glycoprotein synthesis, *Antiviral Res 10*, 11-26.
246. Gruters, R. A., Neefjes, J. J., Tersmette, M., de Goede, R. E. Y., Tulp, A., Huisman, H. G., Miedema, F., and Ploegh, H. L. (1987) Interference with HIV-induced syncytium formation and viral infectivity by inhibitors of trimming glucosidase, *Nature 330*, 74.
247. Karpas, A., Fleet, G. W., Dwek, R. A., Petursson, S., Namgoong, S. K., Ramsden, N. G., Jacob, G. S., and Rademacher, T. W. (1988) Aminosugar derivatives as potential anti-human immunodeficiency virus agents, *Proc Natl Acad Sci U S A 85*, 9229-9233.
248. Ahmed, S. P., Nash, R. J., Bridges, C. G., Taylor, D. L., Kang, M. S., Porter, E. A., and Tyms, A. S. (1995) Antiviral Activity and Metabolism of the Castanospermine Derivative MDL 28,574, in Cells Infected with Herpes Simplex Virus Type 2, *Biochemical and Biophysical Research Communications 208*, 267-273.
249. Courageot, M.-P., Frenkiel, M.-P., Duarte Dos Santos, C., Deubel, V., and Desprès, P. (2000) α-Glucosidase Inhibitors Reduce Dengue Virus Production by Affecting the Initial Steps of Virion Morphogenesis in the Endoplasmic Reticulum, *Journal of Virology 74*, 564-572.
250. Wu, S.-F., Lee, C.-J., Liao, C.-L., Dwek, R. A., Zitzmann, N., and Lin, Y.-L. (2002) Antiviral Effects of an Iminosugar Derivative on Flavivirus Infections, *Journal of Virology 76*, 3596-3604.
251. Alonzi, D. S., Scott, K. A., Dwek, R. A., and Zitzmann, N. (2017) Iminosugar antivirals: the therapeutic sweet spot, *Biochemical Society transactions 45*, 571-582.
252. Ratner, L. E. E., and Heyden, N. V. (1993) Mechanism of Action of N-Butyl Deoxynojirimycin in Inhibiting HIV-1 Infection and Activity in Combination with Nucleoside Analogs, *AIDS Research and Human Retroviruses 9*, 291-297.
253. Fischer, P. B., Karlsson, G. B., Butters, T. D., Dwek, R. A., and Platt, F. M. (1996) N-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry

correlates with changes in antibody recognition of the V1/V2 region of gp120, *Journal of Virology 70*, 7143-7152.

254. Mehta, A., Carrouée, S., Conyers, B., Jordan, R., Butters, T., Dwek Raymond, A., and Block Timothy, M. (2003) Inhibition of hepatitis B virus DNA replication by imino sugars without the inhibition of the DNA polymerase: Therapeutic implications, *Hepatology 33*, 1488-1495.
255. Chang, J., Warren, T. K., Zhao, X., Gill, T., Guo, F., Wang, L., Comunale, M. A., Du, Y., Alonzi, D. S., Yu, W., Ye, H., Liu, F., Guo, J.-T., Mehta, A., Cuconati, A., Butters, T. D., Bavari, S., Xu, X., and Block, T. M. (2013) Small molecule inhibitors of ER α-glucosidases are active against multiple hemorrhagic fever viruses, *Antiviral Research 98*, 432-440.
256. Miller, J. L., Spiro, S. G., Dowall, S. D., Taylor, I., Rule, A., Alonzi, D. S., Sayce, A. C., Wright, E., Bentley, E. M., Thom, R., Hall, G., Dwek, R. A., Hewson, R., and Zitzmann, N. (2016) Minimal In Vivo Efficacy of Iminosugars in a Lethal Ebola Virus Guinea Pig Model, *PLOS ONE 11*, e0167018.
257. Caputo, A. T. (2015) Structural and biochemical characterisation of the endoplasmic reticulum α-glucosidases In *Department of Biochemistry*, University of Oxford.
258. Aricescu, A. R., Lu, W., and Jones, E. Y. (2006) A time- and cost-efficient system for high-level protein production in mammalian cells, *Acta crystallographica. Section D, Biological crystallography 62*, 1243-1250.
259. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system, *Nature protocols 8*, 2281-2308.
260. Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., and O'Connor-Giles, K. M. (2014) Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila, *Genetics 196*, 961-971.
261. Bauer, D. E., Canver, M. C., and Orkin, S. H. (2015) Generation of Genomic Deletions in Mammalian Cell Lines via CRISPR/Cas9, *Journal of visualized experiments : JoVE*, 10.3791/52118.
262. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic acids research 32*, W668-673.
263. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, *Biopolymers 89*, 392-400.
264. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M., and Svergun, D. I. (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *Journal of Applied Crystallography 50*, 1212-1225.
265. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *Journal of Applied Crystallography 36*, 1277-1282.
266. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *Journal of Applied Crystallography 25*, 495-503.
267. Panjkovich, A., and Svergun, D. I. (2016) Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis, *Physical chemistry chemical physics : PCCP 18*, 5707-5719.
268. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering, *Journal of Applied Crystallography 42*, 342-346.

269. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophysical journal 89*, 1237-1250.
270. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering, *Journal of Applied Crystallography 36*, 860-864.
271. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophysical journal 76*, 2879-2886.
272. Webb, M. R. (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems, *Proceedings of the National Academy of Sciences of the United States of America 89*, 4884-4887.
273. Tickle, I. J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A.,, and Vonrhein, C., Bricogne, G. (2018) STARANISO, *Cambridge, United Kingdom: Global Phasing Ltd.*
274. Kabsch, W. (2010) XDS, *Acta Crystallographica Section D: Biological Crystallography 66*, 125-132.
275. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta crystallographica. Section D, Biological crystallography 67*, 235-242.
276. Keegan, R. M., and Winn, M. D. (2007) Automated search-model discovery and preparation for structure solution by molecular replacement, *Acta Crystallographica Section D 63*, 447-457.
277. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *Journal of Applied Crystallography 40*, 658-674.
278. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D: Biological Crystallography 66*, 486-501.
279. Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., Roversi P, Sharff A., Smart O.S., Vonrhein C., Womack T.O. (2016) BUSTER version 2.10.3, *Cambridge, United Kingdom: Global Phasing Ltd.*
280. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta crystallographica. Section D, Biological crystallography 66*, 12-21.
281. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank, *Nature structural biology 10*, 980.
282. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. .
283. Russell, P. K., Nisalak, A., Sukhavachana, P., and Vivona, S. (1967) A plaque reduction test for dengue virus neutralizing antibodies, *Journal of immunology (Baltimore, Md. : 1950) 99*, 285-290.
284. Butters, T. D., van den Broek, L. A. G. M., Fleet, G. W. J., Krulle, T. M., Wormald, M. R., Dwek, R. A., and Platt, F. M. (2000) Molecular requirements of imino sugars for the selective control of N-linked glycosylation and glycosphingolipid biosynthesis, *Tetrahedron: Asymmetry 11*, 113-124.
285. Kantardjieff, K. A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals, *Protein Science : A Publication of the Protein Society 12*, 1865-1871.
286. Keegan, R. M., and Winn, M. D. (2008) MrBUMP: an automated pipeline for molecular replacement, *Acta crystallographica. Section D, Biological crystallography 64*, 119-124.

287. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement, *Journal of Applied Crystallography 41*, 641-643.
288. Miyazaki, T., Ichikawa, M., Iino, H., Nishikawa, A., and Tonozuka, T. (2015) Crystal structure and substrate-binding mode of GH63 mannosylglycerate hydrolase from Thermus thermophilus HB8, *Journal of structural biology 190*, 21-30.
289. Miyazaki, T., Nishikawa, A., and Tonozuka, T. (2016) Crystal structure of the enzyme-product complex reveals sugar ring distortion during catalysis by family 63 inverting α-glycosidase, *Journal of structural biology 196*, 479-486.
290. Boulaflous, A., Saint-Jore-Dupas, C., Herranz-Gordo, M.-C., Pagny-Salehabadi, S., Plasson, C., Garidou, F., Kiefer-Meyer, M.-C., Ritzenthaler, C., Faye, L., and Gomord, V. (2009) Cytosolic N-terminal arginine-based signals together with a luminal signal target a type II membrane protein to the plant ER, *BMC plant biology 9*, 144.
291. Boraston, Alisdair B., Bolam, David N., Gilbert, Harry J., and Davies, Gideon J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition, *Biochemical Journal 382*, 769.
292. Alfonso, P., Pampín, S., Estrada, J., Rodríguez-Rey, J. C., Giraldo, P., Sancho, J., and Pocoví, M. (2005) Miglustat (NB-DNJ) works as a chaperone for mutated acid β-glucosidase in cells transfected with several Gaucher disease mutations, *Blood Cells, Molecules, and Diseases 35*, 268-276.
293. Tierney, M., Pottage, J., Kessler, H., Fischl, M., Richman, D., Merigan, T., Powderly, W., Smith, S., Karim, A., Sherman, J., and et al. (1995) The tolerability and pharmacokinetics of N-butyl-deoxynojirimycin in patients with advanced HIV disease (ACTG 100). The AIDS Clinical Trials Group (ACTG) of the National Institute of Allergy and Infectious Diseases, *Journal of acquired immune deficiency syndromes and human retrovirology : official publication of the International Retrovirology Association 10*, 549-553.
294. Kiappes, J. L., Hill, M. L., Alonzi, D. S., Miller, J. L., Iwaki, R., Sayce, A. C., Caputo, A. T., Kato, A., and Zitzmann, N. (2018) ToP-DNJ, a Selective Inhibitor of Endoplasmic Reticulum α-Glucosidase II Exhibiting Antiflaviviral Activity, *ACS Chemical Biology 13*, 60-65.
295. Tian, J., Liu, J., Li, J., Zheng, J., Chen, L., Wang, Y., Liu, Q., and Ni, J. (2018) The interaction of selenoprotein F (SELENOF) with retinol dehydrogenase 11 (RDH11) implied a role of SELENOF in vitamin A metabolism, *Nutrition & Metabolism 15*, 7.
296. Calles-Garcia, D., Yang, M., Soya, N., Melero, R., Ménade, M., Ito, Y., Vargas, J., Lukacs, G. L., Kollman, J. M., Kozlov, G., and Gehring, K. (2017) Single-particle electron microscopy structure of UDP-glucose:glycoprotein glucosyltransferase suggests a selectivity mechanism for misfolded proteins, *Journal of Biological Chemistry 292*, 11499-11507.
297. Vagin, A., and Teplyakov, A. (2007) MOLREP: an Automated Program for Molecular Replacement, *Journal of Applied Crystallography 30*, 1022-1025.
298. Grant, T. D., Luft, J. R., Carter, L. G., Matsui, T., Weiss, T. M., Martel, A., and Snell, E. H. (2015) The accurate assessment of small-angle X-ray scattering data, *Acta Crystallographica Section D 71*, 45-56.
299. Chilingaryan, Z., Yin, Z., and Oakley, A. J. (2012) Fragment-Based Screening by Protein Crystallography: Successes and Pitfalls, *International Journal of Molecular Sciences 13*, 12857-12879.
300. Wearsch, P. A., Peaper, D. R., and Cresswell, P. (2011) Essential glycan-dependent interactions optimize MHC class I peptide loading, *Proc Natl Acad Sci U S A 108*, 4950-4955.

301. Zhang, W., Wearsch, P. A., Zhu, Y., Leonhardt, R. M., and Cresswell, P. (2011) A role for UDP-glucose glycoprotein glucosyltransferase in expression and quality control of MHC class I molecules, *Proceedings of the National Academy of Sciences 108*, 4956.
302. Neerincx, A., Hermann, C., Antrobus, R., van Hateren, A., Cao, H., Trautwein, N., Stevanović, S., Elliott, T., Deane, J. E., and Boyle, L. H. (2017) TAPBPR bridges UDP-glucose:glycoprotein glucosyltransferase 1 onto MHC class I to provide quality control in the antigen presentation pathway, *eLife 6*, e23049.
303. Heaton, N. S., Moshkina, N., Fenouil, R., Gardner, T. J., Aguirre, S., Shah, P. S., Zhao, N., Manganaro, L., Hultquist, J., Noel, J., Sachs, D., Hamilton, J., Leon, P. E., Chawdury, A., Tripathy, S., Melegari, C., Campisi, L., Hai, R., Metreveli, G., Gamarnik, A. V., García-Sastre, A., Greenbaum, B., Simon, V., Fernandez-Sesma, A., Krogan, N., Mulder, L. C. F., van Bakel, H., Tortorella, D., Taunton, J., Palese, P., and Marazzi, I. (2016) Limiting influenza virus, HIV and dengue virus infection by targeting viral proteostasis, *Immunity 44*, 46-58.
304. Huang, S. (2011) Small Molecule Regulator of ENTPD5, and ER Enzyme in the PTEN/AKT Pathway, In *Graduate School of Biomedical Sciences*, UT Southwestern Medical Center.
305. Shen, Z. (2011) The Endoplasmic Reticulum Udpase ENTPD5 Promotes Cancer Cell Growth and Survival in the PI3K/PTEN.
306. Zebisch, M., Krauss, M., Schafer, P., and Strater, N. (2012) Crystallographic evidence for a domain motion in rat nucleoside triphosphate diphosphohydrolase (NTPDase) 1, *Journal of molecular biology 415*, 288-306.
307. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, In *Mol Syst Biol*, p 539.
308. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server, *Nucleic acids research 42*, W320-W324.
309. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature methods 8*, 785-786.
310. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03: a server for profile--profile sequence alignments, *Nucleic acids research 33*, W284-288.
311. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, *Current protocols in bioinformatics Chapter 5*, Unit-5.6.
312. Zebisch, M., Krauss, M., Schafer, P., Lauble, P., and Strater, N. (2013) Crystallographic snapshots along the reaction pathway of nucleoside triphosphate diphosphohydrolases, *Structure 21*, 1460-1475.
313. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic acids research 44*, W344-W350.
314. Satoh, T., Toshimori, T., Noda, M., Uchiyama, S., and Kato, K. (2016) Interaction mode between catalytic and regulatory subunits in glucosidase II involved in ER glycoprotein quality control, *Protein Science*, n/a-n/a.
315. Lin, D. H., Stuwe, T., Schilbach, S., Rundlet, E. J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F. M., Collins, L. N., Davenport, A. M., Jeon, Y. E., and Hoelz, A. (2016) Architecture of the symmetric core of the nuclear pore, *Science 352*.
316. Braun, N., Fengler, S., Ebeling, C., Servos, J., and Zimmermann, H. (2000) Sequencing, functional expression and characterization of rat NTPDase6, a nucleoside diphosphatase

and novel member of the ecto-nucleoside triphosphate diphosphohydrolase family, *Biochemical Journal 351*, 639-647.

317. Newman, J. (2004) Novel buffer systems for macromolecular crystallization, *Acta Crystallographica Section D 60*, 610-612.
318. Paroutis, P., Touret, N., and Grinstein, S. (2004) The pH of the Secretory Pathway: Measurement, Determinants, and Regulation, *Physiology 19*, 207-215.
319. Carter, S. G., and Karl, D. W. (1982) Inorganic phosphate assay with malachite green: an improvement and evaluation, *Journal of biochemical and biophysical methods 7*, 7-13.
320. Csala, M., Marcolongo, P., Lizák, B., Senesi, S., Margittai, É., Fulceri, R., Magyar, J. É., Benedetti, A., and Bánhegyi, G. (2007) Transport and transporters in the endoplasmic reticulum, *Biochimica et Biophysica Acta (BBA) - Biomembranes 1768*, 1325-1341.
321. Dorner, A. J., and Kaufman, R. J. (1994) The levels of endoplasmic reticulum proteins and ATP affect folding and secretion of selective proteins, *Biologicals : journal of the International Association of Biological Standardization 22*, 103-112.
322. Muller, W. E., Zahn, R. K., Gasic, M. J., Dogovic, N., Maidhof, A., Becker, C., Diehl-Seifert, B., and Eich, E. (1985) Avarol, a cytostatically active compound from the marine sponge Dysidea avara, *Comparative biochemistry and physiology. C, Comparative pharmacology and toxicology 80*, 47-52.
323. Kuchino, Y., Nishimura, S., Schroder, H. C., Rottmann, M., and Muller, W. E. (1988) Selective inhibition of formation of suppressor glutamine tRNA in Moloney murine leukemia virus-infected NIH-3T3 cells by Avarol, *Virology 165*, 518-526.
324. Zebisch, M., and Sträter, N. (2008) Structural insight into signal conversion and inactivation by NTPDase2 in purinergic signaling, *Proceedings of the National Academy of Sciences 105*, 6882-6887.
325. Zimmermann, H., Zebisch, M., and Sträter, N. (2012) Cellular function and molecular structure of ecto-nucleotidases, *Purinergic Signalling 8*, 437-502.
326. Failer, B. U., Braun, N., and Zimmermann, H. (2002) Cloning, Expression, and Functional Characterization of a Ca2+-dependent Endoplasmic Reticulum Nucleoside Diphosphatase, *Journal of Biological Chemistry 277*, 36978-36986.
327. Wang, H., La Russa, M., and Qi, L. S. (2016) CRISPR/Cas9 in Genome Editing and Beyond, *Annual review of biochemistry 85*, 227-264.
328. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity, *Science*.
329. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems, *Science (New York, N.Y.) 339*, 819-823.
330. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G., and Zhang, F. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases, *Nature biotechnology 31*, 827-832.
331. Ishida, N., and Kawakita, M. (2004) Molecular physiology and pathology of the nucleotide sugar transporter family (SLC35), *Pflügers Archiv 447*, 768-775.
332. Warren, W. C., Jasinska, A. J., Garcia-Perez, R., Svardal, H., Tomlinson, C., Rocchi, M., Archidiacono, N., Capozzi, O., Minx, P., Montague, M. J., Kyung, K., Hillier, L. W., Kremitzki, M., Graves, T., Chiang, C., Hughes, J., Tran, N., Huang, Y., Ramensky, V., Choi, O. W., Jung, Y. J., Schmitt, C. A., Juretic, N., Wasserscheid, J., Turner, T. R., Wiseman, R. W., Tuscher, J. J., Karl, J. A., Schmitz, J. E., Zahn, R., O'Connor, D. H., Redmond, E., Nisbett, A., Jacquelin, B., Muller-Trutwin, M. C., Brenchley, J. M., Dione, M., Antonio, M., Schroth,

G. P., Kaplan, J. R., Jorgensen, M. J., Thomas, G. W., Hahn, M. W., Raney, B. J., Aken, B., Nag, R., Schmitz, J., Churakov, G., Noll, A., Stanyon, R., Webb, D., Thibaud-Nissen, F., Nordborg, M., Marques-Bonet, T., Dewar, K., Weinstock, G. M., Wilson, R. K., and Freimer, N. B. (2015) The genome of the vervet (Chlorocebus aethiops sabaeus), *Genome research 25*, 1921-1933.

333. Tang, Y.-D., Liu, J.-T., Fang, Q.-Q., Wang, T.-Y., Sun, M.-X., An, T.-Q., Tian, Z.-J., and Cai, X.-H. (2016) Recombinant Pseudorabies Virus (PRV) Expressing Firefly Luciferase Effectively Screened for CRISPR/Cas9 Single Guide RNAs and Antiviral Compounds, *Viruses 8*.
334. van Diemen, F. R., Kruse, E. M., Hooykaas, M. J. G., Bruggeling, C. E., Schürch, A. C., van Ham, P. M., Imhof, S. M., Nijhuis, M., Wiertz, E. J. H. J., and Lebbink, R. J. (2016) CRISPR/Cas9-Mediated Genome Editing of Herpesviruses Limits Productive and Latent Infections, *PLOS Pathogens 12*, e1005701.
335. Mali, P., Esvelt, K. M., and Church, G. M. (2013) Cas9 as a versatile tool for engineering biology, *Nature methods 10*, 957-963.
336. Alonzi, Dominic S., Neville, David C. A., Lachmann, Robin H., Dwek, Raymond A., and Butters, Terry D. (2008) Glucosylated free oligosaccharides are biomarkers of endoplasmic- reticulum α-glucosidase inhibition, *Biochemical Journal 409*, 571.
337. Meertens, L., Labeau, A., Dejarnac, O., Cipriani, S., Sinigaglia, L., Bonnet-Madin, L., Le Charpentier, T., Hafirassou, M. L., Zamborlini, A., Cao-Lormeau, V.-M., Coulpier, M., Missé, D., Jouvenet, N., Tabibiazar, R., Gressens, P., Schwartz, O., and Amara, A. (2017) Axl Mediates ZIKA Virus Entry in Human Glial Cells and Modulates Innate Immune Responses, *Cell reports 18*, 324-333.
338. Navarro-Sanchez, E., Altmeyer, R., Amara, A., Schwartz, O., Fieschi, F., Virelizier, J. L., Arenzana-Seisdedos, F., and Desprès, P. (2003) Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses, *EMBO reports 4*, 723.
339. Chen, Y.-C., Wang, S.-Y., and King, C.-C. (1999) Bacterial Lipopolysaccharide Inhibits Dengue Virus Infection of Primary Human Monocytes/Macrophages by Blockade of Virus Entry via a CD14-Dependent Mechanism, *Journal of Virology 73*, 2650-2657.
340. Miller, J. L., deWet, B. J. M., Martinez-Pomares, L., Radcliffe, C. M., Dwek, R. A., Rudd, P. M., and Gordon, S. (2008) The Mannose Receptor Mediates Dengue Virus Infection of Macrophages, *PLOS Pathogens 4*, e17.
341. Wright, Addison V., Nuñez, James K., and Doudna, Jennifer A. (2016) Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering, *Cell 164*, 29-44.
342. Hammond, C., Braakman, I., and Helenius, A. (1994) Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control, *Proceedings of the National Academy of Sciences 91*, 913-917.
343. Takeda, Y., Seko, A., Fujikawa, K., Izumi, M., Kajihara, Y., and Ito, Y. (2016) Effects of domain composition on catalytic activity of human UDP-glucose:glycoprotein glucosyltransferases, *Glycobiology 26*, 999-1006.
344. Bredeston, L. M., Marino-Buslje, C., Mattera, V. S., Buzzi, L. I., Parodi, A. J., and D'Alessio, C. (2016) THE CONUNDRUM OF UDP-Glc ENTRANCE INTO THE YEAST ER LUMEN, *Glycobiology*, cww092.