Structural and Biochemical Characterisation of the Endoplasmic Reticulum α-Glucosidases

Alessandro T. Caputo

St Cross College

University of Oxford

Submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2015

To my parents

Abstract

ER α-glucosidases I and II are glycosyl hydrolases that play a key role in eukaryotic glycoprotein folding quality control. Removal of two glucose residues from the *N*-glycan attached to a protein allows association with calnexin or calreticulin which enable folding through associated accessory proteins. Many enveloped viruses utilise the calnexin cycle for the correct folding of their surface glycoproteins. Inhibition of the ER α-glucosidases causes misfolding of these viral glycoproteins and reduction of virion secretion and/or infectivity. Thus, inhibition of these glucosidases is a potential broad-spectrum antiviral strategy with clinical relevance. Inhibition by certain iminosugars, a class of glycomimetics, has given rise to a clinical candidate against dengue virus.

To date there are no high-resolution structures available for either of the mammalian ER α-glucosidases. Presented in this thesis is the work toward the structural and biochemical characterisation of both α-glucosidases. Large-scale production of both murine glucosidases is described along with biochemical characterisation against a number of substrates.

Glucosidase I is a challenging target for biochemical and structural studies. Described is the purification and confirmation of enzymatic activity. Biophysical techniques show that it is well folded and able to bind to iminosugars. Preliminary attempts at crystallisation have resulted in poorly diffracting crystals to 5 Å. This work enables future optimisation toward an atomic resolution structure of glucosidase I.

Glucosidase II plays a complex role in the kinetics of glycoprotein processing in the calnexin cycle. Production of the heterodimeric enzyme has enabled a detailed biochemical characterisation with a range of substrates and inhibitors. Structural characterisation of a large fragment of glucosidase II has enabled new insights into the details of heterodimerisation, substrate specificity, and the basis of iminosugar inhibition. This information can guide future development of more selective inhibitors.

Acknowledgements

There are many people that I must thank for allowing me to get to this point of scientific career. Of course this thesis would not be possible without the support of my supervisor Prof. Nicole Zitzmann. I thank you in so many ways from accepting me to come and work in this lab, to providing me with a truly unique workplace that is likely to be the most harmonious work environment I'll ever be lucky enough to be a part of. But I think the thing I appreciate most from your supervision is the level of trust and freedom you have given me to make mistakes but also learn how to become an independent scientist.

And not far behind Nicole I must thank Dr Pietro Roversi. Without your insight, experience, and crystallographically encyclopaedic mind, I'm not sure this project would have gotten as far as it did. However, let it be known in print that you still owe me a pizza but consider this debt paid in full through the knowledge of macromolecular crystallography that you have so skilfully imparted to me. I can only hope that some day I will approach the level of mastery over something as mind boggling as crystallography that you have.

There are also a number of people that contributed their time and effort to making this work more complete. Firstly, the very helpful Dr Louise Bird and Ms Heather Rada in Prof. Ray Owen's team at the Oxford Protein Production Facility who helped with the cloning and expression trials at the beginning of this project. Dr Dom Alonzi, your scientific and technical knowledge of glycobiology as well as your assistance in measuring the activity of the glucosidases against natural substrates has been invaluable. Dr. J. L. Kiappes, your chemistry knowledge has been essential for many aspects of my project (and highlights my shortcomings as a chemist) as is your friendship and support on a daily basis through the ups and downs of my time in the lab. Ms Snezana Vasiljevic, you may have been sceptical of my choice of this project based on your experience with these beasts, but your insight and support both for my morale, but also in the cell culture lab has been essential. Ms Alice Cross, was an essential pair of hands and an extremely sharp mind that was able to produce and characterise the GluII point mutants. I would like to think of you as my protégé, but in reality you're likely to surpass me in the next few years of your own PhD journey. Thanks to Dr Abhinav Kumar for your efficient help with the mass spectrometry of GluII. And of course I can't forget the many people current and past in Nicole's group who have helped immensely

with particular mentions to Emma Dixon, Steve Woodhouse, and Ben Oestringer. And I can't forget the original glycobiologist, Prof. Raymond Dwek, for all of you support, insight, and interesting conversations.

And of course there are many people in my life that support me and helped me get through the tougher times of my DPhil. Whether they have been the multitude of wonderful people I have gotten to know in Oxford, or the many people I've left on the other side of the world to undertake this opportunity, my heartfelt thanks goes out to them for all of the little things they've done to make this worth it. I must name a few in print: Laura (and the rest of my loyal friends in Melbourne), Emma, Julia, Jono, Kathryn, Erin, and Bill. Special thanks to these last four, my past and present flatmates who have had to put up with my occasional rants and for making sure I've been well fed.

Lastly, a special thanks to my family who have supported and encouraged me throughout my life. My parents, who arrived in Australia with nothing, value education above all else and have passed this onto me. My parents, who were the first generation in their families to have the chance to attend university, have supported and nurtured me to enable me to end up studying at one of the most prestigious universities in the world. My dear parents, who without them, I would not be the person or the scientist that I am now.

Contents

List of Figures

List of Tables

List of Abbreviations, Acronyms and Symbols

*Sc*GluI *S. cerevisiae* glucosidase I

4MU 4-Methylumbelliferone

4MUG 4-Methylumbelliferyl α-D-glucopyranoside

ALG Asparagine-linked glycosyltransferases

asu Asymmetric unit

BiP Binding immunoglobulin protein

bp Base pairs

BVDV Bovine viral diarrhoea virus

CBVS Calcium bond valence sum

CNX Calnexin

CRT Calreticulin

cV Column volumes

DENV Dengue virus

DMEM Dulbecco's modified eagle medium

DMSO Dimethyl sulphoxide

DNJ 1-Deoxynojirimycin

DSF Differential scanning fluorimetry

DTT Dithiothreitol

EC Enzyme commission

EDEM ER degradation-enhancing α-mannosidase-like protein

EDTA Ethylenediaminetetraacetic acid

ERAD ER-associated degradation

ERGIC ER-Golgi intermediate compartment

ESI-MS Electrospray ionisation mass spectrometry

FCS Foetal calf serum

fOS Free oligosaccharides

GANAB Glucosidase alpha neutral AB

GF Gel filtration

GH31 Glycosyl hydrolase 31 family

GH63 Glycosyl hydrolase 63 family

GH Glycosyl hydrolase

Glc D-Glucose

GlcNAc *N*-Acetyl-D-glucosamine

HCV Hepatitis C virus

HEPES 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid

Hsp70 Heat shock protein 70

IMAC Immobilised metal affinity chromatography

IPTG Isopropyl β-D-1-thiogalactopyranoside

LB Lysogeny broth

LDLRa Low-density lipoprotein receptor type-A

MALS Multi-angle laser light scattering

ManI ER α-mannosidase I

Man D-Mannose

MCE Mixed cellulose ester

MCS Multiple cloning site

MOGS Mannosyl-oligosaccharide glucosidase

MRH Mannose 6-phosphate receptor homology

MWCO Molecular weight cut-off

NCS Non-crystallographic symmetry

NP-HPLC Normal phase high performance liquid chromatography

OST Oligosaccharyl transferase

PCR Polymerase chain reaction

PDI Protein disulphide isomerases

PMSF Phenylmethylsulfonyl fluoride

PPI Peptidylprolyl cistrans isomerases

PRKCSH Protein kinase C substrate 80K-H

rpm Revolutions per minute

SAXS Small angle X-ray scattering

SDM Site-directed mutagenesis

SDS-PAGE Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SEC-MALS Size exclusion chromatography multi-angle laser light scattering

SOC Super optimal broth with catabolite repression

SS Signal sequence

TCEP Tris(2-carboxyethyl)phosphine

TLS Translation/libration/screw

UGGT UDP-glucose glycoprotein glucosyltransferase

UPR Unfolded protein response

XGAL 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Notes on nomenclature

The representations of glycan structures are using the Oxford system [1]. The presented colour scheme is adapted from the Consortium for Functional Glycomics (Figure 1). This figure is courtesy of Dr J. L. Kiappes in the Zitzmann laboratory and has been given for this publication with his permission.

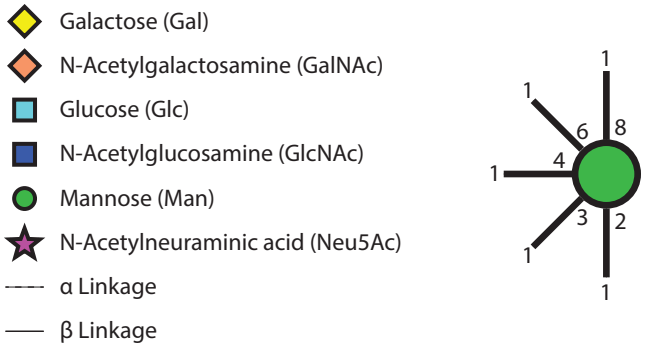


Figure 1: Key to glycan nomenclature.

1

Introduction

Trying to understand the multitude of mechanisms underlying the processing of carbohydrates in living organisms is not a trivial undertaking. Due to the presence and utilisation of carbohydrates in most facets of life as a source of energy, structural components, and signalling molecules, there are many specialised carbohydrates. (Deoxy)ribose in the very backbone of nucleic acids in all organisms is a testament to the ubiquity of carbohydrates in nature. Glycobiology as the study of the various biological processes that have carbohydrates at their core has sought to address the many questions that remain.

Of particular interest, and focussed on in this thesis, are the initial steps of the processing of *N*-linked glycans after they have been attached to proteins in the endoplamsic reticulum (ER) in eukaryotes. The trimming of the glucose residues of the *N*-glycan occurs through the action of two separate glucosidases and controls the immediate fate of these glycoproteins. Structural, biophysical, and biochemical techniques have been employed to probe the molecular mechanisms of these enzymes.

1.1 Glycosylation of proteins in eukaryotes

The use of complex carbohydrate structures as signalling molecules is an area of biology that has yet to be completely understood. The conjugation of various sugars to both lipids and proteins to a level of complexity that sometimes seems almost random but follows strict underlying molecular mechanisms are only now starting to be appreciated. Post-translational modification of proteins to form glycoproteins commonly occurs through asparagine residues (*N*-linked glycosylation) or through serine and threonine residues (*O*-glycosylation). Less common co- or post-translational modifications of proteins involving carbohydrates also include linkage through tryptophan, phosphoserine residues or addition to the C-terminus of proteins to glycolipids such as glycosylphophatidylinositol. The general function of these glycoconjugates tends to be as a signal for the state, localisation, or destination of the glycoprotein. The elaboration and final array of glycan species, known as glycoforms, can be diverse and tissue- or cell-specific, as it is not a template-based process like DNA replication, for example.

1.1.1 *N*-linked glycosylation

N-linked glycosylation is a co-translational modification that happens to most secreted or organelle-targeted proteins in eukaryotes. The purposes of the glycosylation are numerous. It is thought that *N*-glycosylation restricts the number of conformations that the glycopeptide can adopt thereby assisting the folding process by minimising the stability of the unfolded state [2, 3]. In some cases, abolition of the sites of *N*-glycosylation does not impact folding or function, but there are a number of enzymes and receptors that are severely disrupted if the glycosylation is removed (reviewed in [4]). One possibility explaining the lack of function of proteins that have an *N*-glycan attachment site removed is that they then fall out of the scope of the folding quality control mechanisms present in the ER lumen [5]. Initially, theories of the biosynthesis of *N*-linked glycoproteins proposed that assembly of the glycan occurred on the protein. However, it is now known that the precursor *N*-glycan is transferred from an already assembled lipid-linked conjugate [6]. The preassembled precursor contains exactly 3 glucoses (Glc), 9 mannoses (Man), and 2 *N*-acetyl-glucosamines (GlcNAc) (Figure 1.1) [7]. Although classically thought to be a modification exclusive to eukaryotic organisms, *N*-linked glycosylation has also been observed in bacteria [8, 9].

1.1.1.1 Biosynthesis and transfer of *N*-linked precursor glycans to nascent proteins

Biosynthesis of the lipid-linked oligosaccharide precursor occurs on both leaflets of the ER membrane. The assembly of the glycan occurs on dolichol, an unsaturated single-chain lipid. After phosphorylation to a terminal pyrophosphate moiety, a series of glycosyltransferases add the 2 GlcNAc and 5 Man residues from their UDP (GlcNAc) or GDP (Man) conjugated donors (Figure 1.2 left-hand side). At this point,

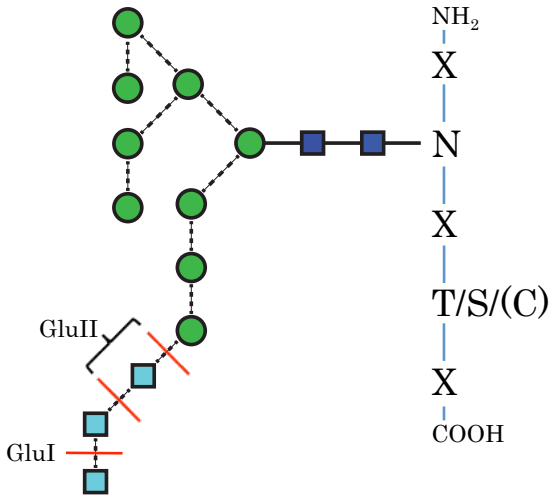


Figure 1.1: The *N*-glycan structure attached to an asparagine residue with the glycosidic bonds that are cleaved by the ER α-glucosidases indicated by the red lines.

the glycan is thought to be flipped across the membrane by a flippase, Rft1 [10]. However, there has been debate as to whether this protein carries out this function [11, 12]. On the lumenal side of the ER, the remaining steps of glycan assembly are carried out by another set of glycosyltransferases; in all of these reactions, the donor is a monosaccharide-conjugated dolichol-phosphate donor (Figure 1.2 right-hand side). The sugar-dolichol donors are synthesised on the cytosolic side (Figure 1.2 top left-hand side) of the membrane and flipped across by unknown components of the system. The details of the asparagine-linked glycosyltransferases (ALG) 1-14, have been reviewed by Breitling and Aebi [13]. Mutations in these genes lead to a range of diseases [14]. A notable exception to using the canonical triglucosylated glycan as an glycan precursor are trypanosomes which transfers a $Man_9GlcNAc_2$ glycan to the nascent protein [15].

Transfer of the lipid-linked oligosaccharide to the acceptor protein is catalysed by the oligosaccharyl transferase (OST). The *en bloc* transfer onto the asparagine residue occurs at the N-X-T/S consensus sequence where X is any amino acid with the exception of proline [16]. Other sequons can be recognised including N-X-C, N-X-V

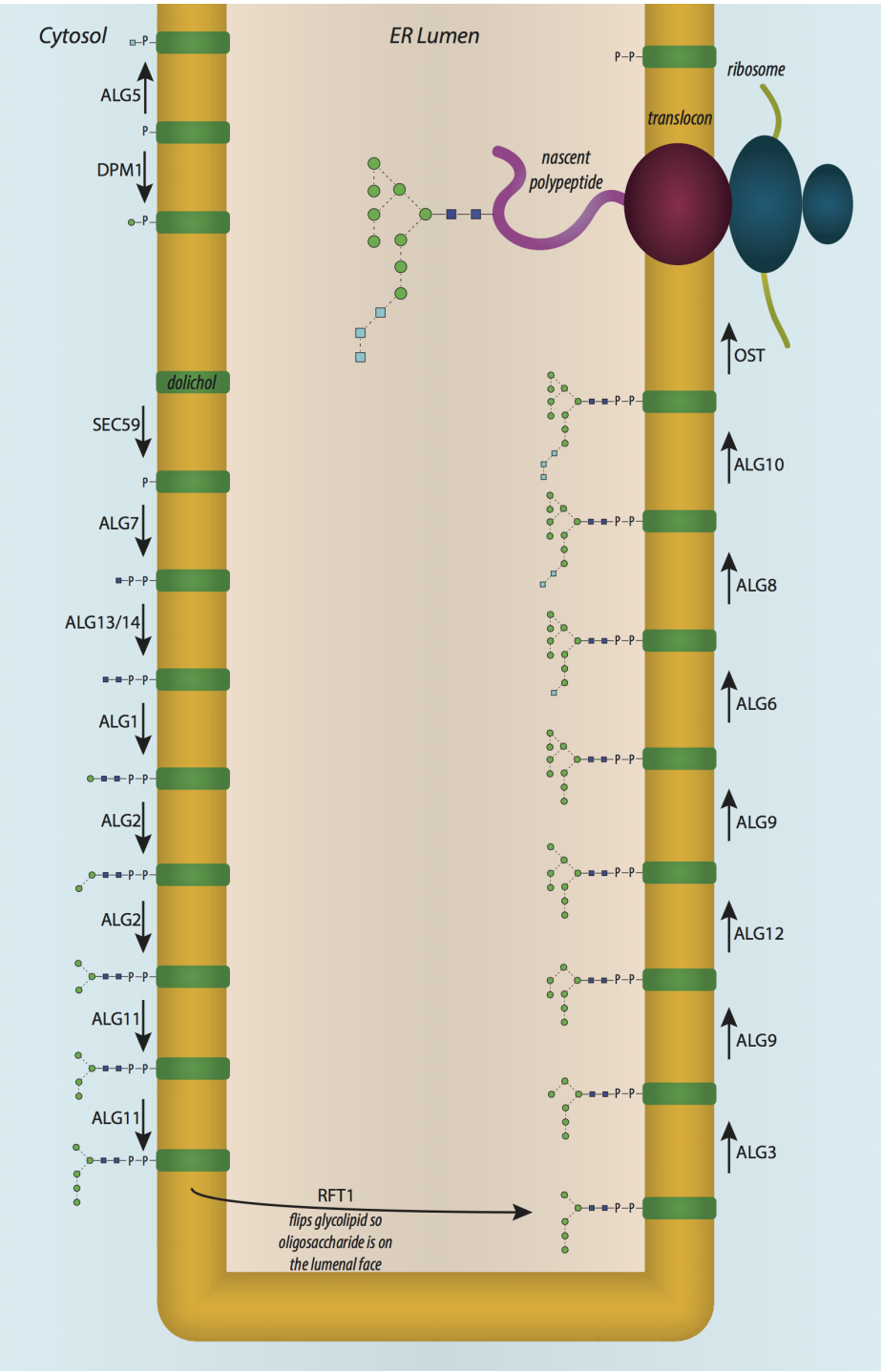


Figure 1.2: Biosynthesis of the lipid-linked oligosaccharide on the ER membrane. Figure courtesy of Dr J. L. Kiappes.

or N-G-X but these only account for a small proportion of glycosylation events [17–19]. The crystal structure of an archaeal OST has been valuable in elucidating the molecular mechanisms of substrate binding and catalysis [20]. Electron microscopy of a mammalian OST shows an intimate complex of the OST with the translocon and associated ribosome [21].

1.2 Endoplasmic reticulum quality control – the calnexin cycle

First described in 1994, the so-called calnexin cycle remains at the centre of protein folding quality control for *N*-linked glycoproteins [5]. Upon further investigation, it became clearer that the order of the molecular events coupled with the identification of calreticulin (CRT), a soluble homologue of calnexin (CNX), have established the canonical pathway [22, 23]. In general, the cycle can be described schematically as in Figure 1.3. Newly translocated and *N*-linked glycoproteins can enter the cycle upon consecutive cleavages of the two outer glucose residues by the ER α-glucosidases I and II (GluI and GluII). The resulting monoglucosylated species can associate with CNX or CRT, which in turn assist in folding though a number of associated accessory proteins (the current list of known proteins is reviewed in [24]). GluII can remove the inner glucose, at which point UDP-glucose glycoprotein glucosyltransferase (UGGT) can carry out its checkpoint role to sense protein misfold and reglucosylate the glycoprotein enabling re-association with CNX or CRT [5]. When properly folded, the glycoprotein can continue its trafficking with its glycan(s) further processed in the Golgi apparatus. Terminally misfolded glycoproteins on the other hand have a mannose removed in the ER by the slow-acting ER α-mannosidase I (ManI) which allows recognition by ER degradation-enhancing α-mannosidase-like

proteins (EDEM) [25, 26]. This commits the misfolded glycoprotein to ER-associated degradation (ERAD) by retro-translocation into the cytosol for ubiquitin-dependent protein degradation [27].

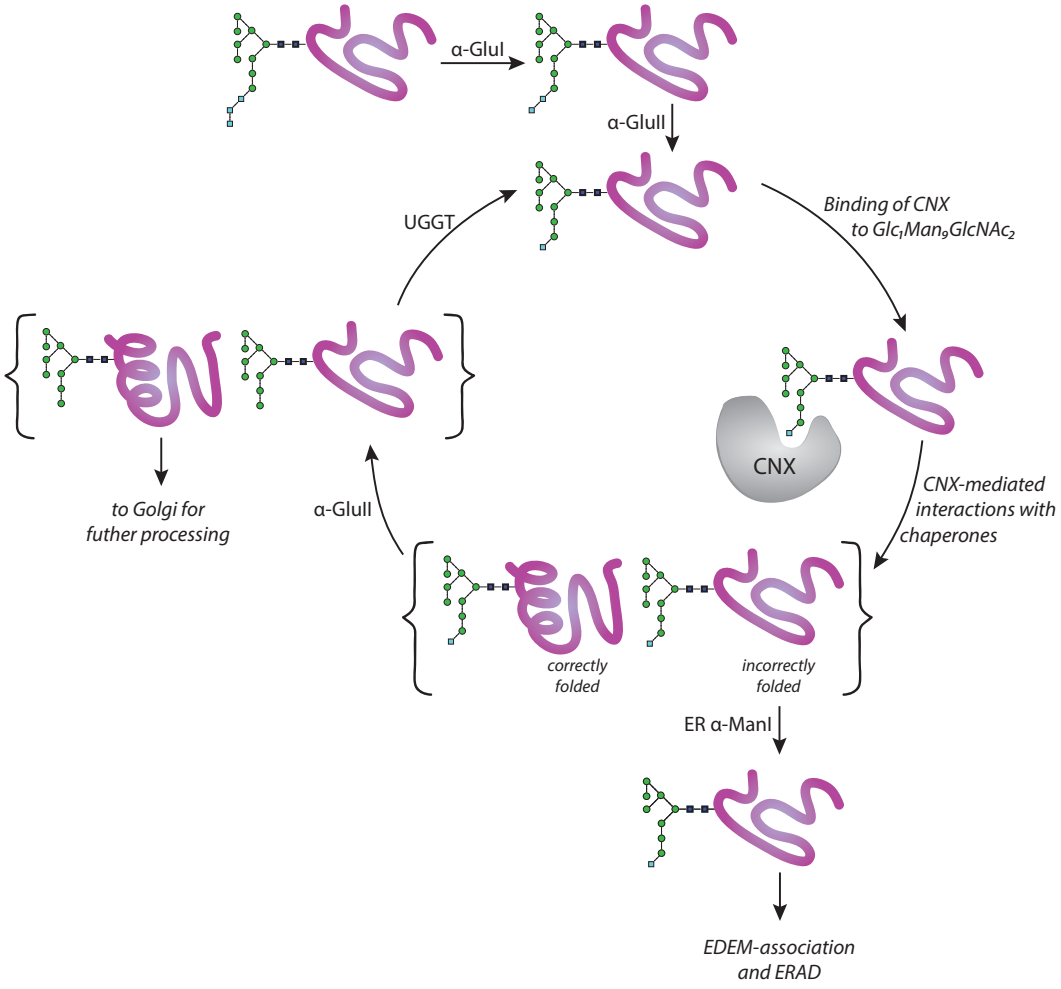


Figure 1.3: Schematic of the calnexin cycle. Figure courtesy of Dr J. L. Kiappes.

For the glycoproteins that achieve their native state, the next step is usually trafficking toward the Golgi apparatus unless an ER-retrieval sequence is present. This can occur through the ER-Golgi intermediate compartment (ERGIC) by binding to the lectin ERGIC-53 [24]. In the Golgi, the glycan is subject to further processing upon the action of Golgi-resident mannosidases and GlcNAc-transferases that convert the glycans to either hybrid or complex forms if accessible [28].

1.2.1 The components of the calnexin cycle

CNX and CRT are lectins that serve as chaperones by binding to the substrate glycoprotein through recognition of the $Glc_1Man_9GlcNAc_2$ glycan. It is thought that they can chaperone in a more classical sense by preventing aggregation by binding to regions of the substrate glycoprotein [29]. CNX/CRT interact through their P-domains with a number of protein disulphide isomerases (PDI) like ERp57, or peptidylprolyl cistrans isomerases (PPI) like cyclophilin B, to actively assist in protein folding [30, 31]. The structural basis of binding CNX/CRT to client monoglucosylated glycoproteins has been elucidated through a crystal structure of CRT bound to a Glc_1Man_3 tetrasaccharide [32]. Upon removal of the final glucose, proteins that have achieved their native state can progress to their final destination. Otherwise the glycoproteins are recognised by UGGT, a central player in the calnexin cycle. It was first discovered that a reglucosylation of the high mannose glycan was occurring in the ER of trypanosomes and eventually this function was narrowed down to UGGT [33, 34]. It appears that recognition of misfolding may involve a molecular determinant at the attachment point of the glycan, which in a folded glycoprotein is not readily accessible [35]. There is likely to be a mechanism for UGGT to sense misold in regions distal to the site of glycosylation, too. Two forms of the protein exist in higher organisms, one of which was until recently thought not to have enzymatic activity [36].

1.2.2 Other components in the ER

Malectin is a newly identified ER lectin that may have a significant role in ER protein folding. First discovered in *Xenopus*, malectin was identified based on sequence homology to carbohydrate binding domains; it binds to Glc_2-containing high-mannose glycans [37]. The levels of malectin increase upon induction of ER stress and it interacts with a component of the OST, ribophorin I, which suggests it may have an

orthogonal role to CNX/CRT or perhaps it sequesters misfolded proteins to prevent secretion [38, 39].

In addition to the calnexin cycle, there are lectin-independent mechanisms in the ER lumen for protein folding. Binding immunoglobulin protein (BiP, also called GRP78) is part of the heat shock protein 70 (Hsp70) family of chaperones. It takes its name from the discovery that immunoglobulins readily bind BiP soon after translocation, thereby avoiding the calnexin cycle [40]. Similarly to CNX, BiP interacts with a number of accessory proteins that assist in correct folding like PDIs and PPIs [41]. BiP is also a key signalling molecule for the unfolded protein response (UPR), the cell's response to stress in the ER [42]. It initiates up-regulation of relevant genes for alleviating the problems associated with the build up of misfolded proteins in the ER.

1.3 Endoplasmic reticulum α-glucosidases

The two exo-glucosidases that enable entry of the glycoprotein into the calnexin cycle share similar names but differ both in structure and mechanism. A classification system for glycosyl hydrolases (GH) has been established that groups the different hydrolases, now containing 134 different families, based on sequence and structure [43–47]. Members in the same family generally have the same catalytic mechanism though not always the same substrate specificities. Further classification based on substrate specificity is contained in the enzyme commission (EC) numbering system by the International Union of Biochemistry and Molecular Biology, where all glycosyl hydrolases fall within the EC 3.2.1.X category.

1.3.1 Glucosidase I

GluI is a ∼90 kDa protein encoded by the mannosyl-oligosaccharide glucosidase (MOGS) gene and is a member of the glycosyl hydrolase 63 family (GH63). GluI is the first enzyme to act on the glycan almost immediately after it is transferred onto the substrate protein by OST [48]. It is specific for α(1,2) glycosidic bonds and hydrolyses them using an inverting mechanism (Figure 1.4) [49]. The hydrolysis occurs through direct displacement where water acts as a nucleophile with an aspartate acting as a general base [50, 51]. GluI is a type II membrane protein with a small cytoplasmic segment after the signal peptide, a transmembrane helix, and two lumenal domains (Figure 1.5) [52–54]. Mechanistic inhibitor probes have been utilised to assess accessible residues in the active site but have not been able to definitively identify residues involved in catalysis over those involved in ligand binding [55–57]. Mutations in the MOGS gene tend to be developmentally fatal with very few cases reported beyond birth indicating its crucial role in early embryogenesis [58, 59].

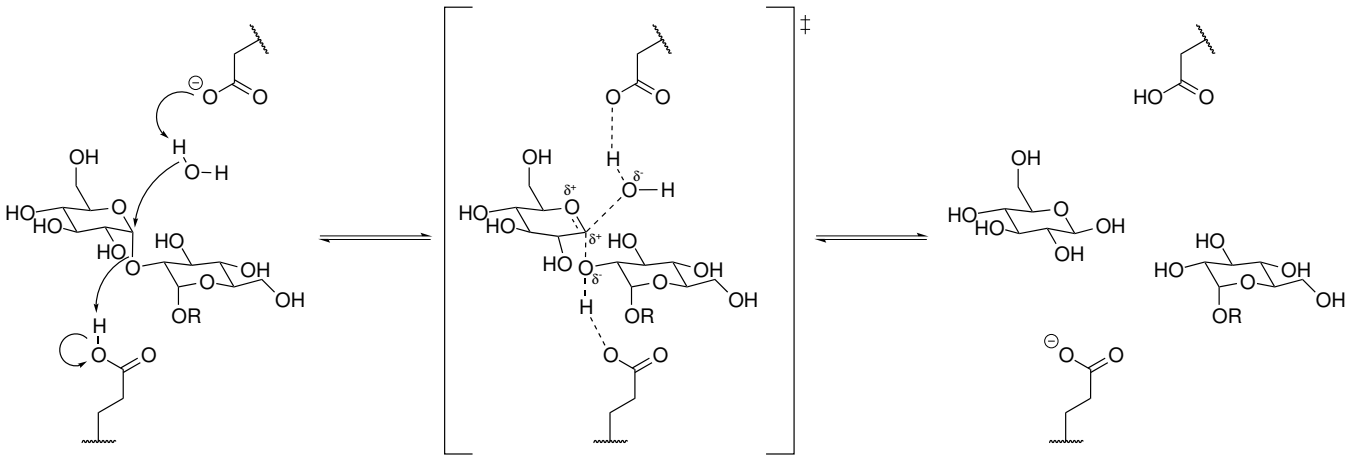


Figure 1.4: Inverting mechanism of hydrolysis of the $Glc_3Man_9GlcNAc_2$-protein substrate. R = $Glc_1Man_9GlcNAc_2$-protein.

Figure 1.5: Schematic of the GluI domain architecture. The protein is organised into three domains with a signal sequence (SS) that targets translocation into the ER lumen.

1.3.1.1 Structure

The first reported crystal structure of a GH63 family enzyme is that of YgjK from *E. coli* that folds as $(\alpha/\alpha)_6$ barrel. However, this enzyme has $\alpha(1,3)$ regiospecificity [60]. More recently the first crystal structure of a GluI has been solved, comprising of the lumenal portion of the *S. cerevisiae* (*Sc*GluI) enzyme [61]. It contains the same fold as YgjK in the catalytic domain with an additional 13-strand β-sandwich domain at the N-terminus. By comparing with the *E. coli* homologue, the catalytic residues were identified and placed in the context of previously inconclusive mutagenesis studies (Figure 1.6A) [57]. In this crystal form, access of ligands to the active site is blocked by the hexahistidine affinity tag from an adjacent molecule in the crystal (Figure 1.6B). As a result, the authors carried out *in silico* docking of different molecules in the active site including Glc_3, the terminal trisaccharide from the enzyme's natural substrate. The favoured docking model accommodates the acute angle formed by the glycosidic bond linkages of the trisaccharide and is comparable to the NMR structure of the glycan [62]. The role of the N-terminal domain is yet to be discovered, but the authors suggest it may be involved in protein-protein interactions such as with OST or GluII. Overall, the crystal structure of *Sc*GluI is currently the best surrogate for understanding the mammalian GluI. However, due to its unfortunate crystal packing, another enzyme or crystal form is needed for understanding its ligand binding and aiding inhibitor design.

1.3.2 Glucosidase II

GluII catalyses the sequential cleavage of the two inner glucose residues of the *N*-linked glycan, $Glc_2Man_9GlcNAc_2$-protein and $Glc_1Man_9GlcNAc_2$-protein being its two separate substrates. The second reaction does not happen immediately after the first and is punctuated by the substrate glycoprotein's association with CNX or CRT

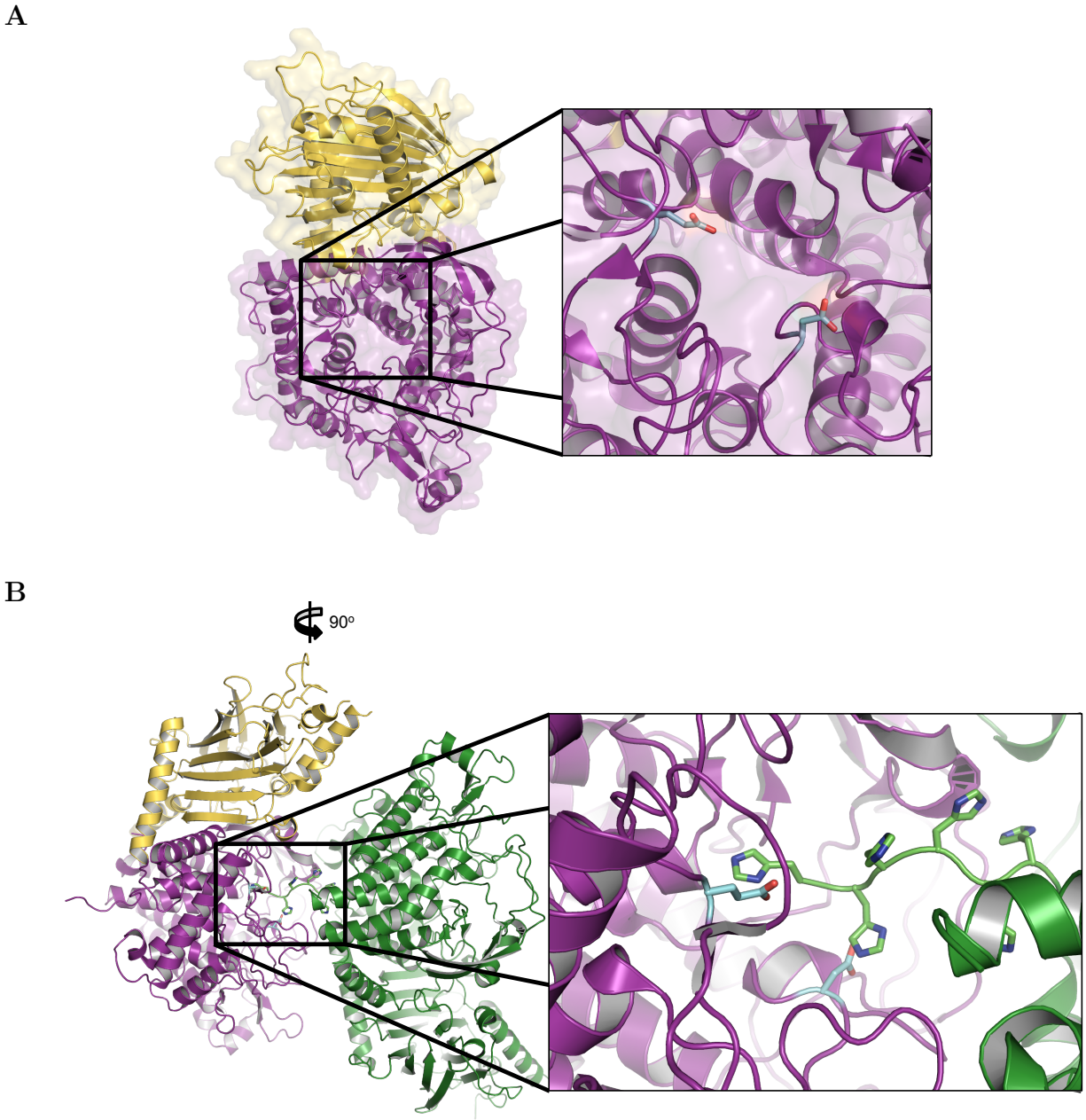


Figure 1.6: Structure of *Sc*GluI. (A) Cartoon representation of the structure. The N-terminal domain is coloured in yellow and the GH63 domain in purple. A close up of the active site highlights the two catalytic aspartate residues. (B) The structure in (A) is rotated 90° and an adjacent crystallographically related copy is shown in green. The hexahistidine tag from the adjacent molecule docks into the active site. Coordinates from PDB ID 4J5T.

initiating the range of molecular events described above. Thus, even though both bonds specifically cleaved by GluII are α(1,3) glycosidic bonds [63], there must be a molecular basis for the discrimination between these two substrates that results in the measurable differences in kinetics between the two cleavages [64–66]. The enzyme consists of two subunits, the α-subunit which carries the catalytic activity and the β-subunit which plays a regulatory role for subcellular localisation of the complex and assists in substrate recognition [67].

The α-subunit is a $\sim$110 kDa glycoprotein encoded by the glucosidase alpha neutral AB (GANAB) gene and is a member of the glycosyl hydrolase 31 family (GH31) [68]. Sequence annotation of the α-subunit segments it into three domains with a signal peptide that is removed in the mature protein (Figure 1.7A). However, with the availability of a number of structurally characterised GH31 members, a more updated domain architecture consistent with the nomenclature proposed in the first GH31 structure is suggested (Figure 1.7B) [69]. GH31 enzymes carry out hydrolysis through an archetypal Koshland double displacement retaining mechanism (Figure 1.8) [50, 70]. In these enzymes, the glycosidic bond is attacked by a nucleophile, usually an aspartate residue, which proceeds to a covalent adduct via an oxo-carbenium-like transition state. A general acid/base, also usually an aspartate, assists in providing the proton in the first step and then acts as a base upon hydrolysis of the adduct by water. This retains the glucose in the α-anomeric conformation in the released monosaccharide. The nucleophile of GH31 enzymes has been localised to the aspartate in the WXDMNE consensus sequence through trapping of a covalent adduct with a 5-fluoro sugar [71]. Due to the conservation observed in this sequence, but without the benefit of any structural information, it was proposed that the glutamate in that sequence may be the general acid/base [72]. However, subsequent GH31 family crystal structures revealed that proton donation/abstraction is effected by the aspartate on the WXGD consensus sequence [69, 73]. Using the same mechanism,

transglycosylation (the formation of a glycosidic bond) can also be carried out and has been observed in some GH31 family members [74].

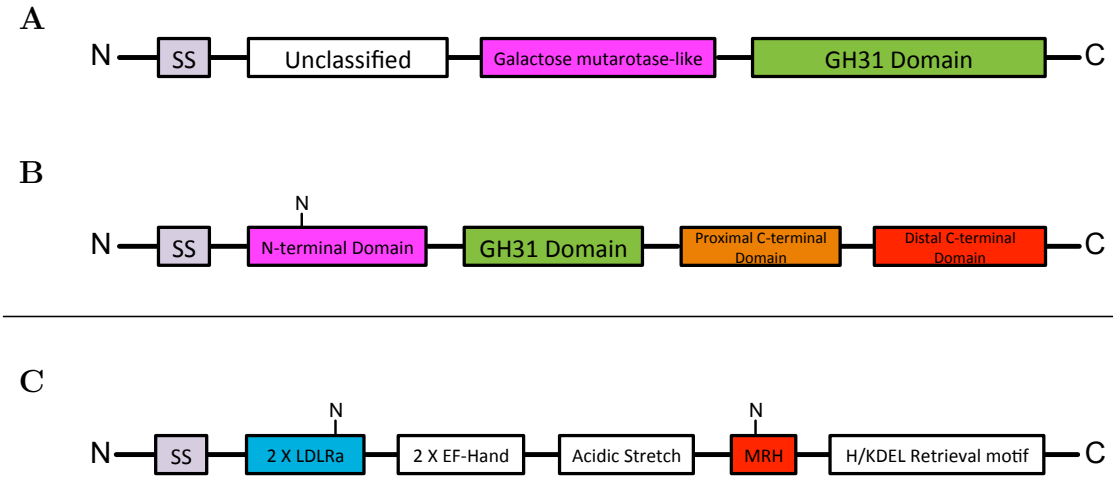


Figure 1.7: Schematic of the GluII domain architecture. (A) Sequence annotation by Pfam of the α-subunit into three domains. (B) Proposed domain architecture using structurally characterised GH31 members. (C) Domain architecture of the β-subunit into five domains with a C-terminal ER retrieval sequence for subcellular targeting. Conserved N-linked glycosylation sites are indicated in the domain in which they occur.

The β-subunit is encoded by the protein kinase C substrate 80K-H (PRKCSH) gene. The name of this gene stems from the initial identification of this protein from human squamous carcinoma cells [75]. Mutations in the PRKCSH gene are responsible for autosomal dominant polycystic liver disease possibly through affecting the folding of a transient receptor potential channel in diseased hepatocytes [76–80]. The glycoprotein is a ∼60 kDa protein with extensive disulphide bonds, and when reduced, it migrates differently (at ∼80 kDa) on SDS-PAGE gels likely due to extensive negative patches of residues [67]. The protein comprises a number of consecutive domains including two tandem low density lipid receptor type-A (LDLRa) folds, two sequential EF-hand motifs, an extended stretch of acidic residues, a mannose 6-phosphate receptor homology (MRH) domain, and a C-terminal ER retrieval motif (H/KDEL) (Figure 1.7C) [81].

The heterodimeric nature of GluII is essential for the proper function for a number of reasons. The HDEL ER retrieval motif is essential for localising the

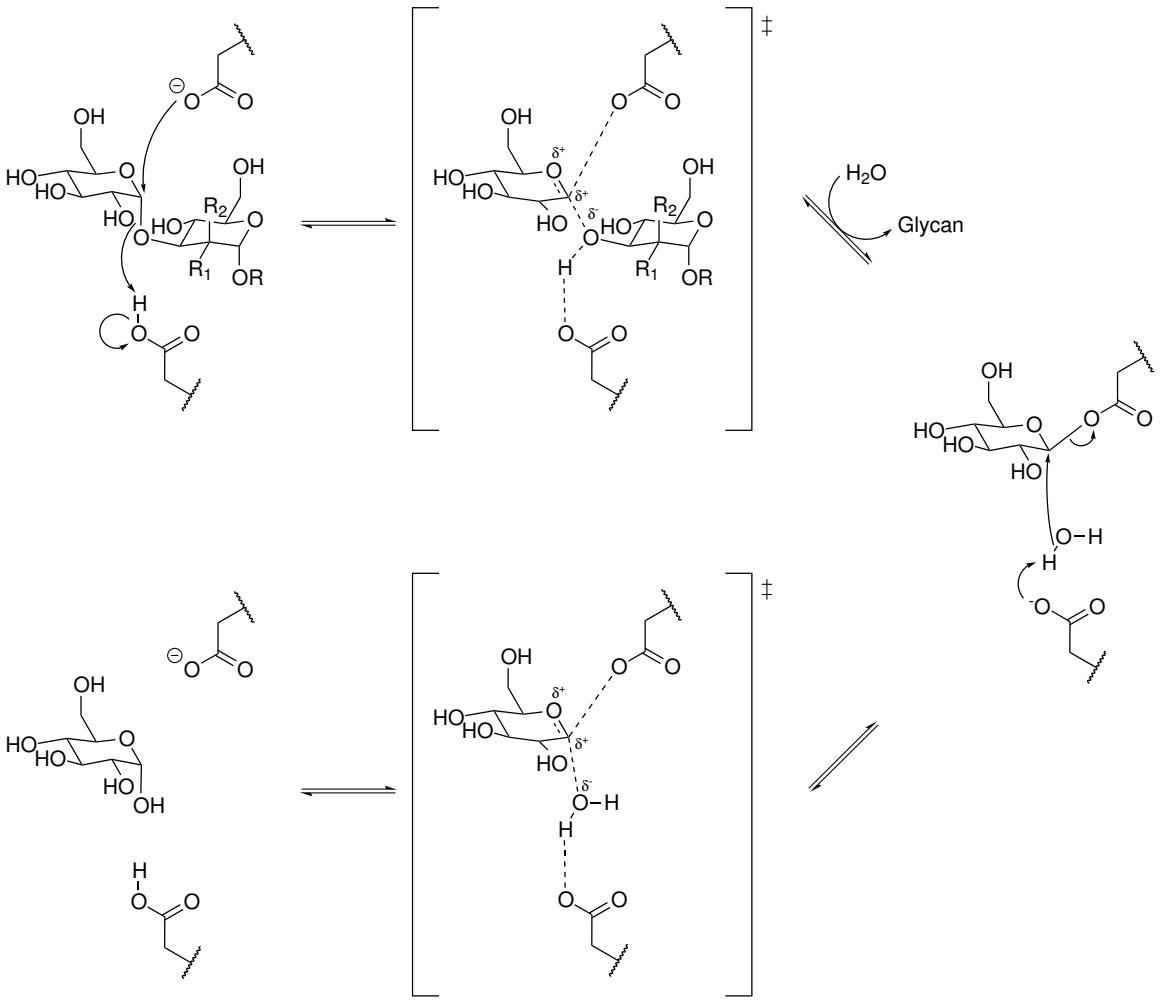


Figure 1.8: Retaining mechanism of hydrolysis of the $Glc_{2/1}Man_9GlcNAc_2$-protein substrate. R = $Man_9GlcNAc_2$-protein, R_1 = OH and R_2 = H when the substrate is $Glc_2Man_9GlcNAc_2$-protein. R = $Man_8GlcNAc_2$-protein, R_1 = H and R_2 = OH when the substrate is $Glc_1Man_9GlcNAc_2$-protein.

soluble GluII complex in the ER lumen, though a noteworthy exception to this is the absence of this motif in the β-subunit of *S. cerevisiae* [67, 82]. The ability of GluII to cleave its substrates is essentially abolished *in vivo* when the β-subunit is removed [82, 83]. The MRH domain on the β-subunit is important for recognising the glycan substrate, most likely through interactions with the mannose arms [84, 85]. Removal of the β-subunit does not render the α-subunit catalytically inactive as it can still cleave small aryl glucosides. There are two specific sites toward each terminus of the β-subunit that bind to the α-subunit [86]. Analysis of the heterodimer in solution

indicates that it forms an extended structure [87]. Furthermore, an extra level of complexity has been suggested whereby GluII is somehow trans-activated and that the substrate glycoprotein must contain at least two glycans to be hydrolysed [88].

1.3.2.1 Structure

Currently the only structural information relating to the catalytic elements of GluII comes from homology modelling based on crystal structures of members of the GH31 family. These include glycosidases from archaeal, bacterial and eukaryotic organisms. The prokaryotic enzymes display an array of differing substrate specificities and include α-xylosidases and transglycosidases. The eukaryotic enzymes are less diverse, with all predominantly exhibiting α(1,4) specificity. An exception in the family is the glucan lyase from a red algae whose catalytic mechanism is different due to a change in the consensus WXDMNE to WXDMNV, which enables the catalytic nucleophile to acquire a dual role by becoming a base as well [89]. The other variant is the isomaltase domain from the intestinal sucrase-isomaltase enzyme that has relaxed substrate specificity including α(1,4) and α(1,6) glycosidic bonds [90]. The structures of intestinal or sugar beet glucosidase in complex with the diabetic therapeutic acarbose, a non-cleavable tetrasachharide, give insight into regiospecificity [91, 92]. The fold of all of these eukaryotic GH31 structures is very similar and at the active site there is also a high degree of similarity. Thus, a structure of GluII is necessary to determine the basis of its specificity.

The only high-resolution structural characterisation available of a portion of GluII is of the MRH domain from *S. pombe*, by both NMR spetroscopy and X-ray crystallography [93, 94]. This domain adopts a flattened β-barrel (Figure 1.9). These two studies show that the explicit binding site to a terminal mannose residue is through H-bonding to a pocket lined with the conserved residues glutamine, arginine,

glutamate, and tyrosine. Concomitant to mannose binding is the movement of a loop containing a conserved tryptophan, which initially was thought to be important to interact with the other mannose arm. The authors amended this proposed binding model based on the crystallographic packing in the mannose-bound structure. They postulate that this loop interacts with another part of GluII (of either subunit) upon binding mannose that may have an allosteric role in the catalytic function of GluII [94].

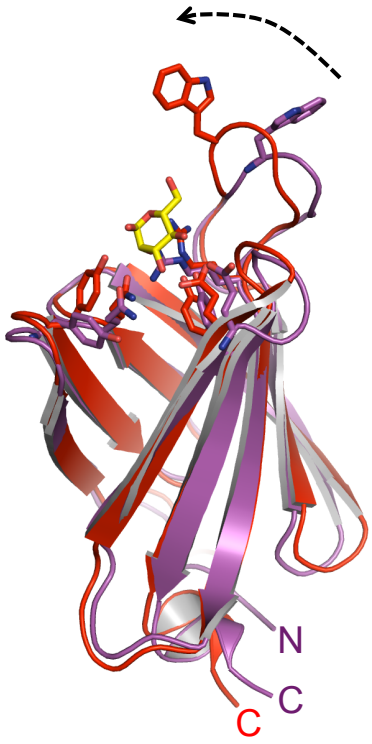


Figure 1.9: Structure of the *S. pombe* β-subunit MRH domain with (red) and without (purple) mannose. Mannose (yellow) binding is mediated by four conserved residues. The apo form shows a shift in a tryptophan containing loop. Coordinates from PDB ID 4XQM (red) and 2LVX (purple).

1.4 Iminosugars

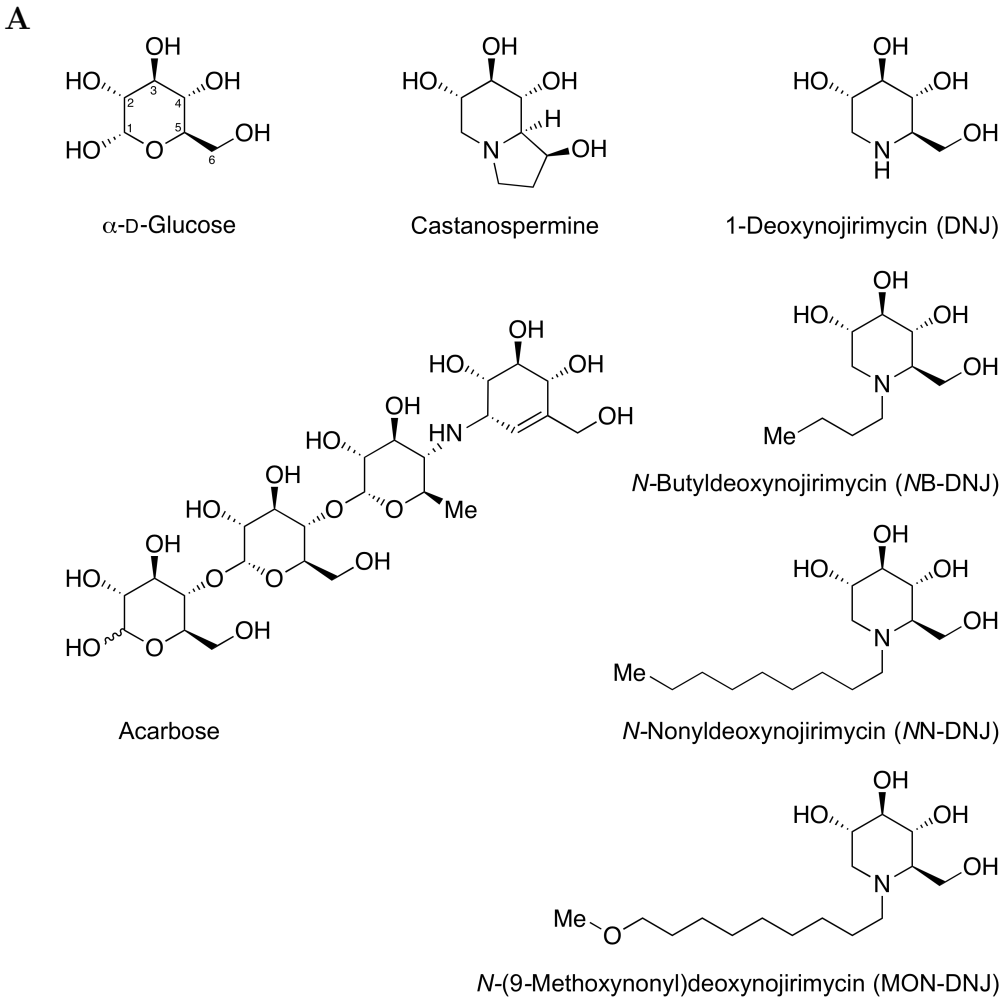
Iminosugars are a class of glycomimetic compounds that typically have been investigated as inhibitors of carbohydrate-processing enzymes. They can take a number

of heterocyclic forms with a nitrogen atom being substituted for the oxygen. The iminosugars relevant to the work described in this thesis are glucose analogues (Figure 1.10A). The prototypical glucose analogue nojirimycin was initially discovered in extracts from *Streptomyces* cultures that had antibiotic activities [95]. Soon after, a more stable derivative was found in mulberry leaves that lacks a hydroxyl at C-1 to give 1-deoxynojirimycin (DNJ) which inhibits α-glucosidases [96, 97]. Later on, another α-glucosidase inhibitor was identified from an extract of *Castanospermum australe* seeds, called castanospermine [98, 99]. *N*-alkylation of DNJ increases bioavailability with typical additions being aliphatic carbon chains [100]. Analogous to glucose mimetics, a range of inhibitors based on the mannose and galactose stereochemistry were synthesised (Figure 1.10B). The mannose-analogue based compounds, kifunensine and swainsonine, have particular uses in engineering specific glycoforms by interrupting the ER ManI and the Golgi α-mannosidase II, respectively [101, 102]. Apart from their use as glycosidase inhibitors, there are also examples of iminosugars that inhibit glycosyltransferases, nucleoside hydrolases, glycogen phosphorylases, and metalloproteases [103–106].

1.4.1 Use as therapeutics

Concomitant with the rise of iminosugars in the laboratory came their applications as potential therapeutics. Iminosugars have been used in the laboratory and in the clinic for the treatment of diabetes mellitus, infectious diseases, and lysosomal storage diseases [107]. Years of iminosugar research have resulted in two drugs that are now on the market. Miglitol (Glyset, 2-hydroxyethyl-DNJ) is licensed as an anti-diabetic drug that inhibits intestinal glucosidases. Miglustat (Zavesca, *N*-Butyl-DNJ) is used for treatment of Gaucher's disease, a lysosomal storage disease.

The use of iminosugars as a therapy for diabetes stems from the observation



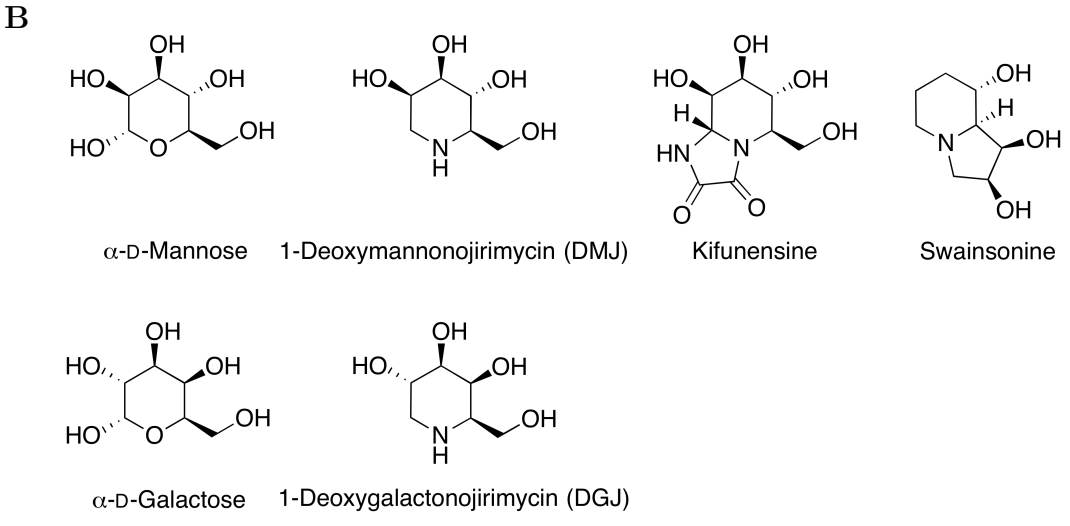


Figure 1.10: Chemical structures of a series of relevant inhibitors. (A) Inhibitors based on D-glucose. Numbering around the glucose ring indicates the referencing scheme throughout this thesis. (B) Inhibitors based on the D-glucose epimers, D-mannose and D-galactose.

that extracts from mulberry leaves can reduce postprandial hyperglycemia [108]. As the identities of the potential glycosidase inhibitory molecules in these extracts were narrowed down to iminosugars, their therapeutic use became apparent. Inhibition of the intestinal glucosidases contributes to the delay of glucose release [109]. Prevention of postprandial hyperglycemia in type 2 diabetes mellitus patients by inhibition of intestinal glucosidases decreases the release of glucose from dietary oligosaccharides like starch. These enzymes are from the GH31 and GH13 families, which have structural and mechanistic similarities. Described above (1.3.2.1) are the structural similarities of GluII with these intestinal enzymes which bind some iminosugars equally well. This homology leads to the off-target side-effects that are observed with the oral use of iminosugars when used for diabetes [110]. Gastrointestinal distress can occur due to oligosaccharide build-up caused by α-glucosidase inhibition in the small intestines that in turn causes osmotic diarrhoea. An inhibitor that is not an iminosugar but whose mode of binding is relevant to compare is acarbose (Figure 1.10A), a natural product licensed for treating type 2 diabetes based on the same mechanism as iminosugars. Acarbose is a relevant molecule for comparison to the iminosugars presented for GluI and GluII as it contains a series of glucose molecules with α(1,4) glycosidic linkages. The characterised binding sites in the intestinal glucosidases can be used with any forthcoming ER α-glucosidase structures to help dissect the molecular basis for substrate specificity.

Some lysosomal storage diseases result in abnormal processing of sphingolipids, a class of glycolipids. In the case of Gaucher's disease, inactive or misfolded glucocerebrosidase, a β-glucosidase, results in the pathological build-up of glucocerebrosides in the lysosome. The most common treatment is enzyme replacement with the recombinant enzyme, but this is a very costly approach. The other is through substrate reduction therapy whereby the partial inhibition of the substrate biosynthetic pathway reduces the pathological levels of the glucosphingolipids [111, 112].

Iminosugars have been successfully utilised for this in the clinic [113]. An alternative approach for these diseases has been proposed that uses iminosugars as scaffolds for the β-glucosidase folding-defective mutants. The principle is to use iminosugars to chaperone the enzyme throughout the post-translational processing in order to traffic it to the lysosome where it can carry out its normal function [114, 115].

1.4.2 Use as antivirals

Inhibition of GluI and GluII by the iminosugars, DNJ, castanospermine, and derivatives are well established [64, 116–118]. Inhibition is thought to occur through mimicry of the oxo-carbenium transition state facilitated by the nitrogen atom in the ring [51, 119]. As the protonated iminosugar is adjacent to the deprotonated catalytic residues, the pK_a of the iminosugar may have implications on its inhibition potency [120–122]. A number of viruses show reduced infectivity upon iminosugar treatment such as vesicular stomatosis virus, Sinbis virus, influenza virus, human cytomegalovirus, hepatitis B virus, and human immunodeficiency virus 1; all are enveloped viruses [123–128]. Inhibition of the ER α-glucosidases affects the folding of the influenza glycoprotein haemagglutinin [5]. The implication of the generated misfold of viral glycoproteins is that they either end up in degradation pathways (*i.e.* ERAD) or are functionally impaired.

The *in cellula* mechanisms of action of iminosugars against a number of members of the *Flaviviridae* family including dengue virus (DENV), bovine viral diarrhoea virus (BVDV), and hepatitis C virus (HCV), show that it can be much more complex. These three viruses are from three genera in the family and can be used to compare and contrast the identified mechanisms of action. For example in dengue virus, iminosugars reduce the viral secretion levels [129, 130]. Whereas in BVDV and HCV, iminosugars reduce both secretion and infectivity of virions [131]. HCV

and BVDV both contain an ion channel that is inhibited by long-chain *N*-alkylated iminosugars leading to decreased infectivity [132–134]. To complicate matters, iminosugars with a DGJ moiety could also be affecting glycolipid processing causing an antiviral effect for certain viruses [Zitamann, unpublished results]. Thus iminosugar treatment has complex consequences and can affect other viral or cellular pathways, eliciting the antiviral effect. Inhibition of the ER α-glucosidases has recently been given clinical context through the identification of MOGS-deficient patients who have survived to childhood and show resistance to viral infections [59]. The list of susceptible viruses *in vitro* has now lengthened to include those from the families *Herpesviridae, Coronaviridae, Paramyxoviridae, Filoviridae, Arenaviridae*, and *Bunyaviridae* (summarised in [135]).

A large number of studies have tried to assess structure-activity relationships of series of antiviral compounds [136–142]. Design and development of most compounds have been focussing on the variation of *N*-substitution but progress has been hampered by the absence of structures of the target enzymes. Furthermore, due to the complex downstream effects of the inhibition of these glucosidases *in vivo*, the finer details of the antiviral mechanisms of action still need to be understood. A DNJ derivative (*N*9'-methoxynonyl DNJ, MON-DNJ, see Figure 1.10) is now in clinical trials against dengue virus (https://clinicaltrials.gov/ct2/show/NCT02061358) reinforcing the relevance of this strategy.

1.5 Aims

The aims of the studies carried out and presented in this thesis are centred around the better understanding of the molecular structure of the two mammalian ER α-glucosidases. In particular, I aimed to address substrate recognition and specificity.

The benefit and longer term goal of these projects are more efficacious and selective antiviral compounds, to be rationally designed using the structural information gathered.

A mammalian GluI structure would aid drug design. An added benefit of the endeavour to clone, express, and purify mammalian GluI for structural studies is that the enzyme can be used for the biochemical screening of any future inhibitory molecules as well as enabling its full biochemical characterisation. Experiments that address this can be found in Chapter 3.

The biophysical characterisation of GluII is used to assess both the solution structure and fold, as less is known for this enzyme. Coupled with a more detailed enzymological dissection of GluII, the aims of Chapter 4 are to better understand the overall shape of the enzyme, its quaternary structure, catalytically important residues, and determinants of substrate specificity.

Any structure of domains of GluII would contribute to a better understanding of its function. A crystal structure of a fragment of GluII comprising the α-subunit bound to a domain of the β-subunit is presented in Chapter 5. Biophysical and biochemical data add to the crystal structure to rationalise the observations at and around the active site. The structure also reveals details of the β-subunit N-terminal interface with the α-subunit which in turn has implications on the positioning and recruitment of the glycan substrate.

2

Materials and methods

All chemicals used were from Sigma-Aldrich and all commercially available enzymes were from New England Biosciences unless stated otherwise. All solutions were made up with 18.2 MΩ cm water from a Milli-Q Plus reverse osmosis water purification system (Millipore). Unless specified otherwise, all graphs were produced in Prism (GraphPad).

2.1 Bioinformatics

Clustal Omega [143] was used to generate multiple sequence alignments and calculate percentage identities. The chosen sequences were sourced from the UniProt server (Table 2.1). Visual representations of the alignments were generated with the ESPript server (http://espript.ibcp.fr) [144]. Disorder prediction was carried out on the RONN server (https://www.strubi.ox.ac.uk/RONN) [145].

2.2 Cloning

2.2.1 High-throughput cloning

All high-throughput cloning was conducted at the Oxford Protein Production Facility at the Research Complex in Harwell under the supervision of Dr Louise Bird. Dr Bird was responsible for designing all primers in the high-throughput cloning campaign after discussion over the chosen construct boundaries.

2.2.1.1 Polymerase chain reaction

Amplification of the genes of interest was achieved by polymerase chain reaction (PCR) using the primers (see Appendix A) and a standard mixture (Table 2.2). The reactions were cycled in a Veriti PCR machine (ABI Biosystems) using a standard program (Table 2.3). PCR products were analysed by agarose gel electrophoresis. Firstly, 5 µl of sample were mixed with 2 µl 5X loading buffer (0.25 % w/v bromophenol blue, 30 % v/v glycerol). A gel composed of 1.6 % w/v agarose dissolved in TBE buffer (100 mM Tris, 90 mM boric acid, 1 mM EDTA) containing 1X SYBR-Safe (Life Technologies) stain was cast and submerged in TBE once set. Samples were

Table 2.1: Sources of gene and protein sequences

	Species	Gene name	Uniprot accession code	NCBI reference sequence
Glucosidase I (Mannosyl-oligosaccharide glucosidase)	*Homo sapiens*	MOGS	Q13724	NP_006293.2
	Mus musculus	*mogs*	Q80UM7	NP_065644.2
	Caenorhabditis elegans	*mogs-1* (F13H10.4)	Q19426	NP_001122771.1
	Saccharomyces cerevisiae	CWH41	P53008	NP_011488.1
Glucosidase II α-subunit (Neutral alpha-glucosidase AB)	*Homo sapiens*	GANAB	Q14697	NP_938149.2
	Mus musculus	*ganab*	Q8BHN3-2	NP_032086.1
	Caenorhabditis elegans	*aagr-3*	G0NPN3	NP_505507.1
	Saccharomyces cerevisiae	ROT2	P38138	NP_009788.3
Glucosidase II β-subunit (Protein kinase C substrate 60.1 kDa protein heavy chain)	*Homo sapiens*	PRKCSH	P14314	NP_001001329.1
	Mus musculus	*prkcsh*	O08795	NP_032951.1
	Caenorhabditis elegans	ZK1307.8	G5EC87	NP_496073.1
	Saccharomyces cerevisiae	GTB1	Q04924	NP_010507.1

loaded into the gel, separated at 100 V for 30 min and visualised under a UV transluminator in a Gel Doc XR+ system (BioRad). Hyperladder 1kb reference markers (BioLine) were used to estimate product size. To avoid background from template DNA in subsequent steps, a DpnI digestion was performed to prevent template DNA from contaminating the newly assembled product. 5 µl of CutSmart buffer containing 10 U DpnI were added to the PCR reactions and incubated at 37 °C for 60 min. Purification of the PCR products was achieved by adding 90 µl of AMPure XP magnetic beads (Beckman Coulter) to bind the large nucleic acids. After 5 min, the plates were placed over a SPRIPlate 96R magnet (Beckman Coulter) and left for 5 minutes for the beads to settle. The unbound components were aspirated and the beads were washed with 200 µl of 70 % ethanol with removal on the magnet as mentioned above. The wash was repeated once and the beads were allowed to air-dry for 10 min. Elution from the beads was done by adding 30 µl of 10 mM Tris pH 8.0.

Table 2.2: High-throughput cloning PCR mixture

Component	Volume (µl)
2X Phusion Flash Master Mix	25
Water	17
Forward Primer (10 µM)	3
Reverse Primer (10 µM)	3
Template (1 ng/µl)	2

Table 2.3: PCR reaction conditions

Temperature (°C)	Time (s)	
98	10	
98	1	
60	5	30 Cycles
72	60	
72	120	
4	hold	

2.2.1.2 In-Fusion assembly

Assembly of the constructs was carried out by mixing 1 µl of the linearised OPPF vector with 5 µl of the purified PCR product. The mixture was made up to 10 µl total volume and added to lyophilised In-Fusion HD EcoDry enzyme mix (Clontech) and incubated for 30 min at 42 °C. 40 µl of TE (10 mM Tris pH 8.0, 1 mM EDTA) were added to stop the reaction and 3 µl of the mix were added to 50 µl One Shot OmniMAX 2 Chemically Competent *E. coli* (Life Technologies). After 30 min on ice, the cells were incubated at 42 °C for 30 s followed by 2 min on ice. 300 µl of Super Optimal broth with Catabolite repression (SOC) were added to the cells and left to incubate at 37 °C for 1 h with agitation. 5 µl and 25 µl of each construct were plated onto Blue-White Lysogeny Broth (LB) Agar containing 100 µg/ml carbenicillin, 0.02% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XGAL) , 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

2.2.1.3 High-throughput culture and plasmid purification

White colonies were picked and cultured in 1.2 ml LB containing 100 µg/ml carbenicillin overnight at 37 °C. Plasmid purification was conducted on a Bio-Robot 8000 (Qiagen) using the Wizard SV 96 kit (Promega) according to the manufacturer's protocol.

2.2.1.4 Construct confirmation

Constructs were analysed for correct insertion by PCR using a generic pOPIN forward primer coupled with the specific reverse primer from each construct and analysed by PCR and gel electrophoresis as detailed in 2.2.1.1.

2.2.2 Mutagenesis

After complications that arose while performing site-directed mutagenesis (SDM) in the OPPF vectors, mutagenesis of both α-glucosidase units was achieved by first sub-cloning the gene into the Litmus28i (New England Biosciences) cloning vector. SDM in this vector was then undertaken before a final sub-cloning step into pHLsec, the mother expression vector of the OPPF series of expression vectors. The subcloning into Litmus28i and the SDM of the point mutants causing the mutations D564E, E567Q, E567D (*Mus musculus* immature sequence numbering) were conducted by Alice Cross under my supervision.

2.2.2.1 Sub-cloning of *Mus musculus* α-glucosidase genes into Litmus28i

Primers to PCR amplify the three *Mus musculus* genes were designed (Table 2.4) to sub-clone into Litmus28i. Primers contained a XhoI site on the forward primers and a NcoI site on the reverse primer to enable cloning into the MCS of the vector. A standard PCR (see 2.2.1.1) was performed using the T_a and the genes in the pOPIN vectors as templates. The correctly sized products and the Litmus28i were digested by the restriction endonucleases XhoI and NcoI overnight at 37 °C. Prior to purification by gel extraction, digested Litmus28i was dephosphorylated by Antarctic Phosphatase for 15 min at 37 °C. Ligation with T4 DNA Ligase of the vector and insert in a 1:3 mole ratio was carried out overnight at 16 °C. The ligated DNA was transformed into NEB5α selecting with carbenicillin (Melford Laboratories Ltd.) and colonies screened by colony PCR with the same primers in Table 2.4 and PCR conditions as above.

Table 2.4: Primers for sub-cloning *Mus musculus* α-glucosidase genes into Litmus28i.

Primer Name	Sequence (5'-3')	T_m (°C)	T_a (°C)
*mogs*_F	ATATCTCGAGTGGGTGCTGGCGTGGCTC	72	71
*mogs*_R	TATCCATGGTTTTTCGAACTGCGGGTGGCTC	71	
*ganab*_F	TATACTCGAGGCTGTGGATAGAAGCAACTTTAAGACC	73	
*prkcsh*_F	ATATCTCGAGGGCGCTGTAGAAGTTAAGAGACC	73	68
His_6_R	TATCCATGGTTAGTGGTGGTGGTGGTGATGC	78	

2.2.2.2 Site-directed mutagenesis

The manufacturer's protocol of the Q5 Site-Direct Mutagenesis kit was followed using the template in Litmus28i and primers (Table 2.5) designed using the NEBaseChange online tool (http://nebasechanger.neb.com/). Colonies that appeared on carbenicillin where assessed by dye-terminator Sanger Sequencing [146] after plasmid purification by Source Bioscience.

Table 2.5: Primers for SDM of *Mus musculus ganab*

Primer name	Sequence (5'-3')	T_m (°C)	T_a (°C)
D564E_F	TGTTTGGAATGAGATGAATGAAC	55	58
D564E_R	TAAAGATTAGGAGCTGAACC	57	
D564N_F	TGTTTGGAATAACATGAATGAACC	55	58
D564N_R	TAAAGATTAGGAGCTGAACC	57	
D567Q_F	TGACATGAATCAACCGTCTGTGTTC	61	62
D567Q_R	TTCCAAACATAAAGATTAGGAGC	59	
D567D_F	TGACATGAATGATCCGTCTGTGTTC	60	59
D567D_R	TTCCAAACATAAAGATTAGGAG	55	
D567V_F	TGACATGAATGTGCCGTCTGTGTTC	56	59
D567V_R	TTCCAAACATAAAGATTAGGAG	55	
D640E_F	GTGGACAGGGGAGAACACTGCCG	70	69
D640E_R	ACAGCTCCAAAGCGCTGG	66	
D640N_F	GTGGACAGGGAACAACACTGCCG	68	69
D640N_R	ACAGCTCCAAAGCGCTGG	66	

2.2.2.3 Sub-cloning of *Mus musculus ganab* and *prkcsh* into pHLsec

Mutated *ganab* constructs were subcloned into pHLsec by Gibson Assembly[147]. Primers (Table 2.6) were designed with the NEBuilder Assembly Tool to create 15 bp overlaps on the PCR product with the AgeI/KpnI ends of the digested vector. Standard PCR (see 2.2.1.1) was performed using the T_a and the mutated genes in Litmus28i as the template. Purified PCR products were assembled into pre-digested pHLsec vector in a 0.06 pmol : 0.02 pmol (insert:vector) as per the manufacturer's protocol. Colonies that appeared on carbenicillin selecting LB agar plates were assessed by dye-terminator Sanger Sequencing after plasmid purification.

Table 2.6: Primers for subcloning *Mus musculus ganab* and *prkcsh* into pHLsec

Primer Name	Sequence (5'-3')	T_a (°C)
*ganab*_F	GGTTGCGTAGCTGAAACCGGTGCTGTGGATAGAAGCAACTTTAAGAC	62
*ganab*_R	GATGGTGGTGCTTGGTACCCTCGAACTGGGGGTGGCT	
*prkcsh*_F	GGTTGCGTAGCTGAAACCGGTGGCGCTGTAGAAGTTAAGAGAC	66
*prkcsh*_R	GATGGTGGTGCTTGGTACCTTATTAGTCCCCATCACTGGGTGC	

2.3 Expression screening and large-scale expression

2.3.1 High-throughput baculoviral-mediated expression in insect cells

These experiments were carried out by Heather Rada at the Oxford Protein Production Facility. Production of the P0 and P1 stock of baculovirus in *Sf*9 along with the expression tests presented were carried out by Heather Rada at the OPPF. In brief, co-transfection of the glucosidase-containing vector with linearised bacmid DNA allows the *in cellula* recombination and the baculoviruses are only able to be formed

due to an essential ORF contained in the expression vector [148, 149]. After a round of amplification, the P1 stock was used to infect cells for expression testing.

2.3.2 High-throughput transient transfection in adherent mammalian cells

HEK293T cells were grown in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10 % foetal calf serum (FCS), 1X non-essential amino acids and 2 mM L-glutamine. All cultures were grown at 37 °C in a 5% CO_2 environment. 1.5 X 10^5 cells/ml were added to each well of a 24-well culture plate and left to grow overnight. 2 µl of GeneJuice transfection reagent (Novagen) were diluted in 60 µl of serum-free DMEM per construct. 1 µg of DNA was added to each well and mixed thoroughly by pipetting and left to incubate for 10 minutes at room temperature. The culture medium was replaced with DMEM with 2 % FCS and the transfection mixture was added drop-wise to the cells. Cells were incubated for 3 days in the same conditions before harvesting the supernatant by centrifugation at 6000 *g* for 15 min. Samples were analysed for expression by western blotting as detailed in 2.5.2.

2.3.3 Transient transfection in suspension mammalian cells

Protein expression in the FreeStyle 293 Expression System (Life Technologies) was carried out according to the manufacturer's protocol with all components from Life Technologies except where specified. FreeStyle 293-F cells were maintained by growing 50 ml cultures in FreeStyle 293 Expression Medium in 500 ml vented Erlenmeyer flasks (Corning) at 37 °C in a 5% CO_2 environment and shaking at 135 rpm. Stock cultures were always split to 1 X 10^5 cells/ml and not allowed to exceed 3 X 10^6 cells/ml. When a new stock was commenced from stored cells in the liquid nitrogen vapour

phase, cells were not used for transfection until at least 5 passages after thawing.

The DNA used for transient transfection was purified from *E. coli* using the GenEluteHP Plasmid Maxiprep Kit (Sigma Aldrich) according to the manufacturer's protocol and sterilised by passing through a 0.22 µm mixed cellulose ester (MCE) filter. Cells were transfected at a density of 1 X 10^6 cells/ml with cells split the day before the transfection. Cells were only transfected when viability was above 95% as assessed by Trypan Blue exclusion. The culture volume was not allowed to exceed 30% of the flask volume. A linearly-scalable transfection protocol using 50 ml of culture as an example is described. A total of 62.5 µg of plasmid DNA was diluted into a total volume of 1 ml OptiPRO SFM in a 50 ml conical tube. 62.5 µl of the FreeStyle MAX transfection reagent were diluted into a total volume of 1 ml OptiPRO SFM in another 50 ml conical tube. The two diluted components were mixed and left to incubate at room temperature for 10 min followed by drop-wise addition to the culture while swirling. The culture was incubated in the same conditions as above for 4 days.

2.4 Protein purification

2.4.1 HEK293F supernatant preparation

Prior to harvesting the cultures, viability and general appearance was checked. The cells tend to clump as the density increases over the course of the transfection but viability generally stayed above 60%. Cells were harvested by centrifugation at 3000 *g* for 15 minutes at room temperature. Generally, 800 ml expressions were performed and the volumes below reflect this example. 88 ml of 10X PBS and 2.2 ml of 2 M imidazole were added making the final concentrations 1X and 5 mM, respectively. The pH of the supernatant was adjusted by drop wise addition of 10 M NaOH to a

final pH of 7.4-7.5. The modified supernatant was clarified by filtering under vacuum through a 0.45 µm MCE filter (Millipore) placed under a 2.7 µm glass microfibre filter (GF/D class, Whatman) to facilitate the speed of filtration.

2.4.2 Immobilised metal affinity chromatography

Buffers used for Immobilised Metal Affinity Chromatography (IMAC) were made by mixing Buffer A (PBS, 5 % (w/v) glycerol) with appropriate volumes of Buffer B (PBS, 5 % (w/v) glycerol, 500 mM imidazole) for the desired imidazole concentration.

2.4.2.1 *Mus musculus* GluI

The clarified supernatant was passed through a 5 ml HisTrap excel column (GE Lifesciences) once and washed with 20 mM imidazole (4 % Buffer B) for 10 column volumes (cV). A 10 cV step elution at 400 mM imidazole (80 % Buffer B) was used to elute the target glucosidase. The buffer was exchanged by dialysing the pooled eluate in 10 kDa MWCO SnakeSkin dialysis tubing (Thermo Scientific) against 1 l of PBS for 2 hours at 4 °C repeated three times in total.

2.4.2.2 *Mus musculus* GluII

The clarified supernatant was passed through a 5 ml HisTrap excel column once and washed with 40 mM imidazole (8 % Buffer B) for 10 cV. A 10 cV step elution at 400 mM imidazole (80 % Buffer B) was used to elute the target glucosidase. The buffer was exchanged by dialysing the pooled eluate in 10 kDa MWCO SnakeSkin dialysis tubing (Thermo Scientific) against 1 l of Strep Wash Buffer (100 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA) for 2 hours at 4 °C repeated three times in total.

2.4.3 StrepII-tag affinity chromatography of *Mus musculus* GluII

The StrepII-tag on the α-subunit was used to isolate the heterodimer of *Mm*GluII from excess β -subunit. The purification was carried out on 10 ml bed of Strep-Tactin Superflow High Capacity (IBA GmbH) in a glass column by gravity according to the manufacturer's protocol with buffer concentrates from the same manufacturer. The dialysis eluate from the IMAC step was passed over the resin followed by 3 cV wash with Strep Wash Buffer taking 1 cV fractions. Elution of the glucosidase was achieved with 3 cV of Strep Elution Buffer (Strep Wash Buffer with 2.5 mM D-desthiobiotin) collected in 0.5 cV fractions. As the *Mm*GluII had a weak affinity for the resin, the second and third wash fractions were combined with all of the elution fractions as excess β-subunit was already removed by these fractions. The pooled fractions were concentrated with 30 kDa MWCO Amicon Ultra-15 Ultrafiltration devices (Millipore) at 3200 *g* kept at 15 °C to a volume below 3 ml.

2.4.4 Trypsin digestion of *Mus musculus*GluII

The concentrated StrepII-tag purification pool was treated with $1/100^{\text{th}}$ of the mass of *Mm*GluII of sequencing grade modified trypsin (Promega) supplemented with 2 mM $CaCl_2$ for 4 h at room temperature. The reaction was stopped by the addition of 1 mM phenylmethylsulfonyl fluoride (PMSF).

2.4.5 Gel filtration chromatography

The final step for the purification of both ER α-glucosidases was by gel filtration (GF) on a Superdex 200 16/600 column (GE Lifesciences). The buffer for *Mm*GluI

was 20 mM HEPES pH 7.5, 100 mM NaCl, 0.1 mM TCEP. The buffer for *Mm*GluII was 20 mM HEPES pH 7.5, 150 mM NaCl.

2.5 Gel electrophoresis and blotting

2.5.1 SDS-PAGE

All sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out using NuPAGE (Life Technologies) 4-12 % Bis-Tris gels in either the MES or MOPS SDS running buffer according to the manufacturer's protocol. Briefly, up to 20 µl of samples were added to 7.5 µl of 4X LDS sample buffer and 3 µl 10X NuPAGE reducing agent, adding water to a total of 30 µl. Samples were incubated at 95 °C for approximately 5 min and half of this mixture was loaded onto the gel. Molecular weight markers used were Novex Sharp unstained protein standards. Gels were run at a constant 200 V for 35 or 50 min for MES or MOPS buffers, respectively. Gels were stained in Instant Blue (Expedeon) for at least 1 h on an orbital shaker, then exchanged into deionised water overnight.

2.5.2 Western blotting

Molecular weight standards for western blots were a mixture of Novex Sharp pre-stained markers mixed with Magic Mark XP standards in a 1:1 ratio. Electrophoresed gels were blotted onto PVDF membranes using the Trans-Blot Turbo Transfer System (Bio-Rad) using the High MW program. The membranes were blocked in TBST (Tris buffered saline, 0.1 % v/v Tween-20) containing 0.2 % w/v I-BLOCK (Applied BioSystems) for 1 h at room temperature or at 4 °C overnight on an orbital shaker. After three 5 min washes with TBST, the primary antibody diluted in TBST was

applied and incubated for 1 h. Following another three 5 min washes in TBST, the HRP-conjugated secondary antibody was applied and incubated for 1 h. The blot was developed with 1 ml Amersham ECL Prime Western Blotting Detection Reagent (GE Life Sciences) for 5 minutes with gentle agitation and imaged on a LAS-1000 Luminescent Image Analyser (Fuji Medical Systems).

2.6 Enzymology

2.6.1 Discontinuous α-Glucosidase II assay

2.6.1.1 4-Methylumbelliferone detection

Cleavage of 4-methylumbelliferyl α-D-glucopyranoside (4MUG) by *Mm*GluII was measured mixing a 50 µl sample with 50 µl of 0.5 mM 4MUG and incubating at 37 °C for 15 min in 96-well black non-binding surface treated microplates (Corning). The reaction was stopped with 100 µl of 0.5 M Na_2CO_3 (pH ≈11.5) and fluorescence was measured at λ_{ex} of 355 nm and λ_{em} of 460 nm on a SpectraMax M5 (Molecular Devices) with 40 scans per well, low photomultiplier tube sensitivity and a emission cut-off filter of 455 nm.

2.6.1.2 Glucose detection assay

Cleavage of glucobiose disaccharides of different glycosidic linkages was performed by mixing 30 µl of *Mm*GluII at 60 nM and 30 µl serially-diluted disaccharide in PCR tubes and incubating at 37 °C for 30 min. Glucose release was quantified against a 7-point D-glucose standard curve using the Amplex Red Glucose/Glucose Oxidase Assay Kit (Life Technologies) according to the manufacturer's protocol using 50 µl of the

previous reaction. Measurements were subtracted from disaccharide only containing wells.

2.6.2 Continuous α-Glucosidase II assay

A quantitative method for measuring the rates of 4MUG hydrolysis was set up similarly to the discontinuous assay in 2.6.1.1. If inhibitors were added, the substrate volume was reduced to 25 µl and 25 µl of inhibitor was added. A *Mm*GluII concentration of 20 nM provided the best signal over the course of the assay. 100 mM potassium phosphate buffer at pH 7.2 was used for all dilutions. 4-Methylumbelliferone (4MU) quantitation was achieved with a five-point standard curve which also contained the enzyme. Continuous measurement on the SpectraMax M5 set to 37 °C was done and by measuring fluorescence every minute for a total 30 min with otherwise identical settings to those in 2.6.1.1. A Michaelis-Menten model ($V = [S]V_{max}/(K_m + [S])$) was fit to the graph of initial velocity versus substrate concentration in order to obtain the values of V_{max}, K_m, and k_{cat}.

2.6.3 Determination of inhibition constants

25 µl amounts of each dilution of inhibitors were added to each column of the assay plate along with 25 µl of 80 nM *Mm*GluII and the enzyme/inhibitor mix was incubated at 37°C for 5 min. A 2-fold serial dilution series of the substrate was made and the reaction was started by its addition to the plate. Hydrolysis was measured as above in 2.6.2. The initial velocity was fit in the linear range of measurements and a competitive inhibition model ($K_{mObs} = K_m(1 + [I]/K_i)$) was fit to the points were K_{mObs} is determined for each concentration of inhibitor, thus allowing the K_i to be determined.

2.6.4 Determination of IC_{50} values

Serial dilution of inhibitors was carried out in half-log steps in the assay plate. 25 µl of 80 nM *Mm*GluII was added to the 50 µl of inhibitor and the enzyme/inhibitor mix was incubated at 37°C for 5 min. 25 µl of 1 mM 4MUG (final concentration 0.25 mM) was added to initiate the reaction. Hydrolysis was measured as above in 2.6.2. The initial velocity was fit in the linear range of measurements and activity was normalised to the uninhibited control in order to plot activity against the inhibitor concentration. A sigmoidal, four-parameter function with a variable slope was used to determine the IC_{50} values.

2.6.5 Cleavage and detection of glycans by normal-phase high-performance liquid chromatography

All experiments in this section were carried out by Dominic Alonzi in the Zitzmann laboratory. In brief, 2-AA-labelled glycans were mixed with varying concentrations of glucosidase at 37 °C and incubated for at least 3 h. The reaction was stopped with the addition of 30 µl acetonitrile. Ultrafiltration in a 10 kDa MWCO device was done at 7,000 *g* for 45 min to separate the glycans from the enzyme. The filtrate was applied to a TSKgel Amide-80 column (Anachem) for NP-HPLC analysis on a Waters Alliance 2695 separations module with an in-line Waters 474 fluorescence detector set at λ_{Ex} =360 nm and λ_{Em} =425 nm. All chromatography was performed at 30 °C. Solvent A was composed of 20 % 100 mM ammonium acetate, pH 3.85, in Milli-Q water and 80 % acetonitrile. Solvent B is composed of 20 % 100 mM ammonium acetate, pH 3.85, in Milli-Q water, 60 % Milli-Q water and 20 % acetonitrile. A linear gradient from 86 % A to 54.7 % A over 31.5 min at 0.8 ml/min was used to separate the glycans.

2.7 Crystallisation and structure determination

2.7.1 Crystallisation

Prior to setting up crystallisation experiments, all solutions were spin-filtered through 0.45 µm Ultrafree PVDF filtration devices (Millipore). For both glucosidases, crystals of the same morphologies could be obtained from both fresh and snap-frozen protein. All crystallisation solutions were obtained from Molecular Dimensions unless otherwise specified. The appropriate concentrations to use for crystallisation were initially determined with the Pre-crystallisation Test (Hampton Research). All crystallisation experiments were set up in 96-well MRC 2-well crystallisation plates or 48-well MRC Maxi crystallisation plates (Swissci). Nanolitre-scale drops were set up with a Mosquito Crystal robot (TTP Labtech). Unless stated, volumes for screening were 100 nl protein and 100 nl precipitant. All experiments were incubated in a Rigaku Gallery HT (Rigaku) imaging system that kept plates at 18 °C and were regularly imaged over a 6-week period.

2.7.1.1 *Mm*GluI

Pure *Mm*GluI at 4 mg/ml was crystallised by vapour diffusion in 30 % w/v polyacrylate 5100, 0.1 M MES pH 6.0 and 5 % v/v ethanol. A ratio of 2:1 protein:precipitant yielded clumps of thin plates. The crystals were cryoprotected by streaking the crystals though the mother liquor supplemented with 20 % v/v glycerol.

2.7.1.2 *Mm*GluII$_{\text{Trypsin}}$

Pure *Mm*GluII$_{\text{Trypsin}}$ at 5.6 mg/ml was crystallised by vapour diffusion in conditions detailed in Table 2.7. A ratio of 3:1 protein:precipitant yielded the largest crystals.

Soaking of D-glucose, D-glucal, and D-mannose was achieved by dissolving solid compounds in the mother liquor of the well that the crystal was in, before placing a single crystal in a drop of this solution. Soaks were performed for up to 2 min before immediately cryocooling as described in the next section. The 2nd orthorhombic crystal form was obtained by placing a crystal obtained in the optimised monoclinic apo drops (Table 2.7) into a solution of 16 % PEG 8000, 0.05 M Morpheus carboxylic acids mix, and 0.1 M Morpheus buffer system 2 pH 7.2, 20 % PEG 400, 10 mM conduritol B-epoxide (Santa Cruz Biotech) for 1 minute before cryocooling. Soaking was carried out by diluting compounds in the same solution and placing a crystal into the solution for at least 30 min.

2.7.2 Data collection

Crystals were harvested in nylon loops and cooled in liquid nitrogen to reduce radiation damage [150]. They were screened at the various macromolecular crystallography beamlines at Diamond Light Source (DLS), Harwell, UK or the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Crystals were stored under liquid nitrogen and data were collected under a gaseous nitrogen stream at 100 K. Automatic strategy in the ISPyB system utilising the EDNA plug-in was used to decide appropriate starting goniometer ϕ angles [151, 152]. Data collection parameters are as stated in Table 2.9.

2.7.3 Data processing and model building

Diffraction images were processed using the autoPROC [153] pipeline, which indexes and integrates with XDS [154], and scales and merges using the CCP4 [155] programs Truncate, Pointless and Aimless. Molecular replacement was achieved with Phaser

Table 2.7: Crystallisation conditions for *Mm*GluII$_{\text{Trypsin}}$

Crystal form	Precipitant	Protein: Precipitant ratio	Concentration of soaked molecule
Trigonal apo	30 % Morpheus Glycerol/PEG 4000 Mix, 0.12 M Morpheus alcohols mix, 0.1 M Morpheus Buffer system 1 pH 6.5	1:1	-
Monoclinic apo	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.1 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	1:1	-
Orthorhombic apo	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.1 M Morpheus ethylene glycols mix, 0.1 M Morpheus buffer system 2 pH 7.25	1:1	-
Glucal soak	30 % Morpheus ethylene glycol/PEG 8000 mix, 0.1 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	15 % w/v
Glucose soak	33 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	2:1	30 % w/v
Mannose soak	31 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	30 % w/v
2$^{\text{nd}}$ Orthorhombic apo	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	-
Castanospermine soak	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	10 mM
DNJ soak	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	50 mM
*N*B-DNJ soak	32 % Morpheus ethylene glycol/PEG 8000 mix, 0.05 M Morpheus carboxylic acids mix, 0.1 M Morpheus buffer system 1 pH 6.25	3:1	10 mM

Table 2.8: Data collection parameters of *Mm*GluI crystals

Beamline	DLS I04-1
Detector	Pilatus P2M
Wavelength (Å)	0.9173
Osc per image (°)	0.5
Number of images	360
Exposure (s)	0.5
Beam Size (µm)	50 x 50
Transmission (%)	47.0

Table 2.9: Data collection parameters of *Mm*GluII$_{\text{Trypsin}}$ crystals

	Trigonal apo	Monoclinic apo	Orthorhombic	Orthorhombic # 2	Glucal
Beamline	DLS I03	DLS I03	DLS I03	ESRF ID30A-1	DLS I02
Detector (Pilatus type)	P3 6M	P3 6M	P3 6M	P3 2M	P6M-F
Wavelength (Å)	0.9796	0.9796	0.9796	0.9660	0.9791
Osc per image (°)	0.15	0.20	0.15	0.05	0.15
Number of images	1200	900	1200	3160	1200
Exposure (s)	0.50	0.20	0.30	0.03	0.04
Beam Size (µm)	50 x 20	80 x 20	80 x 20	100 x 65	82 x 28
Transmission (%)	40.94	40.94	40.94	100	10
	Glucose	Mannose	Castanospermine	DNJ	*N*B-DNJ
Beamline	DLS I02	DLS I02	DLS I04	DLS I04	DLS I04
Detector (Pilatus type)	P6M-F	P6M-F	P6M-F	P6M-F	P6M-F
Wavelength (Å)	0.9795	0.9795	0.9795	0.9795	0.9795
Osc per image (°)	0.10	0.15	0.15	0.10	0.10
Number of images	1800	1200	1200	1800	1800
Exposure (s)	0.055	0.150	0.107	0.150	0.150
Beam Size (µm)	82 x 29	82 x 32	43 x 30	63 x 50	63 x 50
Transmission (%)	100	100	60	60	100

[156], also part of the CCP4 suite. Model building was performed with Coot [157] and Buccaneer, and refinement with BUSTER [158], except for the twinned trigonal form which was refined in Refmac. Model validation was carried out with internal modules of Coot and through the MolProbity server[159]. Ligands not present in the BUSTER libraries and non-standard ligands were generated using the GRADE server (http://grade.globalphasing.org/). All figures were produced in PyMOL [160].

2.8 Mass spectrometry

Samples for mass spectrometry were denatured in 7 M urea for 45 min at room temperature. Disulphides were reduced by the addition of 50 mM DTT for 10 min before diluting to 500 µl with 1% v/v formic acid in mass-spec grade water (both Thermo Scientific). Urea, DTT and other salts were removed by concentrating in a Viva Spin 500 10 kDa MWCO unltrafiltration device. When concentrated to 50 µl, the sample was diluted to 500 µl and concentrated again. This process was repeated five times to ensure any salts were at or below the µM range. The final concentration of the sample was 2 mg/ml which was diluted with an equal volume of 60 % v/v acetonitrile 20% v/v DMSO. The samples were directly infused into a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific) at 5 µl/min. A composite of 3 micoscans in the range of 500 to 6,000 m/z was measured at a resolution of 70,000 in positive ion mode. In-source collision induced dissociation set to 20 eV and a spray voltage of 4 kV enabled a good signal of the intact polypeptides. Spectra were de-convoluted with the supplied Xcalibur software. The data collection and analysis was done under the supervision of Dr Abhinav Kumar in the Zitzmann laboratory.

2.9 Biophysical assays

2.9.1 Protein concentration measurements

Protein concentrations were determined by measuring absorbance at 280 nm (A_{280}) on a NanoDrop 1000 (ThermoFisher). All measurements were made against a buffer blank of the same composition. The extinction coefficients (ϵ_{280}) shown in Table 2.10 were estimated using the ProteinCalculator v3.4[i] which calculates the ϵ_{280} by the number of tryptophans, tyrosines, and oxidised cysteines in the given sequence based on standard values for each of these residues [161].

Table 2.10: Extinction coefficients of recombinant proteins with not post-translational modifications

Construct	M (g/mol)	ϵ (1/Mcm)
*Mm*GluI	88011	167850
*Mm*GluII-α	107940	18741
*Mm*GluII-β	57987	54300
*Mm*GluII heterodimer	165927	241710
*Mm*GluII$_{\text{Trypsin}}$	108718	181750

2.9.2 Circular dichroism

To remove chloride ions and HEPES, purified protein was dialysed three times against 500 ml of 20 mM potassium phosphate pH 7.5 using a 10 kDa MWCO Slide-A-Lyzer (Thermo Scientific) dialysis device for at least 3 hours for each change at 4 °C. The dialysed protein was diluted to 0.1 mg/ml and spectra were recorded on a J-815 Spectropolarimeter (Jasco) set to 20 °C with a peltier thermostat. A 1 mm polarimetric QS quartz cuvette (Hellma) was used. Data collection settings were set to measure a range of 260-185 nm with a 0.5 nm data pitch, 1.0 s DIT, 1.0 nm

[i]Program authored by C. D. Putnam, The Scripps Research Institute: http://protcalc.sourceforge.net/

bandwidth, 20 nm/min scanning speed, and 3 accumulations. Buffer contribution was subtracted by collecting a blank in the same conditions first. Spectra were smoothed using the spectrometer's in-built Savitzky-Golay method [162]. Data were exported and analysed for secondary structure content using the CDSSTR analysis method on the DICHROWEB server [163, 164]. Measurements of the CD spectra of several point mutants, D564E, E567Q, and E567D, were carried out under supervision by Alice Cross.

2.9.3 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) experiments were set up by mixing the protein of interest in the concentration range of 0.025 - 0.1 mg/ml with SYPRO Orange (Molecular Probes) at a concentration of 5X in a final volume of 50 µl in white, polypropylene, non-skirted PCR plates (Stralab) sealed with optically-clear ThermalSeal RT2 film (Alpha Laboratories). All measurements were done in quadruplicate. A 25°C to 80 °C thermal ramp with 1 degree per minute was performed on a MX3005P real time PCR machine (Stratagene) measuring fluorescence with FAM and ROX filters which correspond approximately to λ_{ex} of 494 nm and λ_{em} of 602 nm. Buffer screening was set up in a similar manner, however, a 10X protein/SYPRO Orange mix was prepared and diluted with 45 µl amount of each of the buffer conditions tested Table 2.11. The buffer screen used here was made by Dr Emma Dixon in this group and the experiments were conducted under her direction. An alternative system, 7500 Fast Real-Time PCR System (Applied Biosystems), was used for the experiments on *Mm*GluI in the same manner as above but only using the ROX filter and a final SYPRO Orange concentration of 2.5X.

Table 2.11: DSF buffer screening conditions

Condition Number	Buffer	Salt	Additive
1	20 mM NaOAc pH 4.5	150 mM NaCl	-
2	20 mM NaOAc pH 5.0	150 mM NaCl	-
3	20 mM MES pH 5.5	150 mM NaCl	-
4	20 mM MES pH 6.0	150 mM NaCl	-
5	20 mM MES pH 6.5	150 mM NaCl	-
6	20 mM MOPS pH 7.0	150 mM NaCl	-
7	20 mM HEPES pH 7.5	150 mM NaCl	-
8	20 mM Tris pH 8.0	150 mM NaCl	-
9	20 mM Na_2CO_3 pH 8.5	150 mM NaCl	-
10	20 mM Na_2CO_3 pH 9.0	150 mM NaCl	-
11	20 mM Na_2CO_3 pH 9.5	150 mM NaCl	-
12	20 mM HEPES pH 7.5	-	-
13	20 mM HEPES pH 7.5	50 mM NaCl	-
14	20 mM HEPES pH 7.5	100 mM NaCl	-
15	20 mM HEPES pH 7.5	500 mM NaCl	-
16	20 mM HEPES pH 7.5	150 mM NaCl	10 mM $CaCl_2$
17	20 mM HEPES pH 7.5	150 mM NaCl	5 mM $CaCl_2$
18	20 mM HEPES pH 7.5	150 mM NaCl	10 mM $MgCl_2$
19	20 mM HEPES pH 7.5	150 mM NaCl	5 mM $MgCl_2$
20	20 mM HEPES pH 7.5	150 mM NaCl	10 mM Glucose
21	20 mM HEPES pH 7.5	150 mM NaCl	10 mM Mannose
22	20 mM HEPES pH 7.5	150 mM NaCl	684 mM Glycerol

2.9.4 Multi-angle laser light scattering

All multi-angle laser light scattering (MALS) experiments were conducted by Dr David Staunton, the Facility Manager of the Molecular Biophysics Suite in the Department. Samples were separated on a Superdex 200 10/300 (GE Life sciences) column pre-equilibrated in PBS. Samples were serparated using a Prominence HPLC (Shimadzu) at 0.5 ml/min with an online UV, refractive index and a Dawn HELEOS 8^+ (Wyatt Technologies) multi-angle laser light scattering detector set to 662.3 nm. Peaks were analysed with a Zimm model using a refractive index increment of 0.185 ml/g.

2.9.5 Small angle X-ray scattering

Samples for small angle X-ray scattering (SAXS) were prepared as detailed in 2.4.5, but the final gel filtration step proceeded for an extra 0.5 cV in order to collect enough buffer to use for subsequent subtraction steps. Immediately prior to data collection, the samples were filtered through a 0.22 µm Ultrafree PVDF filtration device (Millipore) to remove any particulates or aggregates that may have formed during shipping. Sample concentrations were determined as described in 2.9.1 and a four-point, two-fold dilution series was made. Experiments were performed at the BM29 BioSAXS beamline (ESRF, Grenoble, France). Samples were kept at 20 °C in a dry block in a sample changer which served as the automated sample delivery into the beam. 30 µl of each sample was flowed through a quartz capillary taking taking 10 X 1 s images. The wavelength was set at 0.992 Å and transmission was at 100 %, with images recorded on a Pilatus 1M detector set to a distance of 2.886 m. Automated image processing followed by buffer subtraction as part of the processing pipeline at the beamline allowed scattering curves to be used for further data processing. Calibration was conducted with measurements on albumin or glucose oxidase in order to derive molecular weights from I(0) values. Generation of composite curves, Guinier fitting and pair distribution functions were calculated using the ATSAS software package using Primus and GNOM [165–167].

3

Structural, biophysical, and biochemical characterisation of *Mus musculus* Endoplasmic Reticulum α-Glucosidase I

The biochemical characterisation of tissue-derived, purified glucosidase I (GluI) has been carried out for a number of different mammalian organisms [52, 63, 168–173]. The isolation of GluI was attempted from mouse and rat livers using established methods [173], but its yield and stability were unsuitable for structural studies de-

spite its good purity. The production of GluI from a number of different fungal species has been achieved using recombinant methods that include removal of the transmembrane domain to facilitate purification and handling [55, 174]. This includes the large-scale expression of the *S. cerevisiae* GluI (*Sc*GluI) whose crystal structure has been determined. However, at present, there is no method of producing large amounts of any mammalian GluI. As is discussed in this and the subsequent chapters, a high-throughput approach was adopted to quickly clone and screen expression and purification of a number of glucosidase constructs at the Oxford Protein Production Facility (OPPF).

3.1 Construct design

GluI is encoded by the mannosyl-oligosaccharide glucosidase (MOGS) gene. As the initial planning of the constructs was carried out prior to the release of the *Sc*GluI structure, the choice of construct boundaries was guided by biochemical studies along with basic bioinformatic analysis of disorder prediction. Apart from the N-terminus, which is likely to be a signal sequence, there are no extended stretches of sequences that have a high degree of disorder for the human GluI (Figure 3.1) indicating that it is likely to be a compact, globular structure. Two lengths of constructs were chosen based on previous attempts of expressing the human GluI in this institute and biochemical studies of *Sc*GluI [56, 57, 175]. The first length corresponds to what is likely to be the entire lumenal fragment proximal to the transmembrane helix. The second is based on endoproteolytic fragments observed *in vitro* that correspond to the removal of approximately 70 residues N-terminal to the first construct.

A large number of GluI sequences similar (>25%) to the human GluI sequence are available but due to practical limitations, our campaign was limited to

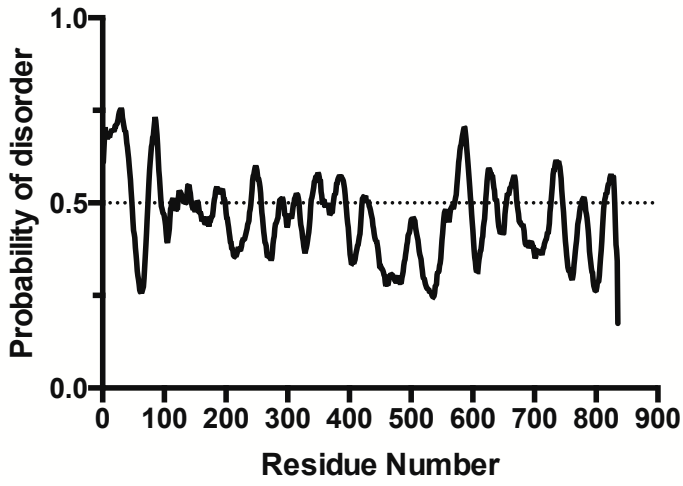


Figure 3.1: Disorder prediction of the gene products of the *H. sapiens* MOGS using the RONN server[145]. Values above 0.5 are predicted disordered regions of the protein.

four different species. Human and *S. cerevisiae* GluI are the two natural choices based on the clinical relevance of the former and the wealth of biochemical information for the latter. The mouse sequence was chosen due to its very high degree of similarity to the human one (Table 3.1) and the *C. elegans* sequence shows the potential for a more compact structure based on smaller lengths in the alignment around regions corresponding to loops (Appendix C).

Table 3.1: Percentage identity matrix of the MOGS gene products of selected genes

	H. sapiens	*M. musculus*	*C. elegans*	*S. cerevisiae*
H. sapiens	100			
M. musculus	87	100		
C. elegans	37	37	100	
S. cerevisiae	32	33	30	100

A total of 24 constructs (Table 3.2) were designed with the following parameters:

Construct lengths	35-end, 138-end (*S. cerevisiae* numbering)
Species	*H. sapiens*, *M. musculus*, *C. elegans*, *S. cerevisiae*
Affinity tags	Hexahistidine, StrepII, FLAG, HALO7
Expression locations	Intracellular, Secreted
Expression systems	Insect cells (*Sf*9), Mammalian cells (HEK 293-T and HEK 293-F), *P.pastoris*

Table 3.2: Glucosidase I expression constructs

Species	Construct reference	Vector	Construct boundary (aa numbering)
C. elegans	A1	pOPINE-3C-HALO7	61-797
	B1		131-797
	C1	pOPINE-FLAG	61-797
	D1		131-797
H. sapiens	E1	pOPINE-3C-HALO7	62-837
	F1		138-837
	G1	pOPINE-FLAG	62-837
	H1		138-837
M. musculus	A2	pOPINE-3C-HALO7	60-835
	B2		137-835
	C2	pOPINE-FLAG	60-835
	D2		137-835
S. cerevisiae	E2	pOPINE-3C-HALO7	35-834
	F2		91-834
	G2	pOPINE-FLAG	35-834
	H2		91-834
H. sapiens	A3	pOPINGS	62-837
	B3		138-837
M. musculus	C3	pOPINGS	60-835
	D3		137-835
C. elegans	E3	pOPINGS	61-797
	F3		131-797
S. cerevisiae	G3	pOPINGS	35-834
	H3		91-834

3.2 Cloning

The 24 constructs were produced in parallel utilising the OPPF platform. The high-throughput PCR reaction was completely successful yielding all 24 products with the correct size of ∼2400 bp (Figure 3.2). After purifying the PCR products, In-Fusion mediated Gibson assembly was used to place the constructs into their respective pOPIN vectors. The successfully assembled constructs were assessed by blue/white selection and once purified, an analytical PCR using a gene and a vector specific primer was used to confirm the presence of the correct construct.

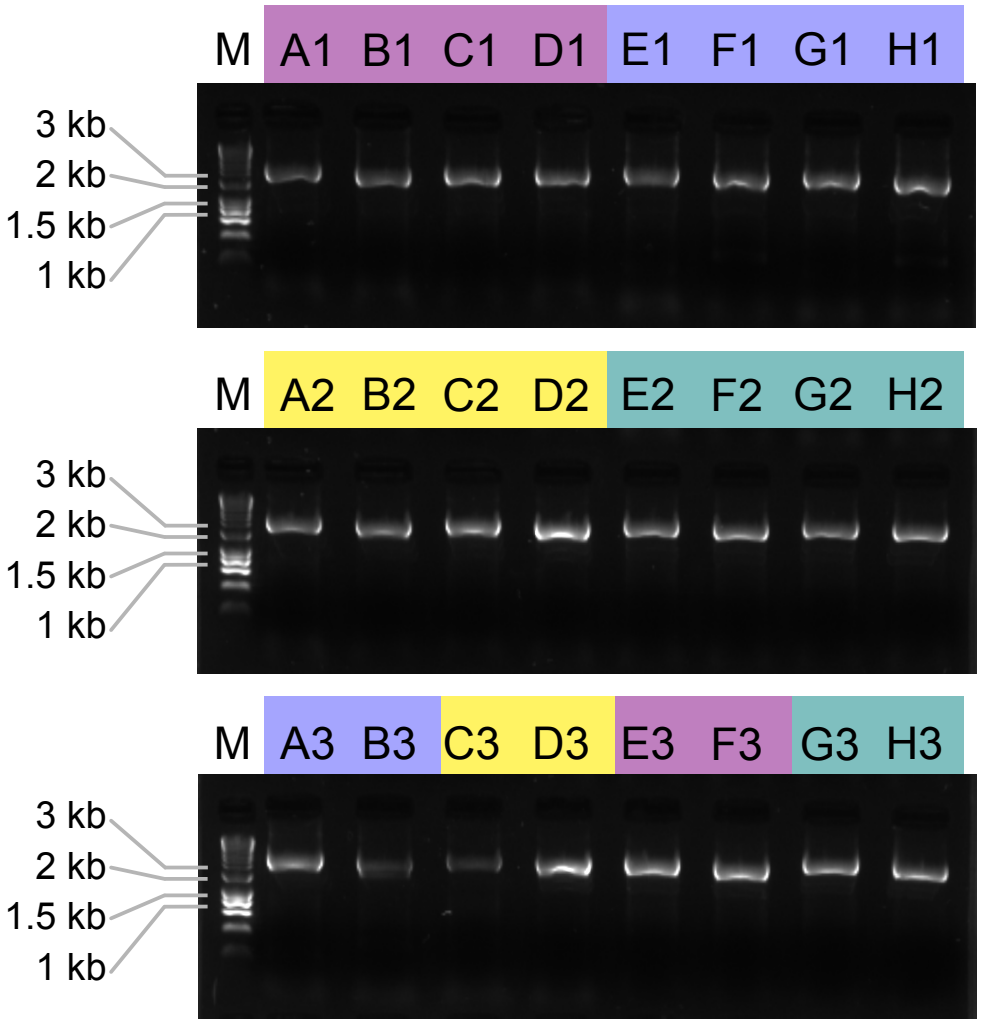


Figure 3.2: Agarose gel of products from high-throughput PCR reaction. 'M' indicates a HyperLadder 1kb marker with selected bands indicated in kilobases (kb)

3.3 Expression screening

Following the confirmation of the correct assembly of the 24 constructs, expression screening in insect and mammalian cells was carried out and is summarised in Table 3.3. Despite the vectors allowing ready expression in *E. coli*, this was not attempted as previous attempts in this institute were not successful and generally resulted in insoluble material located in inclusion bodies. Production of recombinant baculoviruses that contained the constructs were used to infect *Sf*9 cells and both secreted and intracellular fractions were assessed by western blotting and small-scale IMAC. None of the constructs showed any visible expression with the exception of one of the mouse and one of the yeast constructs that were however, barely detectable. Similar results were obtained when transient transfection of HEK 293-T cells was performed. The conditions for expression of *Sc*GluI in *P. pastoris* have been reported previously [176]. The published protocol for expression in stably transformed *P. pastoris* clones was replicated both for the yeast and the similar constructs from the other three species, but did not result in any secreted expression. A final attempt using transient transfection in suspension HEK 293-F cells was carried out. A very weak band on the western blot was observed for the mouse construct C3 (Figure 3.3). A small-scale transfection followed by an IMAC step resulted in successful purification of a large amount of GluI from the supernatant of HEK 293-F cells. This construct, henceforth termed *Mm*GluI, was carried forward for large-scale expression, biochemical, and structural characterisation.

3.4 Expression and purification of *Mm*GluI

The manufacturer's recommended conditions for the transfection of construct C3 proved to be highly effective in producing *Mm*GluI. A typical transfection of 800 ml

Table 3.3: Summary of expression screening for GluI

Species	Construct Reference	*P. pastoris*	Insect Cells		Mammalian cells	
			Intracellular	Secreted	HEK293T Secreted	HEK293F Secreted
C. elegans	A1	Not Tested	×	Not Tested	×	Not Tested
	B1	Not Tested	×	Not Tested	×	Not Tested
	C1	Not Tested	×	Not Tested	×	Not Tested
	D1	Not Tested	×	Not Tested	×	Not Tested
H.sapiens	E1	Not Tested	×	Not Tested	×	Not Tested
	F1	Not Tested	×	Not Tested	×	Not Tested
	G1	Not Tested	×	Not Tested	×	Not Tested
	H1	Not Tested	×	Not Tested	×	Not Tested
M. musculus	A2	Not Tested	×	Not Tested	×	Not Tested
	B2	Not Tested	×	Not Tested	×	Not Tested
	C2	Not Tested	×	Not Tested	×	Not Tested
	D2	Not Tested	×	Not Tested	×	Not Tested
S. cerevisiae	E2	Not Tested	×	Not Tested	×	Not Tested
	F2	Not Tested	×	Not Tested	×	Not Tested
	G2	Not Tested	×	Not Tested	×	Not Tested
	H2	Not Tested	×	Not Tested	×	Not Tested
H. sapiens	A3	×	Not Tested	×	×	×
	B3	Not Tested	Not Tested	×	×	Not Tested
M. musculus	C3	×	Not Tested	Weak expression	×	✓
	D3	Not Tested	Not Tested	×	×	Not Tested
C. elegans	E3	×	Not Tested	×	×	×
	F3	Not Tested	Not Tested	×	×	Not Tested
S. cerevisiae	G3	×	Not Tested	Weak expression	×	×
	H3	Not Tested	Not Tested	×	×	Not Tested

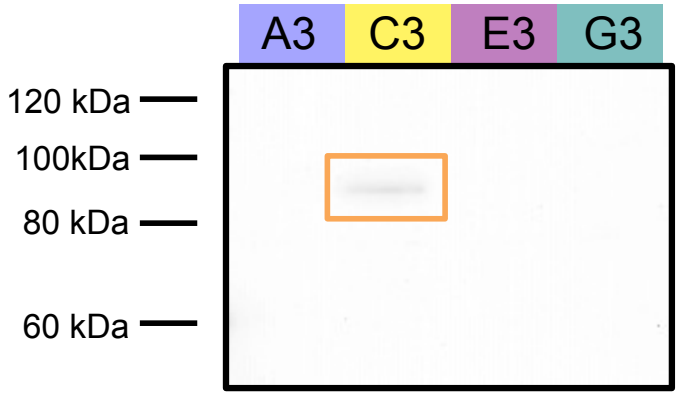


Figure 3.3: Expression of secreted GluI in HEK 293-F cells. Western blot analysis of secreted samples probed with an α-hexahistidine-HRP monoclonal antibody. Orange box highlights the band at ~90 kDa.

of cells was incubated for four days before harvesting. A small amount of reducing agent (0.1 mM TCEP) was included throughout the purification as a simple homology model (expanded upon in 3.9.1) indicated a free cysteine close to the surface of the model. Facile automated processing of large volumes was achieved on an ÄKTA design chromatography system with a step elution of 500 mM imidazole (Figure 3.4A). After removal of the imidazole by dialysis in order to prevent aggregation, the concentrated sample was subjected to GF chromatography. Typical final yield of purified *Mm*GluI is about 2 mg/l of culture. Two distinct peaks were observed on a Superdex 200 column (Figure 3.4B) despite only one molecular weight species being visible by SDS-PAGE both under reducing (Figure 3.4C) and non-reducing conditions (gel not shown); both peaks gave identical gel band sizes and intensities. This likely indicates a concentration-dependent dimerisation and is investigated and presented below in 3.5.3.

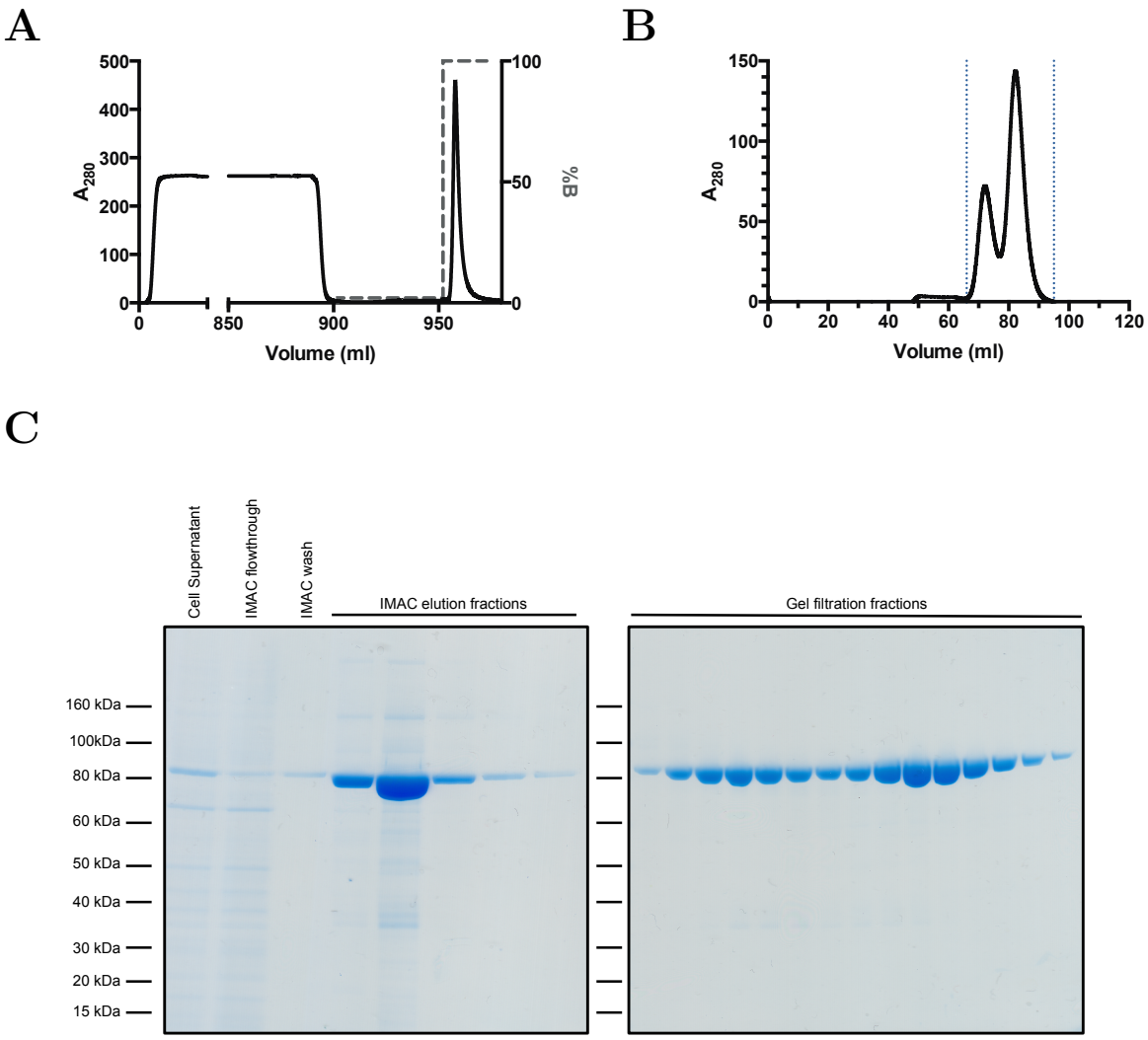


Figure 3.4: Purification of *Mm*GluI. (A) IMAC chromatogram of purification on a 5 ml HisTrap excel column from 800 ml of cell supernatant where buffer B is PBS containing 500 mM imidazole. (B) A gel filtration chromatogram on a Superdex 200 16/600 column. (C) SDS-PAGE analysis of all fractions during a typical purification. 1 cV fractions from the IMAC elution are shown and 1.5 ml fractions within the indicated range formed by the blue lines represent the gel filtration fractions.

3.5 Characterisation

3.5.1 Circular dichroism

To ascertain if the purified *Mm*GluI was correctly folded and if it resembled the *Sc*GluI structure in secondary structure content, the CD spectrum of *Mm*GluI was collected. To minimise interference by absorption of chloride ions, the protein was

dialysed into a low molarity of potassium phosphate at the same pH immediately before collection of the spectrum. The spectrum shows an archetypal double minimum between 210 and 220 nm (Figure 3.5) for α-helical proteins. A more detailed analysis using the CDSSTR secondary structure matching algorithm gives a quantitative deconvolution of the proportions of different secondary structural elements (Table 3.4). A comparison of these values to those from the crystal structure of *Sc*GluI shows a high degree of similarity particularly for α-helices and β-strands.

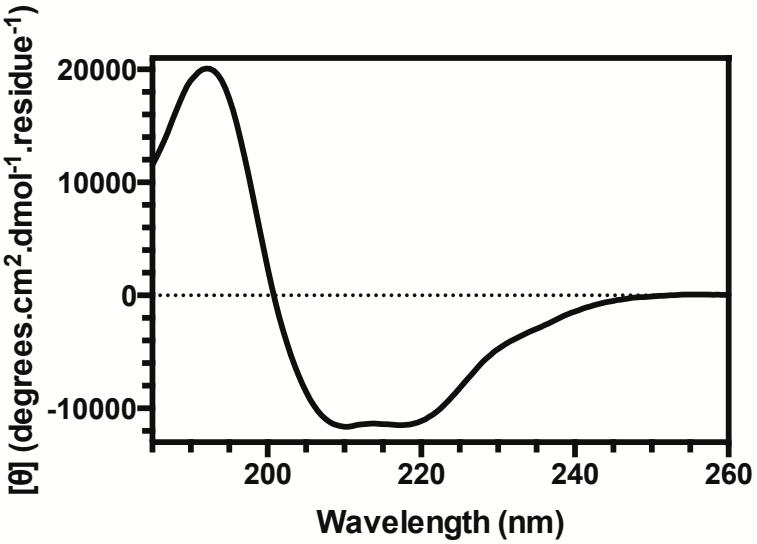


Figure 3.5: CD spectrum of *Mm*GluI at 0.1 mg/ml carried out in 20 mM potassium phosphate pH 7.5. The applied HT voltage of the plot is below 600 V across the wavelengths measured.

Table 3.4: Secondary structure analysis of *Mm*GluI using the CDSSTR method on the DICHROWEB server expressed as percentages of total protein secondary structural elements

Sample	Helix	Strand	Turn	Disordered	NRMSD[a]
*Mm*GluI	34	20	19	27	0.019
*Sc*GluI (PDB: 4J5T)[b]	36.1	18.2	23.1	22.5	N/A

[a] Normalised root mean square deviation
[b] Calculated using DSSP [177] method on 2Struc server (http://2struc.cryst.bbk.ac.uk/twostruc)

3.5.2 Differential scanning fluorimetry

A DSF assay (also called ThermoFluor) for determining the melting temperature (T_m) is a rapid and scalable method that can determine relative T_ms of a protein in multiple buffer/ligand conditions. DSF has advantages over other more precise techniques like differential scanning calorimetry or circular dichroism in measuring T_m, as it allows a much greater throughput for measuring different conditions. DSF assays are carried out in a real-time PCR machine and thus it is possible to concurrently measure fluorescence over a range of accurate temperatures. The principle of the assay is that the dye, SYPRO Orange, is in a quenched state in the aqueous phase. As the temperature increases, the protein begins to unfold and exposes the hydrophobic core. If a protein exhibits a two-state transition, the plot of the fluorescence intensity with respect to temperature can be fitted to a sigmoidal function where the T_m is the temperature halfway along the transition or where the first derivative of that curve is at its maximum. As the temperature continues to increase, aggregation occurs which excludes the SYPRO Orange from the protein causing the characteristic drop in fluorescence at the higher temperatures [178].

A preliminary study to assess the stability of *Mm*GluI was carried out to ascertain if a rapid inhibitor binding assay was possible. The suitable concentration to measure the melting curve was determined to be 0.05 mg/ml which gives a T_m of 49.0 $\pm$ 0.1 °C (Figure 3.6A). Binding of inhibitors can cause an increase in the stability of an enzyme [179]. Indeed, increasing concentrations of *N*B-DNJ increased the T_m of *Mm*GluI by over 10 °C (Figure 3.6B). However, even though this curve has a sigmoidal shape, it is very broad indicating non-standard, reversible binding. This could be due to some form of cooperativity or multiple *N*B-DNJ binding sites that are present. This warrants further investigation.

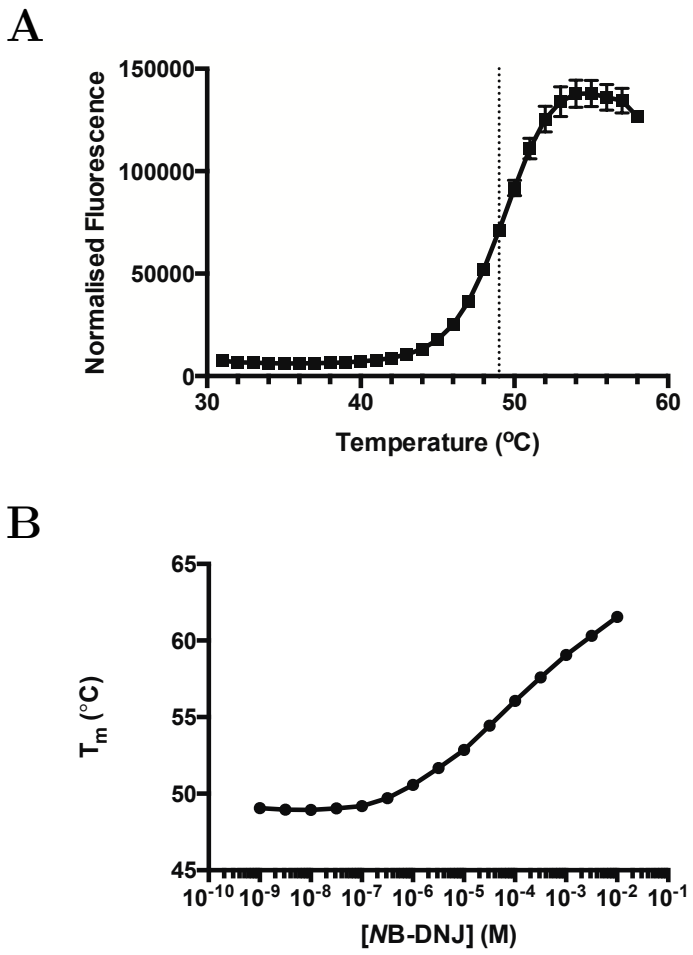


Figure 3.6: Differential scanning fluorimetry of *Mm*GluI. (A) Determination of the T_m of *Mm*GluI based on a sigmoidal fit of the normalised fluorescence. (B) Plot of T_m at various concentrations of *N*B-DNJ. All measurements were done in triplicate and error bars indicate standard deviations.

3.5.3 Multi-angle laser light scattering

In order to confirm the dimerisation observed when purifying *Mm*GluI by GF, size exclusion chromatography multi-angle laser light scattering (SEC-MALS) was utilised. SEC-MALS allows the calculation of the absolute molecular mass by combining the information of the light scattering measured at multiple angles, refractive index and UV absorption [180]. The assignment of a molecular mass of a particular peak can be used to infer the oligomeric state of that peak. In the case of *Mm*GluI, both peaks collected from the preparative column were subjected to SEC-MALS analysis

(Figure 3.7A and B, for the early and late peak, respectively). In both cases two peaks are observed with the later peak corresponding to about 90 kDa, the theoretical molecular weight of *Mm*GluI. The earlier peak of both contains molecules whose mass spans across the broad range 130-170 kDa. Taken together, these experiments indicate the re-establishment of a monomer/dimer equilibrium from the GF-purified monomer and dimer peaks upon their re-injection onto the column. Although not fully investigated, there is most likely a concentration-dependent dimerisation occurring at the concentrations at which this enzyme is being handled during purification (0.5-4 mg/ml) but whether this is occurring in a physiological setting is difficult to know. The dimeric state appears to be either very weakly associated or that the kinetics are very rapid. This can be seen in both traces by the poor fitting, that is, the slope over the peak (green lines) to a single molecular weight species. Fast kinetics may result in a ratio of peaks that are identical in both runs but the difference observed may be the result of beginning each run with differences in protein concentrations. A more thorough experiment would be to conduct a number of separations across a protein concentration dilution series to more thoroughly assess this behaviour.

3.6 Activity against fluorescently-labelled *N*-glycans

The only reliable assay for measuring activity of GluI is by using substrates longer than disaccharides with the most common being the full tri-glucosylated glycan. Historically, tritium-labelled glycans were the most common detectable substrates, though developments in fluorescent labelling have become more prevalent. 2-anthranilic acid (2AA), 2-aminobenzamide (2AB), pyridylamino (PA), and boron-dipyrromethene (BODIPY) labels have all been used successfully [181–183]. Isolation or chemical synthesis of the full $Glc_3Man_9GlcNAc_2$ to use as a substrate is very labour intensive; an easier method of obtaining the similar $Glc_3Man_7GlcNAc_2$ or $Glc_3Man_5GlcNAc_2$

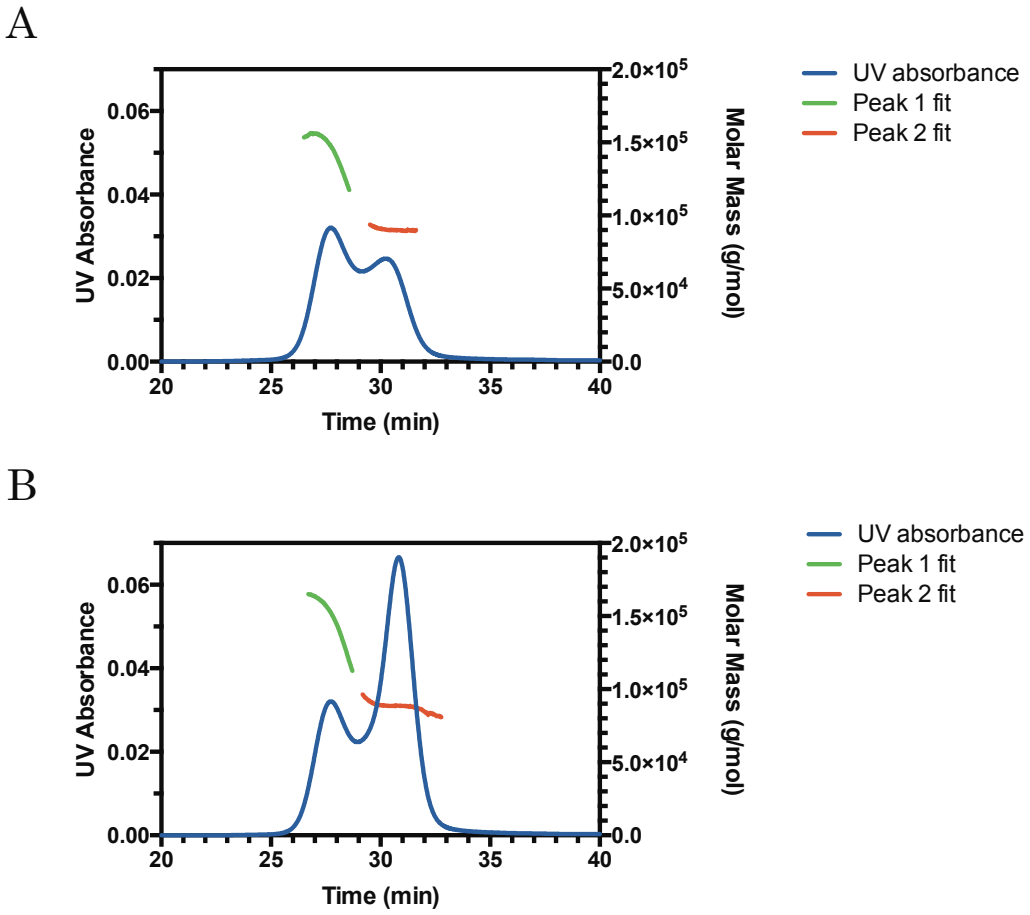


Figure 3.7: SEC-MALS analysis of *Mm*GluI on a Superdex 200 10/300 column in PBS at 0.5 ml/min. (A) Dimer pooled peak and (B) Monomer pooled peak.

glycans is by isolating the free oligosaccharides (fOS) from iminosugar-treated cells. These particular sugars arise from endoglycosidic release of the glycans before the degradation of glycoproteins that were retrotranslocated into the cytoplasm [184, 185]. The intact chitobiose core isolated for these assays is due to isolation of fOS species from the non-proteasomal pathway by an enzyme with PNGase-like activity. The isolated and labelled $Glc_3Man_7GlcNAc_2$-2AA substrate was used to assess if the purified *Mm*GluI is active. An initial experiment was conducted overnight at 37 °C and the analysis of the glycans by normal phase high performance liquid chromatography (NP-HPLC) showed a shift in retention time of about 0.8 glucose units (Figure 3.8). A titration of the enzyme showed that a concentration of 5 µg/ml provides a linear rate

of hydrolysis of the substrate (Figure 3.9). Thus, in an *in vitro* setting, *Mm*GluI is an active enzyme against a very similar substrate to that encountered in physiological conditions.

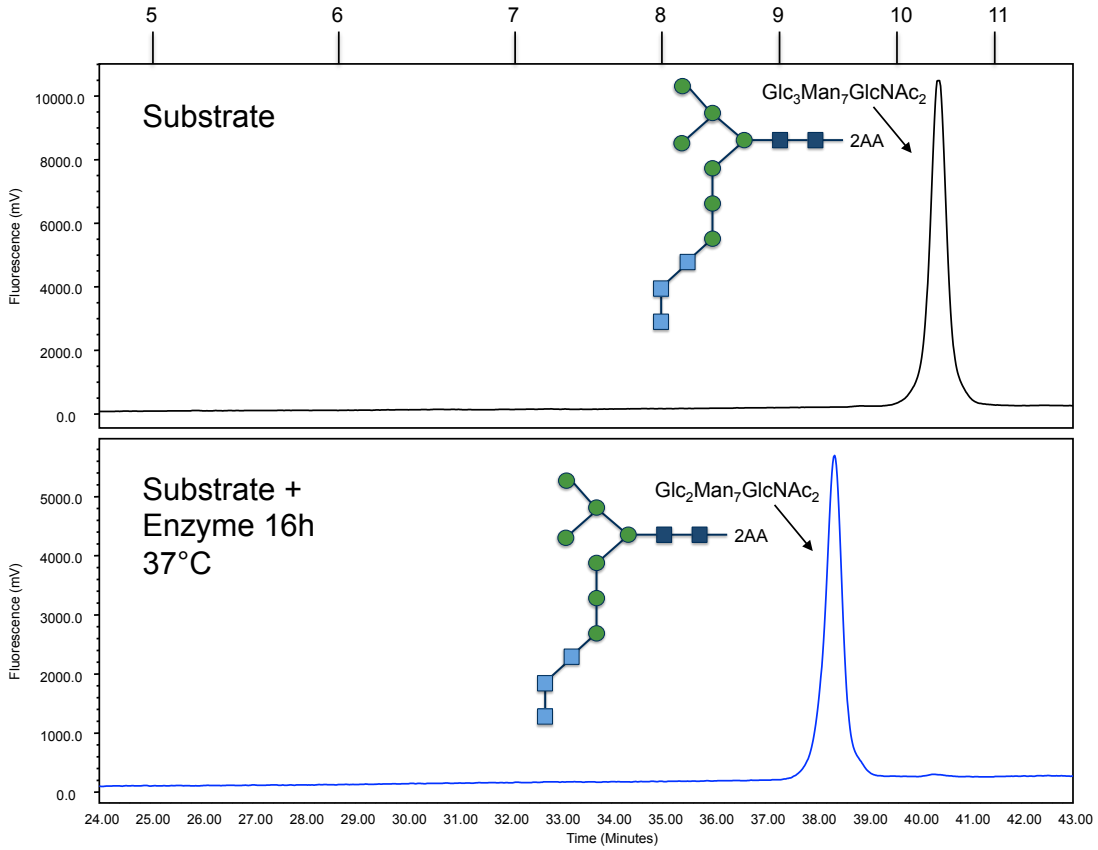


Figure 3.8: Cleavage of fluorescently labelled glycans by *Mm*GluI. Normal phase HLPC traces of $Glc_3Man_7GlcNAc_2$-2AA glycan before and after an overnight treatment with *Mm*GluI. The numbering above the chromatograms indicate glucose units.

3.7 Crystallisation

Use of the Department's nanolitre-dispensing robot allowed many commercial screens to be tested. All crystallisation screening attempts using the purified recombinant enzyme were duplicated after the addition of 1 mM *N*B-DNJ as the DSF experiments showed a marked increase in the stability of the protein upon inhibitor binding. The initial hits from several screens were not promising as only small clusters of needles

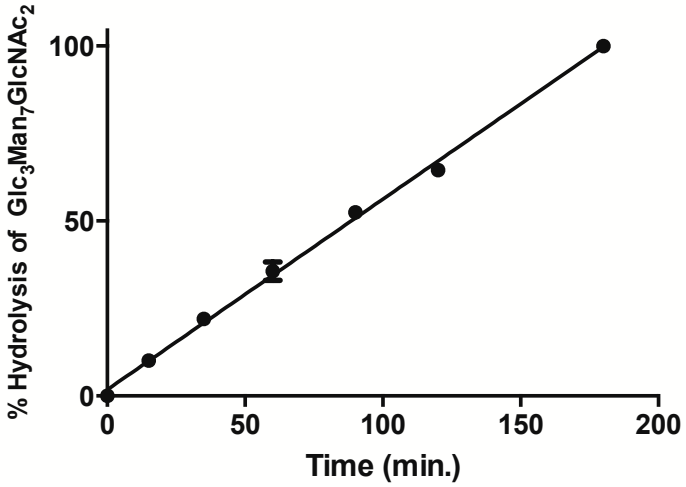


Figure 3.9: Time-course of the hydrolysis of fluorescently labelled $Glc_3Man_7GlcNAc_2$-2AA glycan over three hours with 5 µg/ml *Mm*GluI. Error bars represent the standard deviation from an experiment conducted in triplicate.

were observed. The most promising looking of these was from the MIDAS screen (Figure 3.10A), a screen that expands the crystallisation parameter space by utilising non-standard organic polymer precipitants [186]. Optimisation of the condition over the three components (polyacrylate 5100 concentration, ethanol concentration, and pH) yielded clusters of plates (Figure 3.10B). Careful manipulation of the crystals allowed these clusters to be prised apart and single, thin plates could be successfully mounted into nylon loops for X-ray diffraction measurements (Figure 3.10C).

3.8 X-ray diffraction and data processing

Crystals were screened for diffraction at the I04-1 beamline at DLS with the most promising crystal to date showing diffraction to ∼5 Å. A standard strategy was applied to collect a complete dataset from the crystal. The crystal exists in a face-centred monoclinic (C2) spacegroup with the dimensions and angles shown in Table 3.5. The integrated, scaled and merged dataset was cut at a resolution of 5.06 Å based on assessment of the mean I/σ above 1 and a cross correlation of the two halves of the

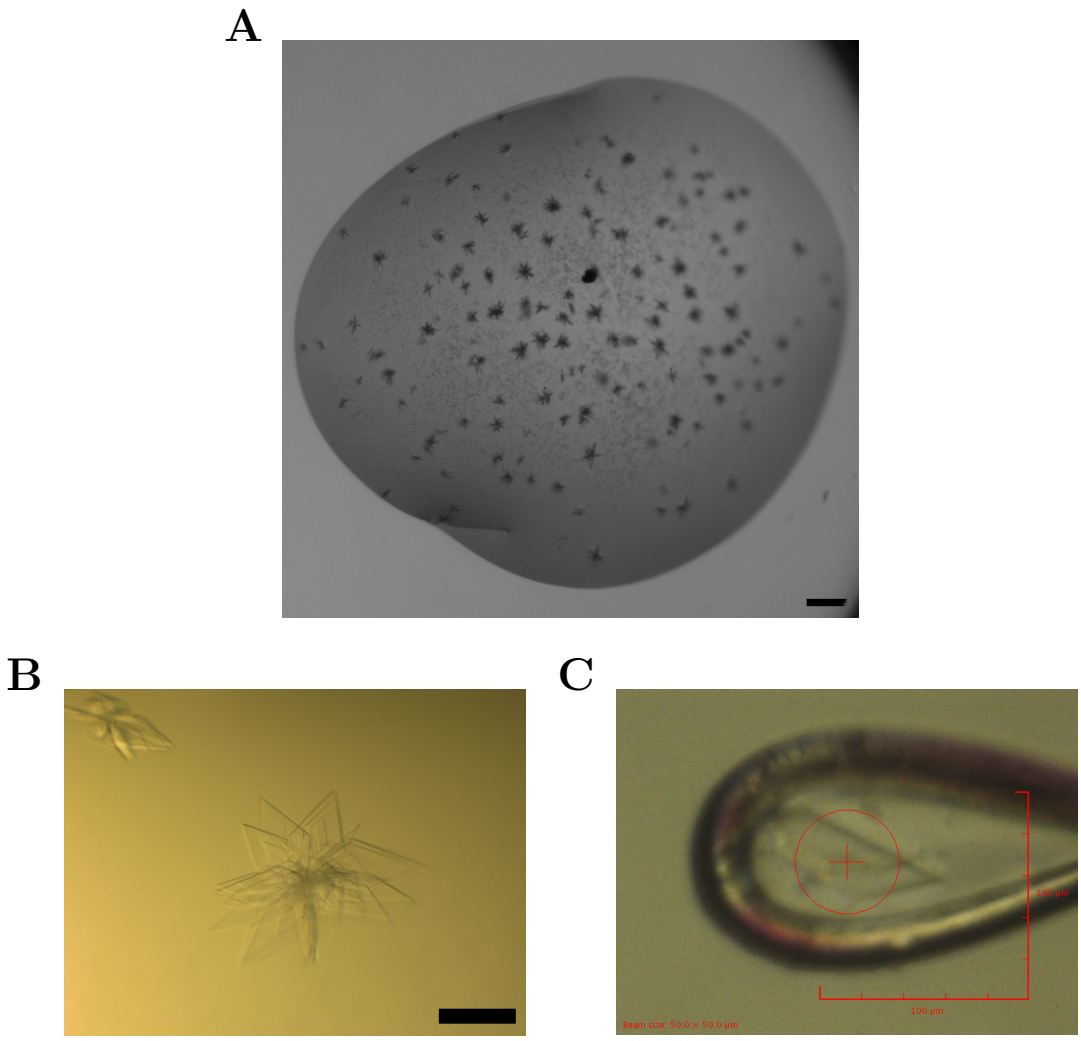


Figure 3.10: Photographs of *Mm*GluI crystals. (A) Initial hit from MIDAS screen. (B) Optimised crystallisation conditions produces clusters of plates in the drop. (C) An isolated plate mounted in a Nylon loop imaged on I04-1 at DLS. Scale bar indicates 100 µm.

dataset ($CC_{1/2}$) above 0.5 in the highest resolution shell. This dataset is complete in all resolution shells with a good degree of multiplicity. The quoted R_{meas} is quite high in the high-resolution shell and can also be seen in the R factor that is related to the precision of the measured reflections, the R_{pim}. However, the resolution cut-off can be justified by the $CC_{1/2}$ which is acceptable and is in line with the current debate on the best metric for data resolution [187]. Overall these statistics point to an overall low-resolution dataset characterised by weak diffraction that may hamper phasing and/or model building efforts.

Table 3.5: Diffraction data processing statistics

Resolution range (Å)	119.7-5.06 (5.33-5.06)
Cell dimensions:	
a, b, c (Å)	240.2, 104.7, 136.2
α, β, γ (°)	90, 94.6, 90
Spacegroup	C2
Total reflections	48223 (7097)
Unique reflections	14254 (2047)
R_{meas} [a]	0.42 (1.21)
R_{pim} [b]	0.29 (0.84)
$\langle I/\sigma(I)\rangle$	3.8 (1.4)
Completeness	99.8 (99.7)
Multiplicity	3.5 (3.4)
$CC_{1/2}$	0.956 (0.598)

Values in parentheses represent the high-resolution data shell

[a] $R_{meas} = \Sigma_h \Sigma_i |\sqrt{n_h/(n_h-1)}| I_{hi} - \langle I_h \rangle | \Sigma_h \Sigma_i I_h$

[b] $R_{pim} = \Sigma_h \Sigma_i |\sqrt{1/(n_h-1)}| I_{hi} - \langle I_h \rangle | \Sigma_h \Sigma_i I_h$

3.9 Phasing

To guess the number of copies of *Mm*GluI in the asymmetric unit (asu), the calculation of the Matthew's coefficient (V_M), the solvent content, and probability for each possibility were carried out (Table 3.6). Based on this analysis, the most likely number of molecules is 4 copies per asu but 3 copies could not be ruled out. *Mm*GluI has a sequence identity to the yeast enzyme of 33% which indicates a likelihood of structural homology. Thus the model for *Sc*GluI (PDB ID: 4J5T) was used as search model for molecular replacement. The initial model was prepared by passing it through Chainsaw, a CCP4 program that mutates and trims side chains from the model based on sequence alignment, removes insertion loops from the model and corrects the numbering. Initial searches were conducted searching for 3 copies. The search results were promising with translational NCS observed that is consistent with a peak at 0.34, 0, 0 in the Patterson map. However, the peak at 0, ϕ, 180 in the self-rotation function could never be explained by the result (Figure 3.11). Therefore, a solution has not been achieved to date but will be aided by optimisation of the crystallisation in order

to obtain better diffraction.

Table 3.6: Calculated Matthew's coefficients based on the number of molecules in the asymmetric subunit.

Molecules/asu	V_M (Å^3/Da)	% solvent	Probability
2	4.85	74.65	0.01
3	3.23	61.98	0.20
4	2.42	49.31	0.65

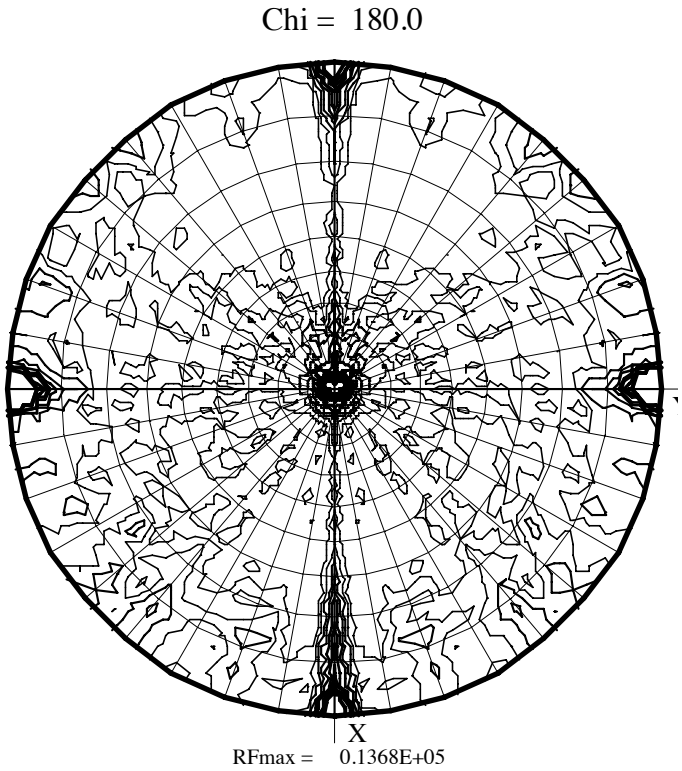


Figure 3.11: Self-rotation function of *Mm*GluI dataset at κ=180°.

3.9.1 Homology modelling

While the structural characterisation attempts were being carried out, a homology model was generated to try to assess the conservation in and around the active site. This was done using the Modeller set of scripts to generate structural alignments of the two GH63 domains that have been structurally characterised [188]. A structure-based sequence alignment was then carried out using the *Mm*GluI sequence which

allowed the utilisation of spatial restraints to generate a set of 25 models [189]. The generated models were ranked based on their energy scoring and the top-scoring ones were assessed visually. There were substantial variations in the scoring due to an incorrectly threaded loop in a chemically infeasible way in a proportion of the models. A comparison of the generated models to the two search templates shows agreement of the overall fold with differences occurring at a few of the loops (Figure 3.12 left panel). A close-up of the active site for the best scoring model and the *Sc*GluI template shows a good agreement of the active site lining residues but a key difference in a loop proximal to the active site (Figure 3.12 right panel).

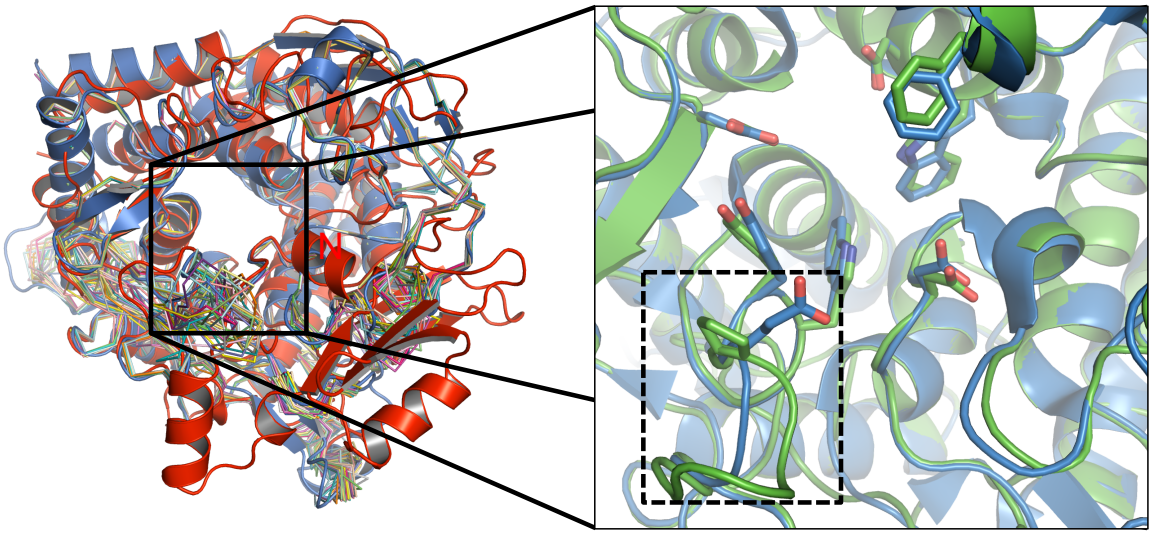


Figure 3.12: Homology modelling of the *Mm*GluI GH63 domain. Modelling carried out using the GH63 domain containing PDBs 4J5T (blue) and 3D3I (red). 25 generated models represented in fine ribbons. A close up of the active site of 4J5T and the best scoring model (green) highlighting the residues lining the active site. A difference in a loop between the template and generated model is highlighted in the dashed box.

3.10 Discussion

The work presented in this chapter represents a significant step toward the first mammalian GluI structure. The success of the cloning and screening campaign has allowed many avenues to be pursued in the further biochemical, biophysical, and structural

characterisation of GluI.

The production of the 24 constructs presented in this chapter was fast but the following experiments for screening of expression were flawed leading to a protracted time-frame to production of large quantities of any of the GluI constructs. No signal in the western blot analyses was observed for any of the 24 constructs despite the hexahistidine-tagged controls performed as expected. As this indicated no expression in the systems tested at the OPPF or in *P. pastoris*, re-evaluation of construct choices was considered. However, *Mm*GluI was eventually identified as the only expressing construct in HEK293-F. Upon scaling to larger cultures, this showed the weakness in the initial screening methodology. The inability to detect GluI by western blotting is due to it not being able to transfer efficiently to either nitrocellulose or PVDF membranes. In conclusion, the screening effort highlights the possible plight of screening for expression based on a single methodology. If the behaviour of the target is not fully understood, it could be fraught with false negatives. At the time the screen was carried out, an efficient and scalable orthogonal assay like that described in the following chapter for *Mm*GluII was not available; an activity assay may have identified other suitable *Mm*GluI constructs in other expression systems.

The purification of *Mm*GluI was straight-forward with the only factor optimised being the presence of a reducing agent throughout the purification. This was in order to reduce the chance of intermolecular disulphides forming. The apparent dimeric behaviour of *Mm*GluI when separating with the GF matrix, Superdex 200, is a novel observation. The only other reported cases vary from a tetrameric form for *Bos taurus* GluI purified from bovine mammary glands to only a monomer for *Sc*GluI [170, 176]. It is very unlikely that this dimerisation is caused by covalent bonding between the two molecules as is evident by both the SDS-PAGE and SEC-MALS analyses. The re-establishment of two peaks observed during purification when each of

those peaks are applied to the same separation matrix in the SEC-MALS experiment suggests a concentration-dependent phenomenon. Whether there is any physiological relevance to this observation or whether it is purely an irrelevant *in vitro* artefact has yet to be further investigated. Detailed characterisation of this interface might shed light on its potential physiological significance.

The availability of larger amounts of *Mm*GluI has now allowed many different avenues to be pursued in studying the interaction of this enzyme with both substrates and inhibitors using biophysical methods. The ability of a known inhibitor of this enzyme, *N*B-DNJ, to increase the thermal stability of *Mm*GluI comes as no surprise. Indeed the effect of inhibitors on the change in T_m of enzymes has been well characterised [190, 191]. However, in order to determine accurate K_d values using this technique, the thermodynamic properties of both the unfolding and binding events must be known [179, 192]. As a higher-throughput method for quickly comparing inhibitor binding, it is a practical method. The DSF platform can now be used to perform a buffer screen to ascertain if the current buffer conditions can be optimised, as a greater success in crystallising proteins has been correlated to a buffer-induced increase in stability[178, 193]. Analysis of the thermodynamics of iminosugar binding would be invaluable in understanding whether binding occurs in a similar manner to that of other hydrolases and also as an investigation of the effect of DNJ alkylation. The quantities of enzyme required for techniques like isothermal titration calorimetry are now available, making the biophysical characterisation of these binding events possible.

The confirmation of the activity of *Mm*GluI utilises a well-established and sensitive assay [194]. Furthermore, as the substrate used in this assay resembles the physiological one, it verifies the identity of *Mm*GluI as the correct enzyme. However, this assay lacks throughput, a quality desirable for inhibitor development. The ability

to produce this enzyme in large amounts allows future investigations into alternative high-throughput assays.

The poor quality of the *Mm*GluI crystals remains the main hurdle for the structural characterisation at high-resolution. One of the major challenges faced with all of the initial hits from the crystallisation screening was the formation of clusters of needles. A number of strategies have been attempted to suppress the nucleation, including use of oils over the drops or reservoirs and additive screening to search for molecules that promote crystal growth in a more regular manner. Two methods that may improve crystals are through modification of the surface of the protein: reductive methylation of lysine residues is commonly used to increase the likelihood of protein-protein interactions by introducing some hydrophobicity to the surface of the protein [195]. Another strategy along similar lines is to reduce surface crystallisation entropic costs by mutating long-side-chain residues on the surface [196]. Correlation of the homology models and the crystal contacts might provide a means of introducing mutations to promote better crystal contacts that would lead to better diffraction quality and resolution. There may also be inherent issues with the construct that may require optimisation of its borders or removal by limited proteolysis of loops that may be hindering uniform crystal growth.

The recent structure of *Sc*GluI that was published during the course of this work elucidated the overall architecture of the enzyme and identified the catalytic residues [61]. The *in silico* calculations of the possible modes of binding of inhibitors and small fragments of the *N*-glycan was also performed in that study. However, it remains to be seen if this is an adequate model for the mammalian enzymes. It is argued by the authors of that study that the sequence at and around the active site is highly conserved, though one of the highlighted differences is at E707 (*Sc*GluI numbering), which shows a change to an arginine in the human, rat, and mouse

sequences. In the homology model, two loops proximal to the active site including this arginine-containing one do not match the alignment of that in the study. Whether there is any significance in the differences in these loops is difficult to comment on in otherwise very similar active sites. If the differences cause changes to substrate binding towards the core of the glycan, a side-by-side comparison of the kinetics of hydrolysis would likely help to answer these questions. The validity of the homology modelling can be questioned when the degree of homology is tenuous, but in this case, it is likely to be justified by the high conservation of this fold and its function. The differences between the *Sc*GluI structure and the homology model in this loop are likely to be small, but in the absence of biochemical or structural experiments the homology model remains speculative at the moment.

The *Sc*GluI structure has enabled the *in silico* probing of binding of the glucotriose component of the *N*-glycan. The structure hints at the possible reason why the aryl-glucosides (*p*-nitrophenyl α-glucopyranoside and 4-methylubelliferyl α-glucopyranoside) are not cleaved by this enzyme. The predicted angle formed by the trisaccharide is very acute, thus this may be a possible explanation of why these unnatural substrates may not be accommodated in the active site. The structure of the *N*-glycan shows an acute angle imposed by the α(1,2) linkage between the two terminal glucose residues [62]. A similar mode of binding must be adopted in mammalian GluI enzymes, too. A mammalian GluI crystal structure with the glycan or a reduced oligosaccharide would be invaluable in proving this mode of binding and would have implications in the design of selective inhibitors.

4

Biophysical and biochemical characterisation of Endoplasmic Reticulum α-Glucosidase II

Very few studies have been published describing the recombinant production of glucosidase II (GluII), thus at the beginning of this work, there was little information that provided a solid starting point for design of constructs and recombinant production of this enzyme. Therefore, a broad approach was applied to generate a range of

constructs to test for expression in a number of heterologous expression systems. Initial trials of isolating GluII from rat or mouse livers – a method which has been used by multiple groups previously[118, 173, 197] – gave poor yields and not very pure protein. As these criteria are important for the techniques presented in this dissertation, a recombinant approach was necessary. As presented in the previous chapter, a high-throughput platform available at the Oxford Protein Production Facility (OPPF) enabled a quick turnaround of cloning and screening for suitable expression of 24 constructs in parallel.

4.1 Construct design

GluII is comprised of two subunits that are encoded by the two genes glucosidase alpha neutral AB (GANAB) and protein kinase C substrate 80K homology (PRCKCSH). As the campaign was limited to 24 constructs, this figure was split into two, 12 constructs for each of the two subunits. A number of variables were considered for testing. These included species, construct boundaries, affinity tags, expression location, and expression systems. A combination of four species, three affinity tags and both intra- and extracellular expression was chosen. The choice at OPPF was between expression in mammalian cells, insect cells, or *E. coli*. Previous unsuccessful attempts at expressing eukaryotic proteins in bacteria in this institute dissuaded us from trying expression in *E. coli*.

The choice of the four different species was made based on a few criteria. Human and murine sequences were chosen because of the clinical relevance of the two species in the drug development pipeline of the antivirals being researched in our group. The high sequence identity (>85%) between human and mouse genes (Table 4.1) indicates a very high degree of homology. The *Caenorhabditis elegans*

orthologue was chosen based on a high sequence homology to the human proteins (46 and 37 % for α- and β-subunits, respectively) and a more thorough phylogenetic and biochemical study of the nematodes[198]. This suggests a high probability of structural homology between nematode and mammalian enzyme. The *C. elegans* GANAB and PRKCSH orthlogues are also slightly smaller and seem to have less extensive loops when compared in multiple alignments possibly indicating a more amenable target for crystallisation. A natural choice for a fungal orthologue would have been that from *Schizosaccharomyces pombe* due to an abundant amount of biochemical and genetic literature on that enzyme. However, the enzyme from *Saccharomyces cerevisiae* was chosen instead due to its still high apparent homology and its slight sequence divergence in the β-subunit, which could increase the chances of success.

Table 4.1: Percentage identity matrix of the GANAB and PRKCSH gene products of selected organisms carried out by Clustal Omega.

GANAB	*H. sapiens*	*M. musculus*	*C. elegans*	*S. cerevisiae*
H. sapiens	100			
M. musculus	92	100		
C. elegans	46	46	100	
S. cerevisiae	38	38	36	100
PRKCSH	***H. sapiens***	***M. musculus***	***C. elegans***	***S. cerevisiae***
H. sapiens	100			
M. musculus	87	100		
C. elegans	37	37	100	
S. cerevisiae	27	26	27	100

As the α/β interface has been associated with two distinct regions of the β-subunit [86], but not anywhere specific for the α-subunit, full-length constructs were prioritised. The construct boundaries were kept at what has been annotated as the mature sequences in the sequence databases based on both experimental and bioinformatical evidence. Disorder prediction based on sequence is not discouraging for the α-subunit, but more so for the β-subunit (Figure 4.1). A region central to the

β-subunit appears likely to be disordered and corresponds to a large stretch of acidic residues. Nevertheless, the full-length β-subunit was pursued, in case the binding to the α-subunit orders the putatively disordered regions .

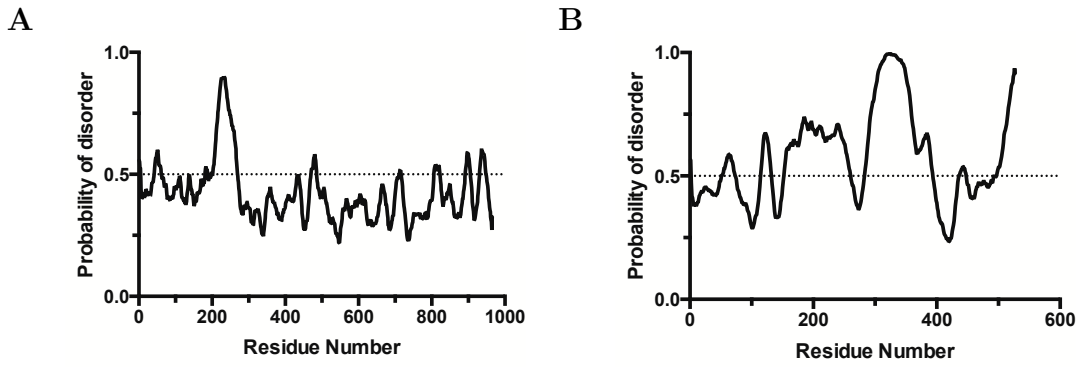


Figure 4.1: Disorder prediction of the gene products of the *H. sapiens* (A) GANAB and (B) PRKCSH using the RONN server[145]. Values above 0.5 predict disordered regions of the protein.

The pOPIN vectors used all contain C-terminal hexahistidine affinity tags, but as the aim was to obtain a heterodimer, the α-subunit constructs also contained an extra affinity tag. The expression location was the final consideration. As both intracellular and secreted expression were attempted, the intracellular constructs retained their native sequences. Two different secretion signals were attempted, the μ-phosphatase (in the pOPING and pOPINGS vectors) and the haemagglutinin sequences (in the pOPINHA and pOPINHA-FLAG vectors). The final 24 constructs and their reference numbers are detailed in Table 4.2.

Therefore, a total of 24 constructs were designed with the following parameters:

Construct length	mature full-length sequences (after predicted signal sequences)
Species	*H. sapiens*, *M. musculus*, *C. elegans*, *S. cerevisiae*
Affinity tag	Hexahistidine, StrepII, FLAG, HALO7
Expression location	Intracellular, Secreted
Expression system	Insect cells, Mammalian cells, *P.pastoris*

Table 4.2: Glucosidase II expression constructs.

Species	Gene	Construct reference	Vector	Construct boundary (aa numbering)
C. elegans	GANAB	A4	pOPINE-3C-HALO7	2-925
		B4	pOPINGS	22-925
		C4	pOPINHA-FLAG	22-925
	PRKCSH	D4	pOPINE	2-508
		E4	pOPING	23-504
		F4	pOPINHA	23-504
H. sapiens	GANAB	G4	pOPINE-3C-HALO7	2-967
		H4	pOPINGS	33-967
		A5	pOPINHA-FLAG	33-967
	PRKCSH	B5	pOPINE	2-529
		C5	pOPING	15-525
		D5	pOPINHA	15-525
M. musculus	GANAB	E5	pOPINE-3C-HALO7	2-967
		F5	pOPINGS	33-967
		G5	pOPINHA-FLAG	33-967
	PRKCSH	H5	pOPINE	2-522
		A6	pOPING	15-518
		B6	pOPINHA	15-518
S. cerevisiae	GANAB	C6	pOPINE-3C-HALO7	2-955
		D6	pOPINGS	23-955
		E6	pOPINHA-FLAG	23-955
	PRKCSH	F6	pOPINE	2-703
		G6	pOPING	24-703
		H6	pOPINHA	24-703

4.2 Cloning

The initial PCR results were highly successful (Figure 4.2), producing correctly sized PCR products on the first attempt. Following the purification, In-Fusion mediated Gibson assembly into the chosen pOPIN vectors was carried out. The overall result was successful insertion of the 24 constructs into the pOPIN vectors as assessed by blue/white selection[199] coupled with an analytical PCR utilising a gene and a vector specific oligonucleotide to confirm the presence of the correct construct.

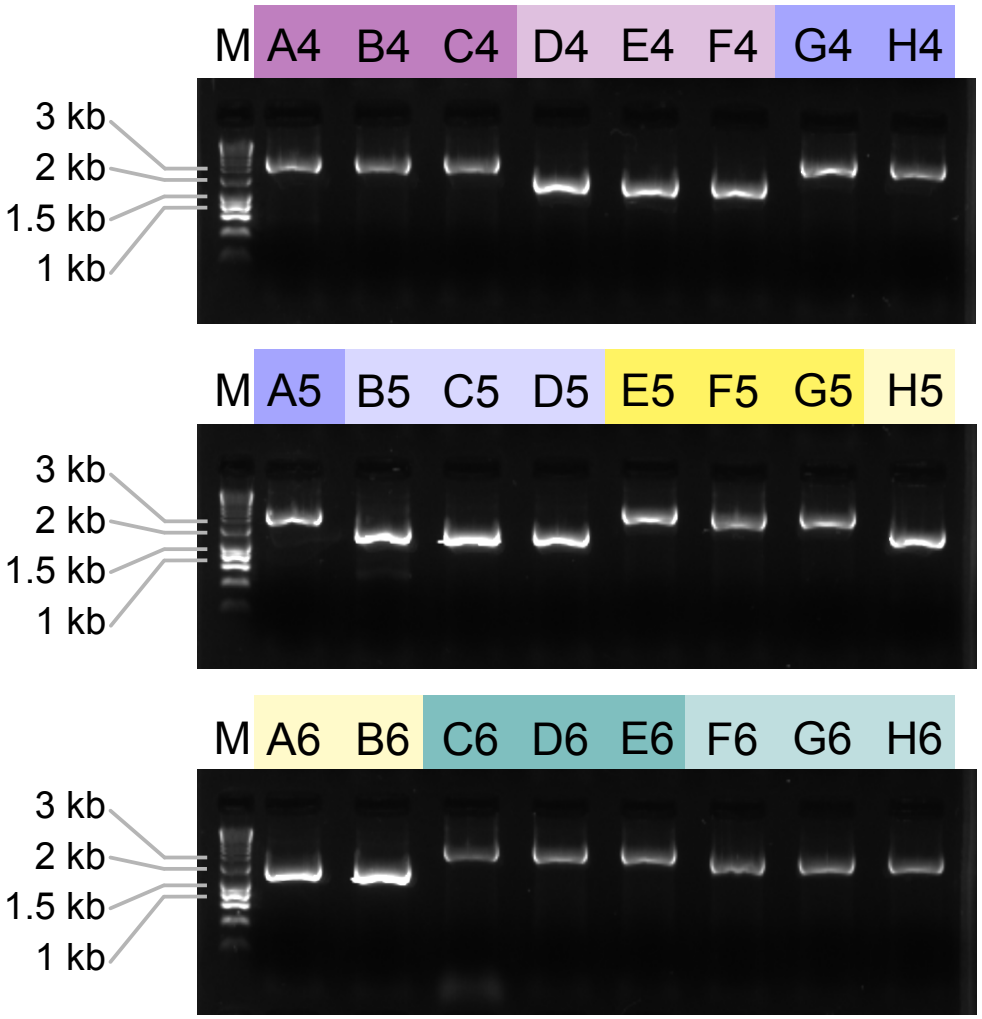


Figure 4.2: Agarose gel of products from high-throughput PCR reaction. 'M' indicates a HyperLadder 1kb marker with selected bands indicated in kilobases (kb).

4.2.1 Mutagenesis

Site directed mutagenesis (SDM) directly in the pOPIN vectors was unfortunately not possible, as very low SDM success rates in these vectors were anecdotally reported by the staff at the OPPF facility and many users of the mother vector, pHLsec. Therefore, all mutagenesis had to be undertaken in a cloning vector, in this case Litmus28i. The routine sub-cloning into Litmus28i of the α-subunit (*M. musculus ganab*) was followed by the SDM with the NEB kit. Upon confirmation of successfully mutated clones by dye-terminator Sanger sequencing (Figure 4.3), the mutants were sub-cloned back into an expression vector, pHLsec. Although the D640E mutant looks successful, an error in the priming of the forward mutagenesis oligonucleotide caused an insertion of 119 bp immediately after the sequence of the primer. This mutant was not carried forward into expression trials.

4.3 Expression screening

Screening for expression was again carried out in a high-throughput fashion. The main assessment of successful expression was by western blot analysis using a HRP-conjugated antibody against affinity tags, primarily against the C-terminal hexahistidine tag present on all constructs. GluII exists as a heterodimer so both single expression and co-expression was tested in case the subunits are unable to be produced in isolation from one another.

4.3.1 Insect cell expression

The constructs for secreted expression of both the singly infected and co-infected cells after 72 hours show consistent expression of the β-subunit and for the *S. cerevisiae*

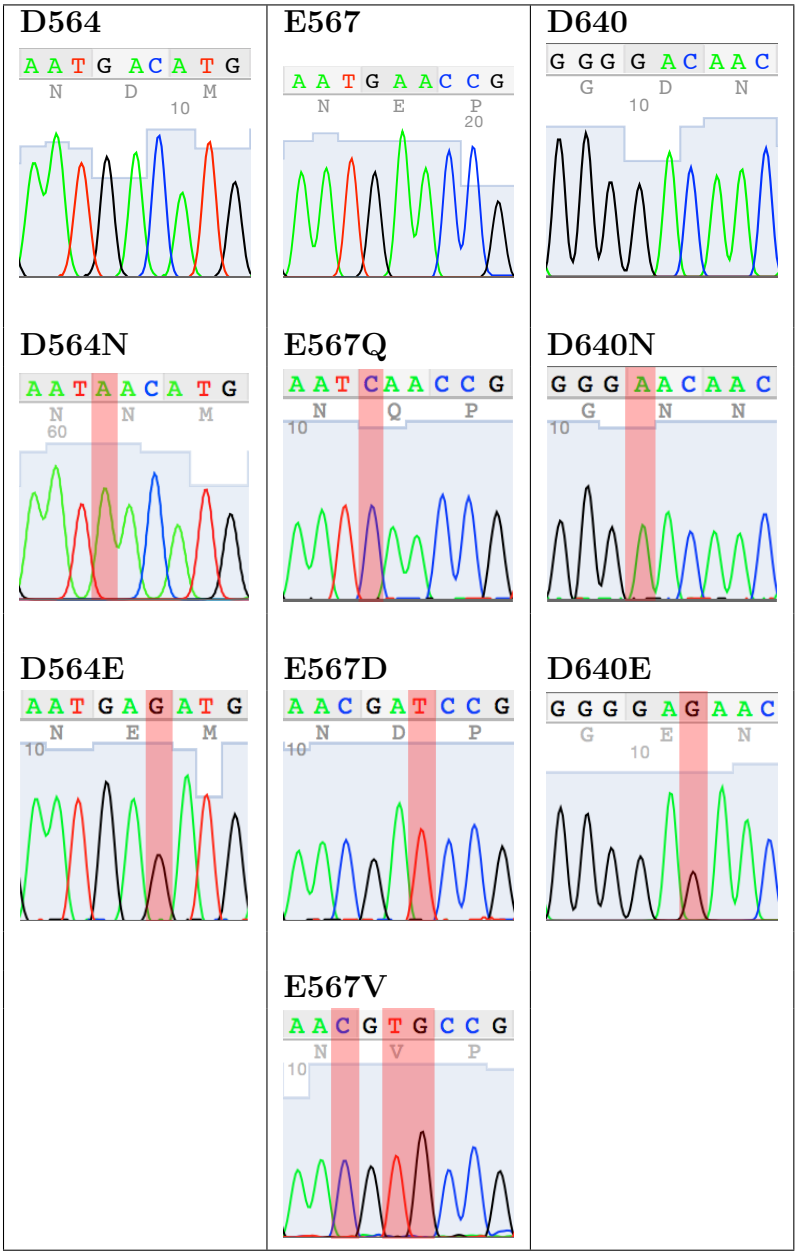


Figure 4.3: Capillary electrophoretograms of dye-terminator Sanger sequencing of *Mm*GluII α-subunit point mutants compared to the wild-type sequencing. Red boxes indicate the introduced mutations.

α-subunit for two of the three constructs (Figure 4.4 A and B, respectively). Despite constructs D4, B5, H5 and F6 lacking secretion signal sequences, these samples appeared in the supernatant as well, possibly as a result of cell death and lysis. These same constructs showed intracellular expression that is visible by Coomassie Brilliant Blue staining upon enrichment with small-scale IMAC (Figure 4.4C).

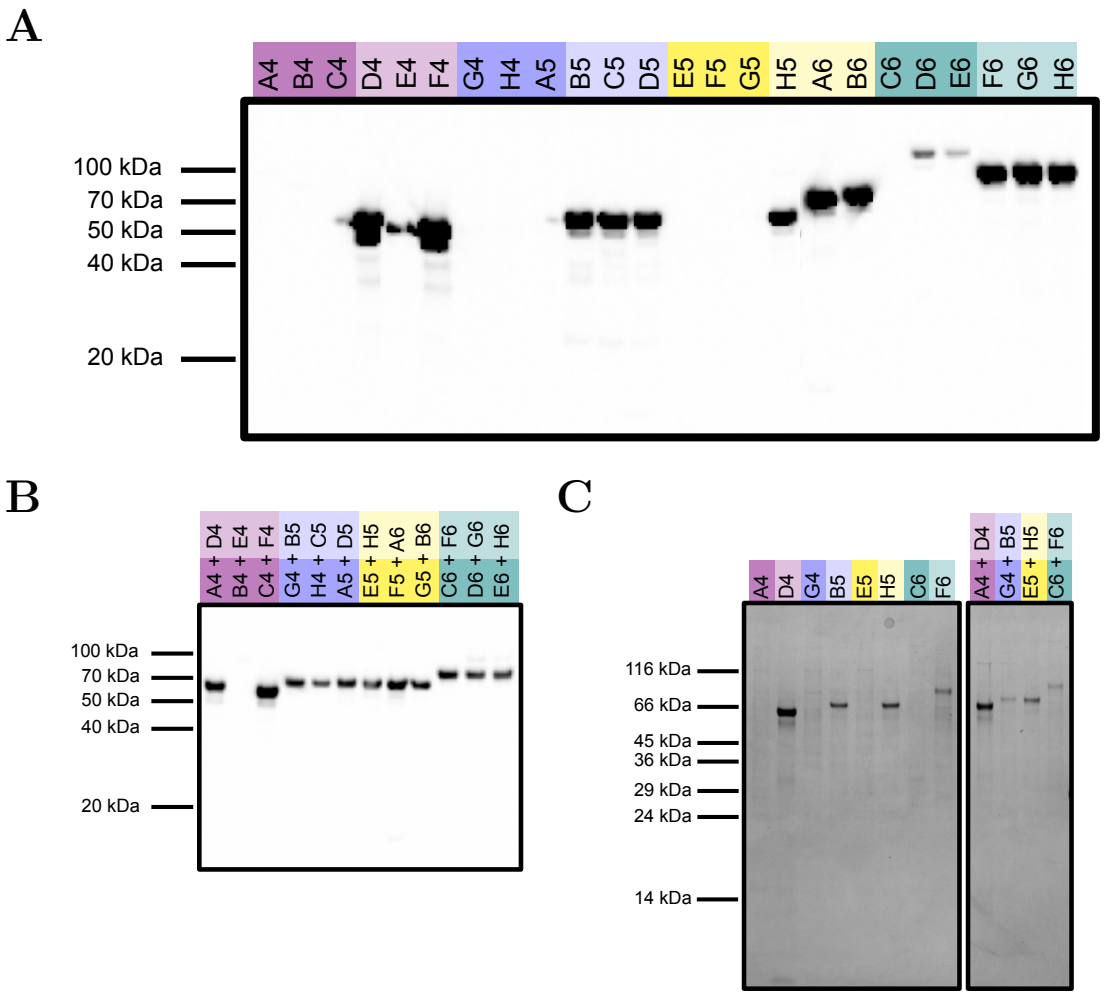


Figure 4.4: Analysis of baculovirus-infected *Sf*9 cells with glucosidase II constructs. Western blot analysis using an anti-His_6-HRP antibody of (A) single or (B) co-infections. (C) SDS-PAGE analysis of intracellular expression purified by small-scale IMAC.

4.3.2 Mammalian cell expression

The final aspect of the high-throughput screening was transfection into HEK293T cells (Figure 4.5). Overall, the secreted expression of the constructs followed the trend seen in the insect cells in the previous section.

After the partial success of expressing the α-subunit in the high-throughput campaign in *Sf*9 or HEK293T cells, a final attempt was made in HEK293F cells, a suspension cell line that the manufacturers claim can produce five times as much secreted protein than adherent HEK293 cell lines. To improve chances of identifying

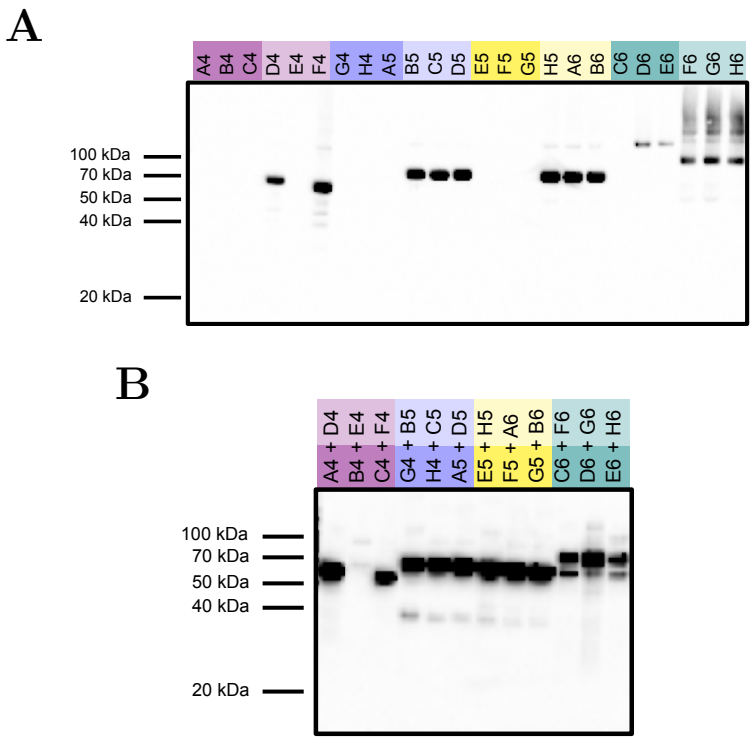


Figure 4.5: Analysis of the transfections of HEK293T cells with glucosidase II constructs. Western blot analysis using an anti-His_6-HRP antibody of (A) single or (B) co-infections.

successfully expressing constructs, a fluorescence-based activity assay for the cleavage of 4MUG was included in the analysis. Similarly to all the previous screens, the β-subunit expresses well as does the *S. cerevisiae* α-subunit (Figure 4.6 A and C). One of the *C. elegans* constructs, C4, has a very strong band that does not correspond to the full-length construct but is below 80 kDa indicating some proteolysis from the N-terminus and very low activity against 4MUG. Due to the lower-throughput of the tissue culture facility available in this institute, the maximum reasonable amount of cultures to be simultaneously assessed was limited to about 16. The co-transfection experiment was limited to 5 pairs of constructs in order to simultaneously attempt transfection of a larger $\alpha : \beta$ ratio as the α-subunit has been the more challenging subunit to express. The western blot of this experiment shows only a slightly visible band corresponding to the α-subunits of *M. musculus* (F5) and *S. cerevisiae* (D6) (Figure 4.6B orange boxes). However, the activity assay is more telling with very

strong signal for these samples (Figure 4.6D).

4.3.3 Summary

The expression screening is summarised in Table 4.3. The main result of the screening was that there are two successfully expressing constructs for producing the murine and fungal heterodimeric GluII. A multitude of different β-subunit constructs readily expressed. The expression trials in *P. pastoris* are not presented in detail in this thesis because they could not produce secreted α-subunit. The *C. elegans* α-subunit produced was similar to that seen in Figure 4.6A for construct C4 which was heavily proteolysed and difficult to access as it only expressed intracellularly. Moving forward, the prioritised construct was of the murine glucosidase II (*Mm*GluII) using the constructs F5 (α) and A6 (β).

4.4 Expression and purification

4.4.1 Expression optimisation of *Mm*GluII constructs

The initial results from the expression of *Mm*GluII indicated that the ratios of F5 and A6 that were transfected had an effect on the amount of secreted *Mm*GluII. An experiment was designed to assay the optimal ratio of the two constructs testing four different molar ratios from those in Figure 4.6B. The length of time after transfection was also assessed to find the optimal time to harvest the culture. The condition that produced most *Mm*GluII as assessed by 4MUG cleavage was at 7 days with the 4:1 F5:A6 ratio, but as the 1:1 mole ratio produced a similar amount of activity in a shorter time frame (Figure 4.7A), this ratio was used for all subsequent co-transfections. The 1:1 mole ratio corresponds to a 1.2:1 F5:A6 mass ratio. An-

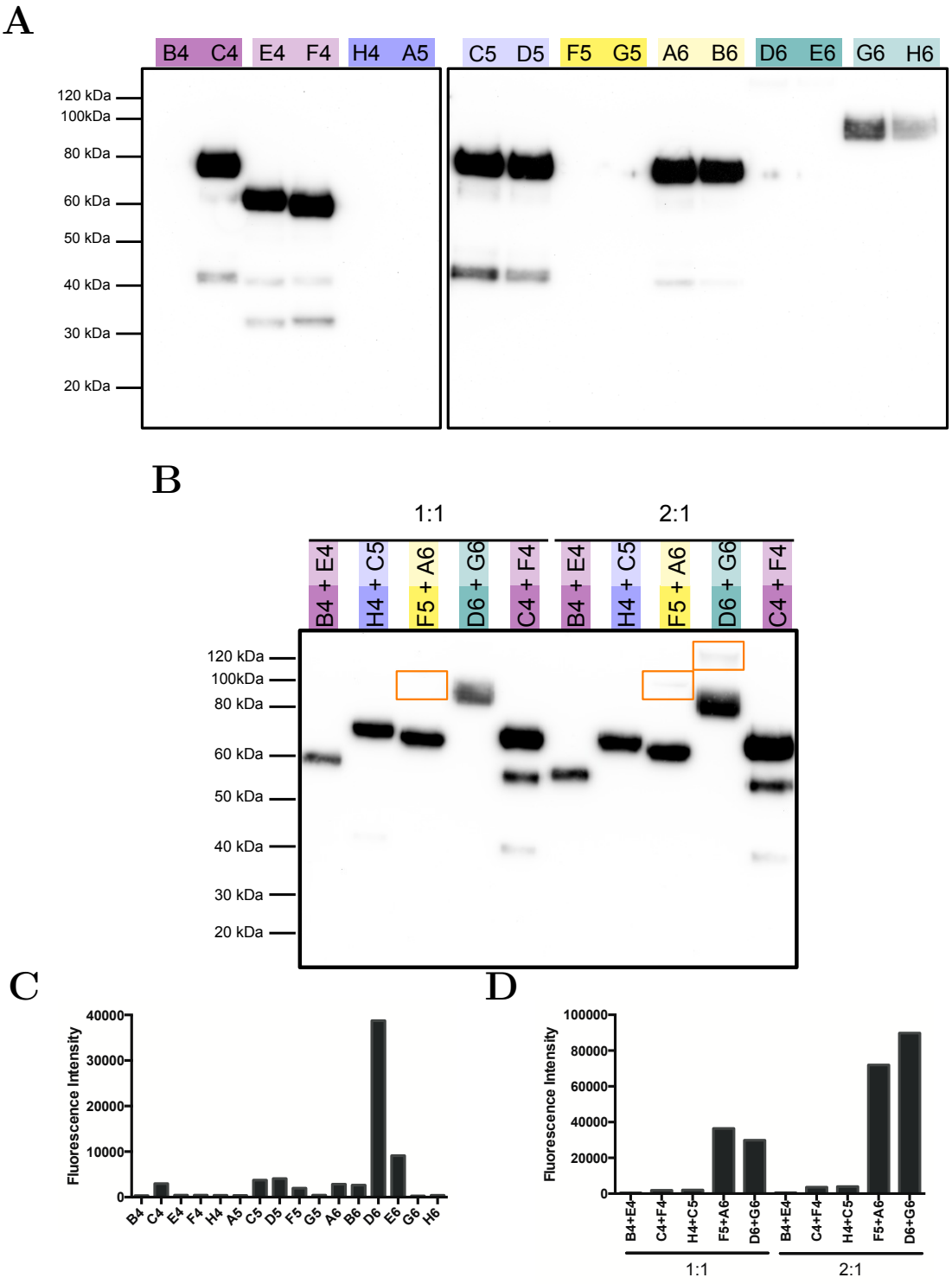


Figure 4.6: Analysis of the transfections of HEK293F cells with glucosidase II constructs. Western blot analysis using an anti-His_6-HRP antibody of (A) single or (B) co-infections. Orange boxes highlight weak bands for α-subunit expression. Cleavage of 4MUG substrate as analysed by a discontinuous fluorescence measurement of (C) single or (D) co-infections. Co-transfections that indicate ratios refer to mass ratios of the $\alpha : \beta$ DNAs that were transfected.

Table 4.3: Summary of expression screening for Glucosidase II

Species	Gene	Construct reference	Protein	*P. pastoris*	Insect cells		Mammalian cells	
					Intracellular	Secreted	HEK293T Secreted	HEK293F Secreted
C. elegans	GANAB	A4	α-subunit	Not Tested	×	×	×	Not Tested
		B4		✓[a]	Not Tested	×	×	×
		C4		Not Tested	Not Tested	×	×	✓
	PRKCSH	D4	β-subunit	Not Tested	✓	✓	✓	Not Tested
		E4		✓	Not Tested	✓	×	✓
		F4		Not Tested	Not Tested	✓	✓	✓
H. sapiens	GANAB	G4	α-subunit	Not Tested	×	×	×	Not Tested
		H4		×	Not Tested	×	×	×
		A5		Not Tested	Not Tested	×	×	×
	PRKCSH	B5	β-subunit	Not Tested	✓	✓	✓	Not Tested
		C5		✓	Not Tested	✓	✓	✓
		D5		Not Tested	Not Tested	✓	✓	✓
M. musculus	GANAB	E5	α-subunit	Not Tested	×	×	×	Not Tested
		F5		×	Not Tested	×	×	×
		G5		Not Tested	Not Tested	×	×	×
	PRKCSH	H5	β-subunit	Not Tested	✓	✓	✓	Not Tested
		A6		✓	Not Tested	✓	✓	✓
		B6		Not Tested	Not Tested	✓	✓	✓
S. cerevisiae	GANAB	C6	α-subunit	Not Tested	×	×	×	Not Tested
		D6		×	Not Tested	✓	✓	✓[a]
		E6		Not Tested	Not Tested	✓	✓	✓[a]
	PRKCSH	F6	β-subunit	Not Tested	✓	✓	✓	Not Tested
		G6		✓	Not Tested	✓	✓	✓
		H6		Not Tested	Not Tested	✓	✓	✓
C. elegans	GANAB+PRKCSH	A4+D4	$\alpha + \beta$	Not Tested	β only	β only	β only	Not Tested
		B4+E4		Not Tested	Not Tested	×	×	β only
		C4+F4		Not Tested	Not Tested	β only	β only	β only
H. sapiens	GANAB+PRKCSH	G4+B5	$\alpha + \beta$	Not Tested	β only	β only	β only	Not Tested
		H4+C5		Not Tested	Not Tested	β only	β only	β only
		A5+D5		Not Tested	Not Tested	β only	β only	Not Tested
M. musculus	GANAB+PRKCSH	E5+H5	$\alpha + \beta$	Not Tested	β only	β only	β only	Not Tested
		F5+A6		Not Tested	Not Tested	β only	β only	✓[a]
		G5+B6		Not Tested	Not Tested	β only	β only	Not Tested
S. cerevisiae	GANAB+PRKCSH	C6+F6	$\alpha + \beta$	Not Tested	β only	β only	✓	Not Tested
		D6+G6		Not Tested	Not Tested	✓	✓	✓[a]
		E6+H6		Not Tested	Not Tested	✓	✓	Not Tested

[a] Indicates activity against 4MUG

other variable to consider was the ratio of the transfection reagent to the amount of DNA. All initial experiments were carried out with the conservative 1:1:1 volume:mass(F5):mass(A6) ratio. The recommended ratio is 1.25:1 volume:mass ratio for a single transfection so this was tested with respect to the 1:1 ratio. As the much larger amounts of DNA added to the cells in the former experiment in this section did not seem very effective, it was hypothesised that this was deleterious to the cells due to the much larger volumes added to the culture. Thus a second variable was added to see if a smaller amount of DNA that was still in the same molar ratio was more effective. As hypothesised the lower amount of DNA caused a larger amount of *Mm*GluII to be produced (Figure 4.7B) with little difference from the contribution of the amount of transfection reagent for these two conditions.

4.4.2 Purification of full-length *Mm*GluII

The optimised transfection conditions are a scalable method of expressing *Mm*GluII in larger quantities. Utilising the two affinity tags, the hexahistidine and StrepII tags, allowed the quick and efficient purification of the heterodimer. Automated purification on an ÄKTA design chromatography system allowed a bulk separation of *Mm*GluII from the other proteins in the cell culture medium by IMAC (Figure 4.8A). As both subunits contain hexahistidine tags, a second affinity purification using the StrepII tag on the α-subunit was necessary to ensure that an excess of the β-subunit was avoided. Unfortunately, the affinity of the StrepII tag to the StrepTactin resin was relatively low and the *Mm*GluII began to elute before the washes were completed (Figure 4.8C). To maximise the yield, the later wash fractions were included in the pool of elution fractions carried forward. A final gel filtration step was employed to ensure monodispersity as evident by the symmetrical peak observed (Figure 4.8B). The final *Mm*GluII was pure with the exception of some slight endoproteolysis of

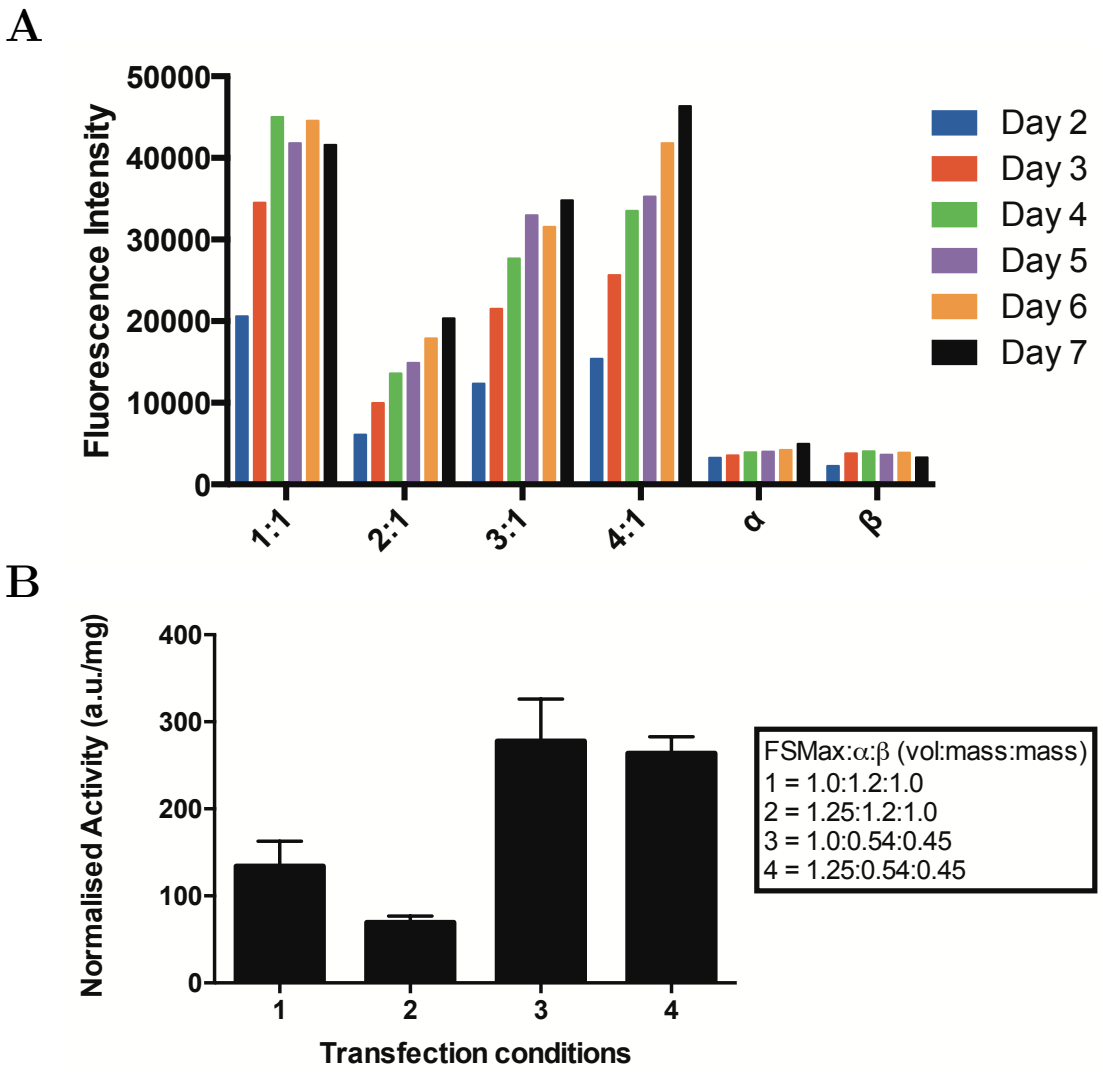


Figure 4.7: Optimisation of transfection conditions as assessed by cleavage of 4MUG substrate and analysed by a discontinuous fluorescence measurement. (A) Optimisation of the molar ratios of the α : β DNA at various time points during the transfection. (B) Optimisation of the FreeStyle MAX (FSMax) transfection reagent:DNA ratios. Activity as indicated after purification by IMAC. Error bars represent the standard deviation from three replicates.

the α-subunit being observed (Figure 4.8C). Typical final yield of purified *Mm*GluII is about 8 mg/l of culture. The presence of post-translational modifications has not been fully investigated especially as there is *N*-linked site on the α-subunit and two on the β-subunit which may have an effect on function or future crystallisation attempts.

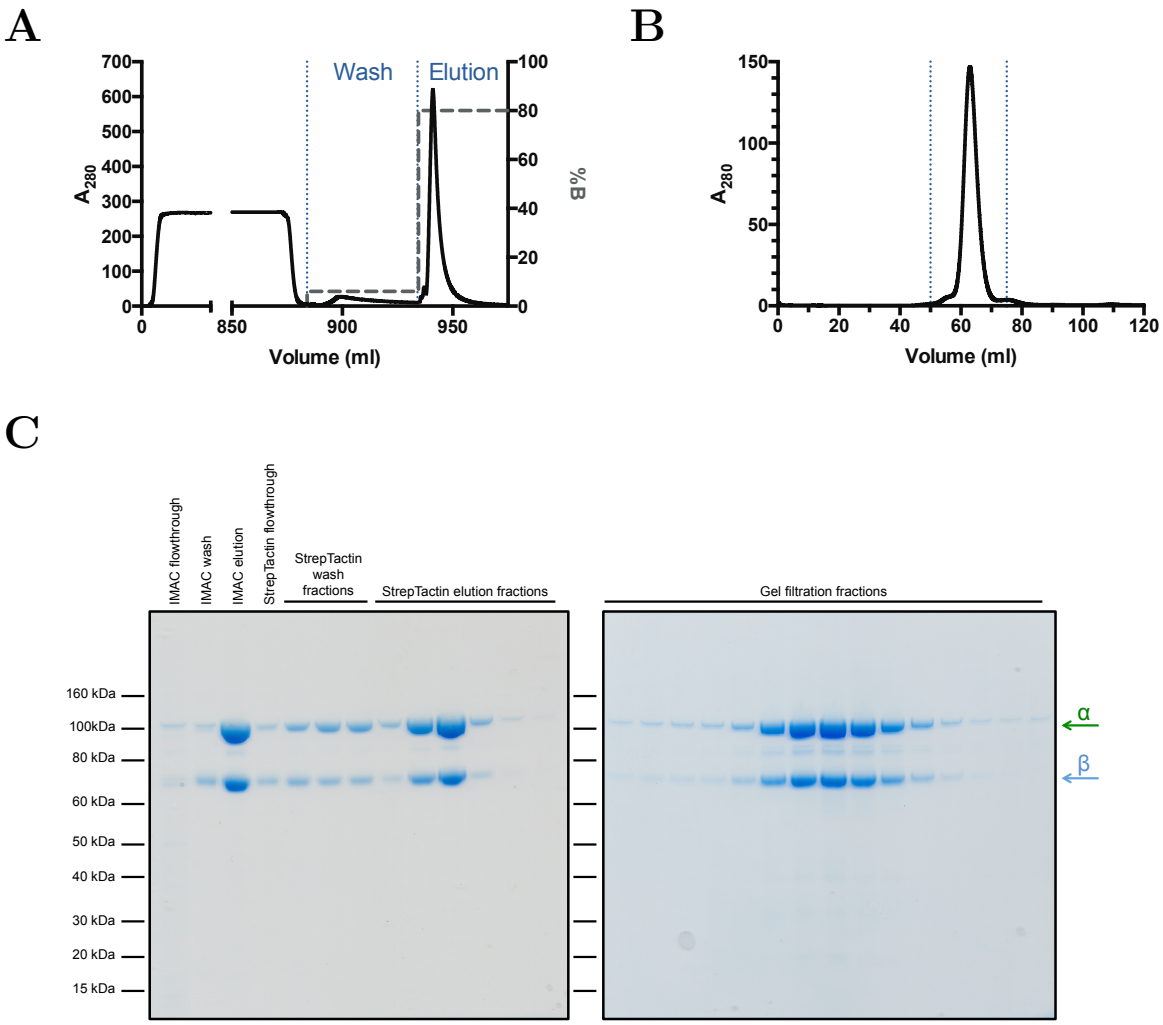


Figure 4.8: Purification of *Mm*GluII. (A) IMAC chromatogram of a purification on a 5 ml HisTrap excel column from 800 ml of cell supernatant where buffer B is PBS containing 500 mM imidazole. The fractions in (C) are indicated within the blue lines. (B) Following a StrepTactin purification, a gel filtration chromatogram on a Superdex 200 16/600 column. The fractions in (C) are indicated within the blue lines. (C) SDS-PAGE analysis of all fractions during a typical purification. Fractions from the StrepTactin purifications are 1 cV and 0.5 cV for the wash and elution, respectively.

4.4.3 Purification of full-length *Mm*GluII point mutants

The point mutants were designed based on the biochemical and structural precedents established for GluII and other GH31 family enzymes [69, 72, 91]. These included the three possible catalytic residues D564, E567 and D640. Prior to the availability of the first crystal structure in this family, the identity of these residues was in question and was purely based on sequence alignments. The authors hypothesised that the

highly conserved $D_{564}MNE_{567}$ sequence could indicate two acidic residues with the requisite ∼5.5 Å distance between these two residues. The advent of the series of GH31 family member crystal structures indicated that the $GD_{640}NTA$ sequence likely contains the catalytic acid. Thus all three of these residues were mutated to either the isosteric carboxamide equivalent (D→N or E→Q) or the equivalent acid (D→E and *vica versa*). A third variety was attempted at E567 to a valine as a member of the GH31 with an alternate lyase mechanism has been described which has this substitution [89].

Expression of the point mutants was carried out in an identical manner to the wild-type albeit on a smaller scale. The removal of the polyhistidine tag from the β-subunit allowed a more straight-forward one-step IMAC purification. A typical purification yielded pure *Mm*GluII (Figure 4.9) that was suitable for the subsequent biochemical and biophysical analyses. Any mutants that were carried forward for crystallisation attempts were subjected to gel filtration chromatography first.

4.5 Characterisation

4.5.1 Circular dichroism spectroscopy

Due to absorbance of the circularly polarised UV light of chloride ions in the buffer, all proteins were exchanged into 20 mM phosphate at pH 7.5. Alongside the heterodimer, two other forms were analysed, the β-subunit, which was purified in a similar manner to the heterodimer and *Mm*$GluII_{Trypsin}$, a shortened form of *Mm*GluII that is discussed further in Chapter 5. The spectra of these three proteins (Figure 4.10A) are overall similar in shape with none containing the archetypal all α-helical double minimum at 208 and 222 nm, all β-sheet single minimum at 218 nm, or the hallmark of a completely unfolded polypeptide, a minimum below 200 nm[200–202]. Therefore,

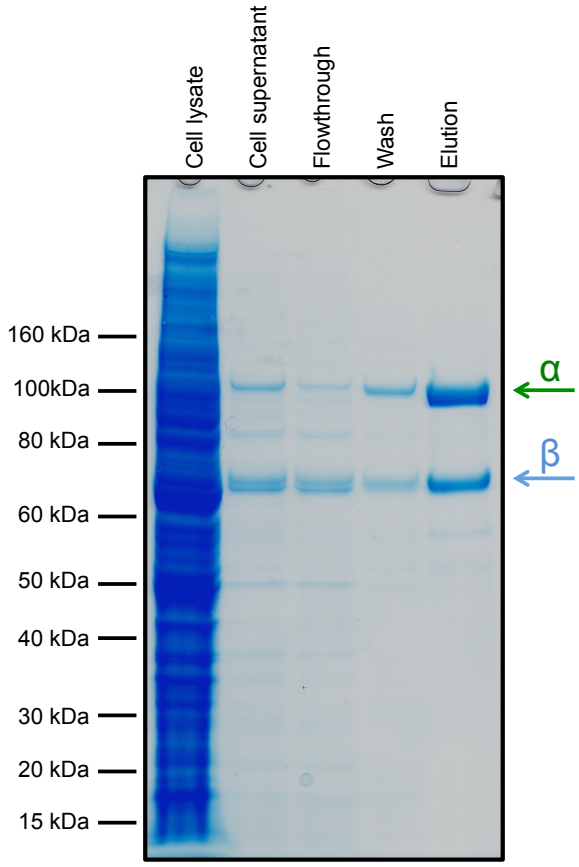


Figure 4.9: Example purification of *Mm*GluII point mutant. SDS-PAGE analysis of the IMAC purification of the D640N point mutant.

these proteins must contain a mixture of secondary structural elements. Deconvolution of the spectra was carried out with the CDSSTR algorithm accessible through the DICHROWEB web interface. CDSSTR uses singular-value decomposition based on a reference set of spectra to reconstruct the spectrum of interest[164]. The resulting deconvolution is presented in Table 4.4. To serve as a control, the spectrum of $Mm\text{GluII}_{\text{Trypsin}}$ was measured as the crystal structure is available (see Chapter 5) to use as a direct comparison. The agreement of the secondary structure elements is very good with the only main deviations being between the amount of α-helices and turns. The confidence in these results allows some conclusions from the spectra of the full-length heterodimer and the β-subunit to be drawn. The RONN analysis of the β-subunit in 4.1, points to a large region of disorder within the sequence.

The observed proportion of disorder with both of these samples (>40%) indicates a large contribution by the β-subunit that exists in solution hinting at likely difficulties in crystallisation attempts of both samples. As isolation of the α-subunit could not be achieved, it is difficult to precisely assign the secondary structure contributions of α and β subunits alone. However, the averaging of the β-subunit and the *Mm*GluII$_{\text{Trypsin}}$ structure, which only contains 20% of the β-subunit sequence, gives a convincing representation of the values observed for the heterodimer.

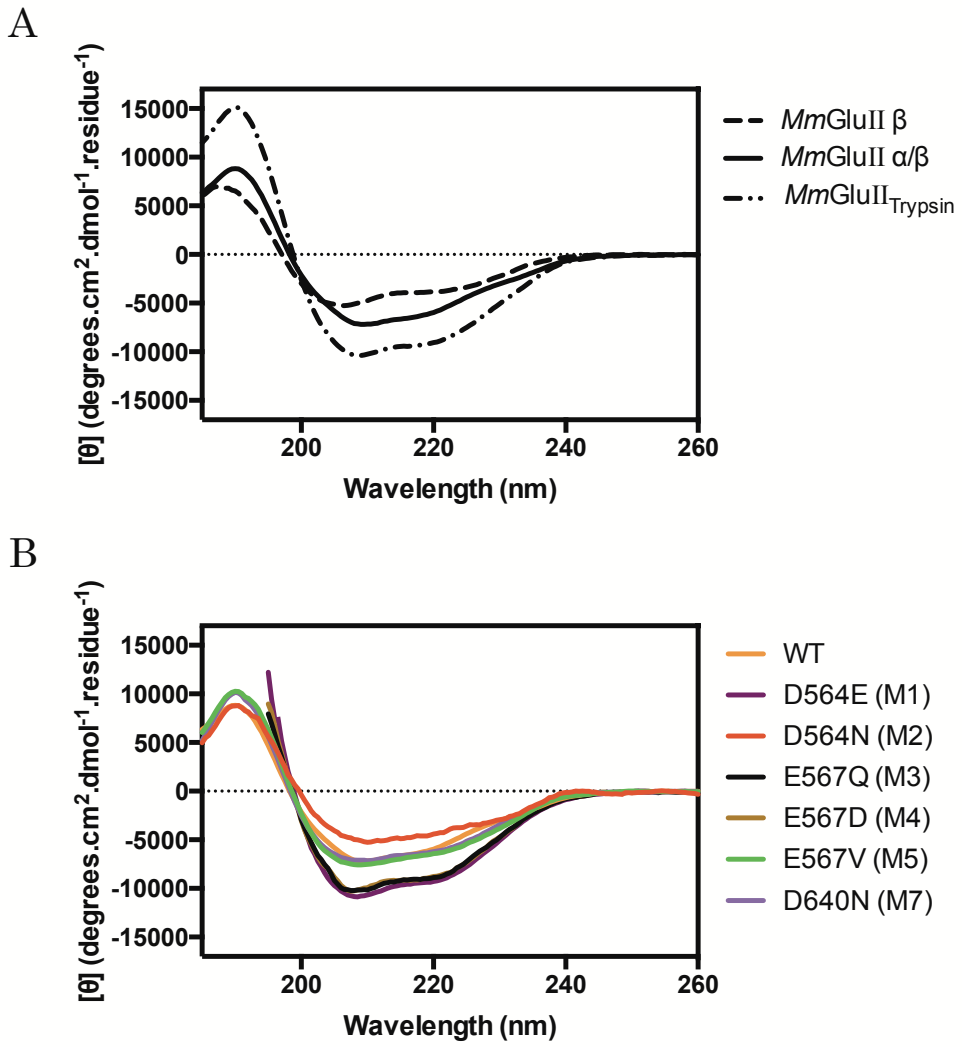


Figure 4.10: CD spectra of *Mm*GluII. (A) Spectra of wild-type heterodimeric *Mm*GluII, β-subunit only and trypsinsed *Mm*GluII (*Mm*GluII$_{\text{Trypsin}}$, see Chap. 5). (B) Spectra of various point mutants of *Mm*GluII. Data is only plotted where the applied voltage on the photomultiplier is below 600 V.

Assessment of whether the point mutants have induced any major folding defects was carried out by CD spectroscopy. The spectra of the six point mutants (Figure 4.10B) show an overall identical shape between them. A challenge of this technique is to obtain a very accurate measurement of protein concentrations. Inaccurate concentration estimates lead to errors in the scaling of the measured ellipticity. Inaccuracy of measuring dilute protein solutions is likely to be the cause of the differences in ellipticity values of the E567 mutants. Removal of the chloride ions was successful in all but three of the point mutants which had measurements excluded below 195 nm where the applied photomultiplier voltage exceeded 600 V.

Table 4.4: Secondary structure analysis of *Mm*GluI using CDSSTR method on DICHROWEB server expressed as percentages of total protein secondary structural elements. *Mm*GluII$_{\text{Trypsin}}$ is discussed further in Chapter 5.

Sample	Helix	Strand	Turn	Disordered	NRMSD[a]
*Mm*GluII heterodimer	10	29	22	40	0.02
*Mm*GluII β-subunit	6	33	18	41	0.019
*Mm*GluII$_{\text{Trypsin}}$	25	24	20	31	0.015
*Mm*GluII$_{\text{Trypsin}}$ structure[b]	18.3	26.4	25	30.4	N/A

[a] Normalised root mean square deviation
[b] Calculated using DSSP method on 2Struc server (http://2struc.cryst.bbk.ac.uk/twostruc) [177]

4.5.2 Differential scanning fluorimetry

A DSF study was conducted to determine the optimal buffer conditions for *Mm*GluII. The first step for optimising the assay is to determine the correct protein concentration that gives adequate signal. Figure 4.11A shows the scans from three different concentrations, where 0.05 mg/ml is the concentration that gives enough signal to observe the two states while being economical in the amount of protein used. The T_m was determined to be 51 °C, which is the basis for the comparisons used in the subsequent experiments.

A screen covering a pH range of 4.5 to 9.5, NaCl concentrations from zero to

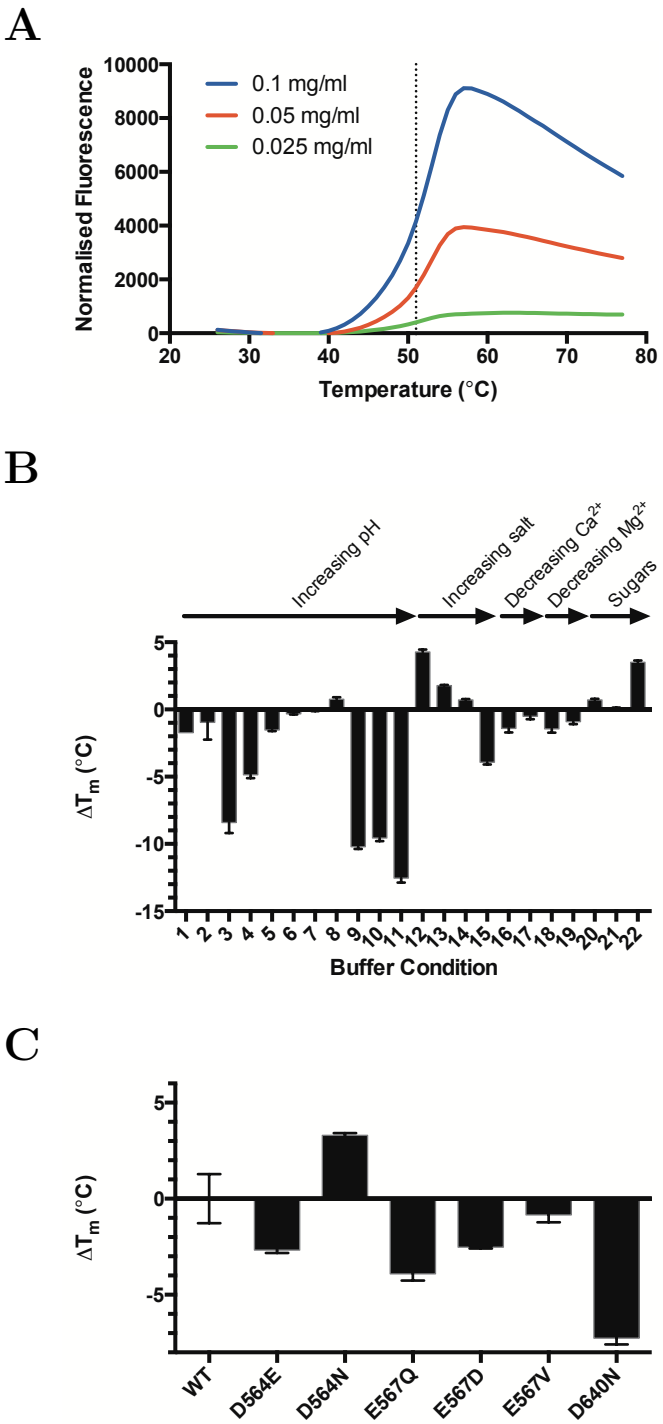


Figure 4.11: DSF analysis of *Mm*GluII. (A) Plots of fluorescence versus temperature in preliminary experiments to determine a suitable concentration of *Mm*GluII to be used in buffer screening. (B)Buffer screening showing T_m relative to *Mm*GluII in 20 mM HEPES pH 7.5, 150 mM NaCl. Conditions are as found in Table 2.11. The protein concentration is kept constant at 0.05 mg/ml. (C) T_m of point mutants relative to *Mm*GluII in 20 mM HEPES pH 7.5, 150 mM NaCl with the protein concentration of 0.05 mg/ml. All samples are statistically significantly different from the WT ($P < 0.01$, two-tailed t-test) with the exception of the E567V mutant ($P = 0.30$). Error bars represent standard deviations from measurements carried out in triplicate.

0.5 M, and a number of additives including divalent cations and monosaccharides was employed to ascertain the solution conditions with the highest stability for *Mm*GluII relative to the condition used in the final step of the purification. Apart from the spurious results in NaOAc buffer at pH 4.0 and 4.5, a general trend for a preference of a pH around 7.0-7.5 was observed (Figure 4.11B). Lower salt concentrations are preferred and of the additives, only glycerol appears to increase stability. What is not immediately apparent is whether the increase in T_m seen at low salt concentrations has any impact on the solubility of the enzyme.

4.5.3 Multi-angle laser light scattering

The elution volume of *Mm*GluII on a Superdex 200 matrix indicates that it exists as a larger order oligomer or that it is not in a standard globular conformation. To determine which of these two scenarios was likely, a SEC-MALS experiment was conducted on the purified *Mm*GluII using the same separation matrix. The principle of the technique is briefly described in the previous chapter (see 3.5.3). The result of the analysis of the peak indicates that the species observed has a molecular weight of about 166 kDa. This indicates that the purified *Mm*GluII is a heterodimer that has a non-globular conformation that causes it to elute earlier from a GF matrix than an equivalent globular protein.

4.5.4 Small angle X-ray scattering

The SEC-MALS experiments correlate well with published observations that GluII exists in an extended conformation [87]. To further explore these observations, small angle X-ray scattering (SAXS) experiments were carried out on the same three proteins as presented above in the CD section (4.5.1). Recording of the scattering curves

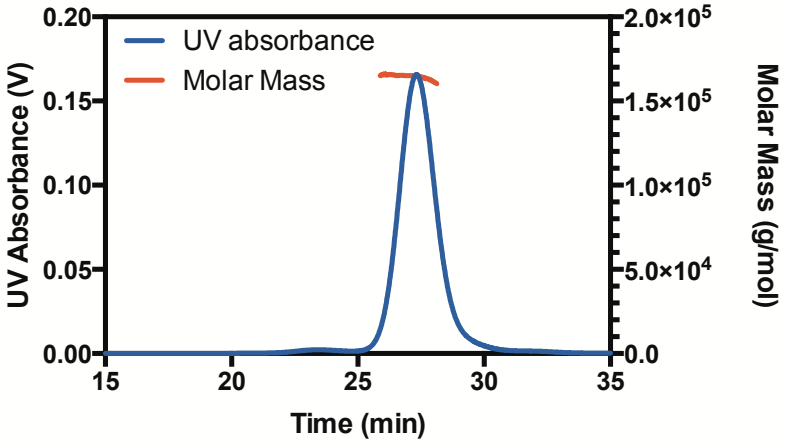


Figure 4.12: SEC-MALS analysis of *Mm*GluII. Chromatogram conducted on a Superdex 200 10/300 column in PBS at 0.5 ml/min. The calculated molecular weight is plotted in red across the peak shown in blue.

at the BioSAXS beamline at ESRF (BM29) gives distinctly different shapes (Figure 4.13A). A Kratky plot enables a quick assessment of the globularity of the three proteins (Figure 4.13B). Both the heterodimer and the trypsinised fragment exhibit characteristic Gaussian curves over the first portion of the graph that levels off at above the x-axis. The pair distribution function of the heterodimer (Figure 4.13C blue line) indicates a long structure in one direction by the center of the peak being at a lower radius with a long tail to the x-intercept (D_{max}).

SAXS from *Mm*GluII$_{Trypsin}$ is consistent with a globular structure that is reinforced by the pair distribution (Figure 4.13C red line) that is fairly Gaussian with a small tail at the higher distances indicative of a slightly elliptical particle.

The β-subunit in isolation has a different Kratky plot (Figure 4.13B green line) characteristic of unfolded or flexible proteins. This is also reflected in the radius of gyration (R_g). The pair distribution function (Figure 4.13C green line) yields the largest D_{max} of the three proteins with an asymmetry that indicates an elongated structure.

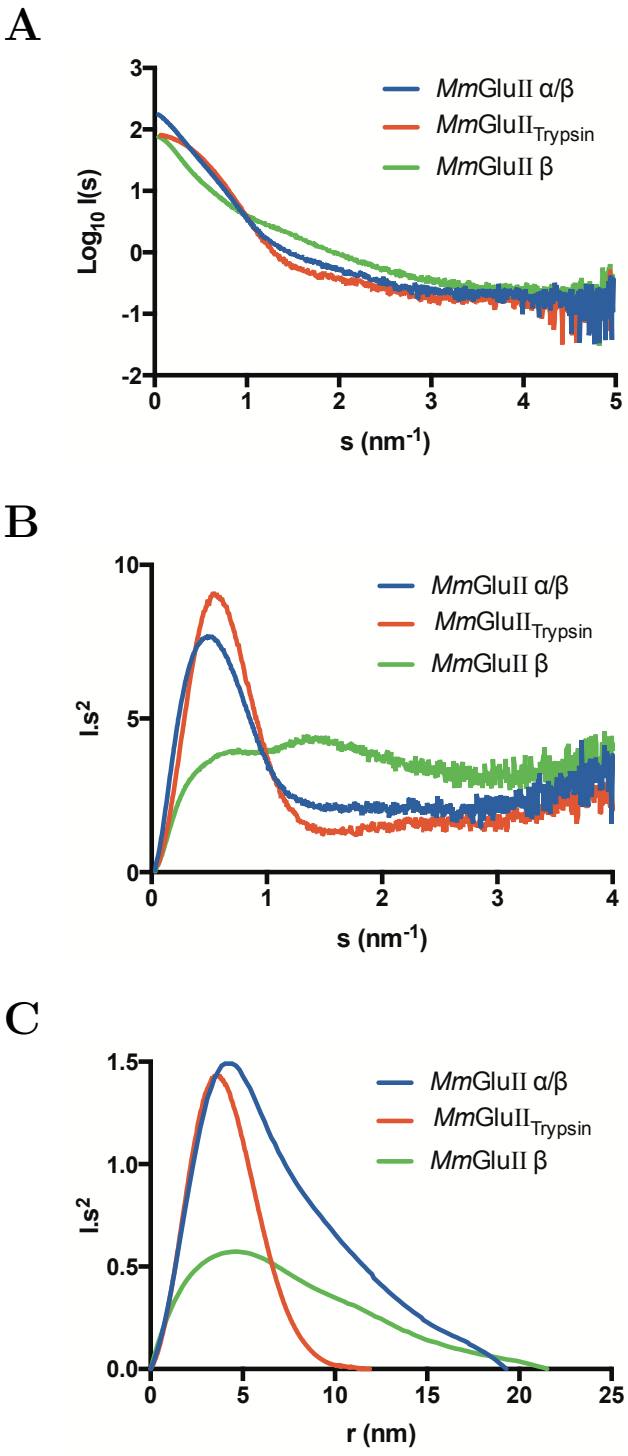


Figure 4.13: SAXS analysis of the full-length heterodimer, the β-subunit in isolation and MmGluII$_{\text{Trypsin}}$ in HEPES buffered saline. (A) Scattering curves. (B) Kratky plot. (C) Pair distribution function. All analyses were carried out in the ATSAS software.

Table 4.5: Summary of SAXS analysis derived parameters.

	R_g (Å)	I_0	D_{max} (Å)
*Mm*GluII heterodimer	55.1	156.75	192.7
*Mm*GluII β-subunit	61.3	77.36	214.7
*Mm*GluII$_{Trypsin}$	31.7	81.2	118.0

4.6 Enzymology

4.6.1 Enzyme activity against *N*-glycans

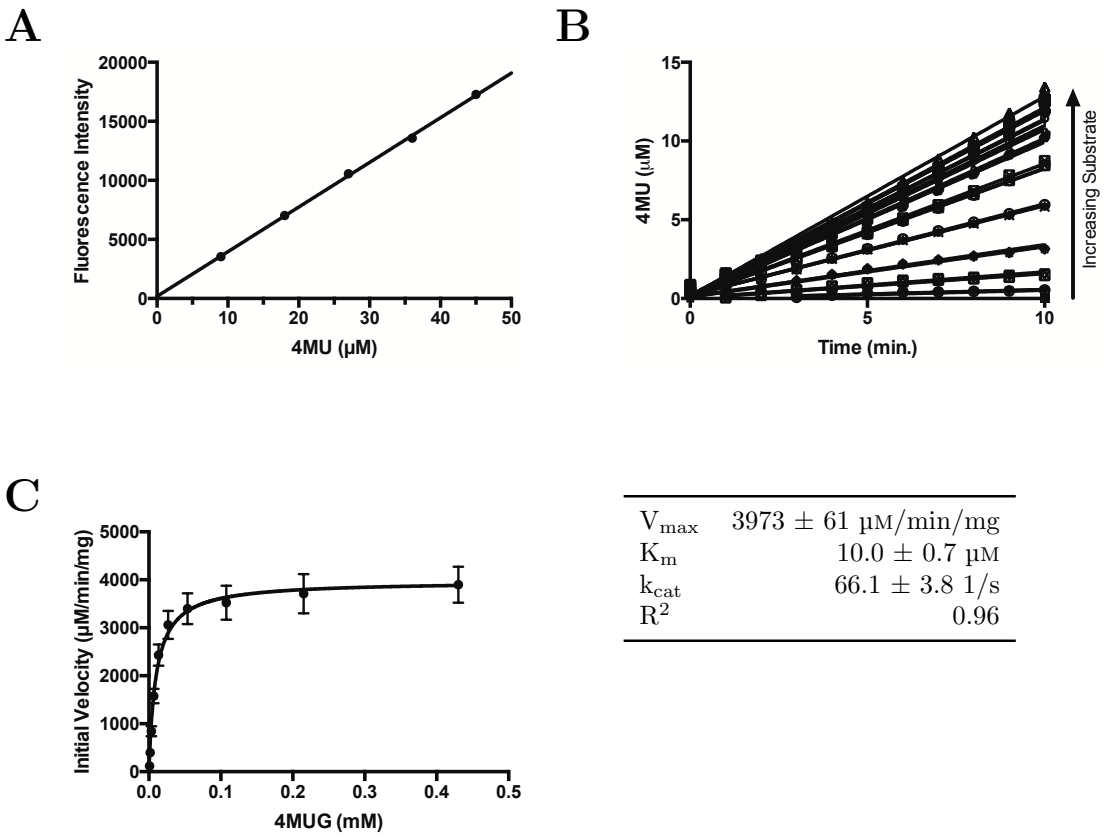
Similar to the previous chapter, fluorescently labelled glycans obtained from fOS were used as substrates for *Mm*GluII. In this case both the mono- and di-glucosylated glycans were tested. The wild-type *Mm*GluII was compared to selected point-mutants as well as the *Mm*GluII$_{Trypsin}$ fragment that is described in the following chapter (Table 4.6). For both the wild-type and *Mm*GluII$_{Trypsin}$, there was a complete digestion to $Man_7GlcNAc_2$ for both substrates. None of the mutants appeared to digest either of the substrates nor did any show any measurable activity against the 4MUG substrate. It must be noted that the mutants do not have any apparent folding defects as assessed by CD spectroscopy.

Table 4.6: Summary of activity of *Mm*GluII, *Mm*GluII$_{Trypsin}$, and the point mutants against different substrates.

	4MUG	$Glc_2Man_7GlcNAc_2$-2AA	$Glc_1Man_7GlcNAc_2$-2AA
Wild-type	YES	YES	YES
*Mm*GluII$_{Trypsin}$	YES	YES	YES
D564N	NO	NO	NO
D564E	NO	N/A	N/A
E567D	NO	N/A	N/A
E567Q	NO	N/A	N/A
E567V	NO	N/A	NO
D640N	NO	NO	NO

4.6.2 Enzyme activity against 4MUG

A convenient and economical alternative to measuring cleavage of 2AA-labelled glycans is the use of simple chromogenic or fluorogenic glucosides. 4MUG has been a known substrate of GluII for almost 30 years [203]. Hydrolysis by GluII liberates the highly fluorescent 4-methylumbelliferone (4MU), which has a greater sensitivity than the common chromogenic substrate, 4-nitrophenyl α-D-glucopyranoside. There are some practical considerations that need to be addressed when utilising 4MU as a reporter fluorophore. A significant parameter that must be accounted for is the pH at which the experiment is conducted as the protonation on the newly formed hydroxyl group affects the spectral properties of the molecule. The ability of 4MUG to absorb at the excitation wavelength was corrected, however, a maximum reduction of 5% of the fluorescent signal was observed at the concentrations used for *Mm*GluII so the overall effect was minimal. A standard curve constructed with standard solutions of 4MU at the pH of the reaction allowed accurate determination of hydrolysed substrate from about 1 to 60 µM (Figure 4.14A). The real-time measurement of hydrolysis gives a more accurate determination of the initial velocity by ensuring the linearity of the data by a runs test (Figure 4.14B). Combined with the ability of measuring the rate with multiple substrate concentrations in replicate under the same experimental conditions, this assay allows the determination of the constants in the Michaelis-Menten equation of the reaction assuming steady-state conditions. The resulting Michaelis-Menten constant, catalytic constant and maximal velocity (Figure 4.14C) are useful for assessing the correct conditions to measure inhibition by small molecules and the effect of any point mutants that remain active.



V_{max}	3973 ± 61 µM/min/mg
K_m	10.0 ± 0.7 µM
k_{cat}	66.1 ± 3.8 1/s
R^2	0.96

Figure 4.14: Enzymatic assay of the cleavage of 4MUG. (A) Standard curve of 4MU in assay buffer. (B) Calculated concentration of 4MU released over time. Plotted are curves for all substrate concentrations measured. (C) Michaelis-Menten plot for 4MUG with the associated kinetic constants derived by non-linear regression analysis. Error bars are standard deviations from three separate experiments conducted in triplicate.

4.6.3 Enzyme activity against disaccharides

An alternative assay was developed in order to assess the ability of *Mm*GluII to cleave disaccharides of different linkages. The 5 stereoisomers of glucobiose with α-glycosidic linkages are: trehalose [α(1,1)], kojibiose [α(1,2)], nigerose [α(1,3)], maltose [α(1,4)], and isomaltose [α(1,6)]. The assay consists of two separate reactions, the first is the hydrolysis of the disaccharides by *Mm*GluII with a thermal denaturing step to inactivate the *Mm*GluII. The second step allows quantification of the released glucose with a coupled glucose oxidase/peroxidase step to oxidise the fluorogenic Amplex Red reporter. The use of the Amplex Red reagent enables a much greater sensitivity than other peroxidase substrates with a limit of detection of glucose concentrations down to

0.5 µM (Figure 4.15A). As the hydrolysis produces two molecules of glucose, this must be accounted for in all calculations. Another consideration is that the glucose oxidase can also oxidise the disaccharide with the most noticeable example being nigerose. This problem was overcome by adjusting the measurements to identically treated disaccharide that did not have *Mm*GluII added. The single time-point measurements were all likely to be initial velocities as the substrate concentrations were all above the upper limit of the standard curve. Thus, the analysis of the steady-state kinetics of the hydrolysis of the five different substrates gave a clear indication of the ability of *Mm*GluII to cleave three of the five disaccharides (Figure 4.15B). Trehalose and isomaltose are unable to be cleaved which is reflected in the low signal and the poor fit to the Michaelis-Menten equation. The ability of *Mm*GluII to cleave nigerose was expected but the cleavage of maltose and kojibiose was not anticipated. Fitting to a Michaelis-Menten model showed cleavage for all three of the substrates (Table 4.7). At increasing nigerose concentrations, nigerose activity decreased. This can be fitted to a substrate inhibition model and concurs with previous observations of GluII having two adjacent substrate binding sites [204].The specificity constants (k_{cat}/K_m) indicate that nigerose is the better substrate but maltose is not too inferior or far-behind.

Table 4.7: Kinetic parameters of disaccharide cleavage by *Mm*GluII. Errors are the 95% confidence interval from two experiments.

	V_{max} (nM/s)	K_m (mM)	k_{cat} (1/s)	k_{cat}/K_m (1/sM)	R^2
Trehalose	0.6	19.1 ± 79.9	0.02 ± 0.06	0	0.68
Kojibiose	42.1	9.0 ± 0.7	1.4 ± 0.07	0.16	0.99
Nigerose	30.6	0.16 ± 0.03	1.0 ± 0.06	6.2	0.95
Maltose	38.5	0.57 ± 0.03	1.3 ± 0.02	2.26	0.99
Isomaltose	1.7	2.6 ± 3.7	0.06 ± 0.03	0.02	0.56

4.6.4 Enzyme inhibition

Two methods for determining the inhibition of *Mm*GluII were optimised in order to assess various inhibitors. The two have their advantages and disadvantages with

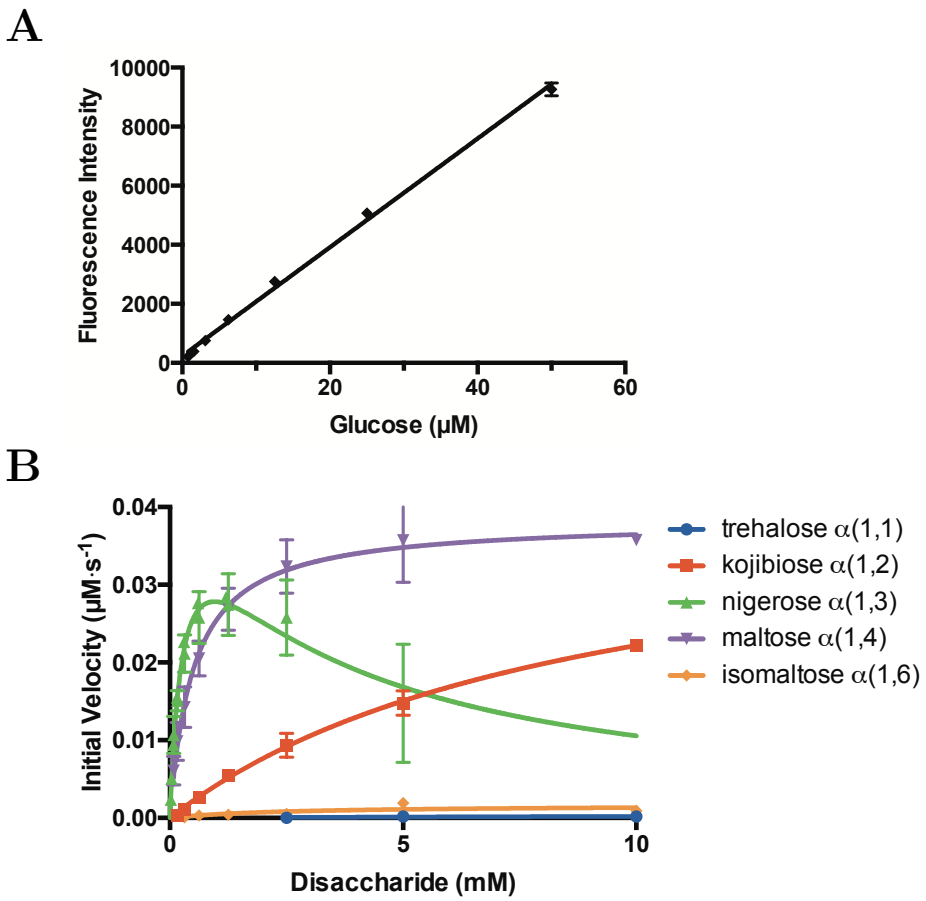


Figure 4.15: Enzymatic assay of the cleavage of variously linked disaccharides. (A) Standard curve used to measure glucose release. (B) Michaelis-Menten curve for the different disaccharides derived by non-linear regression analysis. Enzyme concentration is kept at 60 /nano/Molar/ and reactions are carried out at pH 7.2. Error bars are standard deviations from two separate experiments conducted in duplicate.

respect to their convenience versus the amount of information that can be obtained.

4.6.4.1 Measurements of inhibition constants

The measurement of inhibition constants for *Mm*GluII gives two important pieces of information. Firstly, it enables the discrimination of the type of inhibition. The differences in the V_{max} and K_m at increasing amounts of inhibitor can be used to discern competitive inhibition from non-competitive, un-competitive or some mixture thereof. In the case of competitive inhibition, a change in the observed K_m is observed with no change in V_{max}. Secondly, if competitive inhibitors are being used, it gives a

substrate-independent value for inhibitor binding due to competition of the inhibitor for the same site of binding as the substrate. This is useful as the 4MUG substrate is not very similar to the natural substrate that *Mm*GluII encounters in the cell.

Determination of the initial rates at multiple substrate concentrations versus multiple inhibitor concentrations (Figure 4.16) shows that competitive inhibition is occurring as at different concentrations of inhibitors and the V_{max} approaches the same magnitude. Simultaneous fitting allowed for the fitting of a competitive inhibition model to determine the K_i for the particular inhibitor. Fits to noncompetitive or uncompetitive models had coefficients of determination (R^2) of 0.950, and 0.929, respectively compared to the R^2 of 0.989 for the competitive model indicating this was the more likely model.

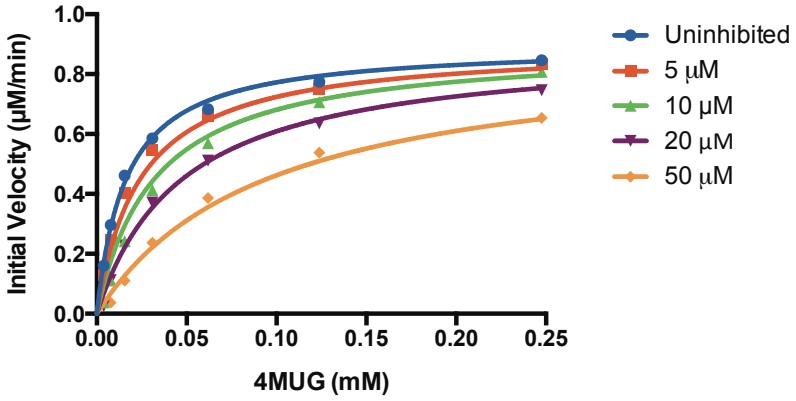


Figure 4.16: Determination of inhibitory constants. An example of the simultaneous fitting of a competitive inhibition model for the inhibition of *Mm*GluII by *N*B-DNJ.

The comparison of the three iminosugars tested (Table 4.8) shows no discernible trend with relation to the presence of *N*-alkylation or its length.

Table 4.8: K_i of a selection of DNJ derivatives against *Mm*GluII. Errors are the 95% confidence interval from two experiments conducted in triplicate.

Inhibitor	K_i (μM)
DNJ	3.1 ± 0.2
*N*B-DNJ	10.4 ± 3.7
*N*N-DNJ	3.7 ± 0.3

4.6.4.2 Measurements of IC_{50} values

An easy method for screening multiple inhibitors simultaneously is by determining relative inhibition. Measurement of the initial velocity of the reaction over a wide range of inhibitor concentrations is compared to the uninhibited sample. Inhibition follows a dose-response relationship, which allows the interpolation of the amount of inhibitor required to reduce activity by 50 % (IC_{50}). The concentration of 4-MUG is well above the K_m, meaning that the enzyme is approaching V_{max}. Even though competitive inhibitors should still have the same V_{max}, over the range measured in this assay, this type of inhibitor would have suitably increased K_m values to see a decrease in the measured initial velocities. This was carried out for a number of different inhibitors of *Mm*GluII. Castanospermine and its prodrug Celgosovir (BuCast) was the third variety of iminosugar tested. The DNJ moiety showed a trend in increased potency based on alkylation length (Figure 4.17). Castanospermine is a less potent inhibitor of *Mm*GluII than any of the DNJ based inhibitors by close to an order of magnitude.

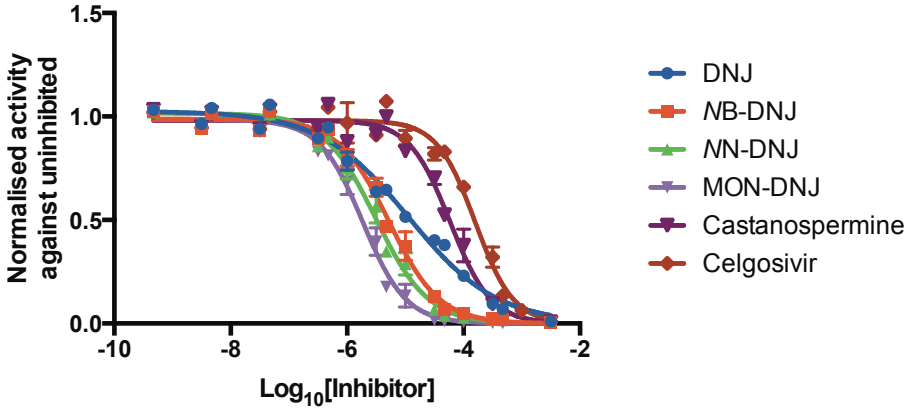


Figure 4.17: IC_{50} plots of multiple iminosugars against *Mm*GluII. Normalised activity against the uninhibited samples was used to fit a four-parameter dose-response curve. Error bars are standard deviations from three separate experiments.

Table 4.9: IC_{50} of a panel of iminosugars against *Mm*GluII. Errors are the 95% confidence interval from three separate experiments.

Inhibitor	IC_{50} (µM)
DNJ	11.4 ± 4.3
*N*B-DNJ	5.21 ± 1.02
*N*N-DNJ	3.11 ± 0.71
MON-DNJ	1.85 ± 0.33
Castanospermine	58.9 ± 14.2
Celgosovir	158.1 ± 37.3

4.7 Discussion

This is not the first report of the recombinant expression of heterodimeric GluII [72, 83]. Nevertheless, it is the first to document the production of the complex on the milligram level and the extensive characterisation of the protein using multiple biochemical and biophysical techniques. We chose to pursue full-length constructs to maximise the ability for the two subunits to interact with each other, another approach would have been to tailor constructs of the β-subunit to the two previously identified binding domains [86]. The disorder prediction alone indicates that dividing up the β-subunit into the two putative interacting domains may be a more promising approach. However, in terms of producing something that is relevant and comparable to other biochemical studies, the full-length construct still constituted the better choice.

The cloning and expression screening follow the lines of that presented in the previous chapter for GluI in which the quick and efficient cloning was then hampered by poor results that were observed when screening for expression. Although a number of expression systems were utilised, the only promising results came from the expression of the β-subunit. In itself this may have been a good start, but was ultimately abandoned as it did not appear to yield a viable target for crystallisation. The apparent lack of expression when only assessing by western blotting was due to poor transfer efficiency of the α-subunit to the membrane. The exact causes of this

problem are not readily apparent but may be due to the size or charge of the protein. A readily available and facile biochemical activity assay of 4MUG cleavage provided a solution to the initial shortcomings of the screening effort. As a result, production of *Mm*GluII was achieved in excess of 8 mg/l of culture. *Sc*GluII has not been fully pursued yet but it remains an interesting target as it may be different in the β-subunit. Although *S cerevisiae* GTB1 has been described to be the genuine partner, it is substantially longer than the other annotated orthologues [205]. The sequence-based annotation of the *Sc*GluII β-subunit protein indicates similar features in the correct order to those in other eukaryotes [206]. Whether it can serve as a suitable surrogate to study relative to human GluII is questionable, thus pursuing *Mm*GluII was prioritised.

Establishing a protocol for the purification of *Mm*GluII was relatively straight forward with only minor hurdles due to the choice of buffer components and adsorption onto ultrafiltration membranes that were readily overcome. The three-step purification is slightly excessive for two reasons: (1) As both subunits contain hexahistidine tags, there is no preference for the heterodimer over either of the other two free subunits. (2) The *Mm*GluII has very low affinity for the StrepTactin resin most likely compounded by poor accessibility in the StrepII tag that has generally low affinity to begin with. The difficulties of mutagenesis in the expression vector were overcome and, at the same time, the affinity tags were rationalised to avoid the duplication that existed. Thus for the point-mutants the hexahistidine tag was removed for the β-subunit and any excess of either subunit was removed by gel filtration due to the size of the complex relative to its individual components.

The utilisation of a DSF-based assay to screen for stability in a range of common buffers to screen across pH, salt concentration and extra additives provides two observations, the preference for low salt, and an optimal pH around that which

is observed in the ER lumen. The cause of the pH sensitivity can not be exclusively linked just to the pH value but also to the varying buffers chosen that may affect the ionic strength of the solution that in turn might decrease protein stability. A more ideal method would be to use the same type of buffer purely as a control and to vary the pH range. If there is a specific buffer effect the same could be set up for a number of buffers to determine more clearly the correct buffer and pH. The implications of the preference for lower salt concentrations may also be linked as the increasing ionic strength may be playing a role in its oligomeric state or specific networks of interactions important for its tertiary structure.

The point-mutants generated behaved in a similar fashion to the wild-type enzyme and were purified in a more streamlined manner. The overall fold does not appear to have been compromised, though the CD spectra are not ideal due to poor removal of chloride ions and incorrectly determined protein concentration. The introduction of these mutations seems to have caused a destabilising effect as demonstrated by the decrease in T_m in all except D564N. Possible future mutations will be informed by the structural efforts presented in the following chapters and, in particular, could focus on understanding the residues involved in substrate specificity.

CD spectroscopy provides some insight into the structure of the GluII complex. The results presented here show similar curve shapes to that from rat tissue derived GluII [207]. Comparison of the CD spectra from full-length, β-subunit, and the trypsinised fragment that is presented in more detail in the following chapter, indicates some features of the enzyme. The spectra indicate a substantial degree of disorder, a property predicted especially for the β-subunit. The role of the structure of the β-subunit may be such that it forms some extended conformation around or on the side of the α-subunit or that the complex is lacking an otherwise unknown binding partner that induces more secondary structure upon binding. This line of ar-

gument is corroborated by adding the two different scattering experiments that were performed. Both the SEC-MALS and the SAXS experiment point to an extended, probably unstructured contribution due to the β-subunit. Analytical ultracentrifugation on the rat GluII reported an extended conformation of the isolated enzyme that can either be fitted to a prolate or oblate particle shape [87]. The former is consistent with the distance distributions calculated from the SAXS measurements and this appears to be predominantly due to the β-subunit. Further investigation is warranted upon the availability of either of the two substrate glycans to see if there is a conformational change that accompanies binding of the substrate. These experiments can shed light on the stoichiometry of glycan binding. One theory is that the glycoprotein needs two N-glycans to be recognised by GluII, in a *trans* activation-like mechanism [88]. A rationalisation with the data above could be made whereby an extended GluII binds one of the glycans through its C-terminal MRH domain and performs the hydrolysis on the other glycan. However, one could also postulate that an extended β-subunit can help in glycan recruitment and positioning of a glycan into the active site. Whichever of these two hypotheses may be the case, data to confirm one or the other are still lacking. But these can be addressed in the future through some of the biophysical methods presented in this chapter.

The various biochemical assays presented in this chapter show that the *Mm*GluII produced is highly active. Somewhat surprising is the ability of the trypsinised fragment to perform both cleavage events despite numerous studies indicating that this is unlikely without the MRH domain of the β-subunit [85, 205, 208, 209]. However, most of these other studies show a decrease in the ability to cleave these substrates in the absence of the β-subunit rather than an absolute abrogation of activity. A likely explanation could be a decreased affinity for the glycans due to the absence of the interactions with the MRH and thus a marked difference in kinetics. It is difficult to imagine a mechanism that is likely to prohibit the cleavage of the glu-

cose residues from the glycan in the absence of the MRH domain, but rather a lower thermodynamic benefit from a smaller enthalpic gain of glycan binding due to fewer intermolecular interactions.

The comparison of the differently linked disaccharide substrates can help formulate some hypotheses on substrate specificity at the catalytic site. An early study on GluII isolated from pig kidneys inferred regioselectivity toward α(1,3) and α(1,4) glycosidic bonds based on the ability of maltose but not kojibiose [α(1,2)] to inhibit 4MUG cleavage [197]. The results from the kinetics presented here confirm and extend the details of that study but also show that kojibiose can also be cleaved albeit with a much lower ($\sim$50X) affinity. The observation of substrate inhibition for nigerose extends previous observations that there are distinct subsites in the active site [119, 204]. Does the ability of *Mm*GluII to cleave these differently linked disaccharides suggest that there is a degree of plasticity that can accommodate these different substrates or that it lacks a mechanism to prevent such substrates from binding in the catalytic site? A tryptophan in one of the loops contributing to the active site has been implicated in increasing the regioselectivity for α(1,3) in some bacterial GH31-containing glucosidases [210–212]. Highlighted in these studies is a tyrosine that at this position confers α(1,4) specificity, a difference that may be marginal enough to accommodate these different substrates but confers an overall preference for one over the other.

The *in vitro* comparison of differently alkylated iminosugars does not appear to follow a strong trend with respect to potency and is similar to previous observations [213]. Although there is a modest correlation between the IC_{50} and in the length of alkylation, the relationship is not apparent with the K_i measurements. It would be difficult to imagine some portion of the active site, likely an exposed and polar pocket, having an adjacent non-polar binding site for an aliphatic carbon chain. The

role of iminosugar alkylation in producing FOS *in cellula* has been attributed to the increased ability of these derivatives to get into the ER lumen and inhibit their targets [194, 214]. There has also been some investigation into the increased bioavailability of alkylated derivatives [140]. However, there has been no definitive proof that the alkylation results in better ER availability. Castanospermine on the other hand shows a lower potency against isolated GluII that is difficult to justify. The major difference to DNJ is the addition of the two extra carbons that form a second ring. It is possible that the rigidification of this molecule might minimise the loss of conformational entropy upon binding compared to the ordering of the C-6 hydroxyl of DNJ. However as the IC_{50} is higher, which likely reflects a lower affinity for *Mm*GluII, there could possibly be a steric hindrance with some residues in the binding site that does not occur with DNJ. There is an inconsistency between the two set of experiments that can be seen when trying to apply the Cheng-Prusoff equation which can be used to interchange K_i and IC_{50} values [215]. With the values presented in this chapter, one value is considered the "true" value, therefore the corresponding value is incorrect by more than an order of magnitude. This points to a flaw in either of the two protocols calling into doubt the significance of these values. Further investigation is warranted to dissect which has the flawed methodology. One possible cause of this observation may be due to the rate at which the iminosugars bind to the the enzyme. Iminosugars have been observed to bind slowly to glycosidases [216, 217]. The initial velocities vary before and after the equilibrium is established between the enzyme and inhibitor. This would lead to incorrect measurement of the either the K_i or IC_{50}. In the case of the IC_{50} measurements, there is a pre-incubation of the enzyme and inhibitor before initiation of the reaction. This was not the case for the K_i measurements which might explain the higher values if the equilibrium is not yet established.

5

Structural characterisation of Endoplasmic Reticulum α-Glucosidase II

A major challenge for the structural characterisation of glucosidase II to date was producing sufficient quantities of the enzyme as was presented and discussed in Chapter 4. A second major challenge was the crystallisation of *Mm*GluII which also proved to be a fruitless exercise to date. As *Mm*GluII is itself an *N*-linked glycoprotein, one avenue that was pursued was to modify the only occupied *N*-glycan site through modification of *N*-glycan processing with kifunensine treatment and/or enzymatic removal during

the purification with the endoglycosidases PNGaseF or EndoH. Ultimately, the most successful strategy was limited proteolysis with trypsin that produced the stable 110 kDa species, *Mm*GluII$_{\text{Trypsin}}$, a readily crystallisable fragment with little missing from the α-subunit.

5.1 Limited proteolysis and purification of *Mm*GluII$_{\text{Trypsin}}$

Different proteases can cleave the rat GluII with it retaining activity against 4MUG[87]. In particular, that study showed that a limited trypsin digestion yielded an approximately 70 kDa fragment with the N-terminus corresponding to the beginning of the GH31 catalytic domain. A time-course experiment was designed to assess whether the same could be observed for the murine protein that is produced as described in the previous chapter. A trypsin:*Mm*GluII sample with the mass ratio of 1:22.5 was incubated over 3 hours and proteolysis that occurred at room temperature was observed by SDS-PAGE (Figure 5.1A). Following incubation overnight, a western blot was done to ascertain which fragments contained the C-terminal polyhistidine tags (Figure 5.1B). Four observations become apparent from this experiment: (1) a large fragment at approximately 70 kDa is present like that observed for the rat glucosidase II. (2) Only one species still contains the C-terminal polyhistidine tag, which is likely to belong to the β-subunit as will be discussed below. (3) Overnight trypsinolysis causes a complete degradation of any detectable species in the western blot. (4) There are some sites of cleavage on the α-subunit that are almost immediately cleaved in the time between adding the protease and taking the time point. The larger fragment was subjected to Edman degradation to determine its new N-terminus, which turned out to be MLDYL. This sequence is similar to that of the

MMDYL observed for the rat enzyme[87], and the annotated beginning of the GH31 domain. Assuming a molecular weight of 60-70 kDa for this fragment as indicated by the gel, it is likely that this fragment belongs to the α-subunit. The ~17 kDa C-terminally hexahistidine-tagged species observed in the western blot was assigned to the β-subunit as it cannot be derived from the α-subunit. Together with the larger α-subunit fragment, this 17 kDa species would add up to a mass too big for the larger α-subunit species considering its N-terminus. Further optimisation was conducted that found a 1:100 trypsin:*Mm*GluII mass ratio for 4 hours at room temperature was a more economical protocol and yielded the same digestion pattern by SDS-PAGE.

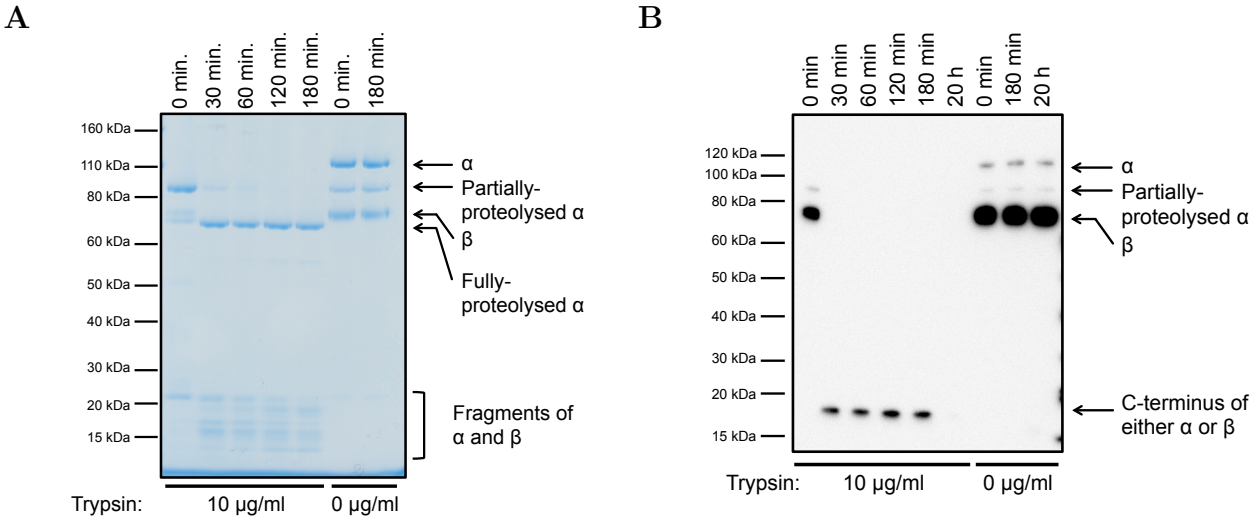


Figure 5.1: Limited proteolysis of *Mm*GluII by trypsin. (A) reducing xsSDS-PAGE and (B) anti-His$_6$ western blot analyses of a time-course experiment.

The production of the *Mm*GluII$_{\text{Trypsin}}$ fragment was easily scaled up to preparative levels and the fragment purified by gel filtration chromatography in an identical manner to the full length protein (Figure 5.2). After gel filtration, four dominant bands appear to co-elute from the column indicating the presence of a complex that is resistant to further proteolysis (Figure 5.2B).

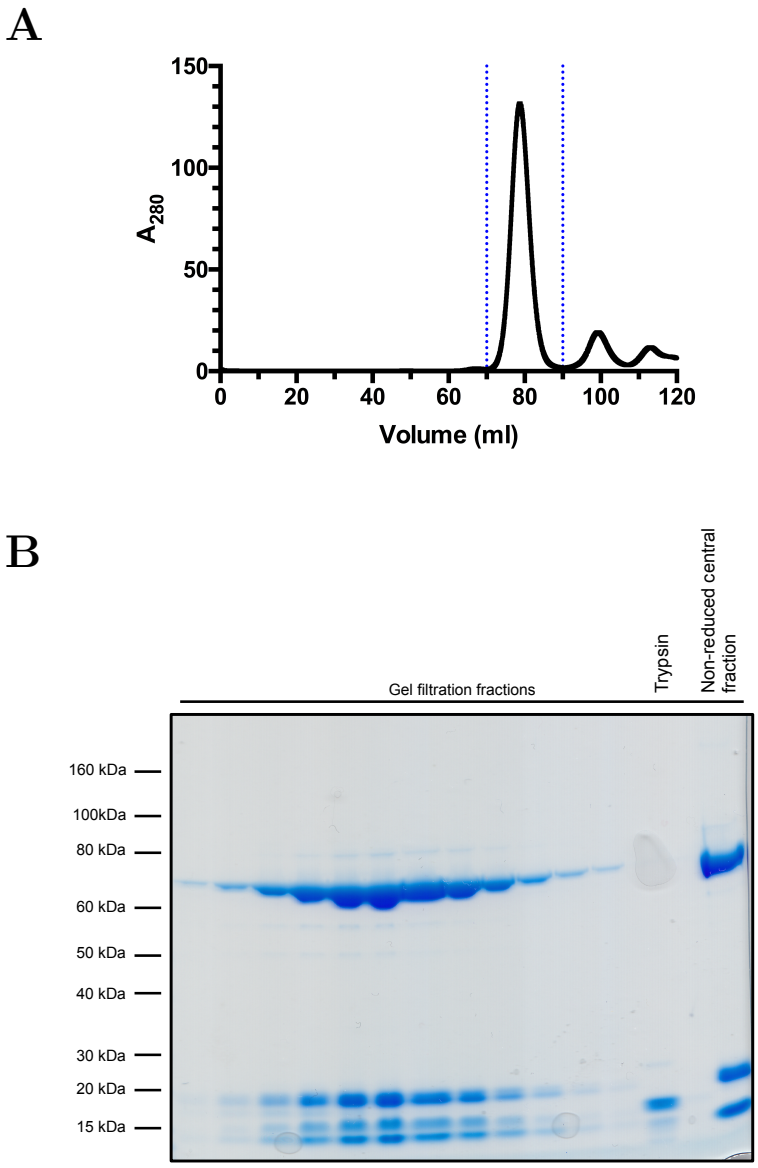


Figure 5.2: Gel filtration purification of the $Mm\mathrm{GluII}_{\mathrm{Trypsin}}$ fragment. (A) Chromatogram after 4 h incubation of *Mm*GluII with trypsin on a Superdex 200 16/600 column. (B) SDS-PAGE analysis of the fractions indicated with the blue boundaries in (A).

5.1.1 Characterisation of *Mm*GluII$_{\text{Trypsin}}$

5.1.1.1 Comparison of kinetic behaviour of *Mm*GluII$_{\text{Trypsin}}$ and full-length *Mm*GluII

A side-by-side comparison of the kinetic constants of *Mm*GluII$_{\text{Trypsin}}$ and the full-length heterodimer was conducted to ascertain if the trypsinolysis had affected the properties of the enzyme. The plots of initial velocity versus 4MUG concentration (Figure 5.3) show an almost identical curve and this is visible in the values derived from fitting the steady-state kinetics to the data. There is a decrease in the V_{max} and k_{cat} but the K_m is the same as seen in Table 5.1. The implication is that *Mm*GluII$_{\text{Trypsin}}$ cannot turn over the 4MUG as efficiently but retains that same ability to bind the substrate. However, the kinetics of binding and dissociation of the substrate must be slightly altered after trypsinolysis. As there is only a minor change in the kinetic parameters of either enzyme for the substrate, we can be confident that trypsinolysis has not altered the enzyme's activity.

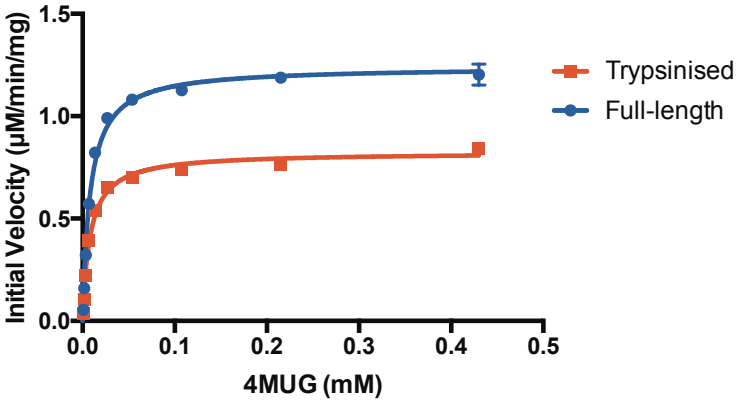


Figure 5.3: Comparison of the Michaelis-Menten plots of *Mm*GluII and *Mm*GluII$_{\text{Trypsin}}$. Error bars represent standard deviations from three separate experiments conducted in triplicate.

Table 5.1: Kinetic parameters of *Mm*GluII and *Mm*GluII$_{\text{Trypsin}}$ against 4MUG. Errors represent the standard error from an experiment conducted in triplicate.

	*Mm*GluII Full-length	*Mm*GluII$_{\text{Trypsin}}$
V_{max} (µM/min/nmol)	1.24 ± 0.01	0.82 ± 0.01
K_m (µM)	10.0 ± 0.7	10.5 ± 0.8
k_{cat} (1/s)	66.1 ± 3.8	46.2 ± 5.0
R^2	0.96	0.95

5.1.1.2 Mass spectrometry

The boundaries of the fragments produced by trypsinolysis are needed to understand the identity of the *Mm*GluII$_{\text{Trypsin}}$ components. Electrospray ionisation mass spectrometry (ESI-MS) of the complex resulted in identification of the masses of the three smaller species present (Figure 5.4). However, this spectrum is difficult to assign without any additional information. Once the crystal structure of *Mm*GluII$_{\text{Trypsin}}$ was being built, the model (see 5.2.4) began to reveal the possible stretches of the protein that were present in the crystal. Combined with this knowledge, it was possible to assign the boundaries of these fragments. The smallest fragment has an *m/z* of 10651.46 which corresponds to the β-subunit fragment G22-K119 (theoretical MW: 10651.4 Da). The second fragment has an *m/z* of 12841.24 which corresponds to the α-subunit fragment A240-K354 (theoretical MW: 12842.2 Da). The third fragment has an *m/z* of 18774.80 which does not match any particular tryptic fragment but most likely the N-terminus of the α-subunit to R185 (theorectical MW: 17584.2 Da), with the mass difference being due to the *N*-linked glycosylation at N97. Due to limitations of the instrument, the fourth 60-70 kDa fragment was not seen, but its N-terminus is known (see 5.1).

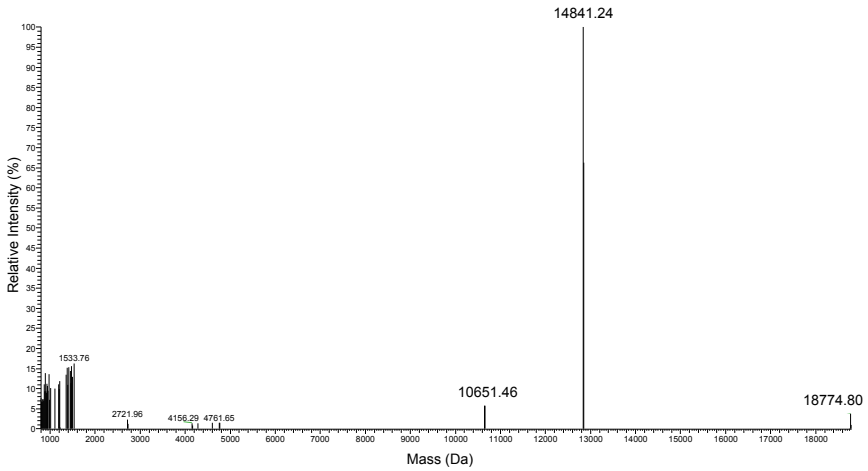


Figure 5.4: Mass spectrum of *Mm*GluII$_{\text{Trypsin}}$ derived from a reduced, desalted sample injected directly into a Q Exactive mass spectrometer. The raw spectrum was deconvoluted in Skyline to produce this spectrum.

5.1.1.3 Multi-angle laser light scattering

*Mm*GluII$_{\text{Trypsin}}$ was run on a gel filtration column (similar to 3.5.3) in order to establish whether the fragments exist as one species in solution. The *Mm*GluII$_{\text{Trypsin}}$ molecule eluted as a single peak from a Superdex 200 matrix and the SEC-MALS analysis shows a size of approximately 107 kDa (Figure 5.5). The addition of the various fragments identified by mass spectrometry, N-terminal sequencing and X-ray crystallography put the molecular weight at 111,741 Da, which closely resembles the figure obtained from the SEC-MALS analysis and supports the notion that the *Mm*GluII$_{\text{Trypsin}}$ complex is comprised of the four identified polypeptides.

5.2 X-ray crystallography

5.2.1 Crystallisation

Screening for suitable conditions to crystallise *Mm*GluII$_{\text{Trypsin}}$ was carried out in a high-throughput manner with access to the Department's crystallisation robotics and

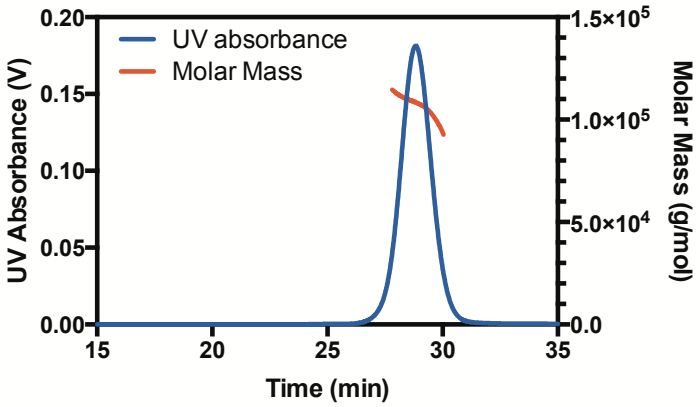


Figure 5.5: SEC-MALS analysis of *Mm*GluII$_{\text{Trypsin}}$. The chromatogram was conducted on a Superdex 200 10/300 column in PBS at 0.5 ml/min. The calculated molecular weight is plotted in red across the peak shown in blue.

commercially available screens. Morpheus HT-96 is a a three-dimensional screen that allows the simultaneous screening of common precipitants, additives, and pH in the range between 6.5 and 8.5 [218]. After a few days, numerous conditions from this screen yielded crystals, of which three diffracted well (<2.5 Å) with no optimisation necessary (Figure 5.6A-C). Two different crystal morphologies were produced: rods (monoclinic and orthorhombic forms) and bipyramids (trigonal crystal form). Two rounds of optimisation yielded the best crystals of both the monoclinic (Figure 5.6D) and the trigonal form, with much larger crystals exceeding 400 µm in their longest dimension. Optimisation was carried out through a standard iterative process of fine tuning the precipitant concentration and pH of the precipitant mixture. Larger drops and larger protein-to-precipitant ratios facilitated much larger crystal growth. A further increase in the diffraction resolution was achieved during the attempts to soak inhibitors and other small molecules into the active site, see 5.4 for details. Briefly, in an attempt to remove the ethylene glycol cryoprotectant molecules bound in the active site, PEG 400 was used to cryoprotect in conjunction with addition of inhibitors. Although the inhibitor tested was not present in the active site of *Mm*GluII$_{\text{Trypsin}}$, a transformation of the lattice into a new orthorhombic form took place, as a result of which diffraction was greatly improved.

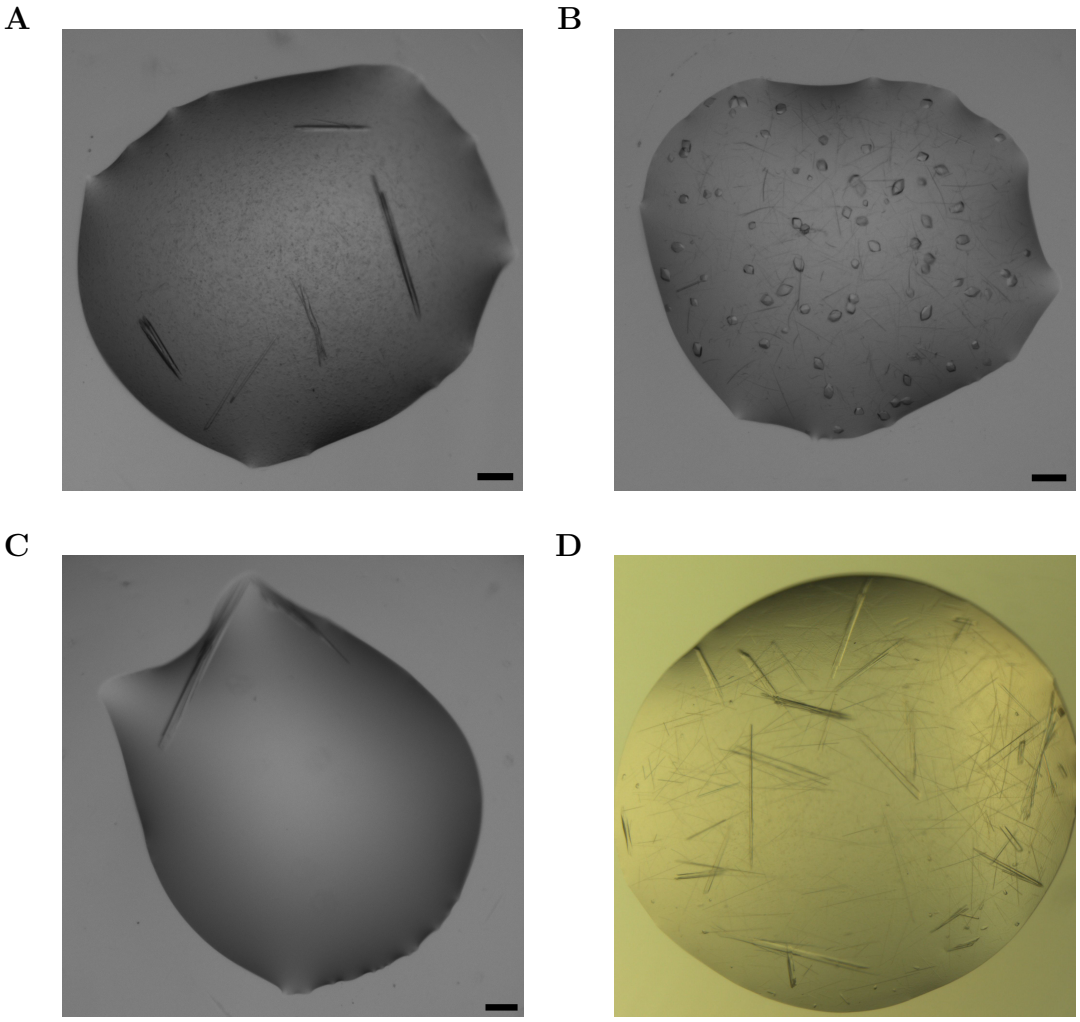


Figure 5.6: Photographs of the various crystal forms. Scale bars indicate 100 µm. The initial monoclinic (A) orthorhombic (B), trigonal (C) and optimised monoclinic (D) crystals.

5.2.2 Data collection and processing

Crystals were mounted in nylon loops and cooled to liquid nitrogen temperatures before measuring diffraction at synchrotron light sources. Diffraction images were collected by first screening a crystal for diffraction at three different ϕ angles to determine the lattice and orientation of the crystal. These images were used to compute the strategy with EDNA[152]. The collected data were automatically processed on the clusters available at the synchrotron. However, the final processing was carried out using the autoPROC pipeline as described in 2.7.3. The data processing statistics

in Table 5.2 pertain to final round of scaling and merging. The general criteria for inclusion of frames was based on visual inspection of the R_{merge} per frame. Spurious frames caused by mechanical problems with the goniometer are easily identifiable and thus excluded from the scaling and merging steps. The criteria for the resolution cut-off was slightly more elaborate. A general cut-off of $\langle I/\sigma(I)\rangle$ of 1 was applied by the script. Assessment of the resolution cut-off from the Aimless output was based on $CC_{1/2}$, completeness, multiplicity, and R_{meas}.

The monoclinic form can readily convert to two different orthorhombic forms. This change was induced when soaking the monosaccharides or inhibitors into the monoclinic crystals. Therefore, crystals measured of this form were processed both in the monoclinic and orthorhombic spacegroups and cells and the correct form was identified based on the R_{meas} values when scaled to the same resolution cut-offs. This is due to the sensitivity of the R_{meas} parameter to the symmetry of the crystal [219]. The monoclinic and two orthorhombic crystal forms have very similar crystal lattices but result in three different observable space groups. When the origins of the crystals are aligned, the similarity of the lattices are observed (Figure 5.7A). The small changes caused by ligand binding or soaking conditions cause very small packing disturbances which result in an overall change in symmetry. The schematic representations of the lattices with the relevant symmetry operations can be used to visualise the operations that can generate the unit cells (Figure 5.7B). The $P2_12_12_1$ form contains a doubling in the cell length in one direction (designated on the c axis) but still contains the same number of molecules in the same volume as the other forms. The $P2_12_12$ form only contains one molecule in the asymmetric subunit which is accounted for by the greater number of symmetry operators that can generate all the molecules in the unit cell. The $P2_12_12_1$ form and the monoclinic form have two molecules in their asymmetric subunits but the NCS is generated from translational symmetry in the former case as opposed to the 2-fold in the monoclinic form. Overall,

Table 5.2: Diffraction data processing statistics of *Mm*GluII$_{\text{Trypsin}}$crystals.

	Monoclinic apo	Orthorhombic apo	Trigonal apo	Orthorhombic # 2 apo	Glucal
Resolution range (Å)	174.5-2.40 (2.53-2.40)	126.94-2.45 (2.58-2.44)	241.6-2.45 (2.60-2.45)	59.1-1.74 (1.83-1.74)	176.4-2.29 (2.30-2.29)
Cell dimensions:					
a, b, c (Å)	63.6, 174.5, 103.5	103.4, 126.4, 173.01	102.9, 102.9, 241.6	104.1, 173.8, 62.8	127.1, 176.4, 103.3
α, β, γ (°)	90, 91.1, 90	90, 90, 90	90, 90, 120	90, 90, 90	90, 90 ,90
Spacegroup	$P2_1$	$P2_12_12_1$	$P3_2$	$P2_12_12$	$P2_12_12_1$
Total reflections	298670 (44392)	464166 (75452)	1097763 (176916)	431415 (62508)	658798 (3622)
Unique reflections	86953 (12708)	85569 (13705)	104907 (16929)	109959 (16202)	101815 (731)
R_{meas} [a]	0.142 (0.710)	0.097 (0.477)	0.264 (2.356)	0.077 (0.663)	0.204 (1.512)
R_{pim} [b]	0.168 (1.165)	0.229 (2.020)	0.117 (1.041)	0.046 (0.392)	0.110 (0.843)
$\langle I/\sigma(I)\rangle$	8.5 (2.4)	7.8 (1.1)	8.6 (1.3)	12.9 (2.4)	8.7 (1.1)
Completeness (%)	99 (99)	99.8 (99.9)	100 (100)	94.7 (96.0)	96.8 (70.6)
Multiplicity	3.4 (3.5)	5.4 (5.5)	10.5 (10.5)	3.9 (3.9)	6.5 (5.0)
$CC_{1/2}$	0.992 (0.753)	0.992 (0.282)	0.991 (0.28)	0.998 (0.802)	0.995 (0.552)
	Glucose	Mannose	Castanospermine	DNJ	*N*B-DNJ
Resolution range (Å)	60.0-2.67 (2.82-2.67)	177.33-2.36 (2.49-2.36)	173.9-1.81 (1.91-1.81)	88.9-1.92 (2.02-1.92)	89.17-2.10 (2.21-2.10)
Cell dimensions:					
a, b, c (Å)	64.0, 174.5, 103.1	126.8, 177.3, 103.5	103.4,173.9, 63.0	103.5, 173.7, 62.9	104.0, 173.5, 62.8
α, β, γ (°)	90, 91.3, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Spacegroup	$P2_1$	$P2_12_12_1$	$P2_12_12$	$P2_12_12$	$P2_12_12$
Total reflections	149865 (21760)	1025554 (102149)	691997 (98264)	578244 (77589)	303989 (17785)
Unique reflections	57704 (8664)	96070 (13852)	103854 (14994)	87597 (12619)	61858 (6235)
R_{meas} [a]	0.182 (1.260)	0.274 (2.092)	0.187 (1.726)	0.238 (1.605)	0.237 (1.397)
R_{pim} [b]	0.090 (0.490)	0.114 (1.062)	0.099 (0.907)	0.126 (0.890)	0.172 (0.864)
$\langle I/\sigma(I)\rangle$	5.4 (1.1)	8.6 (1.2)	8.8 (1.2)	7.2 (1.2)	7.8 (1.2)
Completeness (%)	90.5 (93.6)	100 (100)	100 (100)	100 (100)	92.5 (65.6)
Multiplicity	2.6(2.5)	10.7 (7.4)	6.7 (6.6)	6.6 (6.1)	4.9 (2.9)
$CC_{1/2}$	0.975 (0.307)	0.993 (0.447)	0.996 (0.476)	0.992 (0.415)	0.980 (0.441)

Values in parentheses represent the high-resolution data shell

[a] $R_{meas} = \Sigma_h\Sigma_i|\sqrt{n_h/(n_h-1)}|I_{hi} - \langle I_h\rangle|\Sigma_h\Sigma_i I_h$

[b] $R_{pim} = \Sigma_h\Sigma_i|\sqrt{1/(n_h-1)}|I_{hi} - \langle I_h\rangle|\Sigma_h\Sigma_i I_h$

subtle changes in the packing leads to different observed space groups and unit cells.

A trigonal crystal form was also observed that formed bipyramidal crystals that were the largest crystals by volume. Unfortunately these crystals were twinned. This was detected through the NZ test in Phenix Xtriage, by plotting the data cumulative intensity distribution (Figure 5.8). The shape of the curve in comparison of the typical untwinned distribution (ref line) can be used to detect twinned data which tend to have a more sigmoidal character (blue line), and this was indeed the case for the trigonal $Mm\text{GluII}_{\text{Trypsin}}$ data. Restrained refinement with Refmac with twinning operator k, h, l refined a twinning of fraction of 0.59.

5.2.3 Phasing

In the first orthorhombic ($P2_12_12_1$) crystal form, the Matthews coefficient for two molecules in the asymmetric subunit (asu) was calculated to be 2.96 $Å^3$/Da with a solvent content of 54.25 % and the probablility being 0.94. This strongly indicates that the asu contains 2 copies of $Mm\text{GluII}_{\text{Trypsin}}$. Molecular replacement was attempted on the orthorhombic dataset using the MrBUMP pipeline, which combines a homologue search and model preparation followed by molecular replacement and restrained refinement cycle on all identified hits. In this case, the N-terminal section of human maltase-glucoamylase corresponding to the maltase alone was identified as a good homologue (PDB ID: 3L4Z). The best search result was from a Molrep-made search model fed into Phaser. The final Phaser statistics are summarised in Table 5.3. Furthermore, the search identified translational non-crystallographic symmetry (NCS) which is evident in a peak in the Patterson analysis along the 126 Å axis at 0,0.5,0.01.

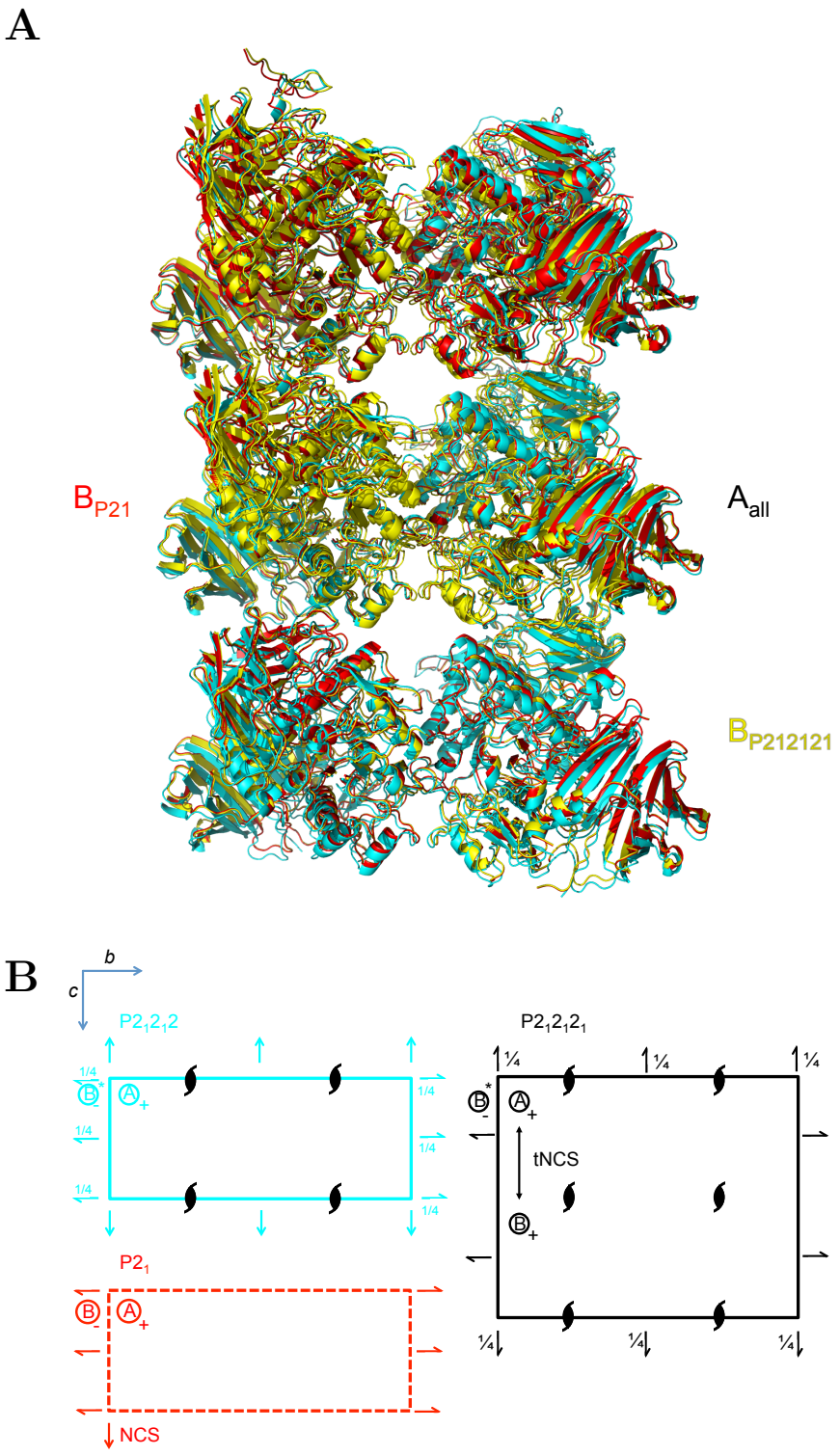


Figure 5.7: Comparison of the monoclinic and two orthorhombic crystal forms. (A) Packing of the *Mm*GluII$_{\text{Trypsin}}$ molecules in the three related crystal from when the origins of each crystal have been translated to coincide. The $P2_1$ form is in red, the $P2_12_12_1$ form is in yellow, and the $P2_12_12$ form is in cyan. (B) A schematic representation of the unit cells along the b-c plane. Symmetry related copies are depicted with the orange asterisk and show the common packing elements between the three forms. Screw (2_1) axes in the plane of the page are indicated by half arrows and normal 2-folds with full arrows. Positive and negative signs indicate positions relative to the plane of the page. Screw axes perpendicular to the page are with the standard ellipse-like symbol.

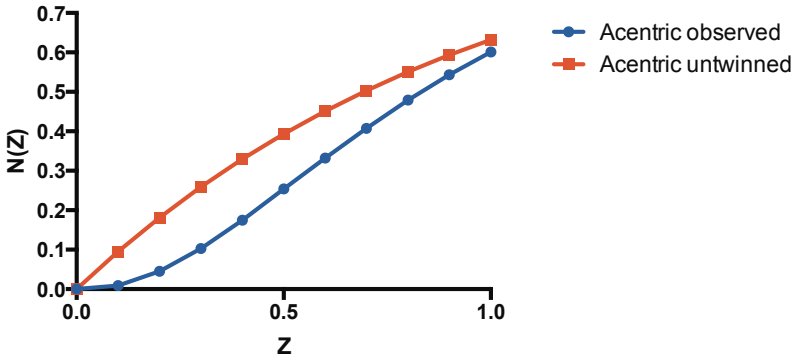


Figure 5.8: Cumulative intensity distribution calculated with Phenix Xtriage on the integrated reflections of the trigonal crystal form.

Table 5.3: Phaser solution statistics for a search with two copies of the 3L4Z search model with the identified translational NCS.

	both copies
Rotation Z-score	9.9
Translation Z-score	15.2
Packing score	22
Log-likelihood Gain	713

5.2.4 Structure building and refinement

The output model from molecular replacement required extensive rebuilding due to low structural homology in the part of the structure upstream from the GH31 domain. An iterative process of autobuilding into density-modified maps by Parrot and Buccaneer followed by manual adjustments and re-assigning sequences in Coot gave the initial model that covers residues G31-P184, P245-N250, and P371-L965 of the α-subunit. Refinement was restrained by the NCS and initial rounds used TLS parametrisation. During this process it became obvious that there was a large amount of density that could not be explained by the α-subunit alone. The N-terminal region of the β-subunit interacts with the α-subunit [86] and structure-based sequence analysis predicted for this region two tandem low-density lipoprotein receptor type-A (LDLRa) domains. Residues 137-161 of the human LDLR (PDB Accession code: 2FCW) were therefore used as a suitable template to build into this density. This was

greatly facilitated by the fold being centred around a Ca^{2+} ion with distinct octahedral coordination by acidic side-chains and backbone carbonyl oxygens. A distinct pattern of disulphides further assisted in assigning the sequence. A re-evaluation of the processing of the original three datasets showed that the monoclinic form was in fact a more appropriate form to work with due to its slightly better resolution and overall processing statistics. The orthorhombic model was used to molecular replace into the monoclinic dataset and was then used for further building and refinement. Once all of the protein was satisfactorily built into the electron density, water and solvent molecules were modelled into the remaining density. The advent of the second, high-resolution orthorhombic form allowed the final building and refinement to be carried out with extra residues becoming resolved in the α-subunits up to R185 and from F30 on the β-subunit. *N*-glycosylation at N97 was modelled with a chitobiose (two GlcNAc residues). The final models were considered complete when the R factors converged to minimum values attainable with acceptable RMSD values for bond lengths and angles around or below 0.01 Å and 1°, respectively. The refinement results are summarised in Table 5.4, Table 5.5, and Table 5.6. All geometry was validated and corrected if necessary based on analyses from the MolProbity server including the Ramachandran analysis (Figure 5.9). The model contains 6 cis-prolines all of which are clearly explained by the density. All except one exist in loops of β-strands with the exception of P489-D490, which places the aspartate into an important position to coordinate a conserved water in the active site. The quality of the modelled ligands can be assured as the B factors of these are similar to the protein.

Table 5.4: Refinement statistics of apo *Mm*GluII$_{\text{Trypsin}}$ model.

	Apo
Resolution range (Å)	59.09-1.74
Observations	109901
Observations in free set	5416
R_{work}[a](%)	14.2
R_{free}[b](%)	16.9
RMSD bond lengths (Å)	0.01
RMSD bond angles (°)	1.07
Ramachandran outliers	0
Ramachandran favoured (%)	97.3
MolProbity clash score	0.72 (99th)
MolProbity score	0.87 (100th)
Good rotamers (%)	96.53
Average B factor (Å^2)	23.0
Residues modelled	Chain A 33-185 Chain A 244-350 Chain A 370-966 Chain B 30-117
Waters modelled	922
Non-protein molecules	2 CA2+, 3 PG4, 3 P6G, 3 EDO, 11 FMT, 2 NAG

Values in parentheses represent the percentiles among structures in the PDB of similar resolution with this score

[a] $R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$

[b] Calculated as R_{work} but with the observations in the free set

PDB small molecule codes. PG4 and P6G: fragments of PEG400. EDO: ethylene glycol. FMT: formate. NAG: *N*-Acetyl-D-glucosamine.

Table 5.5: Refinement statistics of substrate and product *Mm*GluII$_{\text{Trypsin}}$ models.

	Glucal	Glucose	Mannose
Resolution range (Å)	103.10-2.29	60.03-2.67	103.14-2.36
Observations	101524	56122	95813
Observations in free set	5023	2754	4695
R_{work}[a](%)	20.9	21.0	20.0
R_{free}[b](%)	23.5	24.3	22.5
RMSD bond lengths (Å)	0.009	0.01	0.009
RMSD bond angles (°)	1.07	1.12	1.07
Ramachandran outliers	1	1	2
Ramachandran favoured (%)	97.26	97.53	97.47
MolProbity clash score	0.57 (100th)	1.11 (100th)	1.32 (100th)
MolProbity score	0.83 (100th)	0.91 (100th)	0.96 (100th)
Good rotamers (%)	95.97	94.86	94.98
Average B factor (Å^2)	46.8	65.28	44.96
Ligand B factor (Å^2)	51.8	68.5	48.0
Residues modelled	Chain A & C 33-183 Chain A & C 244-350 Chain A & C 371-966 Chain B & D 35-117	Chain A & C 33-183 Chain A & C 244-350 Chain A & C 371-966 Chain B & D 35-117	Chain A & C 33-183 Chain A & C 244-350 Chain A & C 371-966 Chain B & D 35-117
Waters modelled	887	285	588
Non-protein molecules	4 CA2+, 3 EDO, 20 FMT, 4 NAG, 4 BMA, 2 XG3	4 CA2+, 5 EDO, 30 FMT, 2 OXM, 4 NAG, 1 BMA, 2 GLC	4 CA2+, 2 FMT, 4 NAG, 2 BMA, 4 MAN

Values in parentheses represent the percentiles among structures in the PDB of similar resolution with this score

[a] $R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$

[b] Calculated as R_{work} but with the observations in the free set

PDB small molecule codes. EDO: ethylene glycol. FMT: formate. OXM: Oxamate. NAG: *N*-Acetyl-D-glucosamine. BMA: β-D-Mannose. GLC: α-D-Glucose. MAN: α-D-Mannose.

Non-standard codes yet to be determined by the PDB. XG3:

Table 5.6: Refinement statistics of Iminosugar bound *Mm*GluII$_{\text{Trypsin}}$ models.

	Castanospermine	DNJ	*N*B-DNJ
Resolution range (Å)	86.95-1.81	88.90-1.92	42.00-2.10
Observations	103735	87327	61076
Observations in free set	5133	4312	2974
R_{work}[a](%)	16.0	16.5	19.2
R_{free}[b](%)	18.5	19.7	21.3
RMSD bond lengths (Å)	0.01	0.01	0.01
RMSD bond angles (°)	1.05	1.05	1.06
Average B factor (Å^2)	28.55	28.38	31.73
Ligand B factor (Å^2)	17.9	19.1	39.8
Ramachandran outliers	0	0	0
Ramachandran favoured (%)	97.26	97.37	97.37
MolProbity clash score	0.65 (100th)	0.39 (100th)	1.38 (99th)
MolProbity score	0.85 (100th)	0.76 (100th)	0.99 (100th)
Good rotamers (%)	96.53	96.29	96.65
Residues modelled	Chain A 33-185	Chain A 33-185	Chain A 33-185
	Chain A 244-350	Chain A 244-350	Chain A 244-350
	Chain A 370-966	Chain A 370-966	Chain A 370-966
	Chain B 30-117	Chain B 30-117	Chain B 30-117
Waters modelled	886	950	936
Non-protein molecules	2 CA2+, 5 PG4,	2 CA2+, 3 PG4,	2 CA2+, 3 PG4,
	2 P6G, 6 EDO,	4 P6G, 6 EDO,	2 P6G, 4 EDO,
	14 FMT, 2 NAG,	10 FMT, 2 NAG,	8 FMT, 2 NAG,
	1 CST	1 NOJ	1 NBV

Values in parentheses represent the percentiles among structures in the PDB of similar resolution with this score

[a] $R_{work} = \Sigma_{hkl}||F_o| - |F_c||/\Sigma_{hkl}|F_o|$

[b] Calculated as R_{work} but with the observations in the free set

PDB small molecule codes. PG4 and P6G: fragments of PEG400. EDO: ethylene glycol. FMT: formate. NAG: *N*-Acetyl-D-glucosamine. CST: Castanospermine. NOJ: Deoxynojirimycin. NBV: *N*-Butyl-deoxynojirimycin.

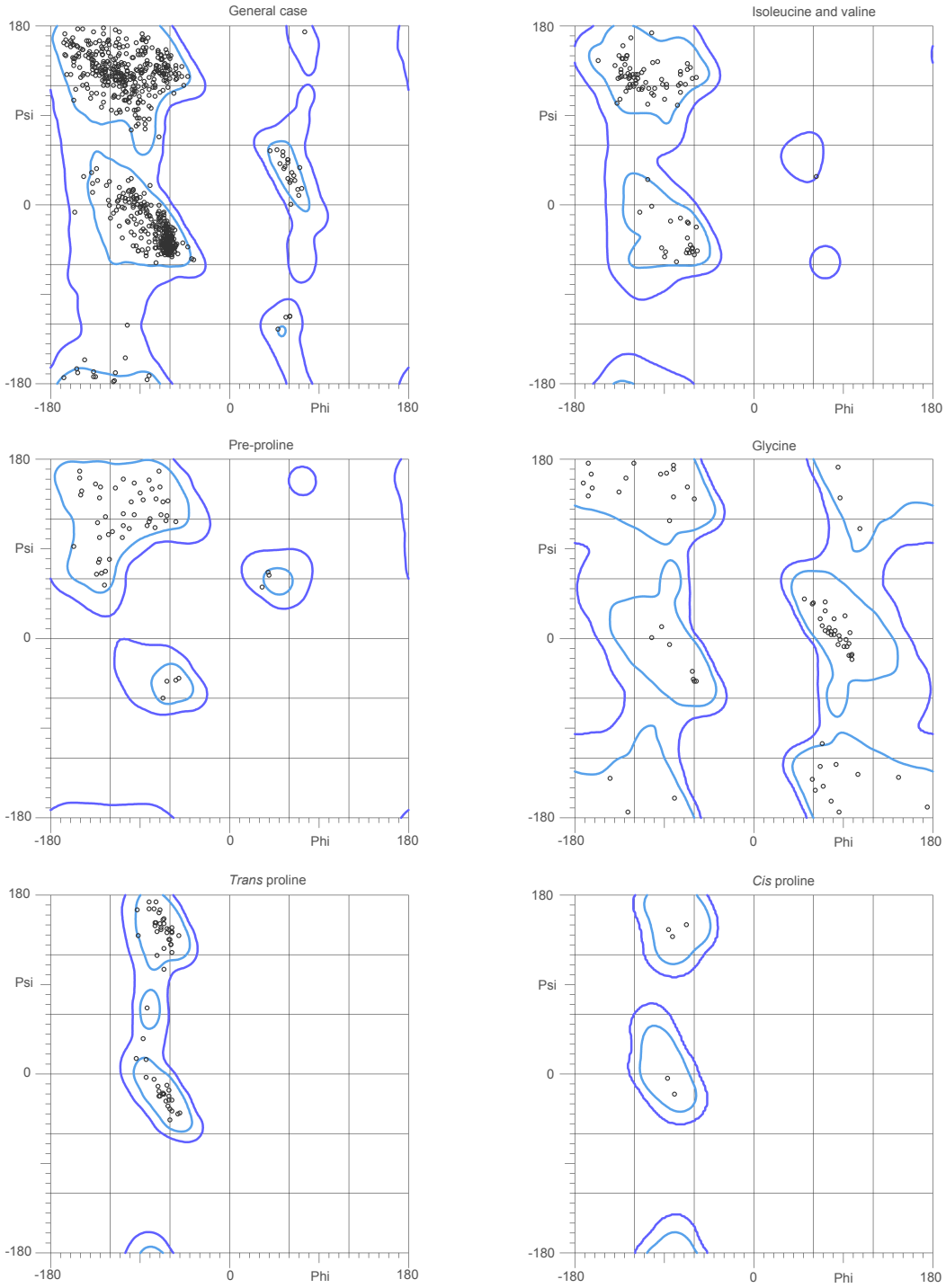


Figure 5.9: Ramachandran plots of the apo *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

5.3 Model analysis

5.3.1 Architecture of *Mm*GluII$_{\text{Trypsin}}$

The crystal structure of *Mm*GluII$_{\text{Trypsin}}$ forms a unique heterodimeric structure composed of the trypsin-resistant portion of the α-subunit (corresponding to the wild-type structure having lost the portions 186-243 and 271-349) and a fragment of the N-terminal section of the β-subunit (Figure 5.10A). The α-subunit is not very different from other reported GH31 enzymes and can be divided into four domains based on the initial description of the architecture of the first reported structure (Figure 5.10B) [69]. Apart from a non-canonical N-terminus which will be described later, the N-terminal domain consists of an eight-stranded antiparallel β-sandwich. The trypsin treatment before crystallisation has removed 2 stretches of sequence from this domain as detailed in the above model building section. The only detectable post-translational modification of this enzyme is the glycosylation that is located in this domain at N97. The GH31 catalytic domain is in agreement with the the other structurally-characterised paralogues and contains an $(\alpha/\beta)_8$ TIM barrel topology with the active site formed by residues at the C-terminal ends of the β-strands. There is an insert of 38 residues between the third strand and third helix that packs against the third and fourth helices of the GH31 domain. A loop from this insert contributes a tryptophan (W525) to the active site. The following two C-terminal domains are termed proximal and distal based on their sequence proximity to the GH31 domain. Both consist of antiparallel β-sandwiches with the proximal domain containing a very short helix-turn-helix motif N-terminal to its β-sandwich. The C-terminal distal domain contributes to the majority of the putative interface between the two subunits and contains a large loop with a turn of helix that packs against the GH31 domain.

The fragment of the β-subunit that is present in the crystal adopts two

tandem LDLRa folds. There is little canonical secondary structure with only a β-hairpin at the start of each motif visible. Each fold is centred around a calcium ion and held together by three highly conserved disulphides.

5.3.1.1 Calcium binding

The two calcium ions modelled into the structure are both octahedrally coordinated by six residues. In both cases, the coordination by the aspartate or glutamate residues occurs in the equatorial plane and the axial coordination is achieved by backbone carbonyl oxygens (Figure 5.11). The biochemical and structural characterisation of these repeats present in human LDLR indicate that calcium is the most likely ion present [220, 221]. To confirm whether the distances observed in the structure that were obtained from unrestrained refinement of the model were genuine, a calcium bond valence sum (CBVS) analysis was undertaken[222]. This bond-valence method uses the distances between the metal and ligand oxygens to determine the CBVS value. An ideal value is 2.0; this is not observed in proteins by the developers of this method, with a mean of 2.23 from approximately 200 structures that were analysed in the PDB. The CBVS values determined from the coordination distances (Table 5.7) are 2.38 and 2.41 for the two sites and lie within the distributions calculated from the PDB and do not coincide with other potential metal ions like Mg^{2+} or Na^{+} adding confidence to our interpretation of modelling calcium ions in the *Mm*GluII$_{\text{Trypsin}}$ model. Alternative methods for identifying calcium binding in this context (*i.e. in cristallo*) would be through the use of X-ray fluorescence or microPIXIE, which is based on a similar principle.

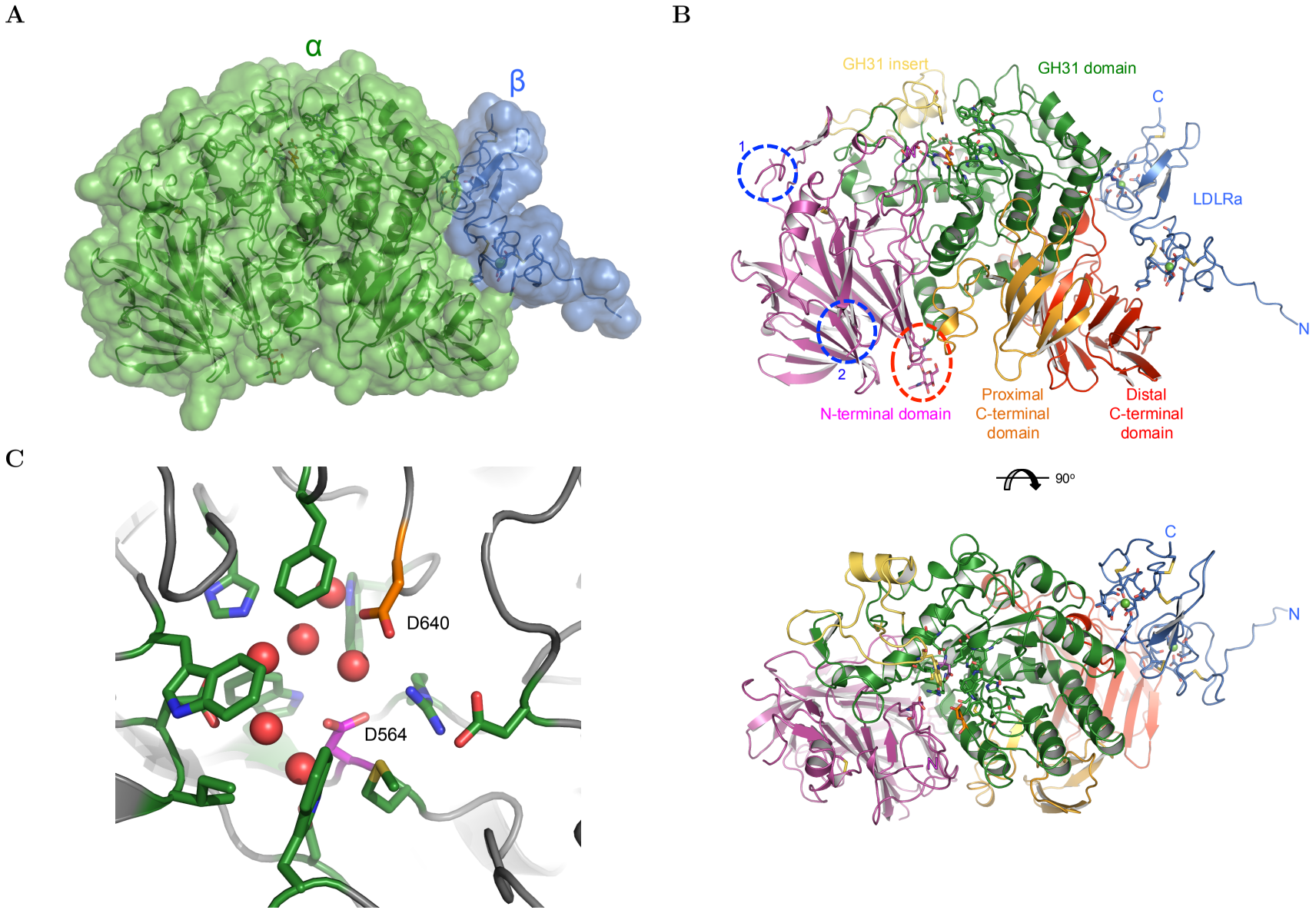


Figure 5.10: Architecture of *Mm*GluII$_{\text{Trypsin}}$ crystal structure. (A). Overview and surface representation of the heterodimer with the α-subunit in green and the β-subunit in blue. (B) Detailed subdivision of the domains of *Mm*GluII$_{\text{Trypsin}}$ view from two perspectives. The α-subunit is composed of 4 domains with a small insert in the GH31 hydrolase domain and theβ-subunit adopting two tandem LDLRa folds. The regions of α-subunit where the two loops were removed by trypsinisation are indicated dashed blue circles (1: 185-243. 2: 351-370). The *N*-linked glycosylation site at N97 with the modelled ManGlcNAc$_2$ is highlighted in the dashed red circle. (C) The active site of *Mm*GluII$_{\text{Trypsin}}$ with the two probable catalytic residues D564 and D640 highlighted in purple and orange, respectively.

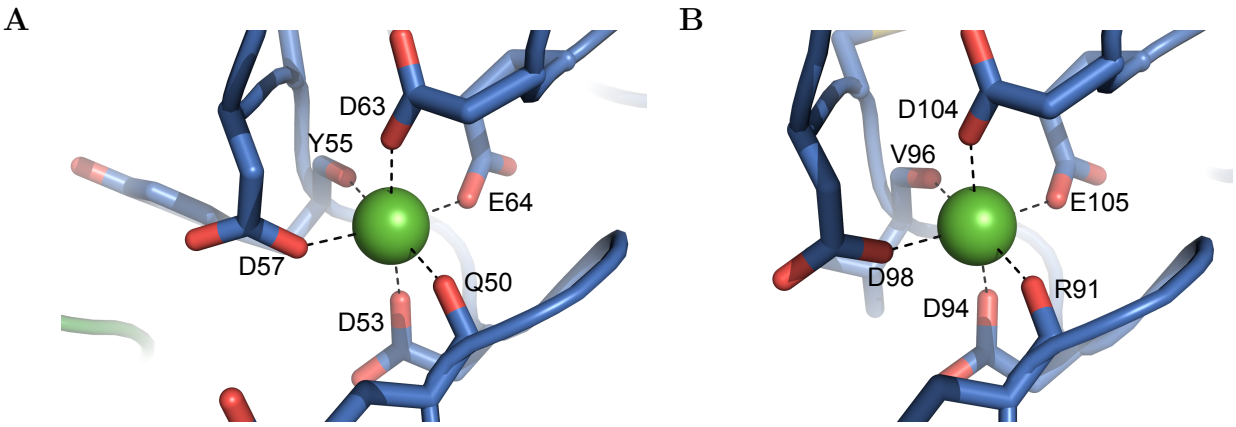


Figure 5.11: Coordination of the two Ca^{2+} ions by the β-subunit for (A) Ca^{2+}998 and (B) Ca^{2+}999

Table 5.7: Distances of calcium ions to coordinating residues and CBVS analysis.

	Ligand oxygen	$d_{Ca^{2+}-O}$ (Å)
Ca^{2+} 998	Q50 O	2.28
	D53 Oδ1	2.29
	Y55 O	2.32
	D57 Oδ2	2.35
	D63 Oδ2	2.31
	E64 Oϵ2	2.31
CBVS[a]		2.38
Ca^{2+} 999	R91 O	2.29
	D94 Oδ1	2.38
	V96 O	2.29
	D98 Oδ2	2.31
	D104 Oδ2	2.32
	E105 Oϵ2	2.25
CBVS[a]		2.41

[a] CBVS $= \Sigma_j(exp(d_0^{Ca} - d_{ij}/b)p_j)$. Where d_0^{Ca} is the bond valence 1.967[223], d_{ij} is measured from the model, b is a 'universal constant' of 0.37 Å[224], and p_j is the occupancy of the ligand which in this case is always 1.0.

5.3.2 The α/β interface

The α/β interface is a unique feature of this structure, a feature that has only been annotated as a cysteine-rich region before. The interface can broadly be described as a network of polar interactions characterised predominantly by salt bridges (Figure 5.12). The majority of these salt bridges are formed by the exposed acidic residues that bind to the calcium ions (D53, D57, D94, D98). The α-subunit contains an array of basic residues (R837, R839, R840, R951 and K952) that take part in these interactions.

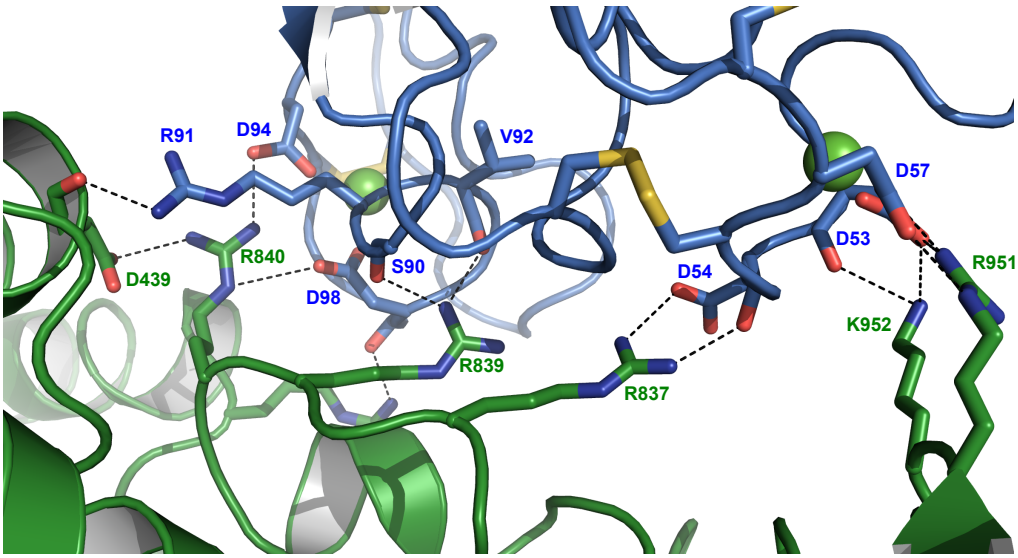


Figure 5.12: The α/β interface. The network of polar interactions between the α-subunit on the bottom (green) and the β-subunit on the top (blue).

5.3.2.1 Analysis of the binding interface

As the crystal itself contains a number of interfaces, analysis of the various possibilities is important to identify the putative α/β interface. The possible interfaces to be assessed are shown in Figure 5.13. In addition to the presented heterodimer (in green and blue), there is a second heterodimer (yellow and purple) that is related along the 2-fold axis with the highlighted interface annotated as α/β'. There also exists

an interface between two symmetry-related α-subunits that is designated α/α" (between green and red α-subunits). An analysis of the interfaces present in the crystal was undertaken on both the PISA [225] and EPPIC [226] servers which use classical thermodynamics or sequence entropy, respectively, to determine the likelihood of a biologically relevant interface (Table 5.8). The PISA analysis identifies both the α/β and α/α" interfaces but not the α/β' one. These results indicate unfavourable solvation energies despite both having a number of polar interactions. When combined with the EPPIC analysis, these two interfaces appear to have a higher likelihood of being genuine and the α/β' interface is classified as only a crystal contact despite a surface area that approaches both the other interfaces. The trigonal crystal form contains the putative α/β and the α/α" interfaces observed in the orthorhombic/monoclinic forms. Combined, these data strongly suggest the α/β interface is likely to represent a genuine interface. The α/α" interface could also be a probable interface, but the *in vitro* data suggest that it is unlikely in physiological circumstances and the only documented example is from analytical ultracentrifugation experiments that suggest a small population of tetrameric GluII [87].

Table 5.8: Analysis of the α/β interface using PISA and EPPIC servers.

		α/β	α/β'	α/α"
PISA results	Interface area (Å^2)	706.9	Not found	843.8
	Solvation energy gain (kcal/mol)	5.8		0.6
	Binding energy (kcal/mol)	-3.1		-4.2
	P-value	0.87		0.65
	Number of H-bonds	13		0
	Number of Salt bridges	16		13
EPPIC results	Interface area (Å^2)	706.7	672.5	894.1
	Interface prediction	Biological	Crystal contact	Biological

Investigation of the putative interface based on the electrostatics of the surfaces reveals the nature of the interface in a more obvious manner. Poisson-Boltzmann methods allow the mapping of the surface charges onto protein models [227]. In the case of *Mm*GluII$_{\text{Trypsin}}$, this was calculated and plotted both onto the heterodimer

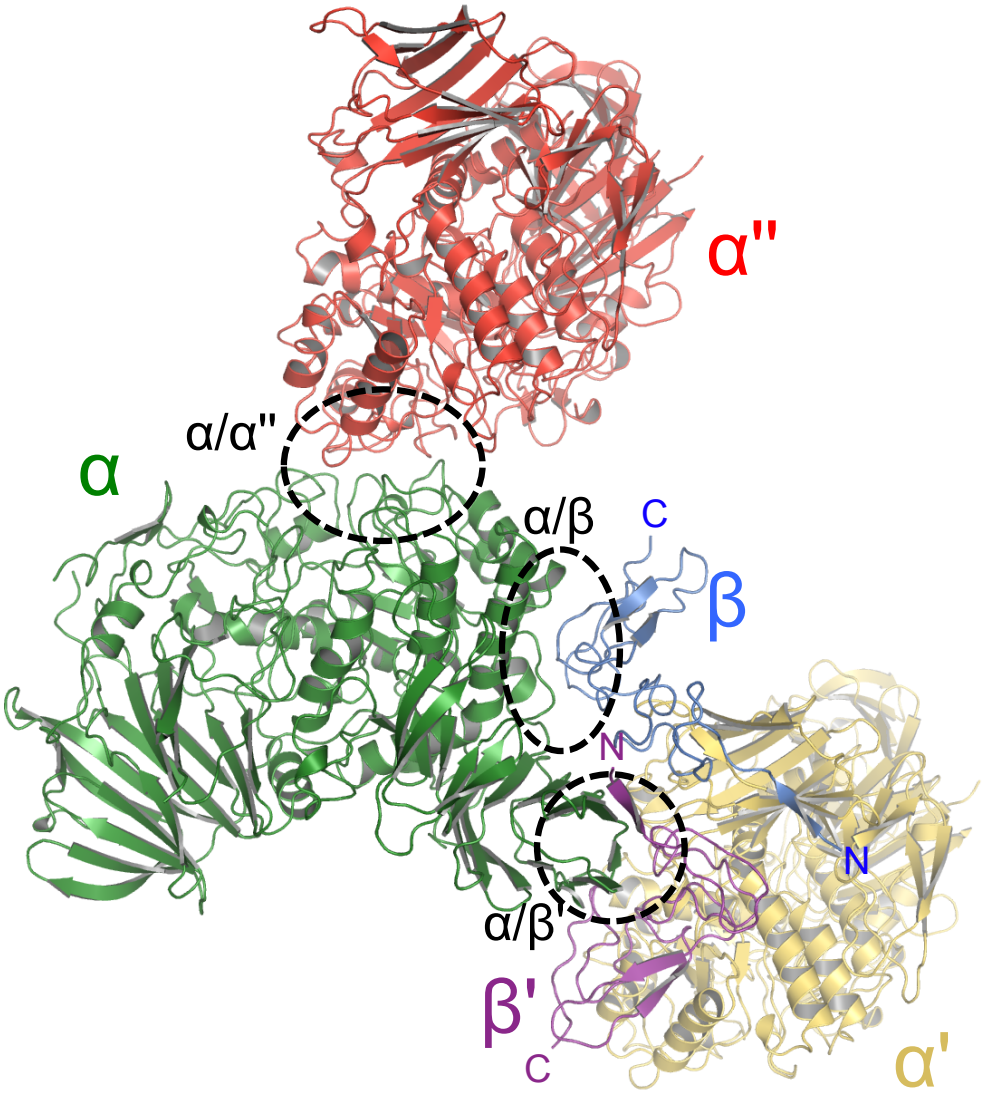


Figure 5.13: Representation of the intermolecular interactions within a reduced portion of the crystal showing three symmetry related copies of the asu.

and also onto the two subunits individually (Figure 5.14). The heterodimer has a predominant negative charge. However, when compared side-by-side, the surfaces of each subunit that interact are starkly opposite in their charge. This is expected by the number of salt bridges that form the interface. To ascertain whether this interface has any evolutionary significance, a surface conservation analysis was conducted using the ConSurf server [228, 229]. The server utilises a multiple sequence alignment (Appendix C) in this case of 29/38 (α/β) orthologues that was generated using Clustal Omega. It can then map the sequence conservation onto the models of two subunits of *Mm*GluII$_{\text{Trypsin}}$ (Figure 5.15). Two regions on the α-subunit appear to be highly conserved, the first is around the active site (Figure 5.15 bottom right panel), the other is at the interface (Figure 5.15 dashed ellipse). The same is seen for the β-subunit (Figure 5.15 dashed ellipse) regarding the interface. These results indicate and support the identity, the likelihood, and the polar nature of the α/β interface.

5.3.3 The N-terminus and substrate specificity

A structural overlay has been performed with all of the published eukaryotic α-glucosidases that have a GH31 domain. Represented here is the comparison of the N-terminal portion of the human intestinal α-glucosidase, which has a specificity for long $\alpha(1,4)$-linked oligosaccharides and has been crystallised with acarbose, a substrate-analogue inhibitor. A closer examination of the active site of *Mm*GluII$_{\text{Trypsin}}$ with respect to the other structurally characterised paralogues gives rise to two observations. The first is that when comparing the active sites there is not much that is remarkably dissimilar between them. This is particularly so at the designated -1 and +1 subsites (Figure 5.16A), which refer respectively to the deeper and closer to surface hexose binding sites on either side of the bond that is hydrolysed by the enzyme, where there are very little observable differences. As *Mm*GluII is capable of cleaving

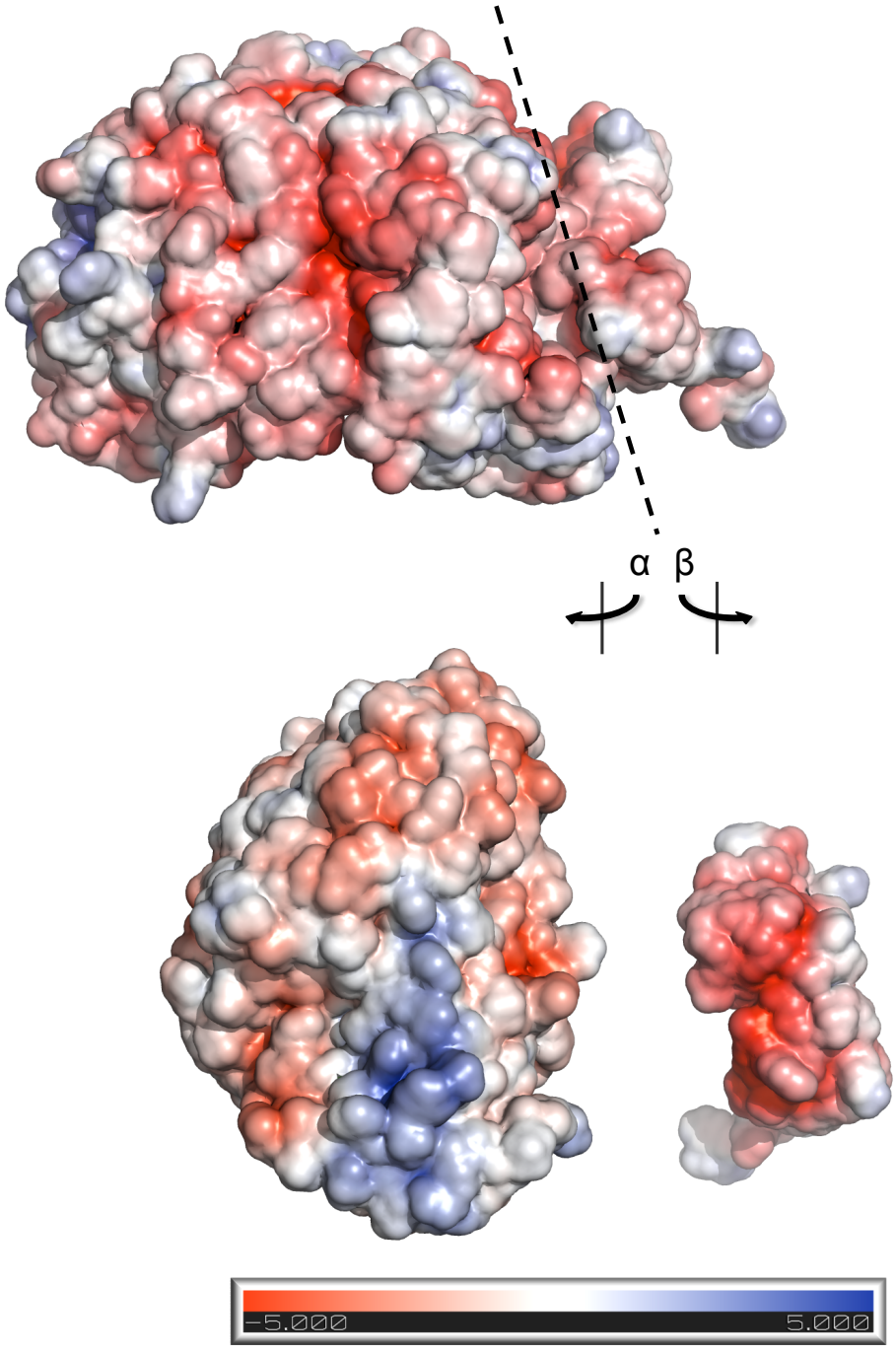


Figure 5.14: Electrostatic potential mapping of the surface of *Mm*GluII$_{\text{Trypsin}}$ using the APBS algorithms part of PyMOL. The top representation is of the overall heterodimer and the bottom representation are the plots over the interface of the two subunits.

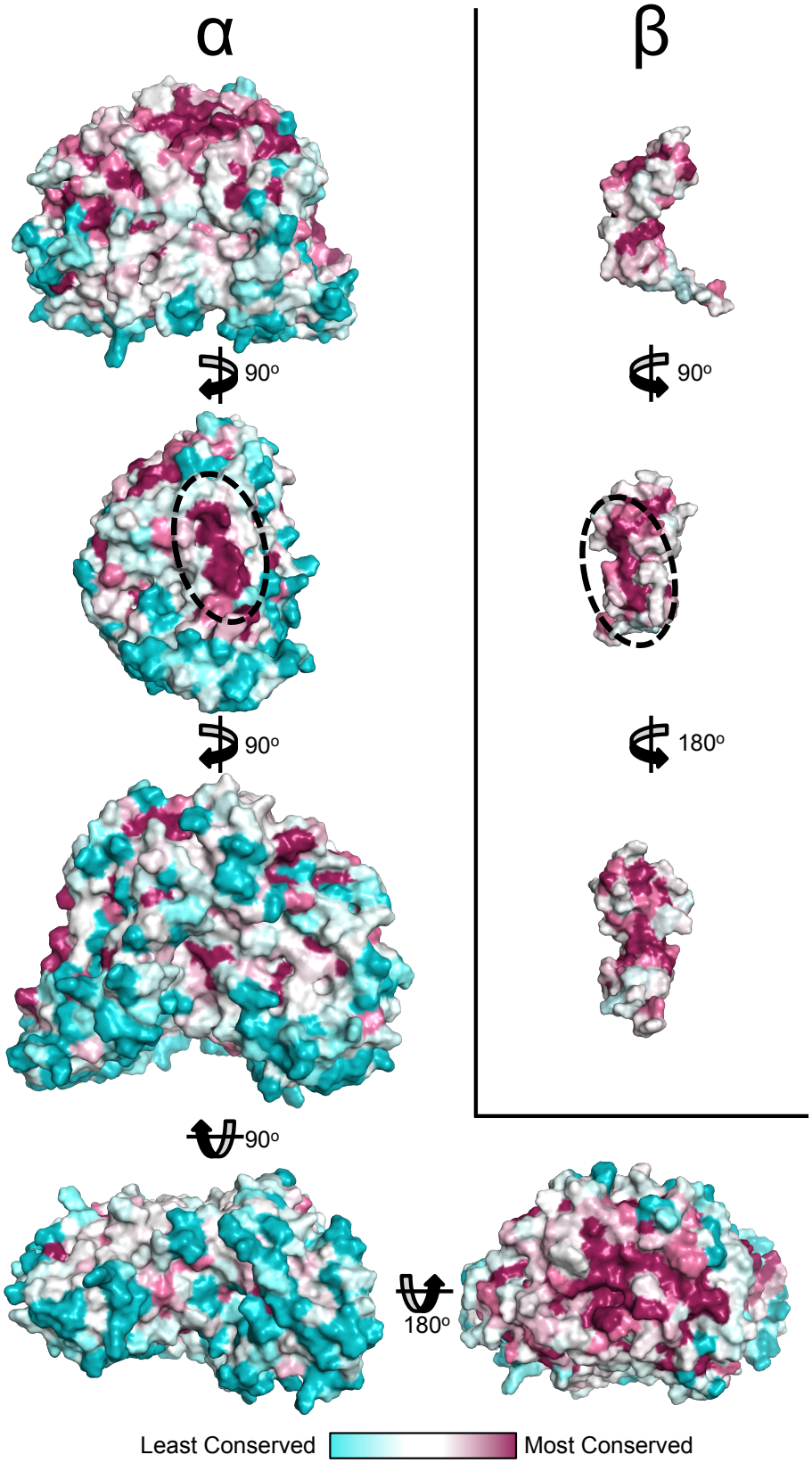


Figure 5.15: Sequence conservation mapped onto the surface of both subunits using the ConSurf server. The region corresponding to the interface on both subunits is highlighted with the dashed ellipses.

disaccharides of α(1,4)-linkages (see 4.6.3), it is reasonable to find that a similar array of interactions can occur. However, there are two key differences that could possibly confer the α(1,3) specificity which are due to W423 (Y to W) and H700 (G/T to H) in comparison to the two maltase-glucoamylase portions. W423 in particular would not be able to accommodate the C-6 hydroxyl group in a α(1,4)-linked disaccharide for the residue in the -1 position. The second observation is on the differences that can be observed proximal to the site of catalysis. The presence of a loop that bulges out in *Mm*GluII$_{\text{Trypsin}}$ (Figure 5.16A, dashed box), which is termed here the specificity loop, is likely to represent a molecular basis for substrate specificity. The loop has two residues, F307 and Q308, which bulge out in a manner that is not observed in other eukaryotic GH31 enzymes. This loop is restricted in its placement by the proximity of the N-terminus of this subunit, a part of sequence which is not found in other α-glucosidases and is very conserved across eukaryotic GluII sequences. The structural overlay shows a clear steric clash of the portion of the acarbose distal to the catalytic site with this specificity loop (Figure 5.16B).

5.4 Ligand binding

The ability to soak compounds into the *Mm*GluII$_{\text{Trypsin}}$ crystals has been optimised in such a way that it is reproducible and efficient. Initial attempts to soak carbohydrates were unsuccessful, which led to soaking very large concentrations of the readily soluble compounds like glucose, mannose, and glucal. Soaking of the iminosugars investigated in the previous chapter was much less successful until a combined change in the cryoprotecting conditions and an increase in pH allowed high-resolution data to be collected reliably.

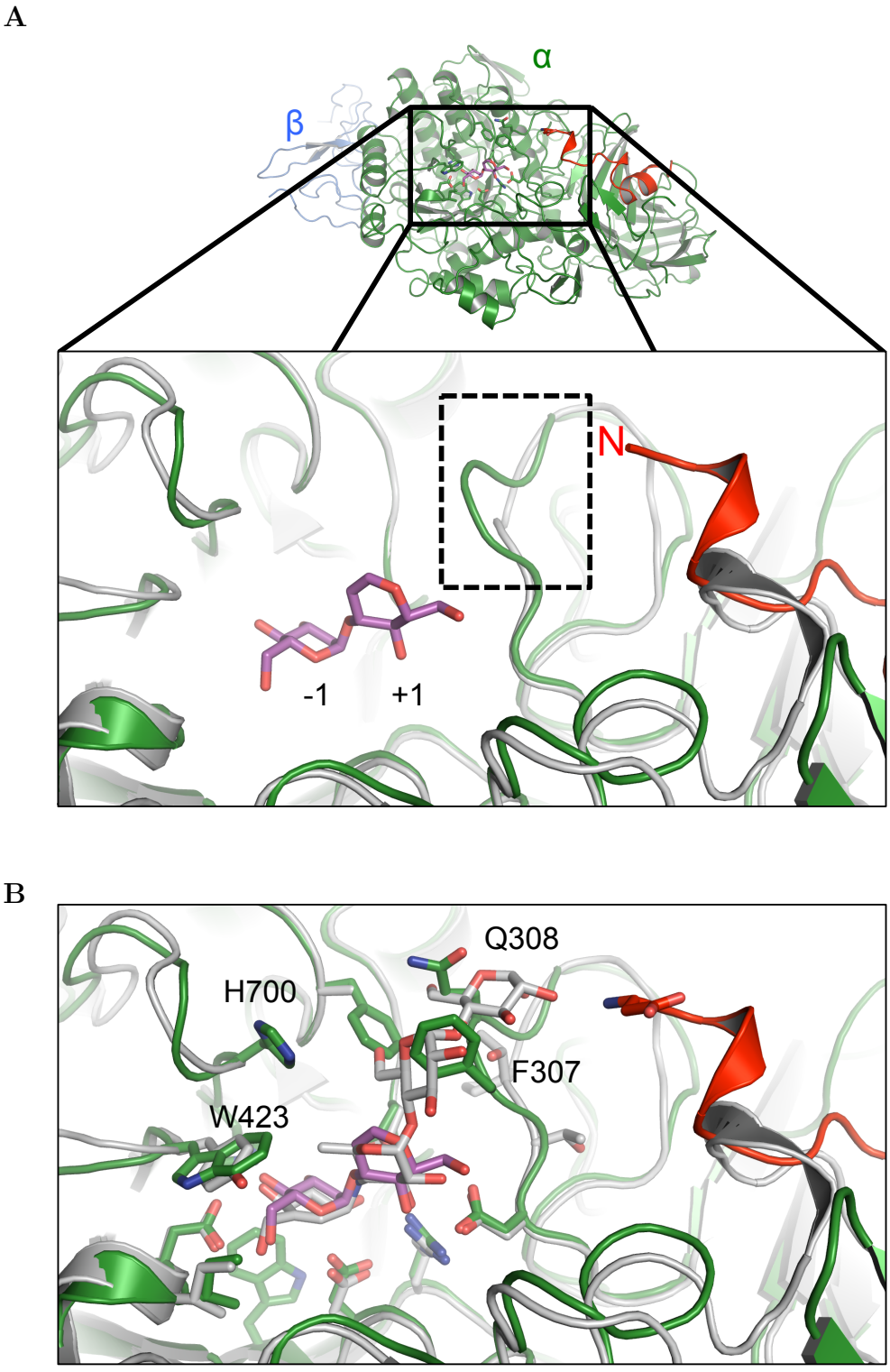


Figure 5.16: Comparison of *Mm*GluII$_{\text{Trypsin}}$ to an intestinal GH31 paralogue. Structural alignment of *Mm*GluII$_{\text{Trypsin}}$ with bound disaccharide analogue (purple) against the maltase domain of *H. sapiens* maltase-glucoamylase (PDB No. 2QMJ). (A) Highlight of the position of the N-terminus in relation to the specificity loop only present in *Mm*GluII$_{\text{Trypsin}}$. (B) Detail of acarbose binding in 2QMJ that is not possible due to steric clashes caused by the specificity loop residues F307 and Q308 in *Mm*GluII$_{\text{Trypsin}}$.

5.4.1 Carbohydrate binding

In an attempt to understand the catalytic cycle and the molecular mode of binding of the species that exist during catalysis, a number of soaks were performed using various carbohydrates. D-Glucal was employed in the hope of trapping an intermediate covalent complex. Coincidentally, a large concentration was used that resulted in the formation of a $\alpha(1,3)$-linked pseudo-disaccharide in the active site. Retaining glycosidases that perform transglycosylation generally follow the Koshland retaining mechanism but the nucleophile is another carbohydrate rather than the water that normally occurs during hydrolysis (Figure 5.17) [51].

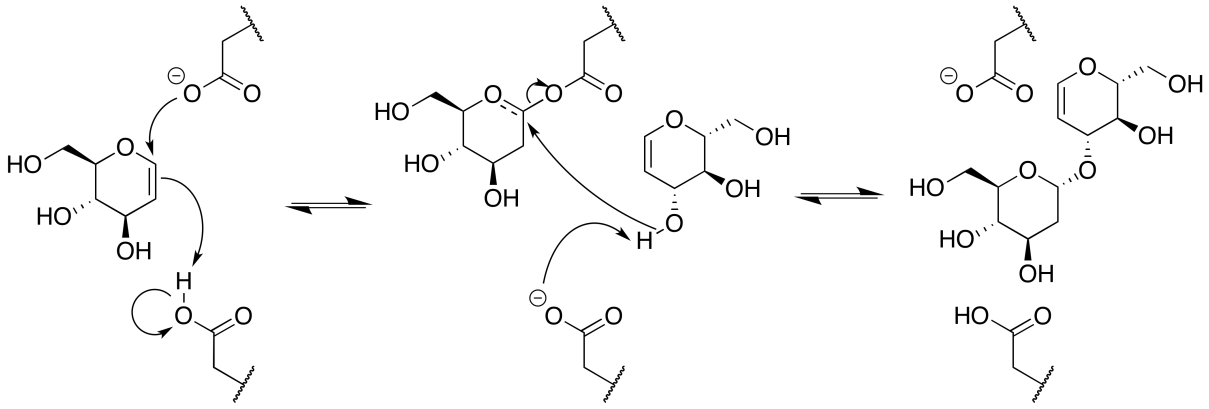


Figure 5.17: Potential mechanism for transglycosylation of D-glucal by another D-glucal in the *Mm*GluII$_{\text{Trypsin}}$ crystals.

The resulting dehydrated pseudo-disaccharide was easily apparent in the initial difference density maps and was modelled in as such (Figure 5.18A). This structure serves as a good comparison for the rest of the soaks as it highlights the many contacts from binding in the -1 and +1 sites. Key interactions are best visualised in the 2-dimensional ligand plot (Figure 5.18B) that show the residues for H-bonding. At the deeper site (-1), these are mediated directly by H698 to the C-3 and C-4 hydroxyls. A series of water-mediated interactions occur with the remaining hydroxyls through the residues D490, D564, D640 and Y637. Interactions at the +1 site are again mediated by conserved waters to D305 and R624. It must be noted that there is a

possibility that the binding mode at this site could be rotated around the glycosidic bond by 180°. As there are two hydroxyls missing from this glucal moiety, the current binding mode suggests a more favourable enthalpic gain of forming the H-bonds that are seen here but would perhaps bind more favourably to H700 identified in the previous section (see Figure 5.16).

On the opposite side of the catalytic cycle is the product bound state. In this case, this is represented by the soak that was performed with D-glucose. Unfortunately, the resolution of this structure is moderate (2.7 Å), so the modelling of the molecule was not very accurate but can be related to the glucal soak. Furthermore, a number of the waters that mediate some of the binding interactions were poorly resolved. The crystal structure does agree with a similar mode of binding to the glucal soak above (Figure 5.18C). The presence of the hydroxyl at the C-2 position enables the interaction that is absent in the previous soak to be determined as a H-bond to R624 (Figure 5.18D), the same arginine that binds to a hydroxyl of the other carbohydrate residue in the +1 position.

The appearance of D-mannose in the -1 site is surprising but could be an artefact due to the high concentration that was used when soaking into the crystals. It is nevertheless of note that the only difference between mannose and glucose is the stereochemistry at the C-2 position which is axial for mannose. This appears to be accommodated though that hydroxyl does not appear to interact with anything while the remaining interactions are as expected (Figure 5.18E and F).

5.4.2 Iminosugar binding

The development of a procedure to soak inhibitors into the *Mm*GluII$_{\text{Trypsin}}$ crystals that also increases the diffraction resolution has now opened up an array of experiments that can be attempted in order to understand the binding of various iminosug-

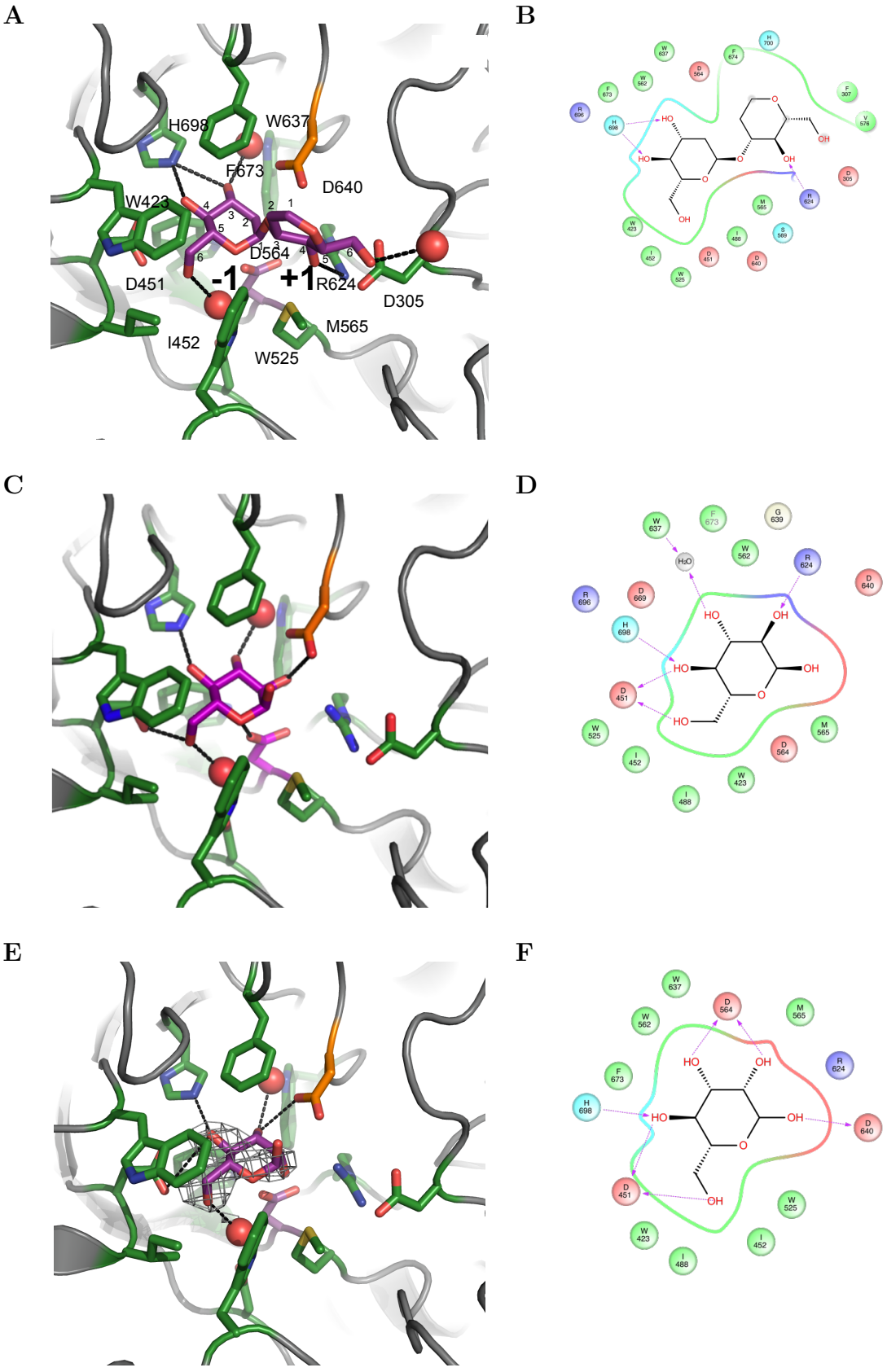


Figure 5.18: Carbohydrate binding in *Mm*GluII$_{\text{Trypsin}}$. (A and B) Structure and 2-dimensional plot of pseudo-disaccharide transglucosylation product of D-glucal soak. (C and D) Structure and 2-dimensional plot of D-glucose soak modelled as the α anomer. (E and F) Structure and 2-dimensional plot of D-mannose soak modelled as the α anomer. Map for the mannose represent unbiased F_o-F_c contoured at 3 σ prior to modelling the ligand. 2-dimensional plots generated in Maestro.

ars as well as any new chemical series. Of comparison here are castanospermine versus DNJ and its alkylated derivative *N*B-DNJ (Figure 5.19).

The three soaks all present very similar molecular interactions similar to those seen in the glucose and glucal soaks. A series of direct H-bonds to the hydroxyls of all but the C-6 hydroxyl are formed by R624, H698 and D451 moving from C-2 through to C-4. Two conserved waters form H-bonds that are coordinated by a number of aspartate and tryptophan residues. The rigidification of the castanospermine by the formation of the five-membered ring appears to be accommodated by a hydrophobic patch of the molecule in a pocket formed by W423, I448 and W525. The alkylated tail of *N*B-DNJ is poorly resolved and has been modelled here based on the weak difference peaks. After modelling in the butyl chain, very little density appeared but neither did any negative difference density appear to contradict the model.

5.5 Discussion

The availability of the trypsinised *Mm*GluII fragment has been an invaluable tool in the structural elucidation of the enzyme. Presented here is the expansion of the initial identification of the catalytically intact fragment that was identified from tissue-derived rat GluII [87]. What was identified as the GH31 catalytic domain is present with the same N-terminus in *Mm*GluII$_{\text{Trypsin}}$. However, other associated fragments are visible as seen by mass spectrometry, light scattering and X-ray crystallography. Not shown in that study is the lower part of the gel, which would likely have confirmed the presence of bands below about 30 kDa. Furthermore, the ability of the enzyme to still retain catalytic activity after treatment by a number of proteases indicates a very compact and stable core of the enzyme. The characterisation of the kinetics of

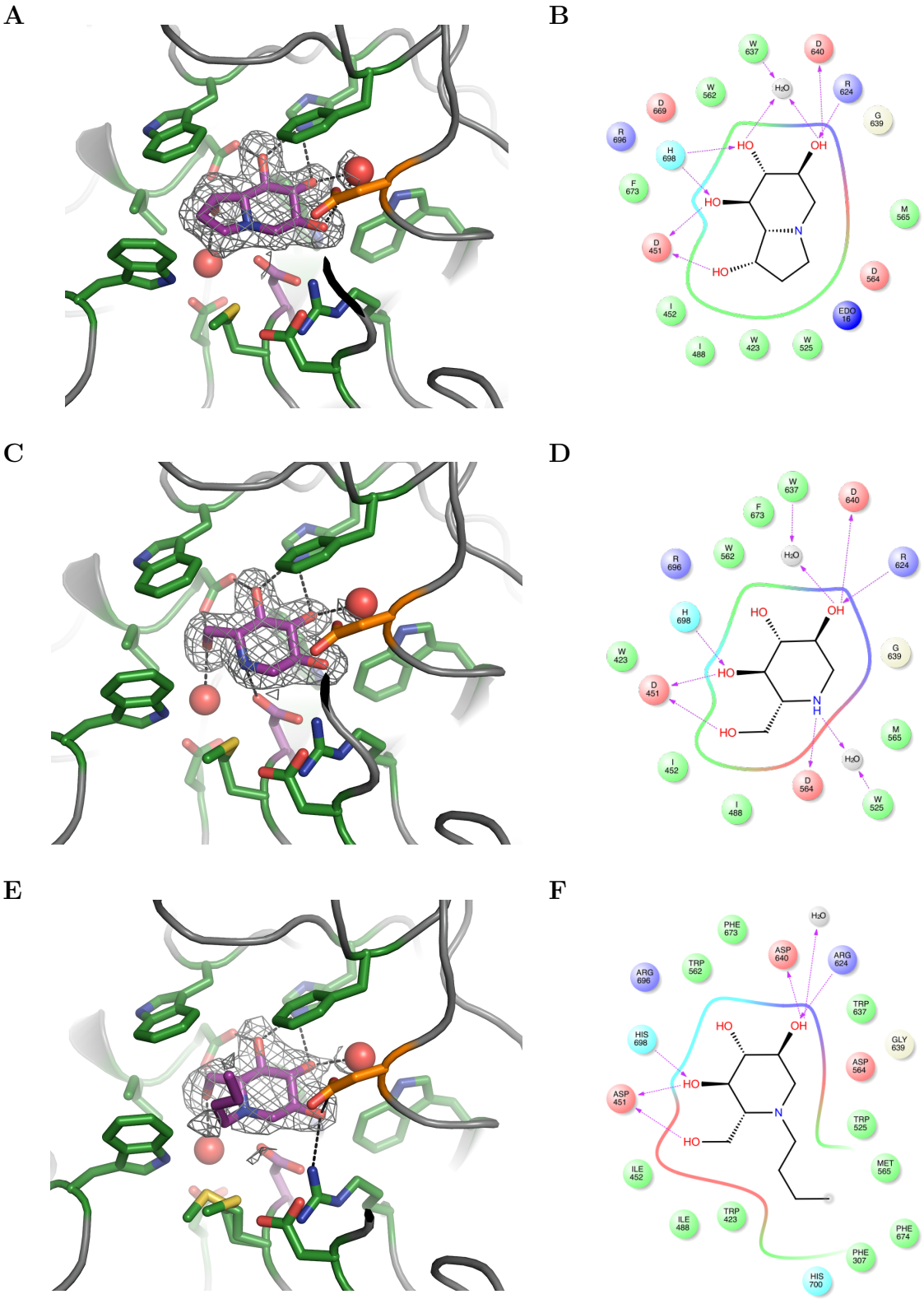


Figure 5.19: Iminosugar binding in *Mm*GluII$_{\text{Trypsin}}$. (A and B) Structure and 2-dimensional plot of castanospermine soak. (C and D) Structure and 2-dimensional plot of 1-deoxynojirimycin soak. (E and F) Structure and 2-dimensional plot of *N*-butyl-deoxynojirimycin soak. Maps represent unbiased F_o-F_c contoured at 3 σ prior to modelling the ligand. 2-dimensional plots generated in Maestro.

4MUG cleavage highlighted an observation that may have otherwise been overlooked if only the change in activity between the full-length and the *Mm*GluII$_{\text{Trypsin}}$ had been quantified. The K_m is identical between the two species but there is a reduction of the k_{cat}, which indicates that there is a small change in the binding site. As the K_m is the same, there must be some compensatory changes in the rate constants for association or dissociation of the enzyme-substrate complex. Whether this results in a dramatic change in the active site is difficult to determine. The comparison of the obtained structure to other GH31-containing enzymes does not indicate that this may be the case. A possible avenue for further investigation is to conduct the kinetic measurements on disaccharide substrates and the determination of inhibition constants on *Mm*GluII$_{\text{Trypsin}}$ in order to assess if any changes have affected binding to other substrates or inhibitors.

The sites cleaved by trypsin are indicative of the validity of the predicted disorder calculated for both subunits. The strategy undertaken here was based on the previous characterisation of proteolysed GluII as well as the adaptation of *in situ* proteolysis, which was unsuccessful, but gave the impetus for further investigation of the trypsinised fragment. The α-subunit must contain two accessible loops that are cleaved, but the overall complex of the remaining three portions is resistant to further proteolysis. The interface between the two subunits with the compactness of the LDLRa domain must have also prevented the destruction of this interface. However, it seems that the rest of the β-subunit does not contain much compact tertiary structure. The exception being the C-terminal domain of the β-subunit, which most likely, is still intact, but is not part of this complex.

The crystallography conducted in this chapter is considered routine with little deviation from established practice. The data reduction of the initial data was cut to the chosen resolution based on the criteria that primarily utilises correlation

coefficients over R factors. These decisions are consistent with the current discourse of data processing practices [230, 231]. R_{meas} is a critical metric for the discriminating between the space groups when comparing data processed to the same resolution [219]. The closely-related monoclinic and orthorhombic forms changed on the addition of compounds or cryo-preservation so comparison of R_{meas} when processing in both of the different symmetries was indispensable in making the correct decision of symmetry. The slight alterations in symmetry are likely reflective of the subtle changes that occurred upon altering the composition of the solution in and around the crystals but do not appear to have any deleterious effects on the enzyme. On the contrary, these changes have led to an increase in the resolution probably through dehydration [232–234]. The end result is that ligand binding can be observed as well as more order at two of the termini formed by trypsinisation. The trigonal form was processed and a molecular replacement solution was obtained. However, due to the difficulties that the twinning introduced, this form was abandoned in favour of the monoclinic/orthorhombic forms.

The resulting structure of the α-subunit is remarkably similar to other members of the GH31 family in terms of domain organisation. The bioinformatical annotation of the α-subunit sequence is divided into three domains (Figure 1.7A) but the structure more closely resembles those of the other published GH31 members [69, 73, 89, 90, 92, 235, 236]. The main variations between the known structures occur in the size of the insertions in the GH31 domain between β3 and α3 which are much smaller in the eukaryotic structures and tend to be the basis of oligomerisation in the bacterial α-glucosidases[73]. Between the eukaryotic structures, there is very little difference in the overall architecture across the domains with the exception of the N-terminus where most of the differences can be seen. The two segments of the maltase-glucoamylase both contain trefoil Type-P domains at the N-termini that are lacking in *Mm*GluII or the sugar beet α-glucosidase. As sequence divergence is oc-

curring at this region of the enzyme, it raises the question of the role of the loops that have been removed by trypsinisation. Might these play a role in binding regions of the β-subunit or other relevant proteins in the ER lumen? Might they have some long-order interactions with distal regions of the glycan or the protein around the glycosylated asparagine? Neither of these questions can be addressed with these structures, but highlight the contribution that a glycan-bound structure would add to our understanding of GluII.

The observed α/β interface identified in the *Mm*GluII$_{\text{Trypsin}}$ structure is likely to be genuine. The N-terminal region of the β-subunit that interacts with the α-subunit is consistent with one of the two distinct α-subunit interacting regions identified on the β-subunit [86]. The nature of the ionic interactions does not appear to be dissimilar to that of the LDL receptor. In the case of LDLR binding to ApoB or ApoE, a similar type of the receptor-ligand interaction has been observed between the LDLRa domains and stretches of basic residues on ApoB [237–239]. The level of conservation on both subunits of *Mm*GluII coupled with the electrostatic mapping indicate that this is a genuine interface. The buffer screening in the previous chapter (4.5.2) support the type of interface. Both the pH and ionic strength have marked affects on the overall stability of the complex that are likely linked to disruption of the α/β interface. A concerted effort is currently being undertaken to validate this interface. Our collaborators are engineering multiple point mutations to remove the arginine and lysine residues to complement into a GluIIα-deficient *Arabidopsis thaliana* line [240]. A GFP-labelled mutant of the α-subunit that removes a number of the basic residues has been engineered. Thus if the interface is disrupted, the mutant α-subunit should not be localised exclusively in the ER lumen. The appearance of tetrameric GluII *in vitro* that is likely caused by artificially high concentrations may be formed by the interaction of two α-subunits seen in the crystals. This sort of interaction might be able to be tested via an enzyme activity experiment whereby

reduced activity should be observed as the interface occurs over both active sites. Regardless, in all of the light scattering experiments, the $Mm\text{GluII}_{\text{Trypsin}}$ species does not appear to form higher order oligomers.

With regard to the other putative region of the β-subunit at its C-terminus that would bind to the α-subunit[86], this is likely to correspond to an interaction of the MRH domain with some location on the α-subunit. Apart from the identified surface conservation that corresponds to the observed α/β interface in the $Mm\text{GluII}_{\text{Trypsin}}$ structure, the only other conserved patch on the surface is around the active site. There may be a lower affinity interface between the β-subunit MRH domain and this region. Although it would appear to be transient as the trypsinolysis did not appear to destroy the MRH domain, it has not been carried through the subsequent purification. A possible route to obtaining a species that shows this interaction would most likely be binding the substrate glycan to an inactive point mutant before trypsinolysis in order to preserve such a complex. There is an extra layer of complexity that must be appreciated when trying to address both the nature of the glycan binding and possibly the remainder of α/β interactions which is due to the two cleavage events the enzyme catalyses. The implication being that after the first cleavage, the glycan likely dissociates to reorient the terminal glucose and perhaps the binding mode of the glycan to make the second cleavage [62, 84, 88]. GluII is at the centre of a series of competing forces: the affinities of the enzyme between the two substrates it can cleave; interactions with malectin for the di-glucosylated form; and interactions with either calnexin or calreticulin for the mono-glucosylated form. Together this illustrates the complexity of the kinetic forces that govern its role in the calnexin cycle.

A closer look at the active site with the various soaks presented here allows a closer dissection of the molecular determinants of substrate specificity. In particu-

lar the glucal soak which has fortuitously produced a pseudo-disaccharide substrate analogue can shed some light onto these matters. The fact that an α(1,3) glycosidic bond was formed indicates that regiospecificity might be conferred by W423 and H700. These represent interesting targets for mutagenesis to further investigate this possibility and have a literature precedent in a *Bacillus* GH31 containing enzyme [212]. By themselves, these two residues do not completely prohibit the binding and cleavage of α(1,4)-linked disaccharides. Although the mode of binding that is present in the glucal soak does not favour these hypotheses, it is possible this is due to the artificial nature of this dehydrated molecule. That is to say, the lack of two OH moieties in the glucal ring at the +1 site may determine the observed mode of binding. Based on the observed binding mode, the direction in which the glycan extends would be at an angle (Figure 5.20 upward blue arrow). However, it is possible that the saccharide in the +1 site could bind 180° to that observed in this structure. The result of such a flip could involve H700 in hydrogen bonding of either the C-4 or C-6 hydroxyls in a α(1,3)-linked disaccharide for the residue in the +1 position. Another consequence of the same flip would be that the glycosidic bond from C-1 of this residue to the subsequent mannose would extend in a linear fashion (Figure 5.20 horizontal blue arrow). This mode of binding would be compatible with the solution NMR structure of $Glc_3Man_9GlcNAc_2$, which indicates a rather linear structure of the glycan [62]. Coupled with the presence of the specificity loop comprising residues F307 and Q308, there appear to be clear determinants for specificity that occur outside the catalytic site. These observations, in addition to the fact that no α(1,4)-linked polysaccharides are found in the ER, lead to the hypothesis that there has been little evolutionary pressure for the active site of GluII to be exquisitely specific for α(1,3) over α(1,4) glycosidic bonds.

The only observed GANAB missense mutations that are not lethal appear in *A. thaliana* [241]. Of the two mutations described, S569F corresponds to the middle

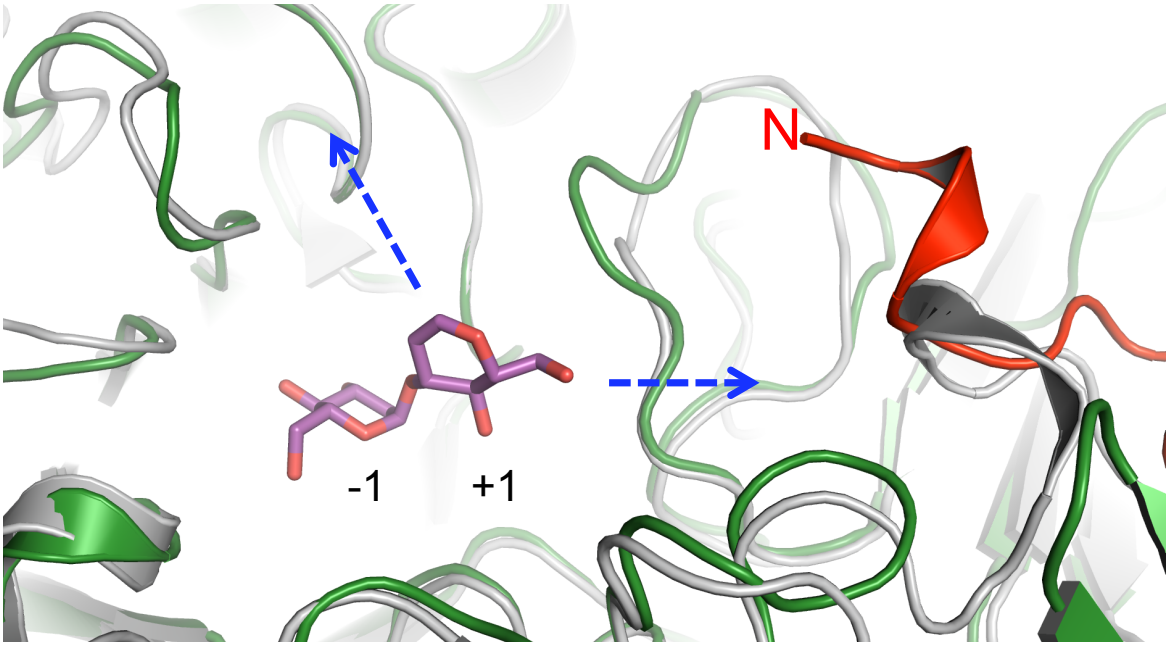


Figure 5.20: Active site of the structure containing the pseudo-disaccharide substrate analogue indicating the possible directions of the non-reducing end of the rest of the glycan.

of one of the α-helices in the GH31 domain, likely conferring a disruption to the active site through an allosteric perturbation. The other, S652F is proximal to the active site where the glycan substrate is hypothesised to extend. Whether this mutation disrupts proper binding through a lowering of the affinity or some other mechanism is difficult to assess without access to a structure of GluII with the glycan.

The other carbohydrate-bound structures give an insight into the binding interactions in the -1 site. The glucose soak mimics one of the two components of the hydrolytic products of this enzyme. As expected glucose makes a series of interactions conserved within the GH31 family as well as preferring the α-anomer despite the β-anomer being more abundant in solution. Perhaps not too surprising is the mannose soak which shows that at high concentrations a less favoured monosaccharide can be forced into the active site. The accommodation of mannose in the active site leads to the consideration of the enzyme's ability to bind the other epimers of glucose. Allose, a rare monosaccharide that is inverted at the C-3 position is unlikely to bind due to the proximity of a phenylalanine that would disfavour a hydroxyl in the axial

position. The epimers at C-4 and C-5, galactose and L-idose, respectively, would also be unfavourable for similar reasons.

A structure of a catalytic intermediate would definitively identify the nucleophile and complete the series of structures of the catalytic cycle. Generally this is done with fluorinated sugars, namely 2-deoxy-2-fluoro-α-D-glucopyranosyl fluoride or 5-fluoro-α-D-glucopyranosyl fluoride. These sugars readily react with the nucleophile due to the efficiency of the fluoride as a leaving group but the second fluoride significantly retards the subsequent hydrolysis of the covalent adduct [242].

A more thorough understanding of the molecular basis of specificity of substrates and various stereochemistries that can be accommodated in the active site are necessary for future inhibitor development. Furthermore, as the active sites are closely related between the different GH31 members in the proteome, this is crucial for the production of a more selective inhibitor. The protocols established here are invaluable in assessing inhibitor binding and the variations observed *in vitro* of the different characterised iminosugars presented in this thesis. More specifically, the approximately 5-fold increase in potency between castanospermine and DNJ are difficult to rationalise based on these structures. The network of H-bonds is almost identical and only the shift of W525 near the five-membered ring of castanospermine could be correlated with the decrease in potency due to a strain induced upon binding. However, this would be contrary to the increase in potency that is concomitant to *N*-alkylation of DNJ which is likely to induce a similar strain on the loop containing W525. The alkylation length is known to be proportional to an increase in potency due to greater bioavailability [100, 243], though its effect on isolated GluII inhibition is less pronounced. The aliphatic chain on the ring N atom may not adopt a single conformation and in the case of *N*B-DNJ this appears to be the case and only further structures with alkylated iminosugars will help clarify this point. A closer inspection

of the water structure in the active sites of the various GH31 enzymes might also provide a clue as to new directions of drug design if any differences are apparent. Another avenue that could be pursued for increasing inhibitor selectivity is tailoring more interactions at the +1 site and beyond where the majority of the differences between the homologous α-glucosidases can be observed.

6

Conclusions and future directions

The work presented in this thesis is a significant step toward a more detailed molecular understanding of the function of the ER glucosidases. The parallel production of a large number of constructs was an invaluable and necessary strategy for maximising the probability of success in identifying suitable constructs to take forward for the biochemical and biophysical characterisation.

6.1 Glucosidase I

A soluble and active mammalian GluI produced at suitable scales made it amenable for the described experiments. Its solution properties indicate the ability to non-covalently form dimers, a phenomenon that is worth further investigation as it may have a relevant but unidentified biological significance. The specific activity for triglucosylated glycans confirms the identity of the enzyme. As such this enzyme is a promising resource for screening future inhibitors and may also be used to develop a more high-throughput assay suitable for a larger screening campaign.

The work towards the structural characterisation indicates promising future avenues of study. The current crystal form is a good basis for future optimisation with a number of viable techniques yet to be utilised. The important questions that a high-resolution structure of *Mm*GluI could address include if there are any differences between *Sc*GluI and *Mm*GluI structure, and what impact these differences have on the enzyme's function, as well as from a pharmacological standpoint, on further inhibitor design efforts. The latter is of importance for minimising the off-target side-effects of intestinal glucosidase inhibition as well as for identifying the basis of any selective inhibitors between the two ER glucosidases.

6.2 Glucosidase II

We have for the first time observed the α/β interface of GluII, which is key to the localisation of the complex in the ER lumen. We are in the process of validating it *in vitro* by monitoring the changes in cellular localisation of a GFP-fusion of the α-subunit mutated in order to abrogate its interaction with the β-subunit. A second interface has been suggested in the literature [86], though in our hands it would appear the affinity is not high enough to survive the trypsinolysis.

This first structure of the GluII active site presented here now provides new testable hypotheses. The subtle differences with respect to other GH31 family enzymes provide the details that can be pursued to explore substrate specificity. The residues that likely confer the regiospecificity at the glycosidic bond are W423 and H700 and with the established mutagenesis protocols this can readily be tested in the near future. Likewise the identified specificity loop can be targeted to assess if its truncation allows acarbose inhibition or malto-oligosaccharides to be cleaved. It is likely due to its subcellular localisation that GluII has only had to evolve to select between α(1,3) and α(1,2) bonds, of which it is indeed more discerning. However, its ability to cleave α(1,4) bonds is not surprising considering its common ancestry to enzymes that cleave these bonds. Together, these hypotheses can be used to address the broader question of whether substrate specificity is due to the molecular features proximal to the active site.

These questions could be more easily answered if a structure of GluII trapped with a physiologically relevant substrate became available. Procuring these glycans is not trivial, but there is ongoing work in the Zitzmann laboratory that will assist this line of investigation. These experiments would also provide more definitive answers to the question of how the β-subunit helps to recognise or place the substrate into the active site. Furthermore, any quaternary structure changes that occur upon glycan binding can be assessed through solution-based techniques like SAXS. Understanding the stoichiometry of glycan binding would help discern which of the *cis-* or *trans-*activation models for cleavage are more likely to be true. Along these lines, the questions on the discrimination between the two cleavage events of GluII are still unanswered.

Designing inhibitors for this very polar catalytic pocket has an obvious caveat in that they generally can be tailored to be very potent and selective but

due to their pharmacokinetic profiles have negligible bioavailability to the ER lumen. Drastic increases in the potency of binding may not necessarily be optimal as this is likely to result in unacceptable cytotoxicity. What must always be considered in the design of antivirals that target these enzymes is that the eventual assessment of their suitability must come from cell-based assays. Potency must be measured in terms of its effect of viral infectivity or secretion and inhibition of the glucosidases through fOS measurements. Information from the atomic to the cellular levels will aid development and optimisation of leads in the future.

An alternative route to designing antivirals that are not based on the iminosugar scaffold could now be undertaken with the information presented in this thesis. Searching through a larger chemical space through fragment-based screening or a more traditional high-throughput screen might provide the chance to identify non-competitive/uncompetitive inhibitors. A different sort of inhibitor might increase the selectivity of inhibitors as well as afford more favourable pharmacological profiles.

In conclusion, the work presented here provides a significant step forward in our understanding of these glucosidases on the molecular level and contributes to the future of broad-spectrum antiviral development.

Appendices

A

Primers for high-throughput cloning

Table A.1: Primers used for high-throughput PCR

Construct	Primer name	Sequence (5'-3')
A1	F	AGGAGATATACCATGTTTATTCTTCTTCATCTATATCTCTACCCAAATCTTG
	R	CAGAACTTCCAGTTTAGTGTCCAAGTTATCACTCATAATCAAAAGG
B1	F	AGGAGATATACCATGTGTAATCAAGATGATCGTCTGCCTGG
	R	CAGAACTTCCAGTTTAGTGTCCAAGTTATCACTCATAATCAAAAGGATTAG
C1	F	AGGAGATATACCATGTTTATTCTTCTTCATCTATATCTCTACCCAAATCTTG
	R	CAGAACTTCCAGTTTAGTGTCCAAGTTATCACTCATAATCAAAAGG
D1	F	AGGAGATATACCATGTGTAATCAAGATGATCGTCTGCCTGG
	R	CAGAACTTCCAGTTTAGTGTCCAAGTTATCACTCATAATCAAAAGGATTAG
E1	F	AGGAGATATACCATGGTGCTGGCGTGGTACCGTGC
	R	CAGAACTTCCAGTTTGTAGTCTTCAGCCATGGCCAGTAAGACAAG
F1	F	AGGAGATATACCATGGAGCAGGGGGACGGTGTGG
	R	CAGAACTTCCAGTTTGTAGTCTTCAGCCATGGCCAGTAAGACAAG
G1	F	AGGAGATATACCATGGTGCTGGCGTGGTACCGTGC
	R	CAGAACTTCCAGTTTGTAGTCTTCAGCCATGGCCAGTAAGACAAG
H1	F	AGGAGATATACCATGGAGCAGGGGGACGGTGTGG
	R	CAGAACTTCCAGTTTGTAGTCTTCAGCCATGGCCAGTAAGACAAG
A2	F	AGGAGATATACCATGTGGGTGCTGGCGTGGCTC
	R	CAGAACTTCCAGTTTGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
B2	F	AGGAGATATACCATGTGTGAACAGGGAGACGGCGTG
	R	CAGAACTTCCAGTTTGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
C2	F	AGGAGATATACCATGTGGGTGCTGGCGTGGCTC
	R	CAGAACTTCCAGTTTGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
D2	F	AGGAGATATACCATGTGTGAACAGGGAGACGGCGTG
	R	CAGAACTTCCAGTTTGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
E2	F	AGGAGATATACCATGTTCGAAGAATATCAAAAGTTCACGAATGAATC
	R	CAGAACTTCCAGTTTGAAGCGTCCAAGGATGTTGACAACAAG
F2	F	AGGAGATATACCATGGCAACGCCTCAGGATAAATTGC
	R	CAGAACTTCCAGTTTGAAGCGTCCAAGGATGTTGACAACAAG
G2	F	AGGAGATATACCATGTTCGAAGAATATCAAAAGTTCACGAATGAATC
	R	CAGAACTTCCAGTTTGAAGCGTCCAAGGATGTTGACAACAAG
H2	F	AGGAGATATACCATGGCAACGCCTCAGGATAAATTGC
	R	CAGAACTTCCAGTTTGAAGCGTCCAAGGATGTTGACAACAAG

Continued on next page...

Table A.1: Primers used for high-throughput PCR

Construct	Primer name	Sequence (5'-3')
A3	F	GCGTAGCTGAAACCGGCGTGCTGGCGTGGTACCGTGC
	R	GGTGGCTCCAGCTAGCGTAGTCTTCAGCCATGGCCAGTAAGACAAG
B3	F	GCGTAGCTGAAACCGGCGAGCAGGGGGACGGTGTGG
	R	GGTGGCTCCAGCTAGCGTAGTCTTCAGCCATGGCCAGTAAGACAAG
C3	F	GCGTAGCTGAAACCGGCTGGGTGCTGGCGTGGCTC
	R	GGTGGCTCCAGCTAGCGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
D3	F	GCGTAGCTGAAACCGGCTGTGAACAGGGAGACGGCGTG
	R	GGTGGCTCCAGCTAGCGTACTCTTCAGCCATGATCAGTAAGACAAGACTG
E3	F	GCGTAGCTGAAACCGGCTTTATTCTTCTTCATCTATATCTCTACCCAAATCTTG
	R	GGTGGCTCCAGCTAGCAGTGTCCAAGTTATCACTCATAATCAAAAGG
F3	F	GCGTAGCTGAAACCGGCTGTAATCAAGATGATCGTCTGCCTGG
	R	GGTGGCTCCAGCTAGCAGTGTCCAAGTTATCACTCATAATCAAAAGGATTAG
G3	F	GCGTAGCTGAAACCGGCTTCGAAGAATATCAAAAGTTCACGAATGAATC
	R	GGTGGCTCCAGCTAGCGAAGCGTCCAAGGATGTTGACAACAAG
H3	F	GCGTAGCTGAAACCGGCGCAACGCCTCAGGATAAATTGC
	R	GGTGGCTCCAGCTAGCGAAGCGTCCAAGGATGTTGACAACAAG
A4	F	AGGAGATATACCATGGGCACCGCCGGGAGCAGTTG
	R	CAGAACTTCCAGTTTAAATTCGATATGAACTTTGAAGCTCGACGTAAG
B4	F	GCGTAGCTGAAACCGGCCAGATGGTCAAACGAGACGATTTTAAAAC
	R	GGTGGCTCCAGCTAGCAAATTCGATATGAACTTTGAAGCTCGACG
C4	F	GACGATGACGACAAGCAGATGGTCAAACGAGACGATTTTAAAAC
	R	GTGATGGTGATGTTTAAATTCGATATGAACTTTGAAGCTCGACG
D4	F	AGGAGATATACCATGGGCCTTCCAAAAATATTAATATTCCTACTGCC
	R	GTGATGGTGATGTTTAAGTTCCTCGTGAACAACTTCCTTGTCAG
E4	F	GCGTAGCTGAAACCGGCAGACAACTGAAACCAGTAAAGGGAGTTCC
	R	GTGATGGTGATGTTTAACAACTTCCTTGTCAGGATCAGCACATG
F4	F	TGCCTGGTGTTCGCGAGACAACTGAAACCAGTAAAGGGAGTTCC
	R	GTGATGGTGATGTTTAACAACTTCCTTGTCAGGATCAGCACATG
G4	F	AGGAGATATACCATGGGCGCGGCGGTAGCGGCAG
	R	CAGAACTTCCAGTTTTCGCAGGTGAATACTCCAATCAGATG
H4	F	GCGTAGCTGAAACCGGCGCTGTGGATAGAAGCAACTTTAAGACCTGTG
	R	GGTGGCTCCAGCTAGCTCGCAGGTGAATACTCCAATCAGATGC

Continued on next page...

Table A.1: Primers used for high-throughput PCR

Construct	Primer name	Sequence (5'-3')
A5	F	GACGATGACGACAAGGCTGTGGATAGAAGCAACTTTAAGACCTGTG
	R	GTGATGGTGATGTTTTCGCAGGTGAATACTCCAATCAGATGC
B5	F	AGGAGATATACCATGGGCCTGTTGCCGCTGCTG
	R	GTGATGGTGATGTTTGAGTTCGTCATGGTCGTCTTCGGTG
C5	F	GCGTAGCTGAAACCGGCGCCGTGGAGGTCAAGAGGCC
	R	GTGATGGTGATGTTTGTCGTCTTCGGTGGGTGCTTCAG
D5	F	TGCCTGGTGTTCGCGGCCGTGGAGGTCAAGAGGCC
	R	GTGATGGTGATGTTTGTCGTCTTCGGTGGGTGCTTCAG
E5	F	AGGAGATATACCATGGGCGCGGCAATAGCGGC
	R	CAGAACTTCCAGTTTTCGAAGATGAATACTCCAGTCGGATGC
F5	F	GCGTAGCTGAAACCGGCGCTGTGGATAGAAGCAACTTTAAGACCTGTG
	R	GGTGGCTCCAGCTAGCTCGAAGATGAATACTCCAGTCGGATGC
G5	F	GACGATGACGACAAGGCTGTGGATAGAAGCAACTTTAAGACCTGTG
	R	GTGATGGTGATGTTTTCGAAGATGAATACTCCAGTCGGATGC
H5	F	AGGAGATATACCATGGGCCTGCTGCTGCTGCTACTAC
	R	GTGATGGTGATGTTTCAGCTCGTCATGGTCCCCATCAC
A6	F	GCGTAGCTGAAACCGGCGCTGTAGAAGTTAAGAGACCCCGGGG
	R	GTGATGGTGATGTTTGTCCCCATCACTGGGTGCTTCTG
B6	F	TGCCTGGTGTTCGCGGCTGTAGAAGTTAAGAGACCCCGGGG
	R	GTGATGGTGATGTTTGTCCCCATCACTGGGTGCTTCTG
C6	F	AGGAGATATACCATGGGCGTCCTTTTGAAATGGCTCG
	R	CAGAACTTCCAGTTTAAAAATAACTTCCCAATCTTCAGTTATGTCAAGCG
D6	F	GCGTAGCTGAAACCGGCACCGACTATCTATTAAAGAAGTGTGCGCAATC
	R	GGTGGCTCCAGCTAGCAAAAATAACTTCCCAATCTTCAGTTATGTCAAGCG
E6	F	GACGATGACGACAAGACCGACTATCTATTAAAGAAGTGTGCGCAATC
	R	GTGATGGTGATGTTTAAAAATAACTTCCCAATCTTCAGTTATGTCAAGCG
F6	F	AGGAGATATACCATGGGCGTGAGCATGTTCTCATTATTTCTGC
	R	GTGATGGTGATGTTTTTCACTTAAATTGAACTTAGGGGGCTCG
G6	F	GCGTAGCTGAAACCGGCCAAAGTCAACGGCATATAGTGGGTGTTCC
	R	GTGATGGTGATGTTTTTCACTTAAATTGAACTTAGGGGGCTCG
H6	F	TGCCTGGTGTTCGCGCAAAGTCAACGGCATATAGTGGGTGTTCC
	R	GTGATGGTGATGTTTTTCACTTAAATTGAACTTAGGGGGCTCG

B

Ramachandran plots of all models

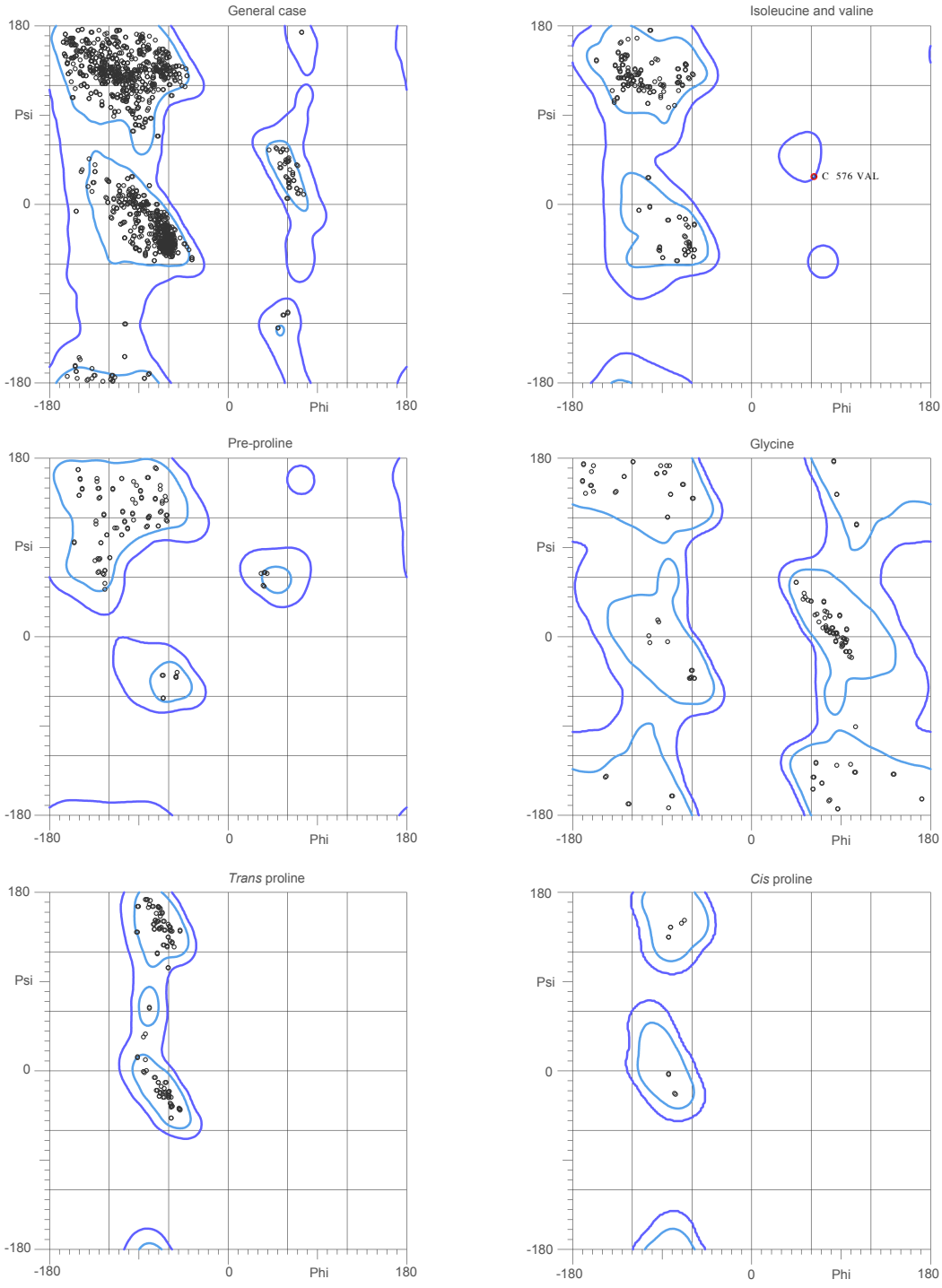


Figure B.1: Ramachandran plots of the glucal soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

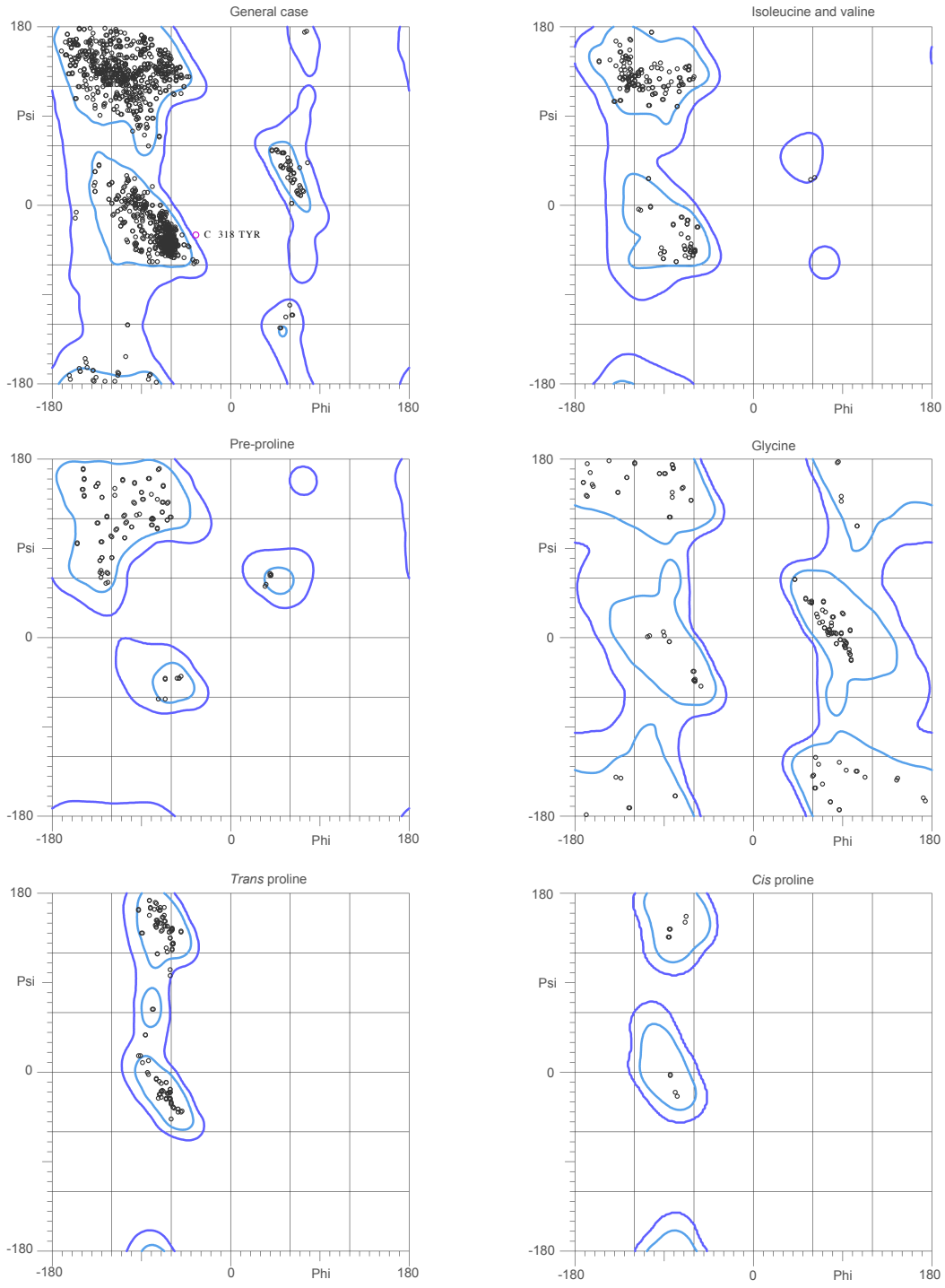


Figure B.2: Ramachandran plots of the glucose soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

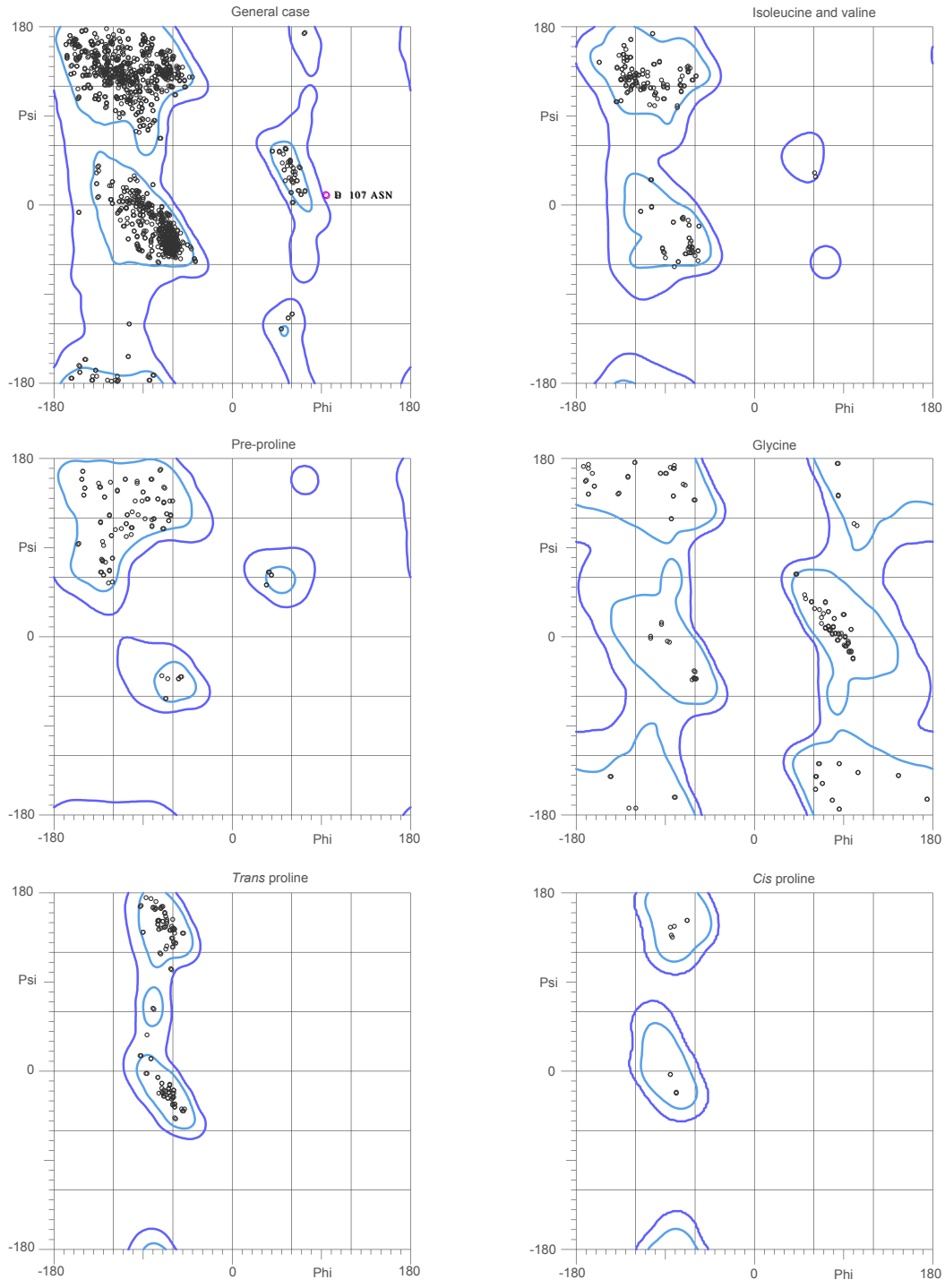


Figure B.3: Ramachandran plots of the mannose soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

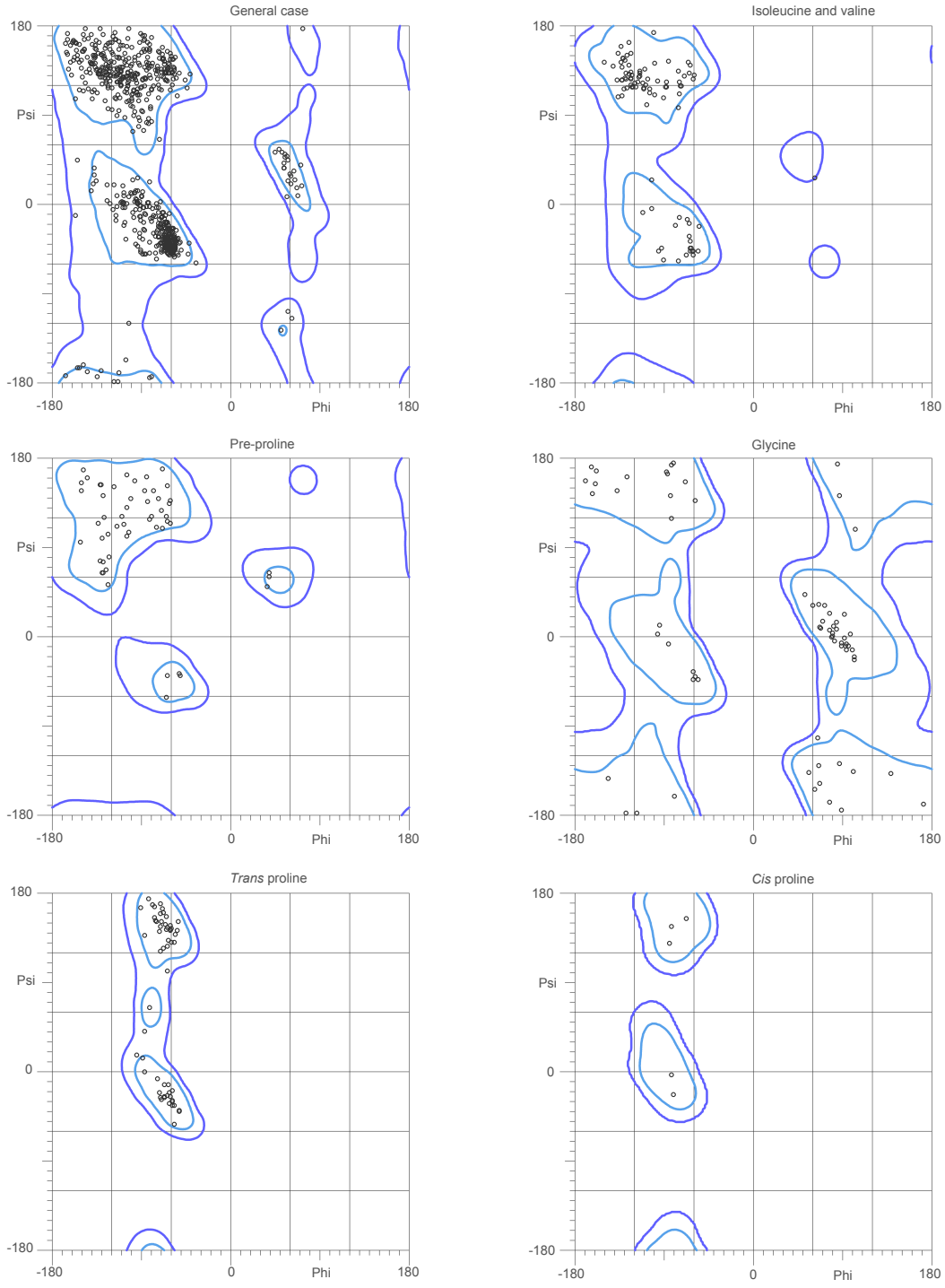


Figure B.4: Ramachandran plots of the castanaospermine soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

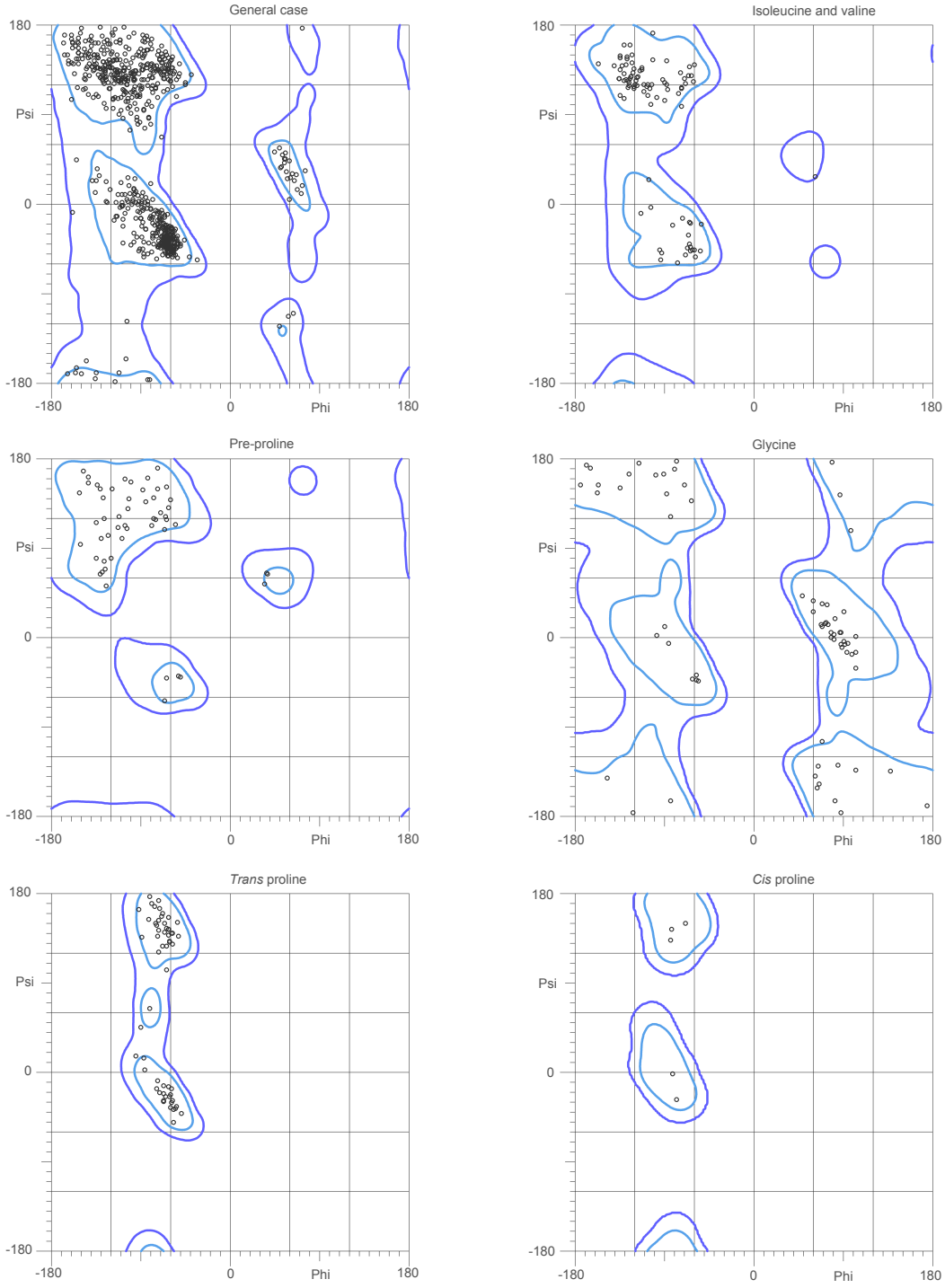


Figure B.5: Ramachandran plots of the DNJ soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

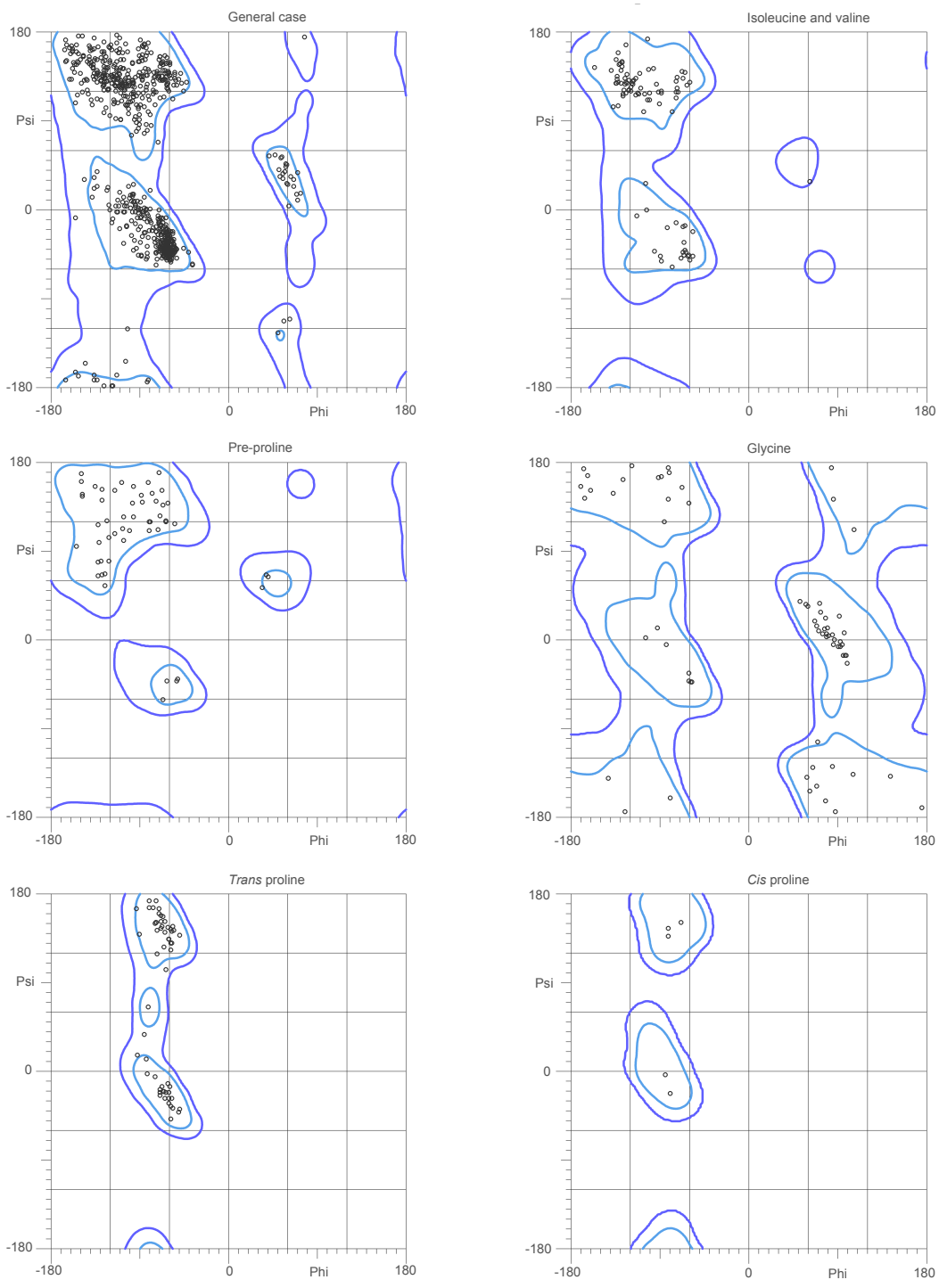
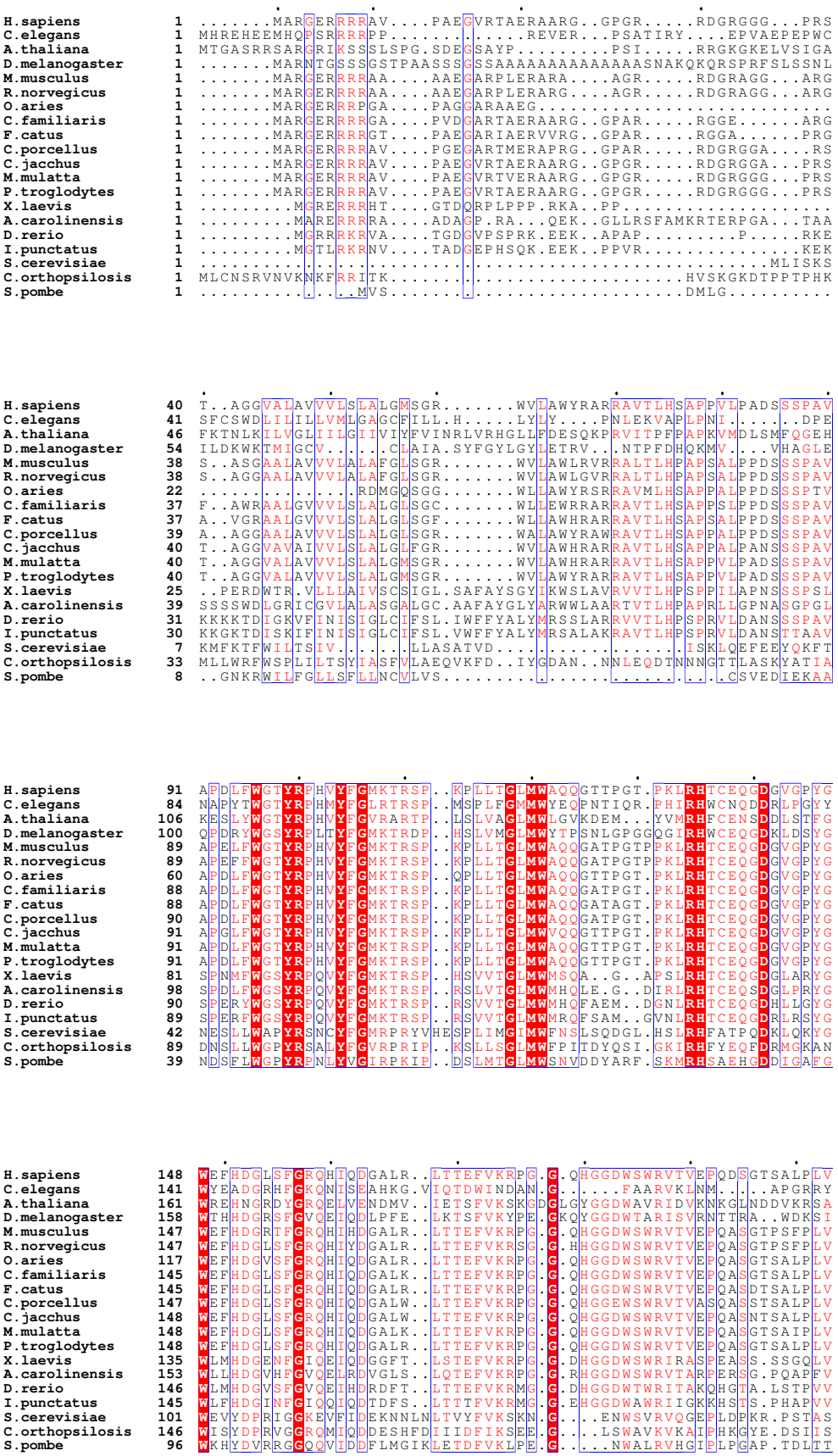
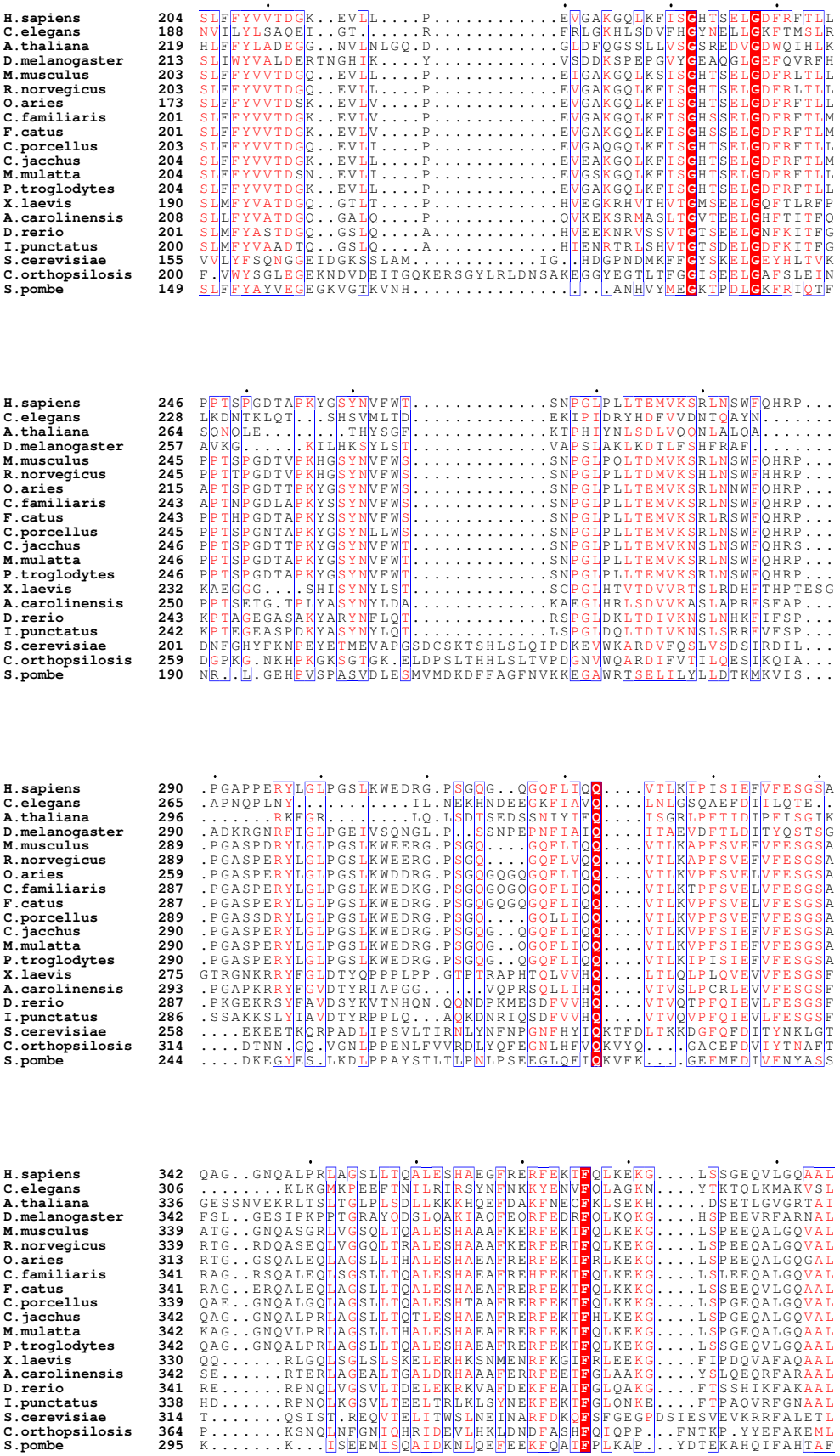


Figure B.6: Ramachandran plots of the *N*B-DNJ soak *Mm*GluII$_{\text{Trypsin}}$ model generated by MolProbity.

C

Multiple sequence alignments

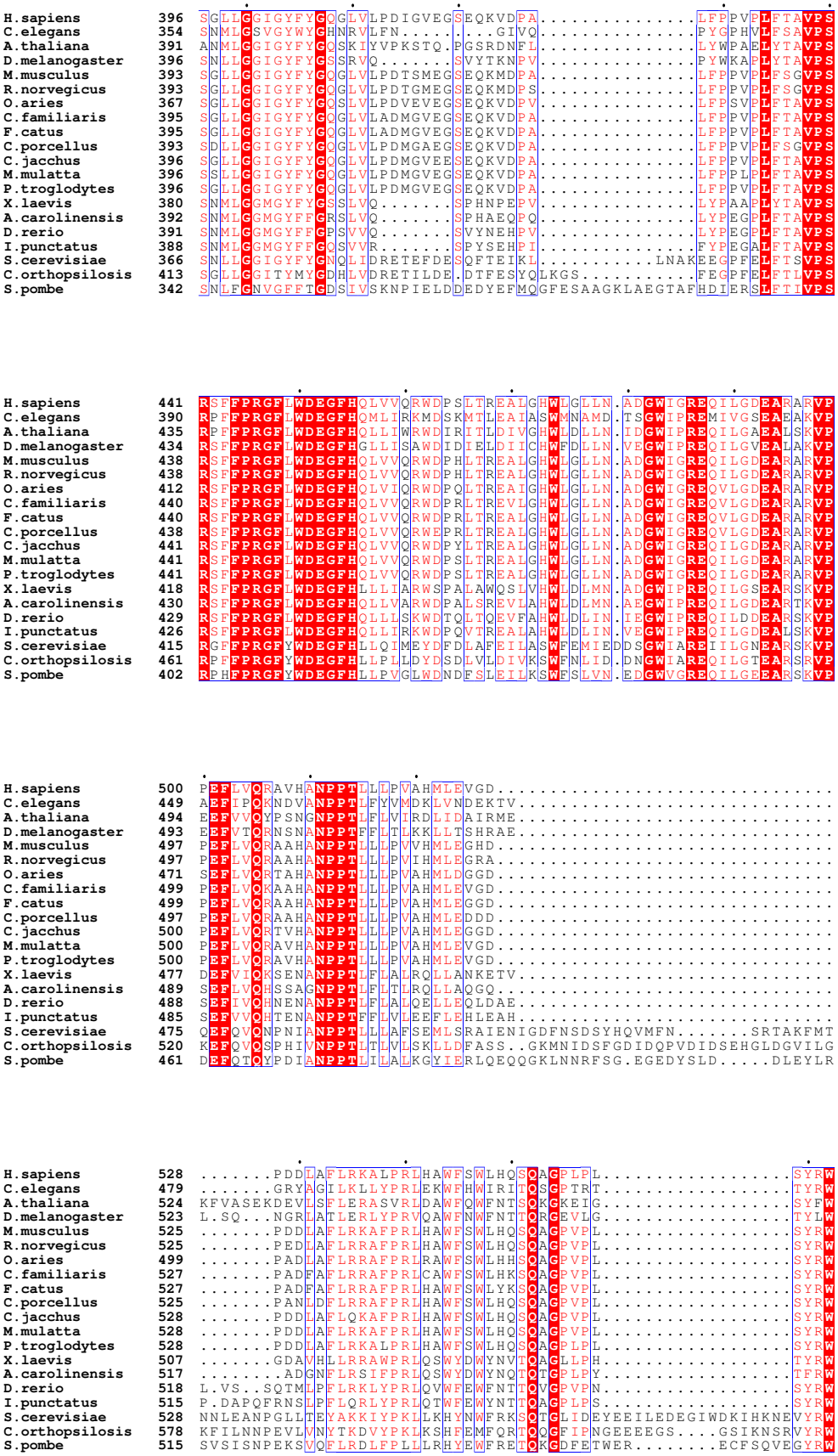

H.sapiens
C.elegans
A.thaliana
D.melanogaster
M.musculus
R.norvegicus
O.aries
C.familiaris
F.catus
C.porcellus
C.jacchus
M.mulatta
P.troglodytes
X.laevis
A.carolinensis
D.rerio
I.punctatus
S.cerevisiae
C.orthopsilosis
S.pombe


H.sapiens 204 SLFFYVVTDGK..EVLL....P.....EVGAKGQLKFISGHTSELGDFRFTLL
C.elegans 188 NVILYLSAQEI..GT.....R.....FRLGKHLSDVFHGYNELLGKFTMSLR
A.thaliana 219 HLFFYLADEGG..NVLNLGQ.D.....GLLDFQGSSLLVSGSREDVGDWQIHLK
D.melanogaster 213 SLIWYVALDERTNGHIK....Y.....VSDDKSPEPGVYGEAQGLGEFQVRFH
M.musculus 203 SLFFYVVTDGQ..EVLL....P.....EIGAKGQLKSISGHTSELGDFRLTLL
R.norvegicus 203 SLFFYVVTDGQ..EVLL....P.....EVGAKGQLKFISGHTSELGDFRLTLL
O.aries 173 SLFFYVVTDSK..EVLV....P.....EVGAKGQLKFISGHTSELGDFRFTLL
C.familiaris 201 SLFFYVVTDGK..EVLV....P.....EVGAKGQLKFISGHSSELGDFRFTLM
F.catus 201 SLFFYVVTDGK..EVLV....P.....EVGAKGQLKFISGHSSELGDFRFTLM
C.porcellus 203 SLFFYVVTDGQ..EVLI....P.....EVGAQGQLKFISGHTSELGDFRFTLL
C.jacchus 204 SLFFYVVTDGK..EVLL....P.....EVEAKGQLKFISGHTSELGDFRFTLM
M.mulatta 204 SLFFYVVTDSN..EVLI....P.....EVGSKGQLKFISGHTSELGDFRFTLL
P.troglodytes 204 SLFFYVVTDGK..EVLL....P.....EVGAKGQLKFISGHTSELGDFRFTLL
X.laevis 190 SLMFYVATDGQ..GTLT....P.....HVEGKRHVTHVTGMSEELGQFTLRFP
A.carolinensis 208 SLLFYVATDGQ..GALQ....P.....QVKEKSRMASLTGVTEELGHFTITFQ
D.rerio 201 SLMFYASTDGQ..GSLQ....A.....HVEEKNRVSSVTGTSEELGNFKITFG
I.punctatus 200 SLMFYVAADTQ..GSLQ....A.....HIENRTRLSHVTGTSDELGDFKITFG
S.cerevisiae 155 VVLYFSQNGGEIDGKSSLAM.....IG..HDGPNDMKFFGYSKELGEYHLTVK
C.orthopsilosis 200 F.VWYSGLEGEKNDVDEITGQKERSGYLRLDNSAKEGGYEGTLTFGGISEELGAFSLEIN
S.pombe 149 SLFFYAYVEGEGKVGTKVNH.....L...ANHVYMEGKTPDLGKFRIQTF

H.sapiens 246 PPTSPGDTAPKYGSYNVFWT.....SNPGLPLLTEMVKSRLNSWFQHRP...
C.elegans 228 LKDNTKLQT..SHSVMLTD.....EKIPIDRYHDFVVDNTQAYN.....
A.thaliana 264 SQNQLE.....THYSGF.....KTPHIYNLSDLVQQNLALQA.....
D.melanogaster 257 AVKG.....KILHKSYLST.....VAPSLAKLKDTLFSHFRAF.....
M.musculus 245 PPTSPGDTVPKHGSYNVFWS.....SNPGLPQLTDMVKSRLNSWFQHRP...
R.norvegicus 245 PPTTPGDTVPKHGSYNVFWS.....SNPGLPLLTDMVKSHLNSWFHHRP...
O.aries 215 APTSPGDTTPKYGSYNVFWS.....SNPGLPLLTEMVKSRLNNWFQHRP...
C.familiaris 243 APTNPGDLAPKYSSYNVFWS.....SNPGLPLLTEMVKSRLNSWFQHRP...
F.catus 243 PPTHPGDTAPKYSSYNVFWS.....SNPGLPLLTEMVKSRLRSWFQHRP...
C.porcellus 245 PPTSPGNTAPKYGSYNLLWS.....SNPGLPLLTEMVKSRLNSWFQHRP...
C.jacchus 246 PPTSPGDTTPKYGSYNVFWT.....SNPGLPLLTEMVKNSLNSWFQHRS...
M.mulatta 246 PPTSPGDTAPKYGSYNVFWT.....SNPGLPLLTEMVKSRLNSWFQHRP...
P.troglodytes 246 PPTSPGDTAPKYGSYNVFWT.....SNPGLPLLTEMVKSRLNSWFQHRP...
X.laevis 232 KAEGGG....SHISYNYLST.....SCPGLHTVTDVVRTSLRDHFTHPTESG
A.carolinensis 250 PPTSETG.TPLYASYNYLDA.....KAEGLHRLSDVVKASLAPRFSFAP...
D.rerio 243 KPTAGEGASAKYARYNFLQT.....RSPGLDKLTDIVKNSLNHKFIFSP...
I.punctatus 242 KPTEGEASPDKYASYNYLQT.....LSPGLDQLTDIVKNSLSRRFVFSP...
S.cerevisiae 201 DNFGHYFKNPEYETMEVAPGSDCSKTSHLSLQIPDKEVWKARDVFQSLVSDSIRDIL...
C.orthopsilosis 259 DGPKG.NKHPKGKSGTGK.ELDPSLTHHLSLTVPDGNVWQARDIFVTILQESIKQIA...
S.pombe 190 NR..L.GEHPVSPASVDLESMVMDKDFFAGFNVKKEGAWRTSELILYLLDTKMKVIS...

H.sapiens 290 .PGAPPERYLGLPGSLKWEDRG.PSGQG..QGQFLIQQ.....VTLKIPISIEFVFESGSA
C.elegans 265 .APNQPLNY.....IL.NEKHNDEEGKFIAVQ.....LNLGSQAEFDIILQTE..
A.thaliana 296 .....RKFGR.....LQ.LSDTSEDSSNIYIPQ.....ISGRLPFTIDIPFISGIK
D.melanogaster 290 .ADKRGNRFIGLPGEIVSQNGL.P..SSNPEPNFIAIQ.....ITAEVDFTLDITYQSTSG
M.musculus 289 .PGASPDRYLGLPGSLKWEERG.PSGQ.....GQFLIQQ.....VTLKAPFSVEFVFESGSA
R.norvegicus 289 .PGASPERYLGLPGSLKWEERG.PSGQ.....GQFLVQQ.....VTLKAPFSVEFVFESGSA
O.aries 259 .PGASPERYLGLPGSLKWDDRG.PSGQGQGQGQFLIQQ.....VTLKVPFSVELVFESGSA
C.familiaris 287 .PGASPERYLGLPGSLKWEDKG.PSGQGQGQGQFLIQQ.....VTLKTPFSVELVFESGSA
F.catus 287 .PGASPERYLGLPGSLKWEDRG.PSGQGQGQGQFLIQQ.....VTLKVPFSVELVFESGSA
C.porcellus 289 .PGASSDRYLGLPGSLKWEDRG.PSGQ.....GQLLIQQ.....VTLKVPFSVEFVFESGSA
C.jacchus 290 .PGASPERYLGLPGSLKWEDRG.PSGQG..QGQFLIQQ.....VTLKVPFSIEFVFESGSA
M.mulatta 290 .PGASPERYLGLPGSLKWEDRG.PSGQG..QGQFLIQQ.....VTLKVPFSIEFVFESGSA
P.troglodytes 290 .PGASPERYLGLPGSLKWEDRG.PSGQG..QGQFLIQQ.....VTLKIPISIEFVFESGSA
X.laevis 275 GTRGNKRRYFGLDTYQPPPLPP.GTPTRAPHTQLVVHQ.....LTLQLPLQVEVVFESGSF
A.carolinensis 293 .PGAPKRRYFGVDTYRIAPGG.....VQPRSQLLIHQ.....VTVSLPCRLEVVFESGSF
D.rerio 287 .PKGEKRSYFAVDSYKVTNHQN.QQNDPKMESDFVVHQ.....VTVQTPFQIEVLFESGSF
I.punctatus 286 .SSAKKSLYIAVDTYRPPLQ...AQKDNRIQSDFVVHQ.....VTVQVPFQIEVLFESGSF
S.cerevisiae 258 ....EKEETKQRPADLIPSVLTIRNLYNFNPGNFHYIQKTFDLTKKDGFQFDITYNKLGT
C.orthopsilosis 314 ....DTNN.GQ.VGNLPPENLFVVRDLYQFEGNLHFVQKVYQ....GACEFDVIYTNAFT
S.pombe 244 ....DKEGYES.LKDLPPAYSTLTLPNLPSEEGLQFIQKVFK....GEFMFDIVFNYASS

H.sapiens 342 QAG..GNQALPRLAGSLLTQALESHAEGFREREEKTFQLKEKG....LSSGEQVLGQAAL
C.elegans 306 .....KLKGMKPEEFTNILRIRSYNFNKKYENVFQLAGKN....YTKTQLKMAKVSL
A.thaliana 336 GESSNVEKRLTSLTGLPLSDLLKKKHQEFDAKFNEGFKLSEKH.....DSETLGVGRTAI
D.melanogaster 342 FSL..GESIPKPPTGRAYQDSLQAKIAQFEQRFEDRFQLKQKG....HSPEEVRFARNAL
M.musculus 339 ATG..GNQASGRLVGSQLTQALESHAAAFKERFEKTFQLKEKG....LSPEEQALGQVAL
R.norvegicus 339 RTG..RDQASEQLVGGQLTRALESHAAAFKERFERTFQLKEKG....LSPEEQALGQVAL
O.aries 313 RTG..GSQALEQLAGSLLTHALESHAEAFRERFEKTFRLKEKG....LSPEEQALGQGAL
C.familiaris 341 RAG..RSQALEQLSGSLLTQALESHAEAFRERFEKTFQLKEKG....LSLEEQALGQVAL
F.catus 341 RAG..ERQALEQLAGSLLTQALESHAEAFRERFEKTFQLKKKG....LSSEEQVLGQAAL
C.porcellus 339 QAE..GNQALGQLAGSLLTQALESHTAAFRERFEKTFQLKKKG....LSPGEQALGQVAL
C.jacchus 342 QAG..GNQALPRLAGSLLTQTLESHAEAFRERFEKTFHLKEKG....LSPGEQALGQVAL
M.mulatta 342 KAG..GNQVLPRLAGSLLTHALESHAEAFRERFEKTFQLKEKG....LSPGEQALGQAAL
P.troglodytes 342 QAG..GNQALPRLAGSLLTQALESHAEAFRERFEKTFQLKEKG....LSSGEQVLGQAAL
X.laevis 330 QQ.....RLGQLSGLSLSKELERHKSNMENRFKGIFRLEEKG....FIPDQVAFAQAAL
A.carolinensis 342 SE.....RTERLAGEALTQALDRHAAAFERRFEETFGLAAKG....YSLQEQRFARAAL
D.rerio 341 RE.....RPNQLVGSVLTDELEKRKVAFDEKFEAIFGLQAKG....FTSSHIKFAKAAL
I.punctatus 338 HD.....RPNQLKGSVLTEELTRLKLSYNEKFEKIFGLQNKE....FTPAQVRFGNAAL
S.cerevisiae 314 T.....QSIST.REQVTELITWSLNEINARFDKQFSFGEGPDSIESVEVKRRFALETL
C.orthopsilosis 364 P.....KSNQLNFGNIQHRIDEVLHKLDNDFASHFQIQPP...FNTKP.YYEFAKEML
S.pombe 295 K.....K...ISEEMISQAIDKNLQEFEEKFQATFPLKAP...YDTEKAHQIFAHTAF


H.sapiens 396 SGLLGGIGYFYGQGLVLPDIGVEGSEQKVDPA.....LFPPVPLFTAVPS
C.elegans 354 SNMLGSVGYWYGHNRVLFN.....GIVQ.....PYGPHVLFSAVPS
A.thaliana 391 ANMLGGIGYFYGQSKIYVPKSTQ.PGSRDNFL.....LYWPAELYTAVPS
D.melanogaster 396 SNLLGGIGYFYGSSRVQ.....SVYTKNPV.....PYWKAPLYTAVPS
M.musculus 393 SGLLGGIGYFYGQGLVLPDTSMEGSEQKMDPA.....LFPPVPLFSGVPS
R.norvegicus 393 SGLLGGIGYFYGQGLVLPDTGMEGSEQKMDPS.....LFPPVPLFSGVPS
O.aries 367 SGLLGGIGYFYGQSLVLPDVEVEGSEQKVDPV.....LFPSVPLFTAVPS
C.familiaris 395 SGLLGGIGYFYGQGLVLADMGVEGSEQKVDPA.....LFPPVPLFTAVPS
F.catus 395 SGLLGGIGYFYGQGLVLADMGVEGSEQKVDPA.....LFPPVPLFTAVPS
C.porcellus 393 SDLLGGIGYFYGQGLVLPDMGAEGSEQKVDPA.....LFPPVPLFSGVPS
C.jacchus 396 SGLLGGIGYFYGQGLVLPDMGVEESEQKVDPA.....LFPPVPLFTAVPS
M.mulatta 396 SSLLGGIGYFYGQGLVLPDMGVEGSEQKVDPA.....LFPPLPLFTAVPS
P.troglodytes 396 SGLLGGIGYFYGQGLVLPDMGVEGSEQKVDPA.....LFPPVPLFTAVPS
X.laevis 380 SNMLGGMGYFYGSSLVQ.....SPHNPEPV.....LYPAAPLYTAVPS
A.carolinensis 392 SNMLGGMGYFFGRSLVQ.....SPHAEQPQ.....LYPEGPLFTAVPS
D.rerio 391 SNMLGGMGYFFGPSVVQ.....SVYNEHPV.....LYPEGPLFTAVPS
I.punctatus 388 SNMLGGMGYFFGQSVVR.....SPYSEHPI.....FYPEGALFTAVPS
S.cerevisiae 366 SNLLGGIGYFYGNQLIDRETEFDESQFTEIKL.....LNAKEEGPFELFTSVPS
C.orthopsilosis 413 SGLLGGITYMYGDHLVDRETILDE.DTFESYQLKGS.....FEGPFELFTLVPS
S.pombe 342 SNLFGNVGFFTGDSIVSKNPIELDDEDYEFMQGFESAAGKLAEGTAFHDIERSLFTIVPS

H.sapiens 441 RSFFPRGFLWDEGFHQLVVQRWDPSLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
C.elegans 390 RPFFPRGFLWDEGFHQMLIRKMDSKMTLEAIASWMNAMD.TSGWIPREMIVGSEAEAKVP
A.thaliana 435 RPFFPRGFLWDEGFHQLLIWRWDIRITLDIVGHWLDLLN.IDGWIPREQILGAEALSKVP
D.melanogaster 434 RSFFPRGFLWDEGFHGLLISAWDIDIELDIICHWFDLLN.VEGWIPREQILGVEALAKVP
M.musculus 438 RSFFPRGFLWDEGFHQLVVQRWDPHLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
R.norvegicus 438 RSFFPRGFLWDEGFHQLVVQRWDPHLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
O.aries 412 RSFFPRGFLWDEGFHQLVIQRWDPQLTREAIGHWLGLLN.ADGWIGREQVLGDEARARVP
C.familiaris 440 RSFFPRGFLWDEGFHQLVVQRWDPRLTREVLGHWLGLLN.ADGWIGREQVLGDEARARVP
F.catus 440 RSFFPRGFLWDEGFHQLVVQRWDPRLTREALGHWLGLLN.ADGWIGREQVLGDEARARVP
C.porcellus 438 RSFFPRGFLWDEGFHQLVVQRWEPRLTREALGHWLGLLN.ADGWIGREQVLGDEARARVP
C.jacchus 441 RSFFPRGFLWDEGFHQLVVQRWDPYLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
M.mulatta 441 RSFFPRGFLWDEGFHQLVVQRWDPSLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
P.troglodytes 441 RSFFPRGFLWDEGFHQLVVQRWDPSLTREALGHWLGLLN.ADGWIGREQILGDEARARVP
X.laevis 418 RSFFPRGFLWDEGFHLLLIARWSPALAWQSLVHWLDLMN.ADGWIPREQILGSEARRSKVP
A.carolinensis 430 RSFFPRGFLWDEGFHQLLVARWDPALSREVLAHWLDLMN.AEGWIPREQILGDEARTKVP
D.rerio 429 RSFFPRGFLWDEGFHQLLLSKWDTQLTQEVFAHWLDLIN.IEGWIPREQILDDEARSKVP
I.punctatus 426 RSFFPRGFLWDEGFHQLLIRKWDPQVTREALAHWLDLIN.VEGWIPREQILGDEALSKVP
S.cerevisiae 415 RGFFPRGFYWDEGFHLLQIMEYDFDLAFEILASWFEMIEDDSGWIAREIILGNEARSKVP
C.orthopsilosis 461 RPFFPRGFYWDEGFHLLPLLDYDSDLVLDIVKSWFNLID.DNGWIAREQILGTEARSRVP
S.pombe 402 RPHFPRGFYWDEGFHLLPVGLWDNDFSLEILKSWFSLVN.EDGWVGREQILGEEARSKVP

H.sapiens 500 PEFLVQRAVHANPPTLLLPVAHMLEVGD.....
C.elegans 449 AEFIPQKNDVANPPTLFYVMDKLVNDEKTV.....
A.thaliana 494 EEFVVQYPSNGNPPTLFLVIRDLIDAIRME.....
D.melanogaster 493 EEFVTQRNSNANPPTFFLTLKKLLTSHRAE.....
M.musculus 497 PEFLVQRAAHANPPTLLLPVVHMLEGHD.....
R.norvegicus 497 PEFLVQRAAHANPPTLLLPVIHMLEGRA.....
O.aries 471 SEFLVQRTAHANPPTLLLPVAHMLDGGD.....
C.familiaris 499 PEFLVQKAAHANPPTLLLPVAHMLEVGD.....
F.catus 499 PEFLVQRAAHANPPTLLLPVAHMLEGGD.....
C.porcellus 497 PEFLVQRAAHANPPTLLLPVAHMLEDDD.....
C.jacchus 500 PEFLVQRTVHANPPTLLLPVAHMLEGGD.....
M.mulatta 500 PEFLVQRAVHANPPTLLLPVAHMLEVGD.....
P.troglodytes 500 PEFLVQRAVHANPPTLLLPVAHMLEVGD.....
X.laevis 477 DEFVIQKSENANPPTLFLALRQLLANKETV.....
A.carolinensis 489 SEFLVQHSSAGNPPTLFLTLRQLLAQGQ.....
D.rerio 488 SEFIVQHNENANPPTLFLALQELLEQLDAE.....
I.punctatus 485 SEFVVQHTENANPPTFFLVLEEFLEHLEAH.....
S.cerevisiae 475 QEFQVQNPNIANPPTLLLAFSEMLSRAIENIGDFNSDSYHQVMFN.....SRTAKFMT
C.orthopsilosis 520 KEFQVQSPHIVNPPTLTLVLSKLLDFASS..GKMNIDSFGDIDQPVDIDSEHGLDGVILG
S.pombe 461 DEFQTQYPDIANPPTLILALKGYIERLQEQQGKLNNRFSG.EGEDYSLD.....DLEYLR

H.sapiens 528 .....PDDLAFLRKALPRLHAWFSWLHQSQAGPLPL.....SYRW
C.elegans 479 .....GRYAGILKLLYPRLEKWFHWIRITQSGPTRT.....TYRW
A.thaliana 524 KFVASEKDEVLSFLERASVRLDAWFQWFNTSQKGKEIG.....SYFW
D.melanogaster 523 L.SQ...NGRLATLERLYPRVQAWFNWFNTTQRGEVLG.....TYLW
M.musculus 525 .....PDDLAFLRKAFPRLHAWFSWLHQSQAGPVPL.....SYRW
R.norvegicus 525 .....PEDLAFLRRAFPRLHAWFSWLHQSQAGPVPL.....SYRW
O.aries 499 .....PADLAFLRRAFPRLRAWFSWLHHSQAGPVPL.....SYRW
C.familiaris 527 .....PADFAFLRRAFPRLCAWFSWLHKSQAGPVPL.....SYRW
F.catus 527 .....PADFAFLRRAFPRLHAWFSWLYKSQAGPVPL.....SYRW
C.porcellus 525 .....PANLDFLRRAFPRLHAWFSWLHQSQAGPVPL.....SYRW
C.jacchus 528 .....PDDLAFLQKAFPRLHAWFSWLHQSQAGPVPL.....SYRW
M.mulatta 528 .....PDDLAFLRKAFPRLHAWFSWLHQSQAGPVPL.....SYRW
P.troglodytes 528 .....PDDLAFLRKALPRLHAWFSWLHQSQAGPLPL.....SYRW
X.laevis 507 .....GDAVHLLRRAWPRLQSWYDWYNVTQAGLLPH.....TYRW
A.carolinensis 517 .....ADGNFLRSIFPRLQSWYDWYNQTQTGPLPY.....TFRW
D.rerio 518 L.VS..SQTMLPFLRKLYPRLQVWFEWFNTTQVGPVPN.....SYRW
I.punctatus 515 P.DAPQFRNSLPFLQRLYPRLQTWFEWYNTTQAGPLPS.....SYRW
S.cerevisiae 528 NNLEANPGLLTEYAKKIYPKLLKHYNWFRKSQTGLIDEYEEILEDEGIWDKIHKNEVYRW
C.orthopsilosis 578 KFILNNPEVLVNYTKDVYPKLKSHFEMFQRTQQGFIPNGEEEEGS....GSIKNSRVYRW
S.pombe 515 SVSISNPEKSVQFLRDLFPLLLRHYEWFRETQKGDFETWER.....ECFSQVEGYRW

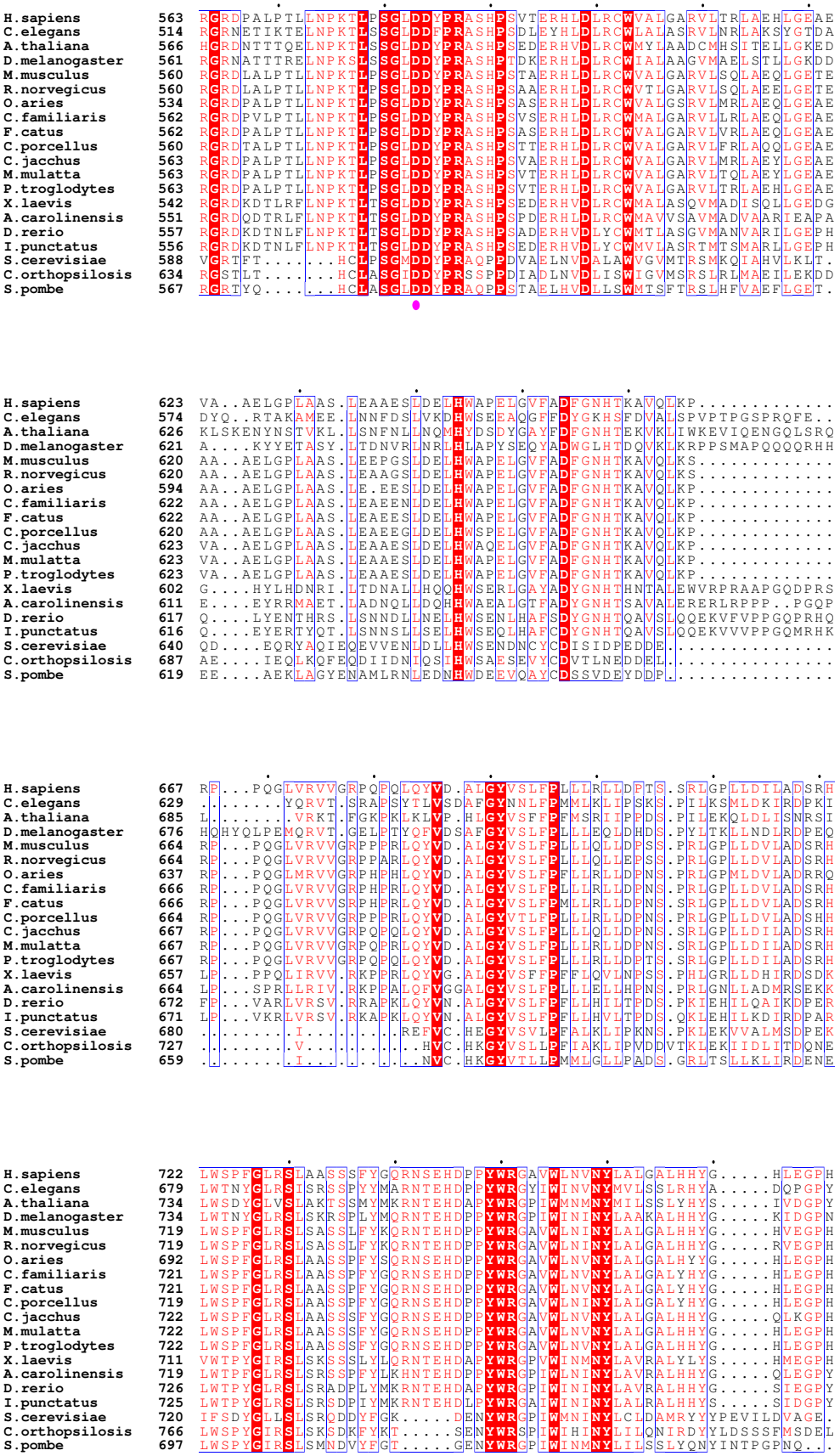
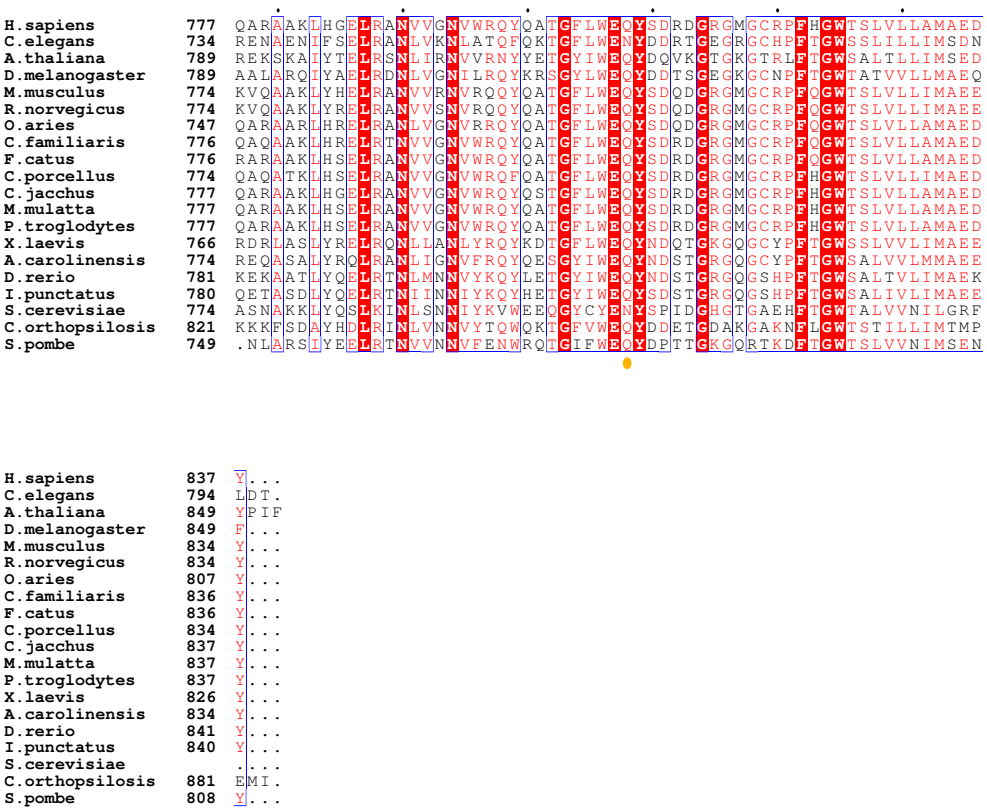
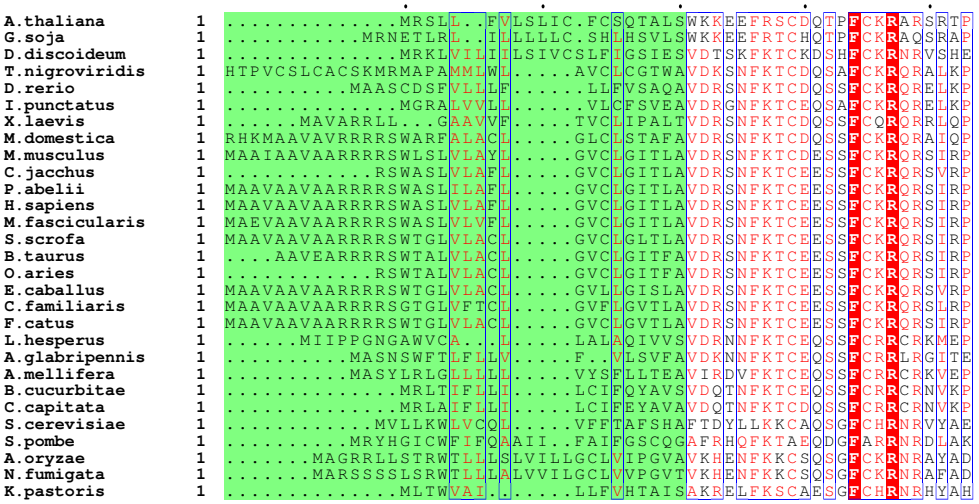
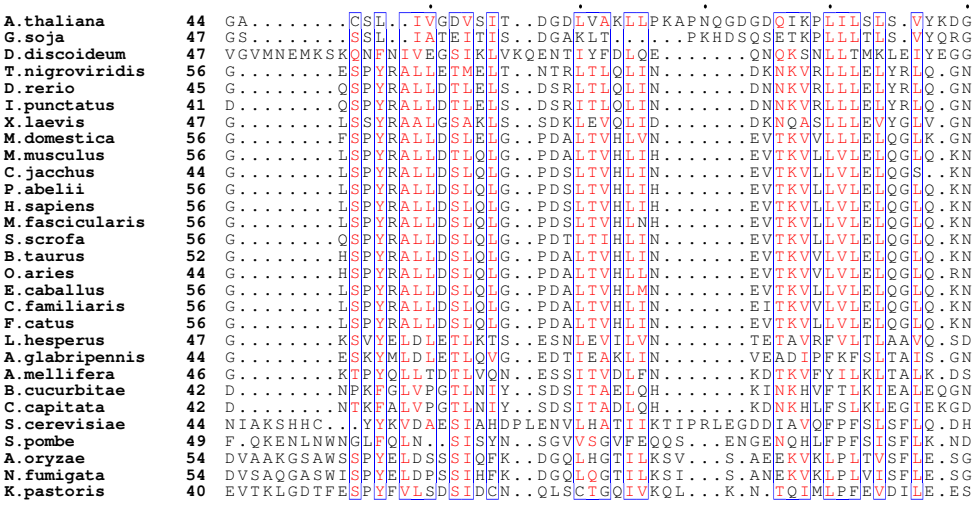
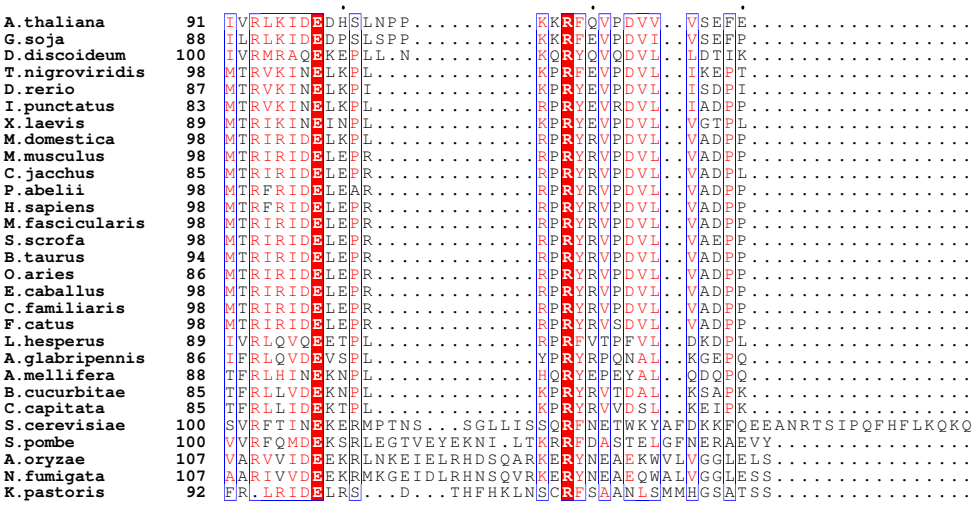

H.sapiens 563
C.elegans 514
A.thaliana 566
D.melanogaster 561
M.musculus 560
R.norvegicus 560
O.aries 534
C.familiaris 562
F.catus 562
C.porcellus 560
C.jacchus 563
M.mulatta 563
P.troglodytes 563
X.laevis 542
A.carolinensis 551
D.rerio 557
I.punctatus 556
S.cerevisiae 588
C.orthopsilosis 634
S.pombe 567
H.sapiens 623
C.elegans 574
A.thaliana 626
D.melanogaster 621
M.musculus 620
R.norvegicus 620
O.aries 594
C.familiaris 622
F.catus 622
C.porcellus 620
C.jacchus 623
M.mulatta 623
P.troglodytes 623
X.laevis 602
A.carolinensis 611
D.rerio 617
I.punctatus 616
S.cerevisiae 640
C.orthopsilosis 687
S.pombe 619
H.sapiens 667
C.elegans 629
A.thaliana 685
D.melanogaster 676
M.musculus 664
R.norvegicus 664
O.aries 637
C.familiaris 666
F.catus 666
C.porcellus 664
C.jacchus 667
M.mulatta 667
P.troglodytes 667
X.laevis 657
A.carolinensis 664
D.rerio 672
I.punctatus 671
S.cerevisiae 680
C.orthopsilosis 727
S.pombe 659
H.sapiens 722
C.elegans 679
A.thaliana 734
D.melanogaster 734
M.musculus 719
R.norvegicus 719
O.aries 692
C.familiaris 721
F.catus 721
C.porcellus 719
C.jacchus 722
M.mulatta 722
P.troglodytes 722
X.laevis 711
A.carolinensis 719
D.rerio 726
I.punctatus 725
S.cerevisiae 720
C.orthopsilosis 766
S.pombe 697

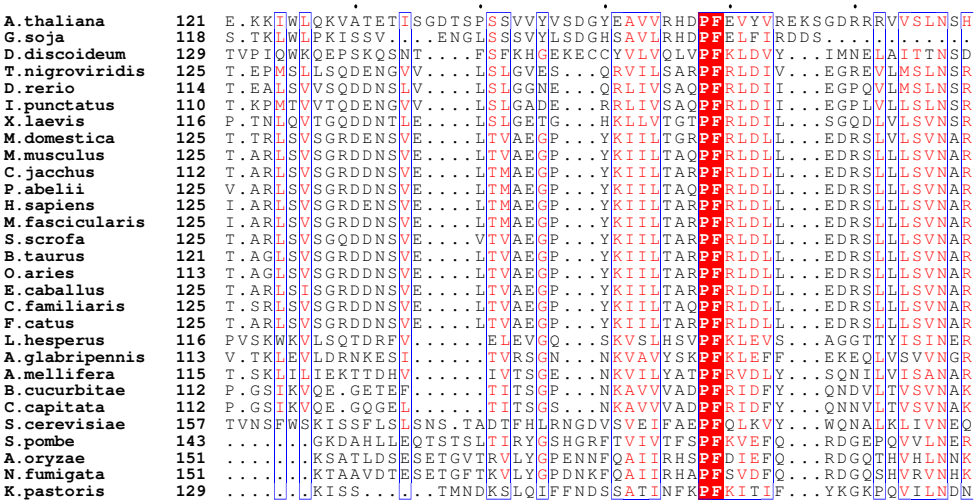
Figure C.1: Multiple sequence alignment of MOGS genes. *Appendix C*

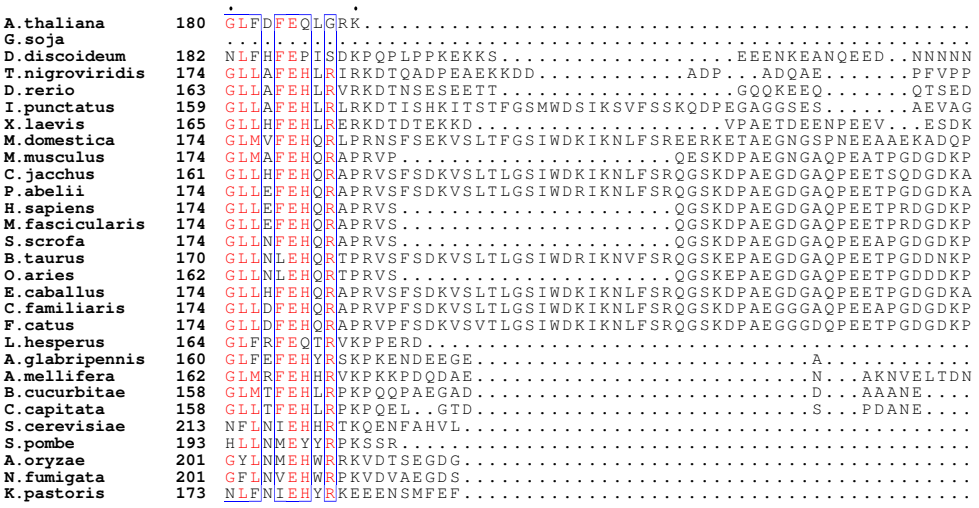


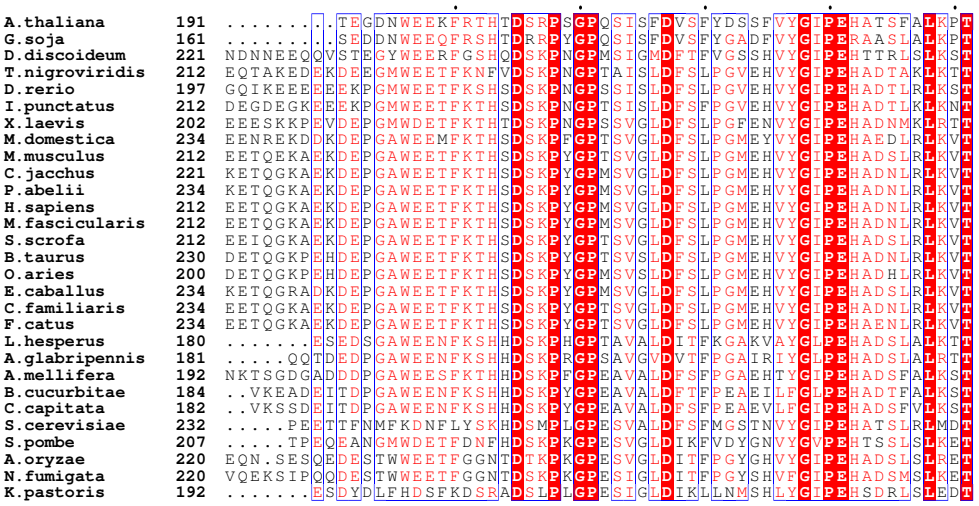

A.thaliana 1
G.soja 1
D.discoideum 1
T.nigroviridis 1
D.rerio 1
I.punctatus 1
X.laevis 1
M.domestica 1
M.musculus 1
C.jacchus 1
P.abelii 1
H.sapiens 1
M.fascicularis 1
S.scrofa 1
B.taurus 1
O.aries 1
E.caballus 1
C.familiaris 1
F.catus 1
L.hesperus 1
A.glabripennis 1
A.mellifera 1
B.cucurbitae 1
C.capitata 1
S.cerevisiae 1
S.pombe 1
A.oryzae 1
N.fumigata 1
K.pastoris 1

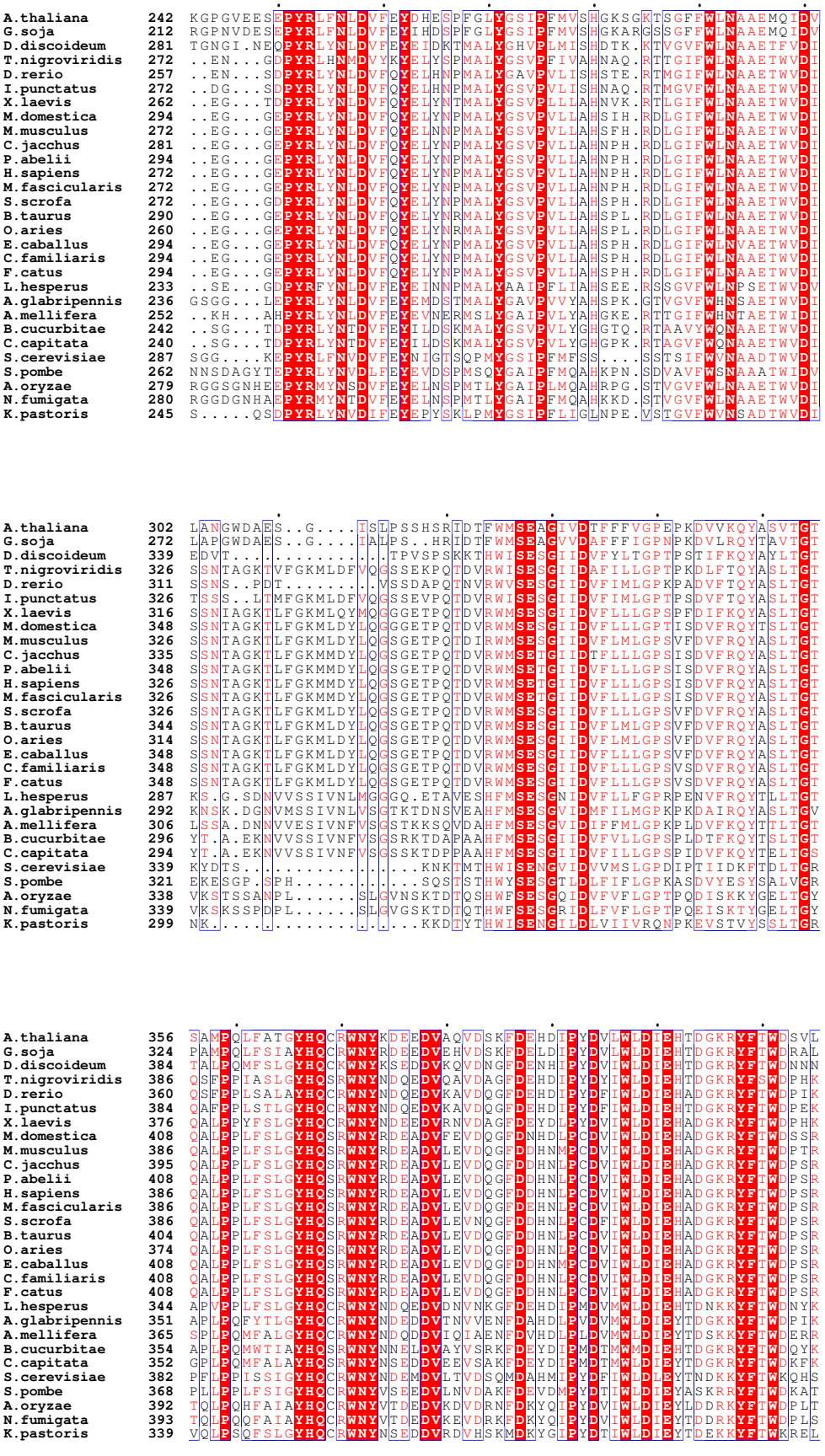

A.thaliana 44
G.soja 47
D.discoideum 47
T.nigroviridis 56
D.rerio 45
I.punctatus 41
X.laevis 47
M.domestica 56
M.musculus 56
C.jacchus 44
P.abelii 56
H.sapiens 56
M.fascicularis 56
S.scrofa 56
B.taurus 52
O.aries 44
E.caballus 56
C.familiaris 56
F.catus 56
L.hesperus 47
A.glabripennis 44
A.mellifera 46
B.cucurbitae 42
C.capitata 42
S.cerevisiae 44
S.pombe 49
A.oryzae 54
N.fumigata 54
K.pastoris 40

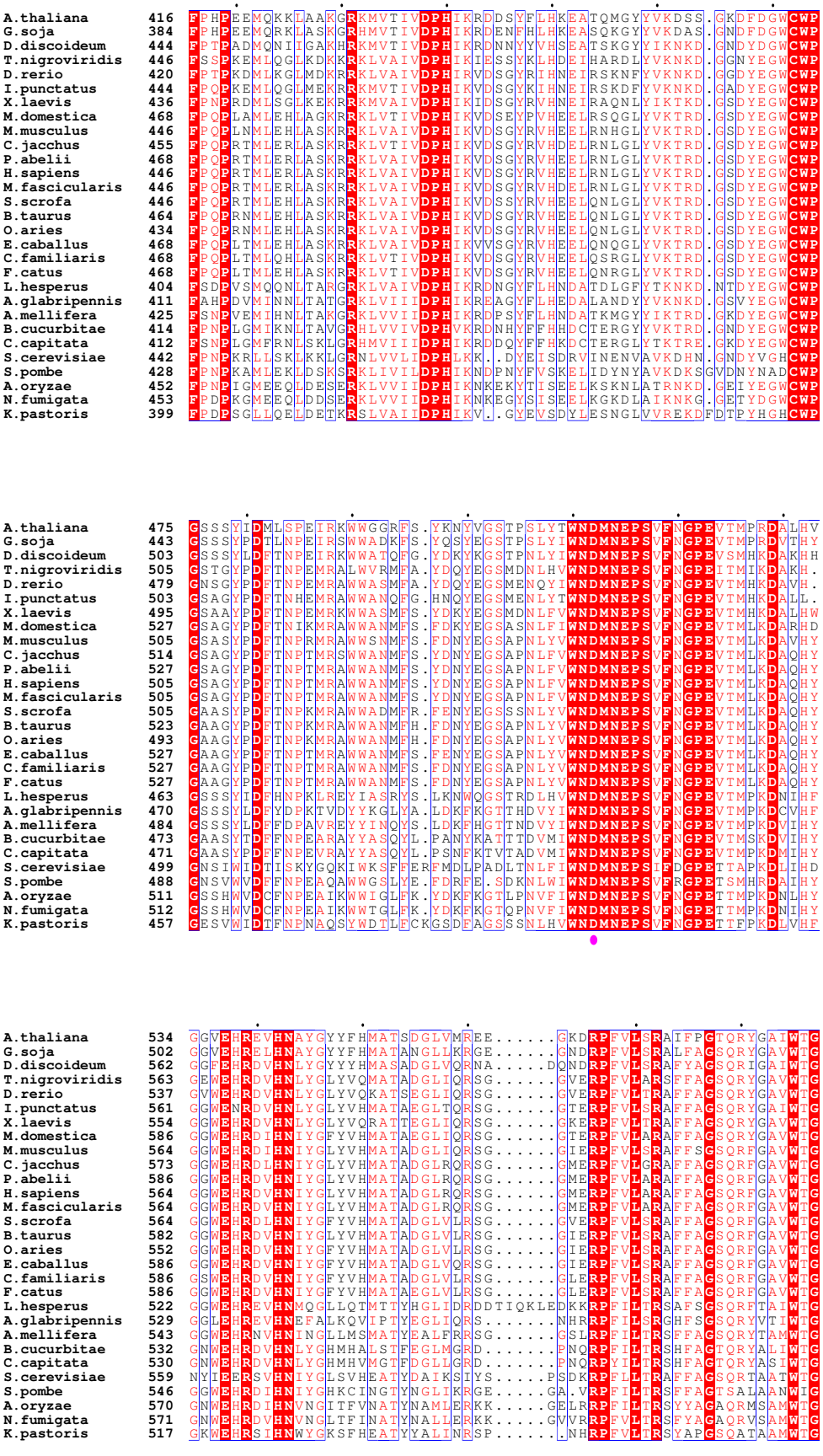

A.thaliana 91
G.soja 88
D.discoideum 100
T.nigroviridis 98
D.rerio 87
I.punctatus 83
X.laevis 89
M.domestica 98
M.musculus 98
C.jacchus 85
P.abelii 98
H.sapiens 98
M.fascicularis 98
S.scrofa 98
B.taurus 94
O.aries 86
E.caballus 98
C.familiaris 98
F.catus 98
L.hesperus 89
A.glabripennis 86
A.mellifera 88
B.cucurbitae 85
C.capitata 85
S.cerevisiae 100
S.pombe 100
A.oryzae 107
N.fumigata 107
K.pastoris 92

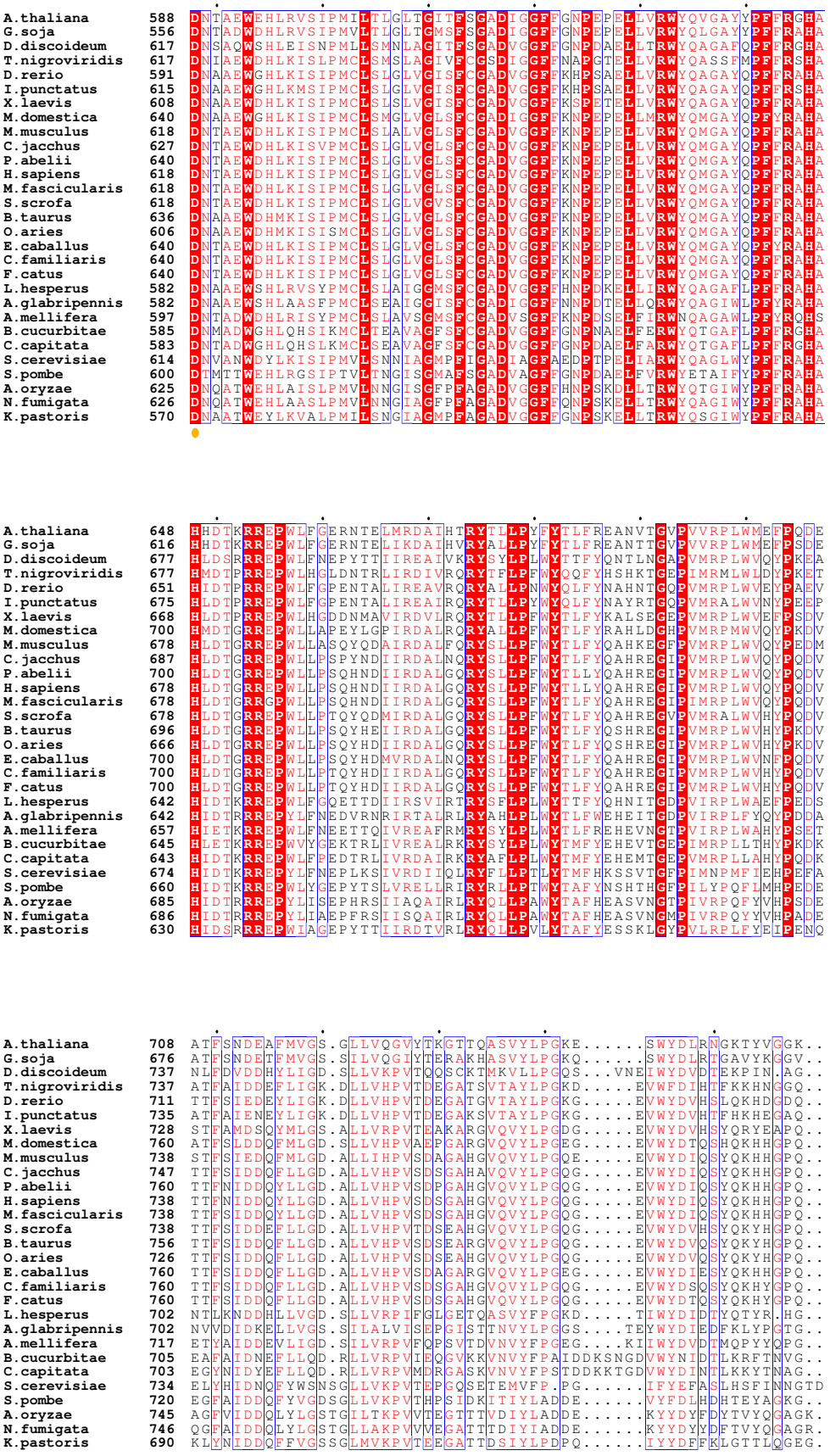

A.thaliana 121 E.KKIWLQKVATETISGDTSPSSVVYVSDGYEAVVRHDPFEVYVREKSGDRRRVVSLNSH
G.soja 118 S.TKLWLPKISSV....ENGLSSSVYLSDGHSAVLRHDPFELFIRDDS.....
D.discoideum 129 TVPIQWKQEPSKQSNT....FSFKHGEKECCYVLVQLVPFKLDVY...IMNELAITTNSD
T.nigroviridis 125 T.EPMSLLSQDENGVV....LSLGVES...QRVILSARPFRLDIV...EGRREVLMSLNSR
D.rerio 114 T.EALSVVSQDDNSLV....LSLGGNE...QRLIVSAQPFRLDII...EGPQVLMSLNSR
I.punctatus 110 T.KPMTVVTQDENGVV....LSLGADE...RRLIVSAQPFRLDII...EGPLVLLSLNSR
X.laevis 116 P.TNLQVTGQDDNTLE....LSLGETG...HKLLVTGTPFRLDIL...SGQDLVLSVNSR
M.domestica 125 T.TRLSVSGRDENSVE....LTVAEGP...YKIILTGRPFRLDLL...EDRSLVLSVNAR
M.musculus 125 T.ARLSVSGRDDNSVE....LTVAEGP...YKIILTAQPFRLDLL...EDRSLLLSVNAR
C.jacchus 112 T.ARLSVSGRDDNSVE....LTMAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
P.abelii 125 V.ARLSVSGRDDNSVE....LTMAEGP...YKIILTAQPFRLDLL...EDRSLLLSVNAR
H.sapiens 125 I.ARLSVSGRDENSVE....LTMAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
M.fascicularis 125 I.ARLSVSGRDDNSVE....LTMAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
S.scrofa 125 T.ARLSVSGQDDNSVE....VTVAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
B.taurus 121 T.AGLSVSGRDDNSVE....LTVAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
O.aries 113 T.AGLSVSGRDDNSVE....LTVAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
E.caballus 125 T.ARLSISGRDDNSVE....LTVAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
C.familiaris 125 T.SRLSVSGRDDNSVE....LTVAEGP...YKIILTAQPFRLDLL...EDRSLLLSVNAR
F.catus 125 T.ARLSVSGRDDNSVE....LTVAEGP...YKIILTARPFRLDLL...EDRSLLLSVNAR
L.hesperus 116 PVSKWKVLSQTDRFV.....ELEVGQ...SKVSLHSVPFKLEVS...AGGTTYISINER
A.glabripennis 113 V.TKLEVLDRNKESI.....TVRSGN...NKVAVYSKPFKLEFF...EKEQLVSVVNGR
A.mellifera 115 T.SKLILIEKTTDHV.....IVTSGE...NKVILYATPFRVDLY...SQNILVISANAR
B.cucurbitae 112 P.GSIKVQE.GETEF.....TITSGP...NKAVVVADPFRIDFY...QNDVLTVSVNAK
C.capitata 112 P.GSIKVQE.GQGEL.....TITSGS...NKAVVVADPFRIDFY...QNNVLTVSVNAK
S.cerevisiae 157 TVNSFWSKISSFLSLSNS.TADTFHLRNGDVSVEIFAEPFQLKVY...WQNALKLIVNEQ
S.pombe 143 .....GKDAHLLEQTSTSLTIRYGSHGRFTVIVTFSPFKVEFQ...RDGEPQVVLNER
A.oryzae 151 .....KSATLDSESETGVTRVLYGPENNFQAIIRHSPFDIEFQ...RDGQTHVHLNNK
N.fumigata 151 .....KTAAVDTESETGFTKVLYGPDNKFQAIIRHAPFSVDFQ...RDGQSHVRVNHK
K.pastoris 129 ....KISS...T.MNDKSLQIFFNDSSATINFKPFKITIF...YKGKPQVILNDN

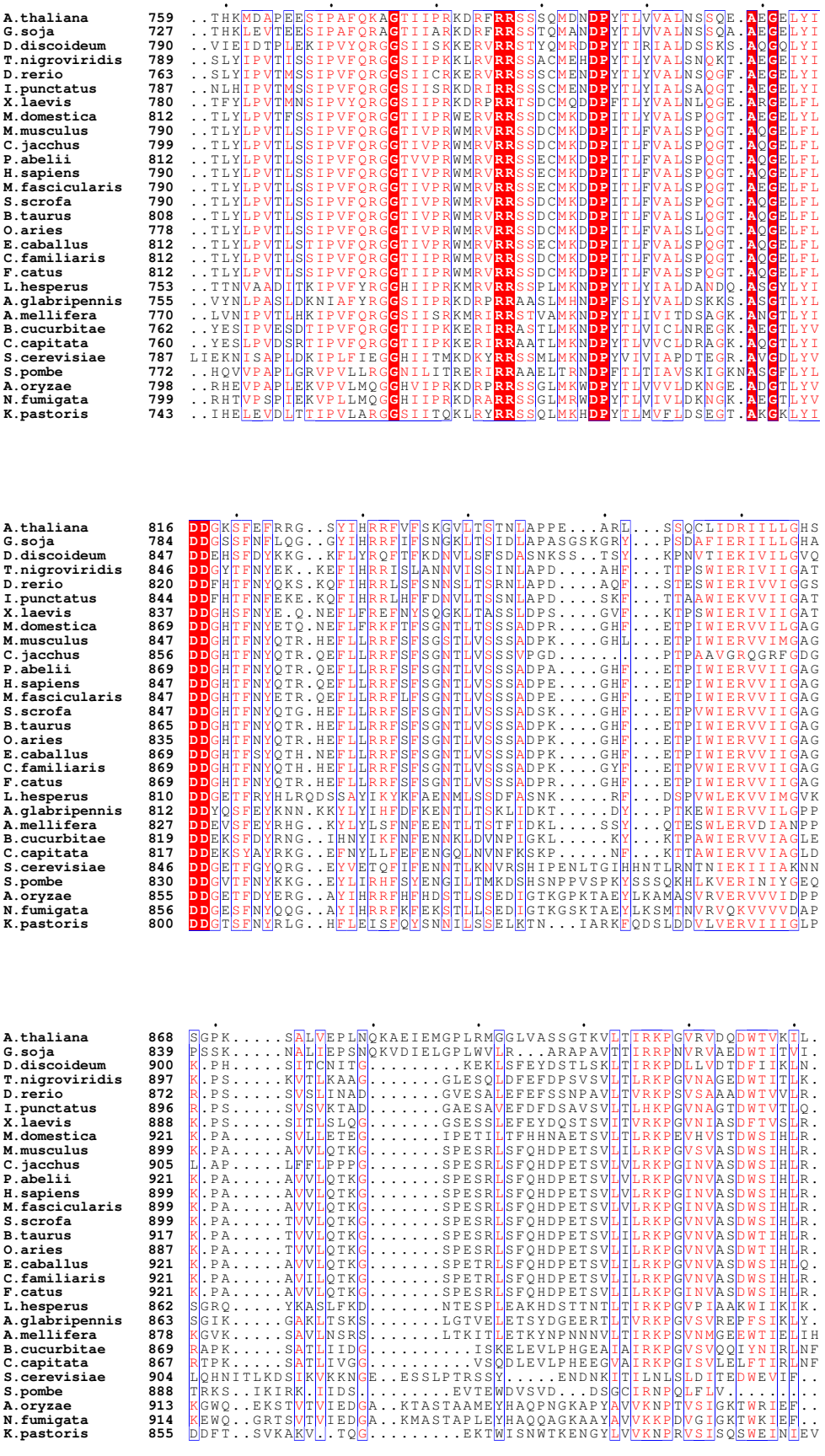

A.thaliana 180 GLFDFEQLGRK
G.soja
D.discoideum 182 NLFHFEPISDKPQPLPPKEKKS.....EEENKEANQEED..NNNNN
T.nigroviridis 174 GLLAFEHLRIRKDTQADPEAEKKDD.....ADP...ADQAE.....PFVPP
D.rerio 163 GLLAFEHLRVRKDTNSESEETT.....GQQKEEQ.....QTSED
I.punctatus 159 GLLAFEHLRLRKDTISHKITSTFGSMWDSIKSVFSSKQDPEGAGGSES.....AEVAG
X.laevis 165 GLLHFEHLRERKDTDTEKKD.....VPAETDEENPEEV...ESDK
M.domestica 174 GLMVFEHQRLPRNSFSEKVSLTFGSIWDKIKNLFSREERKETAEGNGSPNEEAAEKADQP
M.musculus 174 GLMAFEHQRAPRVP.....QESKDPAEGNGAQPEATPGDGDKP
C.jacchus 161 GLLHFEHQRAPRVSFSDKVSLTLGSIWDKIKNLFSRQGSKDPAEGDGAQPEETSQDGDKA
P.abelii 174 GLLEFEHQRAPRVSFSDKVSLTLGSIWDRIKNLFSRQGSKDPAEGDGAQPEETPGDGDKA
H.sapiens 174 GLLEFEHQRAPRVS.....QGSKDPAEGDGAQPEETPRDGDKP
M.fascicularis 174 GLLEFEHQRAPRVS.....QGSKDPAEGDGAQPEETPRDGDKP
S.scrofa 174 GLLNFEHQRAPRVS.....QGSKDPAEGDGAQPEEAPGDGDKP
B.taurus 170 GLLNLEHQRTPRVSFSDKVSLTLGSIWDRIKNVFSRQGSKEPAEGDGAQPEETPGDDNKP
O.aries 162 GLLNLEHQRTPRVS.....QGSKEPAEGDGAQPEETPGDDDKP
E.caballus 174 GLLHFEHQRAPRVSFSDKVSLTLGSIWDKIKNLFSRQGSKDPAEGDGAQPEETPGDGDKA
C.familiaris 174 GLLDFEHQRAPRVPFSDKVSLTLGSIWDKIKNLFSRQGSKDPAEGGGAQPEEAPGDGDKP
F.catus 174 GLLDFEHQRAPRVPFSDKVSVTLGSIWDKIKNLFSRQGSKDPAEGGGDQPEETPGDGDKP
L.hesperus 164 GLFRFEQTRVKPPERD.....
A.glabripennis 160 GLFEFEHYRSKPKENDEEGE.....A.....
A.mellifera 162 GLMRFEHHRVKPKKPDQDAE.....N...AKNVELTDN
B.cucurbitae 158 GLMTFEHLRPKPQQPAEGAD.....D...AAANE....
C.capitata 158 GLLTFEHLRPKPQEL..GTD.....S...PDANE....
S.cerevisiae 213 NFLNIEHHRTKQENFAHVL.....
S.pombe 193 HLLNMEYYRPKSSR.....
A.oryzae 201 GYLNMEHWRRKVDTSEGDG.....
N.fumigata 201 GFLNVEHWRPKVDVAEGDS.....
K.pastoris 173 NLFNIEHYRKEEENSMFEF.....

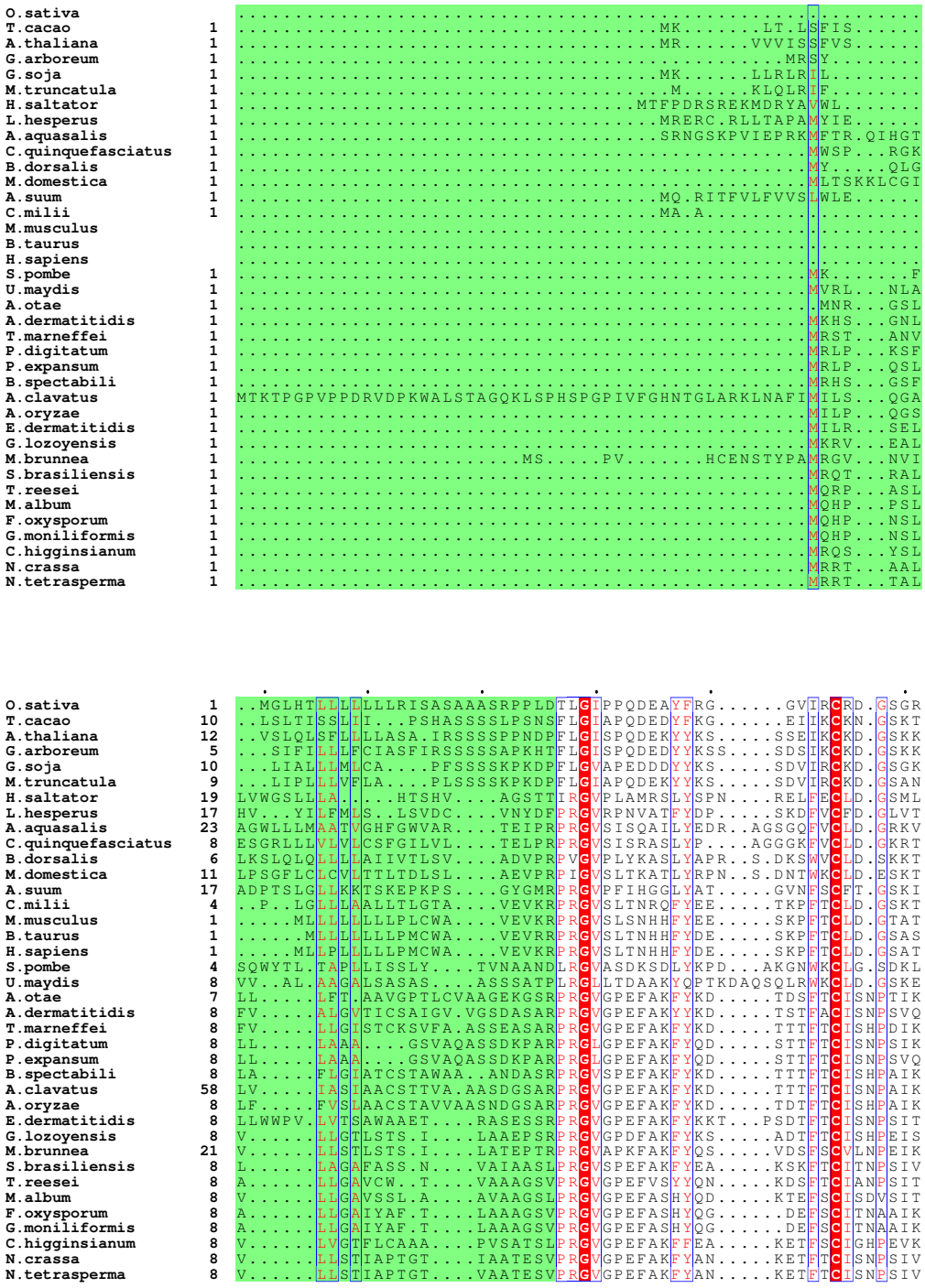

A.thaliana 191 .....TEGDNWEEKFRTHTDSKPSGPQSISFDVSFYDSSFVYGIPEHATSFALKPT
G.soja 161 .....SEDDNWEEQFRSHTDRRPYGPQSISFDVSFYGADFVYGIPERAASLALKPT
D.discoideum 221 NDNNEEQQVSTEGYWEERFGSHQDSKPNGPMSIGMDFTFVGSSHVYGIPEHTTRLSLKST
T.nigroviridis 212 EQTAKEDEKDEEGMWEETFKNFVDSKPNGPTAISIDFSLPGVEHVYGIPEHADTAKLKTT
D.rerio 197 GQIKEEEEEEEKPGMWEETFKSHSDSKPNGPSSISIDFSLPGVEHVYGIPEHADTLRLKST
I.punctatus 212 DEGDEGKEEEKPGMWEETFKTHSDSKPNGPTSISIDFSFPGVEHVYGIPEHADTLKLKNT
X.laevis 202 EEESKKPEVDEPGMWDETFKTHTDSKPNGPSSVGIDFSLPGFENVYGIPEHADNMKLRTT
M.domestica 234 EENREKDDKDEPGAWEEMFKTHSDSKPFGPTSVGIDFSLPGMEYVYGIPEHAEDLRLKVT
M.musculus 212 EETQEKAEKDEPGAWEETFKTHSDSKPYGPTSVGIDFSLPGMEHVYGIPEHADSLRLKVT
C.jacchus 221 KETQGKAEKDEPGAWEETFKTHSDSKPYGPMSVGIDFSLPGMEHVYGIPEHADNLRLKVT
P.abelii 234 KETQGKAEKDEPGAWEETFKTHSDSKPYGPMSVGIDFSLPGMEHVYGIPEHADNLRLKVT
H.sapiens 212 EETQGKAEKDEPGAWEETFKTHSDSKPYGPMSVGIDFSLPGMEHVYGIPEHADNLRLKVT
M.fascicularis 212 EETQGKAEKDEPGAWEETFKTHSDSKPYGPMSVGIDFSLPGMEHVYGIPEHADNLRLKVT
S.scrofa 212 EEIQGKAEKDEPGAWEETFKTHSDSKPYGPTSVGIDFSLPGMEHVYGIPEHADSLRLKVT
B.taurus 230 DETQGKPEHDEPGAWEETFKTHSDSKPYGPTSVSIDFSLPGMEHVYGIPEHADNLRLKVT
O.aries 200 DETQGKPEHDEPGAWEETFKTHSDSKPYGPMSVSIDFSLPGMEHVYGIPEHADHLRLKVT
E.caballus 234 KETQGRADKDEPGAWEETFKTHSDSKPYGPMSVGIDFSLPGMEHVYGIPEHADSLRLKVT
C.familiaris 234 EETQGKAEKDEPGAWEETFKTHSDSKPYGPTSVGIDFSLPGMEHVYGIPEHADNLRLKVT
F.catus 234 EETQGKAEKDEPGAWEETFKTHSDSKPYGPTSVGIDFSLPGMEHVYGIPEHAENLRLKVT
L.hesperus 180 .....ESEDSGAWEENFKSHHDSKPHGPTAVALDITFKGAKVAYGLPEHADSLALKTT
A.glabripennis 181 .....QQTDEDPGAWEENFKSHHDSKPRGPSAVGVDVTFPGAIRIYGLPEHADSLALRTT
A.mellifera 192 NKTSGDGADDDPGAWEESFKTHHDSKPFGPEAVALDFSFPGAEHTYGIPEHADSFALKST
B.cucurbitae 184 ..VKEADEITDPGAWEENFKSHHDSKPYGPEAVALDFTFPEAEILFGLPEHADTFALKST
C.capitata 182 ..VKSSDEITDPGAWEENFKSHHDSKPYGPEAVALDFSFPEAEVLFGIPEHADSFVLKST
S.cerevisiae 232 .....PEETTFNMFKDNFLYSKHDSMPLGPESVALDFSFMGSTNVYGIPEHATSLRLMDT
S.pombe 207 .....TPEQEANGMWDETFDNFHDSKPKGPESVGLDIKFVDYGNVYGVPEHTSSLSLKET
A.oryzae 220 EQN.SESQEDESTWWEETFGGNTDTKPKGPESVGLDITFPGYGHVYGIPEHADSLSLRET
N.fumigata 220 VQEKSIPQQDESTWWEETFGGNTDSKPKGPESIGLDITFPGYSHVFGIPEHADSMSLKET
K.pastoris 192 .....ESDYDLFHDSFKDSRADSIPLGPESIGLDIKLLNMSHLYGIPEHSDRLSLEDT

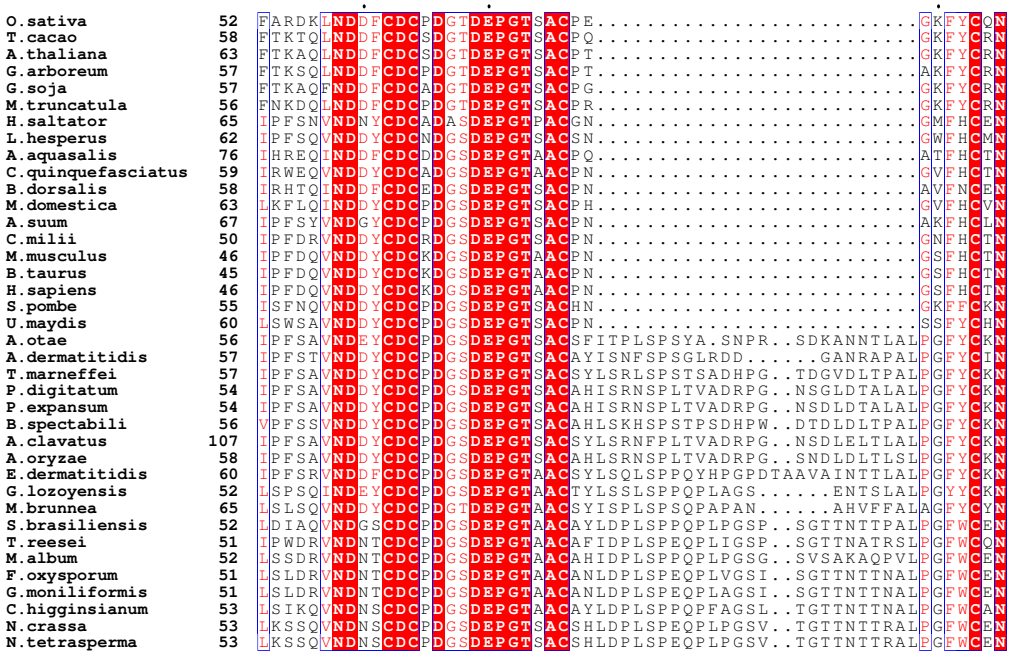

A.thaliana 242 KGPGVEESEPYRLFNLDVFEYDHESPFGLYGSIPFMVSHGKSGKTSGFFWLNAAEMQIDV
G.soja 212 RGPNVDESEPYRLFNLDVFEYIHDSPFGLYGSIPFMVSHGKARGSSGFFWLNAAEMQIDV
D.discoideum 281 TGNGI.NEQPYRLYNLDVFEYEIDKTMALYGHVPLMISHDTK.KTVGVFWLNAAETFVDI
T.nigroviridis 272 ..EN...GDPYRLHNMDVYKYELYSPMALYGSVPFIVAHNAQ.RTTGIFWLNAAETWVDI
D.rerio 257 ..EN...SDPYRLYNLDVFQYELHNPMALYGAVPVLISHSTE.RTMGIFWLNAAETWVDI
I.punctatus 272 ..DG...SDPYRLYNLDVFQYELHNPMALYGSVPVLISHNAQ.RTMGVFWLNAAETWVDI
X.laevis 262 ..EG...TDPYRLYNLDVFQYELYNTMALYGSVPLLLAHNVK.RTLGIFWLNAAETWVDI
M.domestica 294 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHSIH.RDLGIFWLNAAETWVDI
M.musculus 272 ..EG...GEPYRLYNLDVFQYELNNPMALYGSVPVLLAHSFH.RDLGIFWLNAAETWVDI
C.jacchus 281 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHNPH.RDLGIFWLNAAETWVDI
P.abelii 294 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHNPH.RDLGIFWLNAAETWVDI
H.sapiens 272 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHNPH.RDLGIFWLNAAETWVDI
M.fascicularis 272 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHNPH.RDLGIFWLNAAETWVDI
S.scrofa 272 ..EG...GDPYRLYNLDVFQYELYNPMALYGSVPVLLAHSPH.RDLGIFWLNAAETWVDI
B.taurus 290 ..EG...GEPYRLYNLDVFQYELYNRMALYGSVPVLLAHSPL.RDLGIFWLNAAETWVDI
O.aries 260 ..EG...GEPYRLYNLDVFQYELYNRMALYGSVPVLLAHSPL.RDLGIFWLNAAETWVDI
E.caballus 294 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHSPH.RDLGIFWLNVAETWVDI
C.familiaris 294 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHSPH.RDLGIFWLNAAETWVDI
F.catus 294 ..EG...GEPYRLYNLDVFQYELYNPMALYGSVPVLLAHSPH.RDLGIFWLNAAETWVDI
L.hesperus 233 ..SE...GDPYRFYNLDVFEYEINNPMALYAAIPFLIAHSEE.RSSGVFWLNPSETWVDV
A.glabripennis 236 GSGG...LEPYRLYNLDVFEYEMDSTMALYGAVPVVYAHSPK.GTVGVFWHNSAETWVDI
A.mellifera 252 ..KH...AHPYRLYNLDVFEYEVNERMSLYGAIPVLYAHGKE.RTTGIFWHNTAETWIDI
B.cucurbitae 242 ..SG...TDPYRLYNLDVFEYILDSKMALYGSVPVLYGHGTQ.RTAAVYWQNAAETWVDI
C.capitata 240 ..SG...TDPYRLYNTDVFEYILDSKMALYGSVPVLYGHGPK.RTAGVFWQNAAETWVDI
S.cerevisiae 287 SGG....KEPYRLFNVDVFEYNIGTSQPMYGSIPFMFSS....SSTSIFWVNAADTWVDI
S.pombe 262 NNSDAGYTEPYRLYNVDLFEYEVDSPMSQYGAIPFMQAHKPN.SDVAVFWSNAAATWIDV
A.oryzae 279 RGGSGNHEEPYRMYNSDVFEYELNSPMTLYGAIPLMQAHRPG.STVGVFWLNAAETWVDI
N.fumigata 280 RGGDGNHAEPYRMYNTDVFEYELNSPMTLYGAIPFMQAHKKD.STVGVFWLNAAETWVDI
K.pastoris 245 S.....QSDPYRLYNVDIFEYEPYSKLPMYGSIPFLIGLNPE.VSTGVFWVNSADTWVDI
A.thaliana 302 LANGWDAES..G....ISLPSSHSRIDTFWMSEAGIVDTFFFVGPEPKDVVKQYASVTGT
G.soja 272 LAPGWDAES..G....IALPS..HRIDTFWMSEAGVVDAFFFIGPNPKDVLRQYTAVTGT
D.discoideum 339 EDVT.....TPVSPSKKTHWISESGIIDVFYLTGPTPSTIFKQYAYLTGT
T.nigroviridis 326 SSNTAGKTVFGKMLDFVQGSSEKPQTDVRWISESGIIDAFILLGPTPKDLFTQYASLTGT
D.rerio 311 SSNS..PDT.....VSSDAPQTNVRWVSESGIIDVFIMLGPKPADVFTQYASLTGT
I.punctatus 326 TSSS..LTMFGKMLDFVQGSSEVPQTDVRWISESGIIDVFIMLGPTPSDVFTQYASLTGT
X.laevis 316 SSNIAGKTLFGKMLQYMQGGGETPQTDVRWMSESGIIDVFLLLGPSPFDIFKQYASLTGT
M.domestica 348 SSNTAGKTLFGKMLDYLQGGGETPQTDVRWMSESGIIDVFLLLGPTISDVFRQYTSLTGT
M.musculus 326 SSNTAGKTLFGKMLDYLQGSGETPQTDIRWMSESGIIDVFLMLGPSVFDVFRQYASLTGT
C.jacchus 335 SSNTAGKTLFGKMMDYLQGSGETPQTDVRWMSETGIIDTFLLLGPSISDVFRQYASLTGT
P.abelii 348 SSNTAGKTLFGKMMDYLQGSGETPQTDVRWMSETGIIDVFLLLGPSISDVFRQYASLTGT
H.sapiens 326 SSNTAGKTLFGKMMDYLQGSGETPQTDVRWMSETGIIDVFLLLGPSISDVFRQYASLTGT
M.fascicularis 326 SSNTAGKTLFGKMMDYLQGSGETPQTDVRWMSETGIIDVFLLLGPSISDVFRQYASLTGT
S.scrofa 326 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLLLGPSVFDVFRQYASLTGT
B.taurus 344 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLMLGPSVFDVFRQYASLTGT
O.aries 314 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLMLGPSVFDVFRQYASLTGT
E.caballus 348 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLLLGPSVFDVFRQYASLTGT
C.familiaris 348 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLLLGPSVSDVFRQYASLTGT
F.catus 348 SSNTAGKTLFGKMLDYLQGSGETPQTDVRWMSESGIIDVFLLLGPSVSDVFRQYASLTGT
L.hesperus 287 KS.G.SDNVVSSIVNLMGGGQ.ETAVESHFMSESGNIDVFLLFGPRPENVFRQYTLLTGT
A.glabripennis 292 KNSK.DGNVMSSIVNLVSGTKTDNSVEAHFMSESGVIDMFILMGPKPKDAIRQYASLTGV
A.mellifera 306 LSSA.DNNVVESIVNFVSGSTKKSQVDAHFMSESGVIDIFFMLGPKPLDVFKQYTSLTGT
B.cucurbitae 296 YT.A.EKNVVSSIVNFVSGSRKTDAPAAHFMSESGIIDVFVLLGPSPLDTFKQYTSLTGT
C.capitata 294 YT.A.EKNVVSSIVNFVSGSSKTDPPAAHFMSESGIIDVFILLGPSPIDVFKQYTELTGS
S.cerevisiae 339 KYDTS..........KNKTMTHWISENGVIDVVMSLGPDIPTIIDKFTDLTGR
S.pombe 321 EKESGP.SPH.....SQSTSTHWYSESGTLDLFIFLGPKASDVYESYSALVGR
A.oryzae 338 VKSTSSANPL.....SLGVNSKTDTQSHWFSESGQIDVFVFLGPTPQDISKKYGELTGY
N.fumigata 339 VKSKSSPDPL.....SLGVGSKTDTQTHWFSESGRIDLFVFLGPTPQEISKTYGELTGY
K.pastoris 299 NK.....KKDTYTHWISENGILDLVIIVRQNPKEVSTVYSSLTGR
A.thaliana 356 SAMPQLFATGYHQCRWNYKDEEDVAQVDSKFDEHDIPYDVLWLDIEHTDGKRYFTWDSVL
G.soja 324 PAMPQLFSIAYHQCRWNYRDEEDVEHVDSKFDELDIPYDVLWLDIEHTDGKRYFTWDRAL
D.discoideum 384 TALPQMFSLGYHQCKWNYKSEDDVKQVDNGFDENHIPYDVIWLDIEHTDGKRYFTWDNNN
T.nigroviridis 386 QSFPPIASLGYHQSRWNYNDQEDVQAVDAGFDEHDIPYDYIWLDIEHTDGKRYFSWDPHK
D.rerio 360 QSFPPLSALAYHQCRWNYNDQEDVKAVDQGFDEHDIPYDFIWLDIEHADGKRYFTWDPIK
I.punctatus 384 QAFPPLSTLGYHQCRWNYNDQEDVKAVDQGFDEHDIPYDFIWLDIEHTDGKRYFTWDPEK
X.laevis 376 QALPPYFSLGYHQCRWNYNDEEDVRNVDAGFDEYDLPYDVIWLDIEHADGKRYFTWDPHK
M.domestica 408 QALPPLFSLGYHQSRWNYRDEADVFEVDQGFDNHDLPCDVIWLDIEHADGKRYFTWDSSR
M.musculus 386 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPTR
C.jacchus 395 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
P.abelii 408 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
H.sapiens 386 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
M.fascicularis 386 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDEHNLPCDVIWLDIEHADGKRYFTWDPSR
S.scrofa 386 QALPPLFSLGYHQSRWNYRDEADVLEVNQGFDDHNLPCDFIWLDIEHADGKRYFTWDPSR
B.taurus 404 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
O.aries 374 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
E.caballus 408 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNMPCDVIWLDIEHADGKRYFTWDPSR
C.familiaris 408 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
F.catus 408 QALPPLFSLGYHQSRWNYRDEADVLEVDQGFDDHNLPCDVIWLDIEHADGKRYFTWDPSR
L.hesperus 344 APVPPLFSLGYHQCRWNYNDQEDVDNVNKGFDEHDIPMDVMWLDIEHTDNKKYFTWDNYK
A.glabripennis 351 APLPQFYTLGYHQCRWNYNDEDDVTNVVENFDAHDLPVDVMWLDIEYTDGKKYFTWDPIK
A.mellifera 365 SPLPQMFALGYHQCRWNYNDQDDVIQIAENFDVHDLPLDVMWLDIEYTDSKKYFTWDERR
B.cucurbitae 354 APLPQMWTIAYHQSRWNYNNELDVAYVSRKFDEYDIPMDTMWMDIEHTDGKRYFTWDQYK
C.capitata 352 GPLPQMFALAYHQSRWNYNSEDDVEEVSAKFDEYDIPMDTMWLDIEYTDGKRYFTWDKFK
S.cerevisiae 382 PFLPPISSIGYHQCRWNYNDEMDVLTVDSQMDAHMIPYDFIWLDIEYTNDKKYFTWKQHS
S.pombe 368 PLLPPLFSIGYHQCRWNYVSEEDVLNVDAKFDEVDMPYDTIWLDIEYASKRRYFTWDKAT
A.oryzae 392 TQLPQHFAIAYHQCRWNYVTDEDVKDVDRNFDKYQIPYDVIWLDIEYLDDRKYFTWDPLT
N.fumigata 393 TQLPQQFAIAYHQCRWNYVTDEDVKEVDRKFDKYQIPYDVIWLDIEYTDDRKYFTWDPLS
K.pastoris 339 VQLPPSQFSLGYHQCRWNYNSEDDVRDVHSKMDKYGIPYDTIWLDIEYTDEKKYFTWKREL

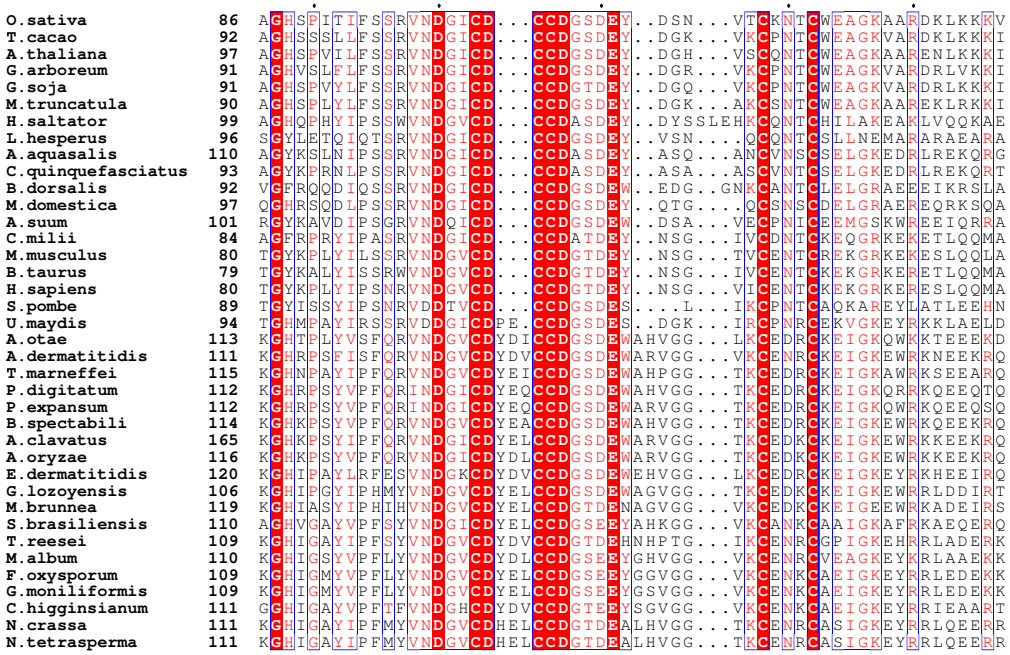

A.thaliana 416 FPHPEEMQKKLAAKGRKMVTIVDPHIKRDDSYFLHKEATQMGYYVKDSS.GKDFDGWCWP
G.soja 384 FPHPEEMQRKLASKGRHMVTIVDPHIKRDENFHLHKEASQKGYYVKDAS.GNDFDGWCWP
D.discoideum 444 FPTPADMQNIIGAKHRKMVTIVDPHIKRDNNYYVHSEATSKGYYIKNKD.GNDYDGWCWP
T.nigroviridis 446 FSSPKEMLQGLKDKKRKLVAIVDPHIKIESSYKLHDEIHARDLYVKNKD.GGNYEGWCWP
D.rerio 420 FPTPKDMLKGLMDKRRKLVAIVDPHIRVDSGYRIHNEIRSKNFYVKNKD.GGDYEGWCWP
I.punctatus 444 FPQPKEMLQGLMEKRRKMVTIVDPHIKVDSGYKIHNEIRSKDFYVKNKD.GADYEGWCWP
X.laevis 436 FPNPRDMLSGLKEKRRKMVAIVDPHIKIDSGYRVHNEIRAQNLYIKTKD.GSDYEGWCWP
M.domestica 468 FPQPLAMLEHLAGKRRKLVTIVDPHIKVDSEYPVHEELRSQGLYVKTRD.GSDYEGWCWP
M.musculus 446 FPQPLNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEELRNHGLYVKTRD.GSDYEGWCWP
C.jacchus 455 FPQPRTMLERLASKRRKLVTIVDPHIKVDSGYRVHDELRNLGLYVKTRD.GSDYEGWCWP
P.abelii 468 FPQPRTMLERLASKRRKLVAIVDPHIKVDSGYRVHEELRNLGLYVKTRD.GSDYEGWCWP
H.sapiens 446 FPQPRTMLERLASKRRKLVAIVDPHIKVDSGYRVHEELRNLGLYVKTRD.GSDYEGWCWP
M.fascicularis 446 FPQPRTMLERLASKRRKLVAIVDPHIKVDSGYRVHDELRNLGLYVKTRD.GSDYEGWCWP
S.scrofa 446 FPQPRTMLEHLASKRRKLVAIVDPHIKVDSSYRVHEELQNLGLYVKTRD.GSDYEGWCWP
B.taurus 464 FPQPRNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEELQNLGLYVKTRD.GSDYEGWCWP
O.aries 434 FPQPRNMLEHLASKRRKLVAIVDPHIKVDSGYRVHEELQNLGLYVKTRD.GSDYEGWCWP
E.caballus 468 FPQPLTMLEHLASKRRKLVAIVDPHIKVVSGYRVHEELQNQGLYVKTRD.GSDYEGWCWP
C.familiaris 468 FPQPLTMLQHLASKRRKLVTIVDPHIKVDSGYRVHEELQSRGLYVKTRD.GSDYEGWCWP
F.catus 468 FPQPLTMLEHLASKRRKLVTIVDPHIKVDSGYRVHEELQNRGLYVKTRD.GSDYEGWCWP
L.hesperus 404 FSDPVSMQQNLTARGRKLVAIVDPHIKRDNGYFLHNDATDLGFYTKNKD.NTDYEGWCWP
A.glabripennis 411 FAHPDVMINNLTATGRKLVIIIDPHIKREAGYFLHEDALANDYYVKNKD.GSVYEGWCWP
A.mellifera 425 FSNPVEMIHNLTAKGRKLVVIIDPHIKRDPSYFLHNDATKMGYYIKTRD.GKDYEGWCWP
B.cucurbitae 414 FPNPLGMIKNLTAVGRHLVVIVDPHVKRDNHYFFHHDCTERGYYVKTRD.GNDYEGWCWP
C.capitata 412 FSNPLGMFRNLSKLGRHMVIIIDPHIKRDDQYFFHKDCTERGLYTKTRE.GKDYEGWCWP
S.cerevisiae 442 FPNPKRLLSKLKKLGRNLVVLIDPHLKK..DYEISDRVINENVAVKDHN.GNDYVGHCWP
S.pombe 428 FPNPKAMLEKLDSKSRKLIVILDPHIKNDPNYFVSKELIDYNYAVKDKSGVDNYNADCWP
A.oryzae 452 FPNPIGMEEQLDESERKLVVIIDPHIKNKEKYTISEELKSKNLATRNKD.GEIYEGWCWP
N.fumigata 453 FPDPKGMEEQLDDSERKLVVIIDPHIKNKEGYSISEELKGKDLAIKNKG.GETYDGWCWP
K.pastoris 399 FPDPSGLLQELDETKRSLVAIIDPHIKV..GYEVSDYLESNGLVVREKDFDTPYHGHCWP
A.thaliana 475 GSSSYIDMLSPEIRKWWGGRFS.YKNYVGSTPSLYTWNDMNEPSVFNGPEVTMPRDALHV
G.soja 443 GSSSYPDTLNPEIRSWWADKFS.YQSYEGSTPSLYIWNDMNEPSVFNGPEVTMPRDVTHY
D.discoideum 503 GSSSYLDFTNPEIRKWWATQFG.YDKYKGSTPNLYIWNDMNEPSVFNGPEVSMHKDAKHH
T.nigroviridis 505 GSTGYPDFTNPEMRALWVRMFA.YDQYEGSMDNLHVWNDMNEPSVFNGPEITMIKDAKH.
D.rerio 479 GNSGYPDFTNPEMRAWWASMFA.YDQYEGSMENQYIWNDMNEPSVFNGPEVTMHKDAVH.
I.punctatus 503 GSAGYPDFTNHEMRAWWANQFG.HNQYEGSMENLYTWNDMNEPSVFNGPEVTMHKDALL.
X.laevis 495 GSAAYPDFTNPEMRKWWASMFS.YDKYEGSMDNLFVWNDMNEPSVFNGPEVTMHKDALHW
M.domestica 527 GSAGYPDFTNIKMRAWWANMFS.FDKYEGSASNLFIWNDMNEPSVFNGPEVTMLKDARHD
M.musculus 505 GSASYPDFTNPRMRAWWSNMFS.FDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAVHY
C.jacchus 514 GSAGYPDFTNPTMRSWWANMFS.YDNYEGSAPNLFVWNDMNEPSVFNGPEVTMLKDAQHY
P.abelii 527 GSAGYPDFTNPTMRAWWANMFS.YDNYEGSAPNLFVWNDMNEPSVFNGPEVTMLKDAQHY
H.sapiens 505 GSAGYPDFTNPTMRAWWANMFS.YDNYEGSAPNLFVWNDMNEPSVFNGPEVTMLKDAQHY
M.fascicularis 505 GSAGYPDFTNPTMRAWWANMFS.YDNYEGSAPNLFVWNDMNEPSVFNGPEVTMLKDAQHY
S.scrofa 505 GAASYPDFTNPKMRAWWADMFR.FENYEGSSSNLYVWNDMNEPSVFNGPEVTMLKDAQHY
B.taurus 523 GAAGYPDFTNPKMRAWWANMFH.FDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAQHY
O.aries 493 GAAGYPDFTNPKMRAWWANMFH.FDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAQHY
E.caballus 527 GAAGYPDFTNPTMRAWWANMFS.FENYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAQHY
C.familiaris 527 GAAGYPDFTNPTMRAWWANMFS.FDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAQHY
F.catus 527 GAAGYPDFTNPTMRAWWANMFS.FDNYEGSAPNLYVWNDMNEPSVFNGPEVTMLKDAQHY
L.hesperus 463 GSSSYIDFHNPKLREYIASRYS.LKNWQGSTRDLHVWNDMNEPSVFNGPEVTMPKDNIHF
A.glabripennis 470 GSSSYLDFYDPKTVDYYKGLYA.LDKFKGTTHDVYIWNDMNEPSVFNGPEVTMPKDCVHF
A.mellifera 484 GSSSYLDFFDPAVREYYINQYS.LDKFHGTTNDVYIWNDMNEPSVFNGPEVTMPKDVIHY
B.cucurbitae 473 GAASYTDFFNPEARAYYASQYL.PANYKATTTDVMIWNDMNEPSVFNGPEVTMSKDVIHY
C.capitata 471 GAASYPDFFNPEVRAYYASQYL.PSNFKTVTADVMIWNDMNEPSVFNGPEVTMPKDMIHY
S.cerevisiae 499 GNSIWIDTISKYGQKIWKSFFERFMDLPADLTNLFIWNDMNEPSIFDGPETTAPKDLIHD
S.pombe 488 GNSVWVDFFNPEAQAWWGSLYE.FDRFE.SDKNLWIWNDMNEPSVFRGPETSMHRDAIHY
A.oryzae 511 GSSHWVDCFNPEAIKWWIGLFK.YDKFKGTLPNVFIWNDMNEPSVFNGPETTMPKDNLHY
N.fumigata 512 GSSHWVDCFNPEAIKWWTGLFK.YDKFKGTQPNVFIWNDMNEPSVFNGPETTMPKDNIHY
K.pastoris 457 GESVWIDTFNPNAQSYWDTLFCKGSDFAGSSSNLHVWNDMNEPSVFNGPETTFPKDLVHF
A.thaliana 534 GGVEHRFVHNAYGYYFHMATSDGLVMREE.....GKDRPFVLSRAIFPGTQRYGAIWTG
G.soja 502 GGVEHRELHNAYGYYFHMATANGLLKRGE.....GNDRPFVLSRALFAGSQRYGAVWTG
D.discoideum 562 GGFEHRDVHNLYGYYYHMASADGLVQRNA.....DQNDRPFVLSRAFYAGSQRIGAIWTG
T.nigroviridis 563 GEWEHRDVHNLYGLYVQMATADGLIQRSG.....GVERPFVLARSFFAGSQRYGAVWTG
D.rerio 537 GVWEHRDVHNLYGLYVQKATSEGLIQRSG.....GVERPFVLTRAFFAGSQRYGAVWTG
I.punctatus 561 GGWENRDVHNLYGLYVHMATAEGLTQRSG.....GTERPFVLSRAFFAGSQRYGAIWTG
X.laevis 554 GGWEHRDVHNLYGLYVQRATTEGLIQRSG.....GKERPFVLTRAFFAGSQRYGAVWTG
M.domestica 586 GGWEHRDIHNIYGFYVHMATAEGLIQRSG.....GTERPFVLARAFFAGSQRYGAVWTG
M.musculus 564 GGWEHRDIHNIYGLYVHMATADGLIQRSG.....GIERPFVLSRAFFSGSQRFGAVWTG
C.jacchus 573 GGWEHRDIHNIYGLYVHMATADGLRQRSG.....GMERPFVLGRAFFAGSQRFGAVWTG
P.abelii 586 GGWEHRDVHNIYGLYVHMATADGLRQRSG.....GMERPFVLARAFFAGSQRFGAVWTG
H.sapiens 564 GGWEHRDVHNIYGLYVHMATADGLRQRSG.....GMERPFVLARAFFAGSQRFGAVWTG
M.fascicularis 564 GGWEHRDVHNIYGLYVHMATADGLRQRSG.....GMERPFVLARAFFAGSQRFGAVWTG
S.scrofa 564 GGWEHRDIHNIYGFYVHMATADGLVLRSG.....GVERPFVLSRAFFAGSQRFGAVWTG
B.taurus 582 GGWEHRDVHNIYGLYVHMATADGLVLRSG.....GIERPFVLSRAFFAGSQRFGAVWTG
O.aries 552 GGWEHRDVHNIYGFYVHMATADGLVLRSG.....GIERPFVLSRAFFAGSQRFGAVWTG
E.caballus 586 GGWEHRDVHNIYGFYVHMATADGLVQRSG.....GIERPFVLSRAFFAGSQRFGAVWTG
C.familiaris 586 GSWEHRDVHNIYGFYVHMATADGLVLRSG.....GLERPFVLSRAFFAGSQRFGAVWTG
F.catus 586 GGWEHRDVHNIYGFYVHMATAEGLVLRSG.....GLERPFVLSRAFFAGSQRFGAVWTG
L.hesperus 522 GGWEHREVHNMQGLLQTMTTYHGLIDRDDTIQKLEDKKRPFILTRSAFSGSQRFTAIWTG
A.glabripennis 529 GGLEHREVHNEFALKQVIPTYEGLIQRS.....NHRRPFILSRGHFSGSQRYVTIWTG
A.mellifera 543 GGWEHRNVHNINGLLMSMATYEALFRRSG.....GSLRPFILTRSFFAGSQRYTAMWTG
B.cucurbitae 532 GNWEHRDVHNLYGHMHALSTFEGLMGRD.....PNQRPFILTRSHFAGTQRYALIWTG
C.capitata 530 GNWEHRDVHNLYGHMHVMGTFDGLLGRD.....PNQRPYILTRSHFAGTQRYASIWTG
S.cerevisiae 559 NYIEERSVHNIYGLSVHEATYDAIKSIYS.....PSDKRPFLLTRAFFAGSQRTAATWTG
S.pombe 546 GGWEHRDIHNIYGHKCINGTYNGLIKRGE.....GA.VRPFILTRSFFAGTSALAANWIG
A.oryzae 570 GNWEHRDIHNVNGITFVNATYNAMLERKK.....GELRPFILTRSYYAGAQRMSAAMWTG
N.fumigata 571 GNWEHRDVHNVNGLTFINATYNALLERKK.....GVVRPFVLTRSFYAGAQRVSAMWTG
K.pastoris 517 GKWEHRSIHNWYGKSFHEATYYALINRSP.....NHRPFVLTRSYAPGSQATAAMWTG

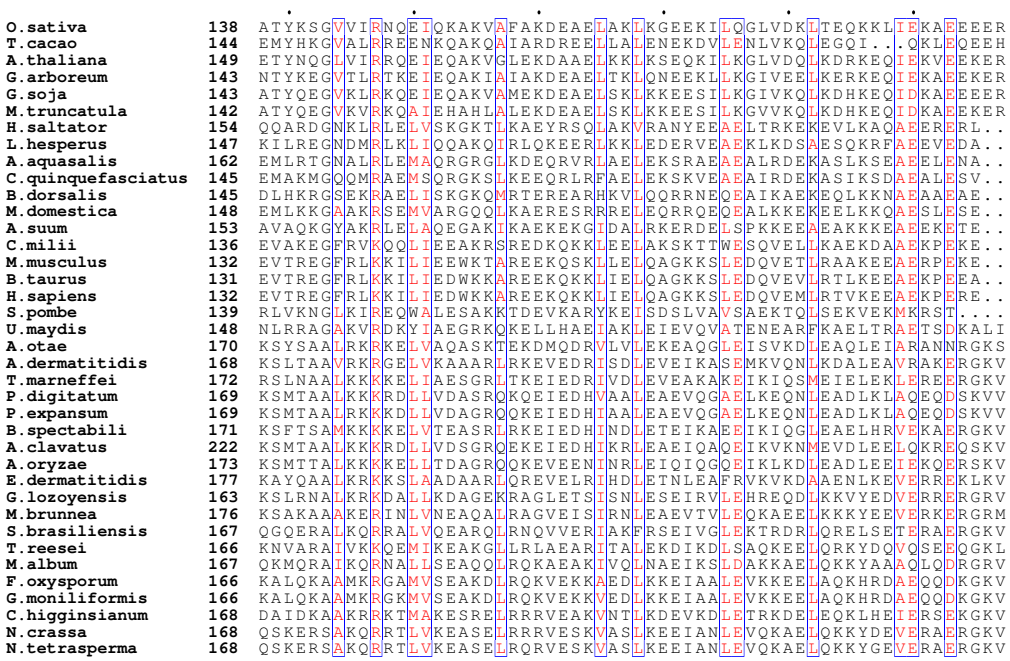

A.thaliana 588 DNTAEWEHLRVSIPMILTLGLTGITFSGADIGGFFGNPEPELLVRWYQVGAYYPFFRGHA
G.soja 556 DNTADWDHLRVSIPMVLTLGLTGMSFSGADIGGFFGNPEPELLVRWYQLGAYYPFFRAHA
D.discoideum 617 DNSAQWSHLEISNPMLLSMNLAGITFSGADVGGFFGNPDAELLTRWYQAGAFQPFFRGHA
T.nigroviridis 617 DNIAEWDHLKISLPMGLSMSLAGIVFCGSDIGGFFNAPGTELLVRWYQASSFMPFFRSHA
D.rerio 591 DNAAEWGHLKISIPMGLSLGLVGISFCGADVGGFFKHPSAELLVRWYQAGAYQPFFRAHA
I.punctatus 615 DNAAEWGHLKMSIPMGLSLGLVGISFCGADVGGFFKHPSAELLVRWYQAGAYQPFFRSHA
X.laevis 608 DNAAEWDHLKISIPMGLSLSLVGISFCGADVGGFFKSPETELLVRWYQAGAYQPFFRAHA
M.domestica 640 DNAAEWGHLKISIPMGLSMGLVGLSFCGADIGGFFKNPEPELLMRWYQMGAYQPFYRAHA
M.musculus 618 DNTAEWDHLKISIPMGLSLALVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
C.jacchus 627 DNTAEWDHLKISVPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
P.abelii 640 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
H.sapiens 618 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
M.fascicularis 618 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
S.scrofa 618 DNTAEWDHLKISIPMGLSLGLVGVSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
B.taurus 636 DNAAEWDHMKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
O.aries 606 DNAAEWDHMKISISMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
E.caballus 640 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFYRAHA
C.familiaris 640 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
F.catus 640 DNTAEWDHLKISIPMGLSLGLVGLSFCGADVGGFFKNPEPELLVRWYQMGAYQPFFRAHA
L.hesperus 582 DNAAEWSHLRVSYPMGLSLAIGGMSFCGADVGGFFHNPDKELLIRWYQAGAFLPFFRAHA
A.glabripennis 582 DNAAEWSHLAASFPMGLSEAIGGISFCGADIGGFFNNPDTELLQRWYQAGIWLPFYRAHA
A.mellifera 597 DNTADWDHLRISYPMGLSLAVSGMSFCGADVSGFFKNPDSELFIRWNQAGAWLPFYRQHS
B.cucurbitae 585 DNMADWGHLQHSIKMGLTEAVAGFSFCGADVGGFFGNPNAELFFRWYQTGAFLPFFRGHA
C.capitata 583 DNTADWGHLQHSLKMGLSEAVAGFSFCGADVGGFFGNPDAELFARWYQTGAFLPFFRGHA
S.cerevisiae 614 DNVANWDYLKISIPMVLSNNIAGMPFIGADIAGFAEDPTPELLIARWYQAGLWYPFFRAHA
S.pombe 600 DTMTTWEHLRGSIPTVLTNGISGMAFSGADVAGFFGNPDAELFVRWYETAIFYPFFRAHA
A.oryzae 625 DNQATWEHLAISLPMVLNNGISGFPFAGADVGGFFHNPSKDLLTRWYQTGIWYPFFRAHA
N.fumigata 626 DNQATWEHLAASLPMVLNNGIAGFPFAGADVGGFFQNPSKELLTRWYQAGIWYPFFRAHA
K.pastoris 570 DNAATWEYLKVALPMILSNGIAGMPFAGADVGGFFGNPSKELLTRWYQSGIWYPFFRAHA
A.thaliana 648 HHDTKRREPWLFGERNTELMRDAIHTRYTLLPYFYTLFREANVTGVPVVRPLWMEFPQDE
G.soja 616 HHDTKRREPWLFGERNTELIKDAIHVRYALLPYFYTLFREANTTGVPVVRPLWMEFPSDE
D.discoideum 677 HLDSRRREPWLFNEPYTTIIREAIVKRYSYLPLWYTTFYQNTLNGAPVMRPLWVQYPKEA
T.nigroviridis 677 HMDTPRREPWLHGLDNTRLIRDIVRQRYTFLPFWYQQFYHSHKTGEPIMRMLWIDYPKET
D.rerio 651 HIDTPRREPWLFGPENTALIREAVRQRYALLPNWYQLFYNAHNTGQPVMRPLWVEYPAEV
I.punctatus 675 HLDTPRREPWLFGPENTALIREAIRQRYTLLPYWYQLFYNAYRTGQPVMRALWVNYPEEP
X.laevis 668 HLDTPRREPWLHGDDNMAVIRDVLRQRYTLLPFWYTLFYKALSPGEPVMRPLWVEFPSDV
M.domestica 700 HMDTGRREPWLLAPEYLGPIRDALRQRYALLPFWYTLFYRAHLDGHPVMRPMWVQYPKDV
M.musculus 678 HLDTGRREPWLLASQYQDAIRDAIFQRYSLLPFWYTLFYQAHKEGFPVMRPLWVQYPEDM
C.jacchus 687 HLDTGRREPWLLPSPYNDIIRDALNQRYSLLPFWYTLFYQAHREGIPVMRPLWVQYPQDV
P.abelii 700 HLDTGRREPWLLPSQHNDIIRDALGQRYSLLPFWYTLLYQAHREGIPVMRPLWVQYPQDV
H.sapiens 678 HLDTGRREPWLLPSQHNDIIRDALGQRYSLLPFWYTLLYQAHREGIPVMRPLWVQYPQDV
M.fascicularis 678 HLDTGRRGPWLLPSQHNDIIRDALGQRYSLLPFWYTLFYQAHREGIPIMRPLWVQYPQDV
S.scrofa 678 HLDTGRREPWLLPTQYQDMIRDALGQRYSLLPFWYTLFYQAHREGVPVMRALWVHYPQDV
B.taurus 696 HLDTGRREPWLLPSQYHEIIRDALGQRYSLLPFWYTLFYQSHREGIPVMRPLWVHYPKDV
O.aries 666 HLDTGRREPWLLPSQYHDIIRDALGQRYSLLPFWYTLFYQSHREGIPVMRPLWVHYPKDV
E.caballus 700 HLDTGRREPWLLPSQYHDMVRDALNQRYSLLPFWYTLFYQAHREGIPVMRPLWVNFPQDV
C.familiaris 700 HLDTGRREPWLLPTQYHDIIRDALGQRYSLLPFWYTLFYQAHREGIPVMRPLWVHYPQDV
F.catus 700 HLDTGRREPWLLPTQYHDIIRDALGQRYSLLPFWYTLFYQAHREGIPVMRPLWVHYPQDV
L.hesperus 642 HIDTKRREPWLFGQETTDIIRSVIRTRYSFLPLWYTTFYQHNITGDPVIRPLWAEFPEDS
A.glabripennis 642 HIDTRRREPYLFNEDVRNRIRTALRIRYAHLPLWYTLFWEHEITGDPVIRPLFYQYPDDA
A.mellifera 657 HIETKRREPWLFNEETTQIVREAFRMRYSYLPLWYTLFREHEVNGTPVIRPLWAHYPSET
B.cucurbitae 645 HLETKRREPWVYGEKTRLIVREALRKRYSYLPLWYTMFYEHEVTGEPIMRPLLTHYPKDK
C.capitata 643 HIDTKRREPWLFPEDTRLIVRDAIRKRYAFLPLWYTMFYEHEMTGEPVMRPLLAHYPQDK
S.cerevisiae 674 HIDTKRREPYLFNEPLKSIVRDIIQIRYFLLPTLYTMFHKSSVTGFPIMNPMFIEHPEFA
S.pombe 660 HIDTKRREPWLYGEPYTSLVRELLRIRYRLLPTWYTAFYNSHTHGFPILYPQFLMHPEDE
A.oryzae 685 HIDTRRREPYLISEPHRSIIAQAIRLRYQLLPAWYTAFHEASVNGTPIVRPQFYVHPSDE
N.fumigata 686 HIDTRRREPYLIAEPFRSIISQAIRLRYQLLPAWYTAFHEASVNGMPIVRPQYYVHPADE
K.pastoris 630 HIDSRRREPWIAGEPYTTIIRDTVRLRYQLLPVLYTAFYESSKLGYPVLRPLFYEIPENQ
A.thaliana 708 ATFSNDEAFMVGS.GLLVQGVYTKGTTQASVYLPGKE.....SWYDLRNGKTYVGGK..
G.soja 676 ATFSNDETFMVGS.SILVQGIYTERAKHASVYLPGKQ.....SWYDLRTGAVYKGGV..
D.discoideum 737 NLFDVDDHYLIGD.SLLVKPVTQQSCKTMKVLLPGQS...VNEIWYDVDTEKPIN.AG..
T.nigroviridis 737 ATFAIDDEFLIGK.DLLVHPVTDEGATSVTAYLPGKD.....EVWFDIHTFKKHNGGQ..
D.rerio 711 TTFSIEDEYLIGK.DLLVHPVTDEGATGVTAYLPGKG.....EVWYDVHSLQKHDGDQ..
I.punctatus 735 ATFAIENEYLIGK.DLLVHPVTDEGAKSVTAYLPGKG.....EVWYDVHTFHKHEGAQ..
X.laevis 728 STFAMDSQYMLGS.ALLVRPVTEAKARGVQVYLPGDG.....EVWYDVHSYQRYEAPQ..
M.domestica 760 ATFSLDDQFMLGD.SLLVHPVAEPGARGVQVYLPGEG.....EVWYDTQSHQKHHGPQ..
M.musculus 738 STFSIEDQFMLGD.ALLIHPVSDAGAHGVQVYLPGQE.....EVWYDIQSYQKHHGPQ..
C.jacchus 747 TTFSIDDQFLLGD.ALLVHPVSDSGAHAVQVYLPGQG.....EVWYDIQSYQKHHGPQ..
P.abelii 760 TTFNIDDQYLLGD.ALLVHPVSDPGAHGVQVYLPGQG.....EVWYDIQSYQKHHGPQ..
H.sapiens 738 TTFNIDDQYLLGD.ALLVHPVSDSGAHGVQVYLPGQG.....EVWYDIQSYQKHHGPQ..
M.fascicularis 738 TTFSIDDQYLLGD.ALLVHPVSDSGAHGVQVYLPGQG.....EVWYDIQSYQKHHGPQ..
S.scrofa 738 TTFSIDDEFLLGD.ALLVHPVTDSEAHGVQVYLPGQG.....EVWYDVHSYQKYHGPQ..
B.taurus 756 TTFSIDDQFLLGD.ALLVHPVSDSEARGVQVYLPGQG.....EVWYDVQSYQKYHGPQ..
O.aries 726 TTFSIDDQFLLGD.ALLVHPVSDSEAHGVQVYLPGQG.....EVWYDVQSYQKYHGPQ..
E.caballus 760 TTFSIDDQFLLGD.ALLVHPVSDAGARGVQVYLPGEG.....EVWYDIESYQKHHGPQ..
C.familiaris 760 TTFSIDDQFLLGD.ALLVHPVSDSGAHGVQVYLPGQG.....EVWYDSQSYQKHYGPQ..
F.catus 760 TTFSIDDQFLLGD.ALLVHPVSDSGAHGVQVYLPGQG.....EVWYDTQSYQKHYGPQ..
L.hesperus 702 NTLKNDDHLLVGD.SLLVRPIFGLGETQASVYFPGKD.....TIWYDIDTYQTYR.HG..
A.glabripennis 702 NVVDIDKELLVGS.SILALVISEPGISTTNVYLPGGS....TEYWYDIEDFKLYPGTG..
A.mellifera 717 ETYAIDDEVLIGD.SILVRPVFQPSVTDVNVYFPGEG....KIIWYDVDTMQPYYQPG..
B.cucurbitae 705 EAFAIDNEFLLQD.RLLVRPVIEQGVKKVNVYFPAIDDKSNGDVWYNIDTLKRFTNVG..
C.capitata 703 EGYNIDYEFLLQD.RLLVRPVMDRGASKVNVYFPSTDDKKTGDVWYDINTLKKYTNAG..
S.cerevisiae 734 ELYHIDNQFYWSNSGLLVKPVTEPGQSETEMVFP.PG.....IFYEFASLHSFINNGTD
S.pombe 720 EGFAIDDQFYVGDSGLLVKPVTHPSIDKITIYLADDE.....VYFDLHDHTEYAGKG..
A.oryzae 745 AGFVIDDQLYLGSTGILTKPVVTEGTTTVDIYLADDE.....KYYDYFDVTVYQGAGK.
N.fumigata 746 QGFAIDDQLYLGSTGLLAKPVVVEGATTTDIYIADDE.....KYYDYYDFTVYQGAGR.
K.pastoris 690 KLYNIDDQFFVGSSGLMVKPVTEEGATTDSIYLPDPQ.....IYYDFFKLGTTLQGEG.

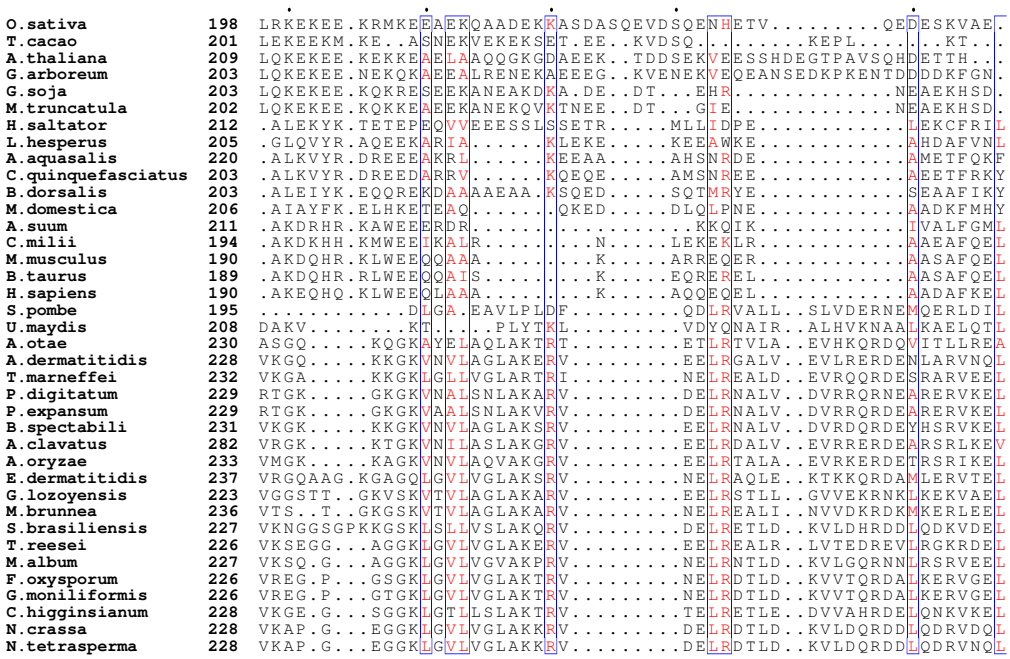

A.thaliana 759 ..THKMDAPEESIPAFQKAGTIIPRKDRFRRSSSQMDNDPYTLVVALNSSQE.AEGELYI
G.soja 727 ..THKLEVTEESIPAFQRAGTIIARKDRFRRSSTQMANDPYTLVVALNSSQA.AEGELYI
D.discoideum 790 ..VIEIDTPLEKIPVYQRGGSIISKKERVRRSTYQMRDDPYTIRIALDSSKS.AQGQLYI
T.nigroviridis 789 ..SLYIPVTISSIPVFQRGGSIIPKKLRVRRSSACMEHDPYTLYVALSNQKT.AEGEIYI
D.rerio 763 ..SLYIPVTMSSIPVFQRGGSIICRKERVRRSSSCMENDPYTLYVALNSQGF.AEGELYI
I.punctatus 787 ..NLHIPVTMSSIPVFQRGGSIISRKDRIRRSSSCMENDPYTLYIALSAQGT.AEGELYI
X.laevis 780 ..TFYLPVTMNSIPVYQRGGSIIPRKDRPRRTSDCMQDDPFTLYVALNLQGE.ARGELFL
M.domestica 812 ..TLYLPVTFSSIPVFQRGGTIIPRWERVRRSSDCMKDDPITLYVALSPQGT.AEGELYL
M.musculus 790 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSDCMKDDPITLFVALSPQGT.AQGELFL
C.jacchus 799 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSDCMKDDPITLFVALSPQGT.AQGELFL
P.abelii 812 ..TLYLPVTLSSIPVFQRGGTVVPRWMRVRRSSECMKDDPITLFVALSPQGT.AQGELFL
H.sapiens 790 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSECMKDDPITLFVALSPQGT.AQGELFL
M.fascicularis 790 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSECMKDDPITLFVALSPQGT.AEGELFL
S.scrofa 790 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSDCMKDDPITLFVALSPQGT.AQGELFL
B.taurus 808 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSDCMKDDPITLFVALSLQGT.AQGELFL
O.aries 778 ..TLYLPVTLSSIPVFQRGGTIVPRWMRVRRSSDCMKDDPITLFVALSLQGT.AQGELFL
E.caballus 812 ..TLYLPVTLSTIPVFQRGGTIVPRWMRVRRSSECMKDDPITLFVALSPQGT.AQGELFL
C.familiaris 812 ..TLYLPVTLSSIPVFQRGGTIIPRWMRVRRSSDCMKDDPITLFVALSPQGT.AQGELFL
F.catus 812 ..TLYLPVTLSSIPVFQRGGTIIPRWMRVRRSSDCMKDDPITLFVALSPQGT.AQGELFL
L.hesperus 753 ..TTNVAADITKIPVFYRGGHIIPRKMRVRRSSPLMKNDPYTLYIALDANDQ.ASGYLYI
A.glabripennis 755 ..VYNLPASLDKNIAFYRGGSIIPRKDRPRRAASLMHNDPFSLYVALDSKKS.ASGTLYL
A.mellifera 770 ..LVNIPVTLHKIPVFQRGGSIISRKMRIRRSTVAMKNDPYTLIVITDSAGK.ANGTLYI
B.cucurbitae 762 ..YESIPVESDTIPVFQRGGTIIPKKERIRRASTLMKNDPYTLVICLNREGK.AEGTLYV
C.capitata 760 ..YESLPVDSRTIPVFQRGGTIIPKKERIRRAATLMKNDPYTLVVCLDRAGK.AQGTLYI
S.cerevisiae 787 LIEKNISAPLDKIPLFIEGGHIITMKDKYRRSSMLMKNDPYVIVIAPDTEGR.AVGDLYV
S.pombe 772 ..HQVVPAPLGRVPVLLRGGNILITRERIRRAAELTRNDPFTLTIAVSKIGKNASGFLYL
A.oryzae 798 ..RHEVPAPLEKVPVLMQGGHVIPRKDRPRRSSGLMKWDPYTLVVVLDKNGE.ADGTLYV
N.fumigata 799 ..RHTVPSPIEKVPLLMQGGHIIPRKDRARRSSGLMRWDPYTLVIVLDKNGK.AEGTLYV
K.pastoris 743 ..IHELEVDLTTIPVLARGGSIITQKLRYRRSSQLMKHDPYTLMVFLDSEGT.AKGKLYI
A.thaliana 816 DDGKSFEFRRG..SYIHRRFVFSKGVLTSTNLAPPE...ARL...SSQCLIDRIILLGHS
G.soja 784 DDGSSFNFLQG..GYIHRRFIFSNGKLTSIDLAPASGSKGRY...PSDAFIERIILLGHA
D.discoideum 847 DDEHSFDYKKG..KFLYRQFTFKDNVLSFSDASNKSS..TSY...KPNVTIEKIVILGVQ
T.nigroviridis 846 DDGYTFNYEK..KEFIHRRISLANNVISSINLAPD....AHF...TTPSWIERIVIIGAT
D.rerio 820 DDFHTFNYQKS.KQFIHRRLSFSNNSLTSRNLAPD....AQF...STESWIERIVVIGGS
I.punctatus 844 DDFHTFNFEKE.KQFIHRRLHFFDNVLTSSNLAPD....SKF...TTAAWIEKVVIIGAT
X.laevis 837 DDGHSFNYE.Q.NEFLFREFNYSQGKLTASSLDPS....GVF...KTPSWIERIVIIGAT
M.domestica 869 DDGHTFNYETQ.NEFLFRKFTFSGNTLTSSSADPR....GHF...ETPIWIERVVILGAG
M.musculus 847 DDGHTFNYQTR.HEFLLRRFSFSGSTLVSSSADPK....GHL...ETPIWIERVVIMGAG
C.jacchus 856 DDGHTFNYQTR.QEFLLRRFSFSGNTLVSSSVPGD.....P...PTPAAVGRQGRFGDG
P.abelii 869 DDGHTFNYQTR.QEFLLRRFSFSGNTLVSSSADPA....GHF...ETPIWIERVVIIGAG
H.sapiens 847 DDGHTFNYQTR.QEFLLRRFSFSGNTLVSSSADPE....GHF...ETPIWIERVVIIGAG
M.fascicularis 847 DDGHTFNYETR.QEFLLRRFLFSGNTLVSSSADPE....GHF...ETPIWIERVVIIGAG
S.scrofa 847 DDGHTFNYQTG.HEFLLRRFSFSGNTLVSSSADSK....GHF...ETPVWIERVVIIGAG
B.taurus 865 DDGHTFNYQTR.HEFLLRRFSFSGNTLVSSSADPK....GHF...ETPIWIERVVIIGAG
O.aries 835 DDGHTFNYQTR.HEFLLRRFSFSGNTLVSSSADPK....GHF...ETPIWIERVVIIGAG
E.caballus 869 DDGHTFSYQTH.NEFLLRRFSFSGNTLVSSSADPK....GHF...ETPIWIERVVIIGAG
C.familiaris 869 DDGHTFNYQTH.HEFLLRRFSFSGNTLVSSSADPK....GYF...ETPVWIERVVIIGAG
F.catus 869 DDGHTFNYQTR.HEFLLRRFSFSGNTLVSSSADPR....GHF...ETPIWIERVVIIGAG
L.hesperus 810 DDGETFRYHLRQDSSAYIKYKFAENMLSSDFASNK.....RF...DSPVWLEKVVIMGVK
A.glabripennis 812 DDYQSFEYKNN.KKYLYIHFDFKENTLTSKLIDKT.....DY...PTKEWIERVVILGPP
A.mellifera 827 DDEVSFEYRHG..KYLYLSFNFEENTLTSTFIDKL....SSY...QTESWLERVDIANPP
B.cucurbitae 819 DDEKSFDYRNG..IHNYIKFNFENNKLDVNPIGKL.....KY...KTPAWIERVVIAGLE
C.capitata 817 DDEKSYAYRKG..EFNYLLFEFENGQLNVNFKSKP.....NF...KTTAWIERVVIAGLD
S.cerevisiae 846 DDGETFGYQRG..EYVETQFIFENNTLKNVRSHIPENLTGIHHNTLRNTNIEKIIIAKNN
S.pombe 830 DDGVTFNYKKG..EYLIRHFSYENGILTMKDSHSNPPVSPKYSSSQKHLKVERINIYGEQ
A.oryzae 855 DDGETFDYERG..AYIHRRFHFHDSTLSSEDIGTKGPKTAEYLKAMASVRERVVVVIDPP
N.fumigata 856 DDGESFNYQQG..AYIHRRFKFEKSTLLSEDIGTKGSKTAEYLKSMTNVRVQKVVVVDAP
K.pastoris 800 DDGTSFNYRLG..HFLEISFQYSNNILSSELKTN...IARKFQDSLDDVLVERVIIIGLP
A.thaliana 868 SGPK.....SALVEPLNQKAEIEMGPLRMGGLVASSGTKVLTIRKPGVRVDQDWTVKIL.
G.soja 839 PSSK.....NALIEPSNQKVDIELGPLWVLR...ARAPAVTTIRRPNVRVAEDWTITVI.
D.discoideum 900 K.PH.....SITCNITG.........KEKLSFEYDSTLSKLTIRKPDLLVDTDFIIKLN.
T.nigroviridis 897 K.PS.....KVTLKAAG.........GLESQLDFEFDPSVSVLTLRKPGVNAGEDWTITLK.
D.rerio 872 R.PS.....SVSLINAD.........GVESALEFEFSSNPAVLTVRKPSVSAAADWTVVLR.
I.punctatus 896 R.PS.....SVSVKTAD.........GAESAVEFDFDSAVSVLTLHKPGVNAGTDWTVTLQ.
X.laevis 888 K.PS.....SITLSLQG.........GSESSLEFEYDQSTSVITVRKPGVNIASDFTVSLR.
M.domestica 921 K.PA.....SVLLETEG.........IPETILTFHHNAETSVLTLRKPEVHVSTDWSIHLR.
M.musculus 899 K.PA.....AVVLQTKG.........SPESRLSFQHDPETSVLILRKPGVSVASDWSIHLR.
C.jacchus 905 L.AP.....LFFLLPPPG.........SPESRLSFQHDPETSVLVLRKPGINVASDWSIHLR.
P.abelii 921 K.PA.....AVVLQTKG.........SPESRLSFQHDPETSVLVLRKPGVNVASDWSIHLR.
H.sapiens 899 K.PA.....AVVLQTKG.........SPESRLSFQHDPETSVLVLRKPGINVASDWSIHLR.
M.fascicularis 899 K.PA.....AVVLQTKG.........SPESRLSFQHDPETSVLVLRKPGINVASDWSIHLR.
S.scrofa 899 K.PA.....TVVLQTKG.........SPESRLSFQHDPETSVLILRKPGVNVASDWSIHLR.
B.taurus 917 K.PA.....TVVLQTKG.........SPESRLSFQHDPETSVLILRKPGVNVASDWTIHLR.
O.aries 887 K.PA.....TVVLQTKG.........SPESRLSFQHDPETSVLILRKPGVNVASDWTIHLR.
E.caballus 921 K.PA.....AVVLQTKG.........SPETRLSFQHDPETSVLILRKPGVNVASDWSIHLQ.
C.familiaris 921 K.PA.....AVILQTKG.........SPETRLSFQHDPETSVLILRKPGVNVASDWSIHLR.
F.catus 921 K.PA.....AVVLQTKG.........SPESRLSFQHDPETSVLILRKPGINVASDWSIHLR.
L.hesperus 862 SGRQ.....YKASLFKD.........NTESPLEAKHDSTTNTLTIRKPGVPIAAKWIIKIK.
A.glabripennis 863 SGIK.....GAKLTSKS.........LGTVELETSYDGEERTLTVRKPGVSVREPFSIKLY.
A.mellifera 878 KGVK.....SAVLNSRS.........LTKITLETKYNPNNNVLTIRKPSVNMGEEWTIELIH
B.cucurbitae 869 RAPK.....SATLIIDG.........ISKELEVLPHGEAIAIRKPGVSVQQIYNIRLNF
C.capitata 867 RTPK.....SATLIVGG.........VSQDLEVLPHEEGVAIRKPGISVLELFTIRLNF
S.cerevisiae 904 LQHNITLKDSIKVKKNGE..ESSLPTRSSY.....ENDNKITILNLSLDITEDWEVIF..
S.pombe 888 TRKS..IKIRK.IIDS.........EVTEWDVSVD...DSGCIRNPQLFLV.....
A.oryzae 913 KGWQ..EKSTVTVIEDGA..KTASTAAMEYHAQPNGKAPYAVVKNPTVSIGKTWRIEF..
N.fumigata 914 KEWQ..GRTSVTVIEDGA..KMASTAPLEYHAQQAGKAAYAVVKKPDVGIGKTWKIEF..
K.pastoris 855 DDFT..SVKAKV..TQG.........EKTWISNWTKENGYLVVKNPRVSISQSWEINIEV

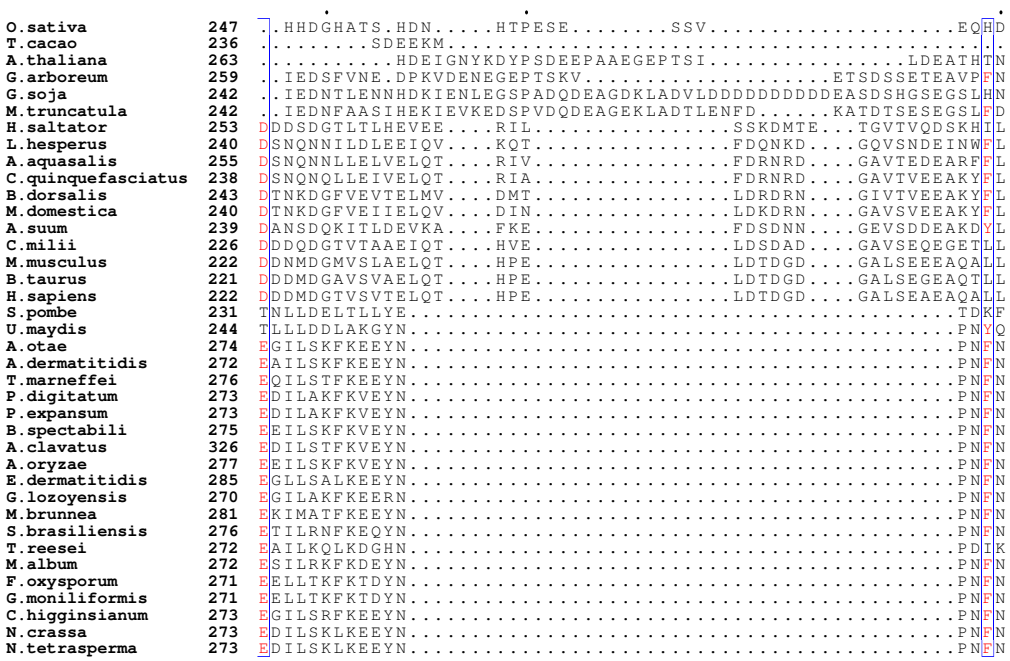

O.sativa
T.cacao 1
A.thaliana 1
G.arboreum 1
G.soja 1
M.truncatula 1
H.saltator 1
L.hesperus 1
A.aquasalis 1
C.quinquefasciatus 1
B.dorsalis 1
M.domestica 1
A.suum 1
C.milii 1
M.musculus
B.taurus
H.sapiens
S.pombe 1
U.maydis 1
A.otae 1
A.dermatitidis 1
T.marneffei 1
P.digitatum 1
P.expansum 1
B.spectabili 1
A.clavatus 1
A.oryzae 1
E.dermatitidis 1
G.lozoyensis 1
M.brunnea 1
S.brasiliensis 1
T.reesei 1
M.album 1
F.oxysporum 1
G.moniliformis 1
C.higginsianum 1
N.crassa 1
N.tetrasperma 1
O.sativa 1
T.cacao 10
A.thaliana 12
G.arboreum 5
G.soja 10
M.truncatula 9
H.saltator 19
L.hesperus 17
A.aquasalis 23
C.quinquefasciatus 8
B.dorsalis 6
M.domestica 11
A.suum 17
C.milii 4
M.musculus 1
B.taurus 1
H.sapiens 1
S.pombe 4
U.maydis 8
A.otae 7
A.dermatitidis 8
T.marneffei 8
P.digitatum 8
P.expansum 8
B.spectabili 8
A.clavatus 58
A.oryzae 8
E.dermatitidis 8
G.lozoyensis 8
M.brunnea 21
S.brasiliensis 8
T.reesei 8
M.album 8
F.oxysporum 8
G.moniliformis 8
C.higginsianum 8
N.crassa 8
N.tetrasperma 8

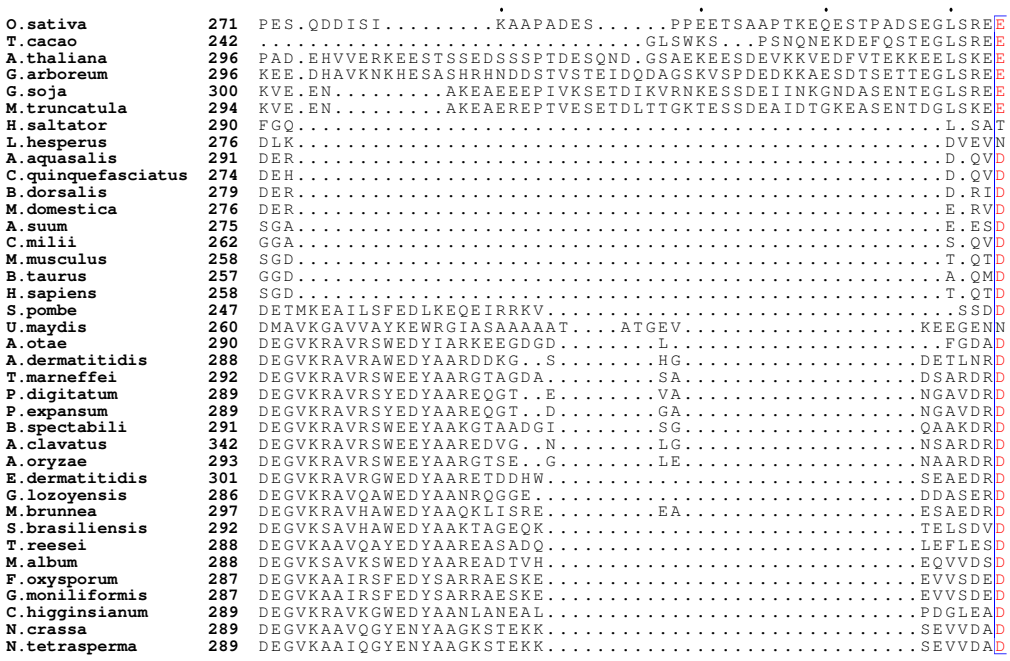

O.sativa 52
T.cacao 58
A.thaliana 63
G.arboreum 57
G.soja 57
M.truncatula 56
H.saltator 65
L.hesperus 62
A.aquasalis 76
C.quinquefasciatus 59
B.dorsalis 58
M.domestica 63
A.suum 67
C.milii 50
M.musculus 46
B.taurus 45
H.sapiens 46
S.pombe 55
U.maydis 60
A.otae 56
A.dermatitidis 57
T.marneffei 57
P.digitatum 54
P.expansum 54
B.spectabili 56
A.clavatus 107
A.oryzae 58
E.dermatitidis 60
G.lozoyensis 52
M.brunnea 65
S.brasiliensis 52
T.reesei 51
M.album 52
F.oxysporum 51
G.moniliformis 51
C.higginsianum 53
N.crassa 53
N.tetrasperma 53

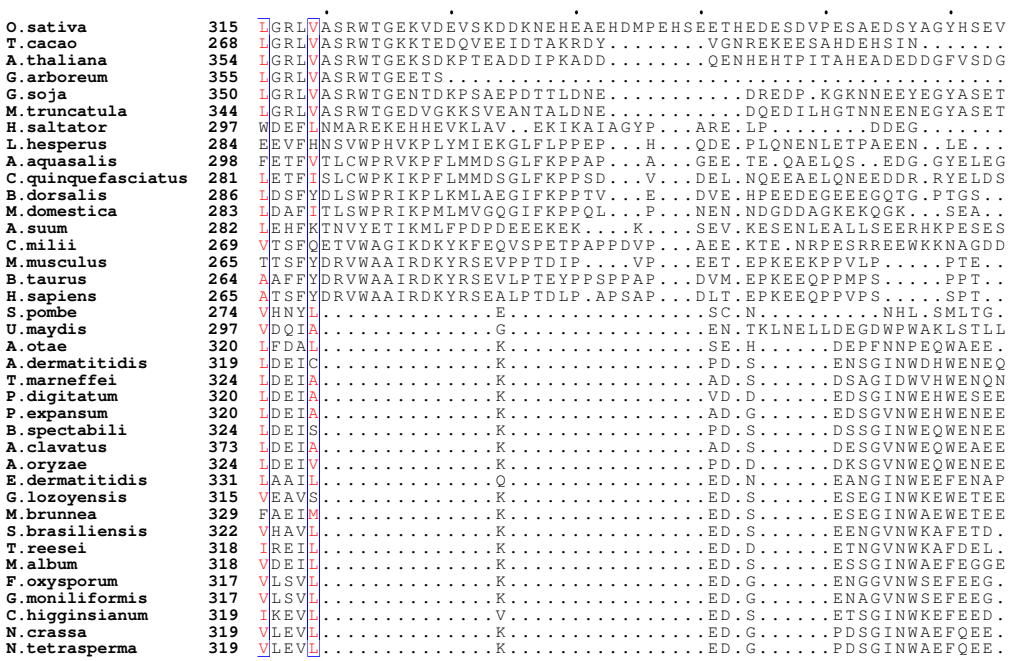

O.sativa 86
T.cacao 92
A.thaliana 97
G.arboreum 91
G.soja 91
M.truncatula 90
H.saltator 99
L.hesperus 96
A.aquasalis 110
C.quinquefasciatus 93
B.dorsalis 92
M.domestica 97
A.suum 101
C.milii 84
M.musculus 80
B.taurus 79
H.sapiens 80
S.pombe 89
U.maydis 94
A.otae 113
A.dermatitidis 111
T.marneffei 115
P.digitatum 112
P.expansum 112
B.spectabili 114
A.clavatus 165
A.oryzae 116
E.dermatitidis 120
G.lozoyensis 106
M.brunnea 119
S.brasiliensis 110
T.reesei 109
M.album 110
F.oxysporum 109
G.moniliformis 109
C.higginsianum 111
N.crassa 111
N.tetrasperma 111

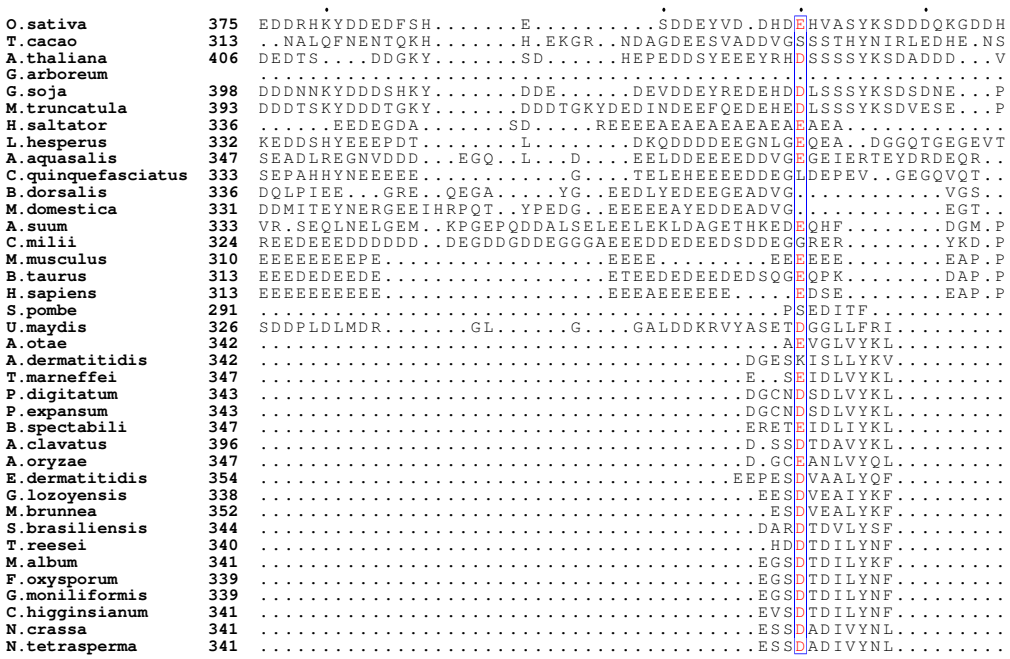

O.sativa 138 ATYKSGVVIRNQEIQKAKVAFAKDEAELAKLKGEEKILQGLVDKLTEQKKLIEKAEEEER
T.cacao 144 EMYHKGVALRREENKQAKQAIARDREELLALENEKDVLENLVKQLEGQI...QKLEQEEH
A.thaliana 149 ETYNQGLVIRRQEIEQAKVGLEKDAAELKKLKSEQKILKGLVDQLKDRKEQIEKVEEKER
G.arboreum 143 NTYKEGVTLRTKEIEQAKIAIAKDEAELTKLQNEEKLLKGIVEELKERKEQIEKAEEKER
G.soja 143 ATYQEGVKLRKQEIEQAKVAMEKDEAELSKLKKEESILKGIVKQLKDHKEQIDKAEEEER
M.truncatula 142 ATYQEGVKVRKQAIEHAHLALEKDEAELSKLKKEESILKGVVKQLKDHKEQIDKAEEKER
H.saltator 154 QQARDGNKLRLELVSKGKTLKAEYRSQLAKVRANYEEAELTRKEKEVLKAQAEERERL..
L.hesperus 147 KILREGNDMRLKLIQQAKQIRLQKEERLKKLEDERVEAEKLKDSAESQKRFAEEVEDA..
A.aquasalis 162 EMLRTGNALRLEMAQRGRGLKDEQRVRLAELEKSRAEAELRDEKASLKSEAEELENA..
C.quinquefasciatus 145 EMAKMGQQMRAEMSQRGKSLKEEQRLRFAELEKSKVEAEAIRDEKASIKSDAEALESV..
B.dorsalis 145 DLHKRGSEKRAELISKGKQMRTEREARHKVLQQRRNEQEAIKAEKEQLKKNAEAAEAE..
M.domestica 148 EMLKKGAAKRSEMVARGQQLKAERESRRRELEQRRQEQEALKKEKEELKKQAESLESE..
A.suum 153 AVAQKGYAKRLELAQEGAKIKAEKEKGIDALRKERDELSPKKEEAEAKKKEAEEKETE..
C.milii 136 EVAKEGFRVKQQLIEEAKRSREDKQKKLEELAKSKTTWESQVELLKAEKDAAEKPEKE..
M.musculus 132 EVTREGFRLKKILIEEWKTAREEKQSKLLELQAGKKSLEDQVETLRAAKEEAERPEKE..
B.taurus 131 EVTREGFRLKKILIEDWKKAREEKQKKLIELQAGKKSLEDQVEVLRTLKEEAEKPEEA..
H.sapiens 132 EVTREGFRLKKILIEDWKKAREEKQKKLIELQAGKKSLEDQVEMLRTVKEEAEKPERE..
S.pombe 139 RLVKNGLKIREQWALESAKKTDEVKARYKEISDSLVAVSAEKTQLSEKVEKMKRST....
U.maydis 148 NLRRAGAKVRDKYIAEGRKQKELLHAEIAKLEIEVQVATENEARFKAELTRAETSDKALI
A.otae 170 KSYSAALRKRKELVAQASKTEKDMQDRVLVLEKEAQGLEISVKDLEAQLEIARANNRGKS
A.dermatitidis 168 KSLTAAVRKRGELVKAAARLRKEVEDRISDLEVEIKASEMKVQNLKDALEAVRAKERGKV
T.marneffei 172 RSLNAALKKKKELIAESGRLTKEIEDRIVDLEVEAKAKEIKIQSMEIELEKLERERERGKV
P.digitatum 169 KSMTAALKKKRDLLVDASRQKQEIEDHVAALEAEVQGAELKEQNLEADLKLAQEQDSKVV
P.expansum 169 KSMTAALRKKKDLLVDAGRQQKEIEDHIAALEAEVQGAELKEQNLEADLKLAQEQDSKVV
B.spectabili 171 KSFTSAMKKKKELVTEASRLRKEIEDHINDLETEIKAEEIKIQGLEAELHRVEKAERGKV
A.clavatus 222 KSMTAALKKKRDLLVDSGRQEKEIEDHIKRLEAEIQAQEIKVKNMEVDLEELQKREQSKV
A.oryzae 173 KSMTTALKKKKELLTDAGRQQKEVEENINRLEIQIQGQEIKLKDLEADLEEIEKQERSKV
E.dermatitidis 177 KAYQAALKRKKSLAADAARLQREVELRIHDLETNLEAFRVKVKDAAENLKEVERREKLKV
G.lozoyensis 163 KSLRNALKRKDALLKDAGEKRAGLETSISNLESEIRVLEHREQDLKKVYEDVERRERGRV
M.brunnea 176 KSAKAAAKERINLVNEAQALRAGVEISIRNLEAEVTVLEQKAEELKKKYEEVERKERGRM
S.brasiliensis 167 QGQERALKQRRALVQEARQLRNQVVERIAKFRSEIVGLEKTRDRLQRELSETERAERGKV
T.reesei 166 KNVARAIVKKQEMIKEAKGLLRLAEARITALEKDIKDLSAQKEELQRKYDQVQSEEQGKL
M.album 167 QKMQRAIKQRNALLSEAQQLRQKAEAKIVQLNAEIKSLDAKKAELQKKYAAAQLQDRGRV
F.oxysporum 166 KALQKAAMKRGAMVSEAKDLRQKVEKKAEDLKKEIAALEVKKEELAQKHRDAEQQDKGKV
G.moniliformis 166 KALQKAAMKRGKMVSEAKDLRQKVEKKVEDLKKEIAALEVKKEELAQKHRDAEQQDKGKV
C.higginsianum 168 DAIDKAAKRRKTMAKESRELRRRVEAKVNTLKDEVKDLETRKDELEQKLHEIERSEKGKV
N.crassa 168 QSKERSAKQRRTLVKEASELRRRVESKVASLKEEIANLEVQKAELQKKYDEVERAERGKV
N.tetrasperma 168 QSKERSAKQRRTLVKEASELRQRVESKVASLKEEIANLEVQKAELQKKYGEVERAERGKV

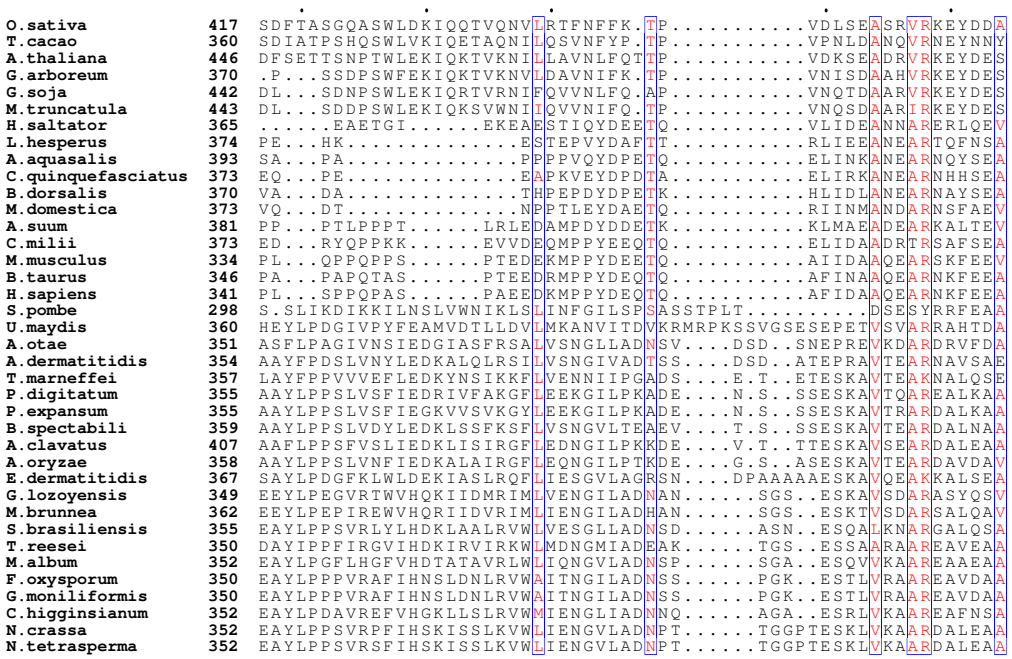

O.sativa 198 LRKEKEE.KRMKEEAEKQAADEKKASDASQEVDSQENHETV.....QEDESKVAE.
T.cacao 201 LEKEEKM.KE..ASNEKVEKEKSET.EE..KVDSQ.....KEPL.....KT..
A.thaliana 209 LQKEKEE.KEKKEAELAAQQGKGDAEEK..TDDSEKVEESSHDEGTPAVSQHDETTH..
G.arboreum 203 LQKEKEE.NEKQKAEEALRENEKAEEEG..KVENEKVEQEANSEDKPKENTDDDDKFGN.
G.soja 203 LQKEKEE.KQKRESEEKANEAKDKA.DE..DT...EHR.....NEAEKHSD.
M.truncatula 202 LQKEKEE.KQKKEAEEKANEKQVKTNEE..DT...GIE.....NEAEKHSD.
H.saltator 212 .ALEKYK.TETEPEQVVEEESSLSSETR.....MLLIDPE.....LEKCFRIL
L.hesperus 205 .GLQVYR.AQEEKARIA.....KLEKE.....KEEAWKE.....AHDAFVNL
A.aquasalis 220 .ALKVYR.DREEEAKRL.....KEEAA.....AHSNRDE.....AMETFQKF
C.quinquefasciatus 203 .ALKVYR.DREEDARRV.....KQEQE.....AMSNREE.....AEETFRKY
B.dorsalis 203 .ALEIYK.EQQREKDAAAEAA.KSQED.....SQTMRYE.....SEAAFIKY
M.domestica 206 .AIAYFK.ELHKETEAQ.....QKED.....DLQLPNE.....AADKFMHY
A.suum 211 .AKDRHR.KAWEEERDR.....KKQIK.....IVALFGML
C.milii 194 .AKDKHH.KMWEEIKAIR.....N.....LEKEKLR.....AAEAFQEL
M.musculus 190 .AKDQHR.KLWEEQQAAA.....K.....ARREQER.....AASAFQEL
B.taurus 189 .AKDQHR.RLWEEQQAIS.....K.....EQREREL.....AASAFQEL
H.sapiens 190 .AKEQHQ.KLWEEQLAAA.....K.....AQQEQEL.....AADAFKEL
S.pombe 195 .....DLGA.EAVLPLDF.....QDLRVALL..SLVDERNEMQERLDIL
U.maydis 208 DAKV.....KT.....PLYTKL.....VDYQNAIR..ALHVKNAALKAELQTL
A.otae 230 ASGQ.....KQGKAYELAQLAKTRT.....ETLRTVLA..EVHKQRDQVITLLREA
A.dermatitidis 228 VKGQ.....KKGKVNVLAGLAKERV.....EELRGALV..EVLRERDENLARVNQL
T.marneffei 232 VKGA.....KKGKLGLLVGLARTRI.....NELREALD..EVRQQRDESRARVEEL
P.digitatum 229 RTGK.....GKGKVNAISNLAKARV.....DELRNALV..DVRRQRNEARERVKEL
P.expansum 229 RTGK.....GKGKVAAISNLAKVRV.....DELRNALV..DVRRQRDEARERVKEL
B.spectabili 231 VKGK.....KKGKVNVLAGLAKSRV.....EELRNALV..DVRDQRDEYHSRVKEL
A.clavatus 282 VRGK.....KTGKVNILASLAKGRV.....EELRDALV..EVRRERDEARSRLKEV
A.oryzae 233 VMGK.....KAGKVNVIAQVAKGRV.....EELRTALA..EVRKERDETRSRIKEL
E.dermatitidis 237 VRGQAAG.KGAGQLGVLVGLAKSRV.....NELRAQLE..KTKKQRDAMLERVTEL
G.lozoyensis 223 VGGSTT..GKVSKVTVLAGLAKARV.....EELRSTLL..GVVEKRNKLKEKVAEL
M.brunnea 236 VTS..T..GKGSKVTVLAGLAKARV.....NELREALI..NVVDKRDKMKERLEEL
S.brasiliensis 227 VKNGGSGPKKGSKLSLLVSLAKQRV.....DELRETLD..KVLDHRDDLQDKVDEL
T.reesei 226 VKSEGG...AGGKLGVLVGLAKERV.....EELREALR..LVTEDREVLRGKRDEL
M.album 227 VKSQ.G...AGGKLGVLVGVAKPRV.....NELRNTLD..KVLGQRNNLRSRVEEL
F.oxysporum 226 VREG.P...GSGKLGVLVGLAKTRV.....NELRDTLD..KVVTQRDALKERVGEL
G.moniliformis 226 VREG.P...GTGKLGVLVGLAKTRV.....NELRDTLD..KVVTQRDALKERVGEL
C.higginsianum 228 VKGE.G...SGGKLGTLLSLAKTRV.....TELRETLE..DVVAHRDELQNKVKEL
N.crassa 228 VKAP.G...EGGKLGVLVGLAKKRV.....DELRDTLD..KVLDQRDDLQDRVDQL
N.tetrasperma 228 VKAP.G...EGGKLGVLVGLAKKRV.....DELRDTLD..KVLDQRDDLQDRVNQL

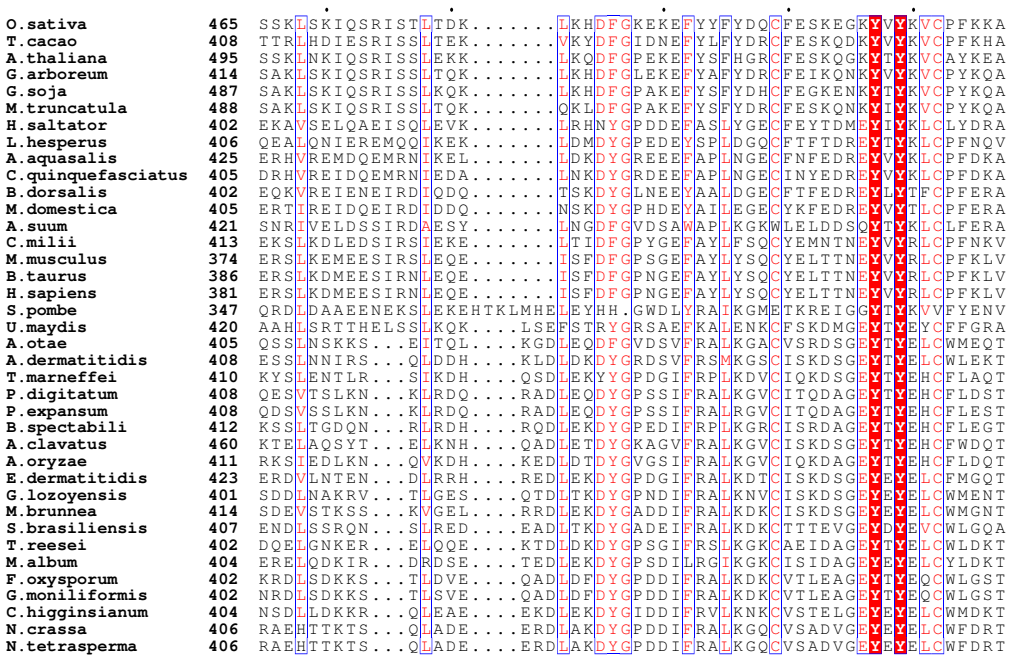

O.sativa 247 .HHDGHATS.HDN.....HTPESE.....SSV.....EQHD
T.cacao 236 .....SDEEKM.....
A.thaliana 263 .....HDEIGNYKDYPSDEEPAAEGEPTSI.....LDEATHTN
G.arboreum 259 .IEDSFVNE.DPKVDENEGEPTSKV.....ETSDSSETEAVPFN
G.soja 242 .IEDNTLENNHDKIENLEGSPADQDEAGDKLADVLDDDDDDDDDDEASDSHGSEGSLHN
M.truncatula 242 .IEDNFAASIHEKIEVKEDSPVDQDEAGEKLADTLENFD.....KATDTSESEGSLFD
H.saltator 253 DDDSDGTLTLHEVEE....RIL.....SSKDMTE...TGVTVQDSKHIL
L.hesperus 240 DSNQNNILDLEEIQV....KQT.....FDQNKD....GQVSNDEINWFL
A.aquasalis 255 DSNQNNLLELVELQT....RIV.....FDRNRD....GAVTEDEARFFL
C.quinquefasciatus 238 DSNQNQLLEIVELQT....RIA.....FDRNRD....GAVTVEEAKYFL
B.dorsalis 243 DTNKDGFVEVTELMV....DMT.....LDRDRN....GIVTVEEAKYFL
M.domestica 240 DTNKDGFVEIIELQV....DIN.....LDKDRN....GAVSVEEAKYFL
A.suum 239 DANSDQKITLDEVKA....FKE.....FDSDNN....GEVSDDEAKDYL
C.milii 226 DDDQDGTVTAAEIQT....HVE.....LDSDAD....GAVSEQEGETLL
M.musculus 222 DDNMDGMVSLAELQT....HPE.....LDTDGD....GALSEEEAQALL
B.taurus 221 DDMDGAVSVAELQT....HPE.....LDTDGD....GALSEGEAQTLL
H.sapiens 222 DDMDGTVSVTELQT....HPE.....LDTDGD....GALSEAEAQALL
S.pombe 231 TNLLDELTLLYE.....TDKF
U.maydis 244 TLLLDDLAKGYN.....PNYQ
A.otae 274 EGILSKFKEEYN.....PNFN
A.dermatitidis 272 EAILSKFKEEYN.....PNFN
T.marneffei 276 EQILSTFKEEYN.....PNFN
P.digitatum 273 EDILAKFKVEYN.....PNFN
P.expansum 273 EDILAKFKVEYN.....PNFN
B.spectabili 275 EEILSKFKVEYN.....PNFN
A.clavatus 326 EDILSTFKVEYN.....PNFN
A.oryzae 277 EEILSKFKVEYN.....PNFN
E.dermatitidis 285 EGLLSALKEEYN.....PNFN
G.lozoyensis 270 EGILAKFKEERN.....PNFN
M.brunnea 281 EKIMATFKEEYN.....PNFN
S.brasiliensis 276 ETILRNFKEQYN.....PNFN
T.reesei 272 EAILKQLKDGHN.....PDIK
M.album 272 ESILRKFKDEYN.....PNFN
F.oxysporum 271 EELLTKFKTDYN.....PNFN
G.moniliformis 271 EELLTKFKTDYN.....PNFN
C.higginsianum 273 EGILSRFKEEYN.....PNFN
N.crassa 273 EDILSKLKEEYN.....PNFN
N.tetrasperma 273 EDILSKLKEEYN.....PNFN

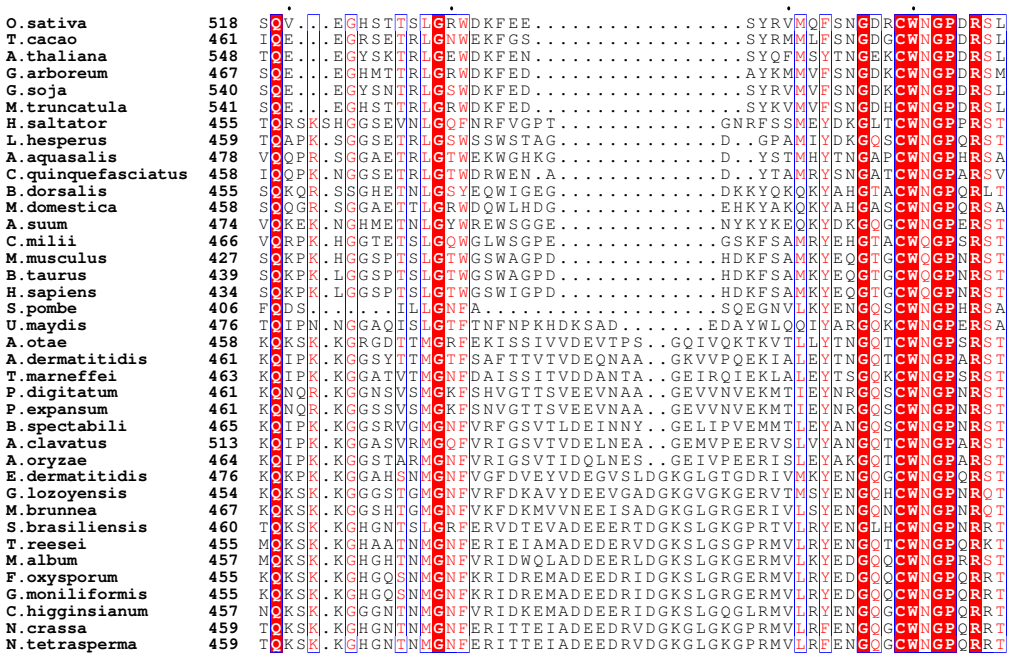

O.sativa 271 PES.QDDISI.....KAAPADES.....PPEETSAAPTKEQESTPADSEGLSREE
T.cacao 242 .....GLSWKS...PSNQNEKDEFQSTEGLSREE
A.thaliana 296 PAD.EHVVERKEESTSSEDSSSPTDESQND.GSAEKEESDEVKKVEDFVTEKKEELSKEE
G.arboreum 296 KEE.DHAVKNKHESASHRHNDDSTVSTEIDQDAGSKVSPDEDKKAESDTSETTEGLSREE
G.soja 300 KVE.EN.....AKEAEEEPIVKSETDIKVRNKESSDEIINKGNDASENTEGLSREE
M.truncatula 294 KVE.EN.....AKEAEREPTVESETDLTTGKTESSDEAIDTGKEASENTDGLSKEE
H.saltator 290 FGQ.....L.SAT
L.hesperus 276 DLK.....DVEVN
A.aquasalis 291 DER.....D.QVD
C.quinquefasciatus 274 DEH.....D.QVD
B.dorsalis 279 DER.....D.RID
M.domestica 276 DER.....E.RVD
A.suum 275 SGA.....E.ESD
C.milii 262 GGA.....S.QVD
M.musculus 258 SGD.....T.QTD
B.taurus 257 GGD.....A.QMD
H.sapiens 258 SGD.....T.QTD
S.pombe 247 DETMKEAILSFEDLKEQEIRRKV.....SSDD
U.maydis 260 DMAVKGAVVAYKEWRGIASAAAAAT....ATGEV.....KEEGENN
A.otae 290 DEGVKRAVRSWEDYIARKEEGDGD.....L.....FGDAD
A.dermatitidis 288 DEGVKRAVRAWEDYAARDDKG..S.....HG.....DETLNRD
T.marneffei 292 DEGVKRAVRSWEEYAARGTAGDA.....SA.....DSARDRD
P.digitatum 289 DEGVKRAVRSYEDYAAREQGT..E.....VA.....NGAVDRD
P.expansum 289 DEGVKRAVRSYEDYAAREQGT..D.....GA.....NGAVDRD
B.spectabili 291 DEGVKRAVRSWEEYAAKGTAADGI.....SG.....QAAKDRD
A.clavatus 342 DEGVKRAVRSWEEYAAREDVG..N.....LG.....NSARDRD
A.oryzae 293 DEGVKRAVRSWEEYAARGTSE..G.....LE.....NAARDRD
E.dermatitidis 301 DEGVKRAVRGWEDYAARETDDHW.....SEAEDRD
G.lozoyensis 286 DEGVKRAVQAWEDYAANRQGGE.....DDASERD
M.brunnea 297 DEGVKRAVHAWEDYAAQKLISRE.....EA.....ESAEDRD
S.brasiliensis 292 DEGVKSAVHAWEDYAAKTAGEQK.....TELSDVD
T.reesei 288 DEGVKAAVQAYEDYAAREASADQ.....LEFLESD
M.album 288 DEGVKSAVKSWEDYAAREADTVH.....EQVVDSD
F.oxysporum 287 DEGVKAAIRSFEDYSARRAESKE.....EVVSDED
G.moniliformis 287 DEGVKAAIRSFEDYSARRAESKE.....EVVSDED
C.higginsianum 289 DEGVKRAVKGWEDYAANLANEAL.....PDGLEAD
N.crassa 289 DEGVKAAVQGYENYAAGKSTEKK.....SEVVDAD
N.tetrasperma 289 DEGVKAAIQGYENYAAGKSTEKK.....SEVVDAD


O.sativa 315 LGRLVASRWTGEKVDEVSKDDKNEHEAEHDMPEHSEETHEDESDVPESAEDSYAGYHSEV
T.cacao 268 LGRLVASRWTGKKTEDQVEEIDTAKRDY.....VGNREKEESAHDEHSIN.....
A.thaliana 354 LGRLVASRWTGEKSDKPTEADDIPKADD.....QENHEHTPITAHEADEDDGFVSDG
G.arboreum 355 LGRLVASRWTGEETS.....
G.soja 350 LGRLVASRWTGENTDKPSAEPDTTLDNE.....DREDP.KGKNNEEYEGYASET
M.truncatula 344 LGRLVASRWTGEDVGKKSVEANTALDNE.....DQEDILHGTNNEENEGYASET
H.saltator 297 WDEFLNMAREKEHHEVKLAV..EKIKAIAGYP...ARE.LP.....DDEG.....
L.hesperus 284 EEVFHNSVWPHVKPLYMIEKGLFLPPEP...H...QDE.PLQNENLETPAEEN..LE...
A.aquasalis 298 FETFVTLCWPRVKPFLMMDSGLFKPPAP...A...GEE.TE.QAELQS..EDG.GYELEG
C.quinquefasciatus 281 LETFISLCWPKIKPFLMMDSGLFKPPSD...V...DEL.NQEEAELQNEEDDR.RYELDS
B.dorsalis 286 LDSFYDLSWPRIKPLKMLAEGIFKPPTV...E...DVE.HPEEDEGEEEGQTG.PTGS..
M.domestica 283 LDAFITLSWPRIKPMLMVGQGIFKPPQL...P...NEN.NDGDDAGKEKQGK...SEA..
A.suum 282 LEHFKTNVYETIKMLFPDPDEEEKEK....K....SEV.KESENLEALLSEERHKPESES
C.milii 269 VTSFQETVWAGIKDKYKFEQVSPETPAPPDVP...AEE.KTE.NRPESRREEWKKNAGDD
M.musculus 265 TTSFYDRVWAAIRDKYRSEVPPTDIP....VP...EET.EPKEEKPPVLP.....PTE..
B.taurus 264 AAFFYDRVWAAIRDKYRSEVLPTEYPPSPPAP...DVM.EPKEEQPPMPS.....PPT..
H.sapiens 265 ATSFYDRVWAAIRDKYRSEALPTDLP.APSAP...DLT.EPKEEQPPVPS.....SPT..
S.pombe 274 VHNYL.....E.....SC.N.....NHL.SMLTG.
U.maydis 297 VDQIA.....G.....EN.TKLNELLDEGDWPWAKLSTLL
A.otae 320 LFDAL.....K.....SE.H.....DEPFNNPEQWAEE.
A.dermatitidis 319 LDEIC.....K.....PD.S.....ENSGINWDHWENEQ
T.marneffei 324 LDEIA.....K.....AD.S.....DSAGIDWVHWENQN
P.digitatum 320 LDEIA.....K.....VD.D.....EDSGINWEHWESEE
P.expansum 320 LDEIA.....K.....AD.G.....EDSGVNWEHWENEE
B.spectabili 324 LDEIS.....K.....PD.S.....DSSGINWEQWENEE
A.clavatus 373 LDEIA.....K.....AD.S.....DESGVNWEQWEAEE
A.oryzae 324 LDEIV.....K.....PD.D.....DKSGVNWEQWENEE
E.dermatitidis 331 LAAIL.....Q.....ED.N.....EANGINWEEFENAP
G.lozoyensis 315 VEAVS.....K.....ED.S.....ESEGINWKEWETEE
M.brunnea 329 FAEIM.....K.....ED.S.....ESEGINWAEWETEE
S.brasiliensis 322 VHAVL.....K.....ED.S.....EENGVNWKAFETD.
T.reesei 318 IREIL.....K.....ED.D.....ETNGVNWKAFDEL.
M.album 318 VDEIL.....K.....ED.S.....ESSGINWAEFEGGE
F.oxysporum 317 VLSVL.....K.....ED.G.....ENGGVNWSEFEEG.
G.moniliformis 317 VLSVL.....K.....ED.G.....ENAGVNWSEFEEG.
C.higginsianum 319 IKEVL.....V.....ED.S.....ETSGINWKEFEED.
N.crassa 319 VLEVL.....K.....ED.G.....PDSGINWAEFQEE.
N.tetrasperma 319 VLEVL.....K.....ED.G.....PDSGINWAEFQEE.


O.sativa 375 EDDRHKYDDEDFSH.....E.....SDDEYVD.DHDEHVASYKSDDDQKGDDH
T.cacao 313 ..NALQFNENTQKH.....H.EKGR..NDAGDEESVADDVGSSSTHYNIRLEDHE.NS
A.thaliana 406 DEDTS....DDGKY.....SD.....HEPEDDSYEEEYRHDSSSSYKSDADDD...V
G.arboreum
G.soja 398 DDDNNKYDDDSHKY.....DDE.....DEVDDEYREDEHDDLSSSYKSDSDNE...P
M.truncatula 393 DDDTSKYDDDTGKY.....DDDTGKYDEDINDEEFQEDEHEDLSSSYKSDVESE...P
H.saltator 336 .....EEDEGDA.....SD.....REEEEAEAEAEAEAEAEAEA.....
L.hesperus 332 KEDDSHYEEEPDT.....L.....DKQDDDDEEGNLGEQEA..DGGQTGEGEVT
A.aquasalis 347 SEADLREGNVDDD...EGQ..L...D....EELDDEEEEDDVGEGEIERTEYDRDEQR..
C.quinquefasciatus 333 SEPAHHYNEEEEE.....G....TELEHEEEEDDEGLDEPEV..GEGQVQT..
B.dorsalis 336 DQLPIEE...GRE..QEGA.....YG..EEDLYEDEEGEADVG.....VGS..
M.domestica 331 DDMITEYNERGEEIHRPQT..YPEDG..EEEEEAYEDDEADVG.....EGT..
A.suum 333 VR.SEQLNELGEM..KPGEPQDDALSELEELEKLDAGETHKEDEQHF.....DGM.P
C.milii 324 REEDEEEDDDDDD..DEGDDGDDEGGGAEEEDDEDEEDSDDEGGRER.....YKD.P
M.musculus 310 EEEEEEEEPE.....EEEE.....EEEEEE.....EAP.P
B.taurus 313 EEEDEDEEDE.....ETEEDEDEEDEDSQGEQPK.....DAP.P
H.sapiens 313 EEEEEEEEEE.....EEEAEEEEEE.....EDSE.....EAP.P
S.pombe 291 .....PSEDITF.....
U.maydis 326 SDDPLDLMDR.....GL.....G....GALDDKRVYASETDGGLLFRI.....
A.otae 342 .....AEVGLVYKL.....
A.dermatitidis 342 .....DGESKISLLYKV.....
T.marneffei 347 .....E..SEIDLVYKL.....
P.digitatum 343 .....DGCNDSDLVYKL.....
P.expansum 343 .....DGCNDSDLVYKL.....
B.spectabili 347 .....ERETEIDLIYKL.....
A.clavatus 396 .....D.SSDTDAVYKL.....
A.oryzae 347 .....D.GCEANLVYQL.....
E.dermatitidis 354 .....EEPESDVAALYQF.....
G.lozoyensis 338 .....EESDVEAIYKF.....
M.brunnea 352 .....ESDVEALYKF.....
S.brasiliensis 344 .....DARDTDVLYSF.....
T.reesei 340 .....HDDTDILYNF.....
M.album 341 .....EGSDTDILYKF.....
F.oxysporum 339 .....EGSDTDILYNF.....
G.moniliformis 339 .....EGSDTDILYNF.....
C.higginsianum 341 .....EVSDTDILYNF.....
N.crassa 341 .....ESSDADIVYNL.....
N.tetrasperma 341 .....ESSDADIVYNL.....


O.sativa 417 SDFTASGQASWLDKIQQTVQNVLRTFNFFK.TP.....VDLSEASRVRKEYDDA
T.cacao 360 SDIATPSHQSWLVKIQETAQNILQSVNFYP.TP.....VPNLDANQVRNEYNNY
A.thaliana 446 DFSETTSNPTWLEKIQKTVKNILLAVNLFQTTP.....VDKSEADRVRKEYDES
G.arboreum 370 .P...SSDPSWFEKIQKTVKNVLDAVNIFK.TP.....VNISDAAHVRKEYDES
G.soja 442 DL...SDNPSWLEKIQRTVRNIFQVVNLFQ.AP.....VNQTDAARVRKEYDES
M.truncatula 443 DL...SDDPSWLEKIQKSVWNIIQVVNIFQ.TP.....VNQSDAARIRKEYDES
H.saltator 365 ......EAETGI......EKEAESTIQYDEETQ.....VLIDEANNARERLQEV
L.hesperus 374 PE...HK.....ESTEPVYDAFTT.....RLIEEANEARTQFNSA
A.aquasalis 393 SA...PA.....PPPPVQYDPETQ.....ELINKANEARNQYSEA
C.quinquefasciatus 373 EQ...PE.....EAPKVEYDPDTA.....ELIRKANEARNHHSEA
B.dorsalis 370 VA...DA.....THPEPDYDPETK.....HLIDLANEARNAYSEA
M.domestica 373 VQ...DT.....NPPTLEYDAETQ.....RIINMANDARNSFAEV
A.suum 381 PP...PTLPPPT......LRLEDAMPDYDDETK.....KLMAEADEARKALTEV
C.milii 373 ED...RYQPPKK......EVVDEQMPPYEEQTQ.....ELIDAADRTRSAFSEA
M.musculus 334 PL...QPPQPPS......PTEDEKMPPYDEETQ.....AIIDAAQEARSKFEEV
B.taurus 346 PA...PAPQTAS......PTEEDRMPPYDEQTQ.....AFINAAQEARNKFEEA
H.sapiens 341 PL...SPPQPAS......PAEEDKMPPYDEQTQ.....AFIDAAQEARNKFEEA
S.pombe 298 S.SLIKDIKKILNSLVWNIKLSLINFGILSPSASSTPLT.....DSESYRRFEAA
U.maydis 360 HEYLPDGIVPYFEAMVDTLLDVLMKANVITDVKRMRPKSSVGSESEPETVSVARRAHTDA
A.otae 351 ASFLPAGIVNSIEDGIASFRSALVSNGLLADNSV....DSD..SNEPREVKDARDRVFDA
A.dermatitidis 354 AAYFPDSLVNYLEDKALQLRSILVSNGIVADTSS....DSD..ATEPRAVTEARNAVSAE
T.marneffei 357 LAYFPPVVVEFLEDKYNSIKKFLVENNIIPGADS....E.T..ETESKAVTEAKNALQSE
P.digitatum 355 AAYLPPSLVSFIEDRIVFAKGFLEEKGILPKADE....N.S..SSESKAVTQARNEALKAA
P.expansum 355 AAYLPPSLVSFIEGKVVSVKGYLEEKGILPKADE....N.S..SSESKAVTRARDALKAA
B.spectabili 359 AAYLPPSLVDYLEDKLSSFKSFLVSNGVLTEAEV....T.S..SSESKAVTEARDALNAA
A.clavatus 407 AAFLPPSFVSLIEDKLISIRGFLEDNGILPKKDE....V.T..TTESKAVSEARDALEAA
A.oryzae 358 AAYLPPSLVNFIEDKALAIRGFLEQNGILPTKDE....G.S..ASESKAVTEARDAVDAV
E.dermatitidis 367 SAYLPDGFKLWLDEKIASLRQFLIESGVLAGRSN....DPAAAAAESKAVQEAKKALSEA
G.lozoyensis 349 EEYLPEGVRTWVHQKIIDMRIMLVENGILADNAN.....SGS..ESKAVSDARASYQSV
M.brunnea 362 EEYLPEPIREWVHQRIIDVRIMLIENGILADHAN.....SGS..ESKTVSDARSALQAV
S.brasiliensis 355 EAYLPPSVRLYLHDKLAALRVWLVESGLLADNSD.....ASN..ESQALKNARGALQSA
T.reesei 350 DAYIPPFIRGVIHDKIRVIRKWLMDNGMIADEAK.....TGS..ESSAARAARAVEAA
M.album 352 EAYLPGFLHGFVHDTATAVRLWLIQNGVLADNSP.....SGA..ESQVVKAARAAEAA
F.oxysporum 350 EAYLPPPVRAFIHNSLDNLRVWAITNGILADNSS.....PGK..ESTLVRAARAVDAA
G.moniliformis 350 EAYLPPPVRAFIHNSLDNLRVWAITNGILADNSS.....PGK..ESTLVRAARAVDAA
C.higginsianum 352 EAYLPDAVREFVHGKLLSLRVWMIENGLIADNNQ.....AGA..ESRLVKAARAVDAA
N.crassa 352 EAYLPPSVRPFIHSKISSLKVWLIENGVLADNPT.....TGGPTESKLVKAARDALEAA
N.tetrasperma 352 EAYLPPSVRSFIHSKISSLKVWLIENGVLADNPT.....TGGPTESKLVKAARDALEAA


O.sativa 465 SSKLSKIQSRISTLTDK.....LKHDFGKEKEFYYFYDQCFESKEGKYVYKVCPFKKA
T.cacao 408 TTRLHDIESRISSLTEK.....VKYDFGIDNEFYLFYDRCFESKQDKYVYKVCPFKHA
A.thaliana 495 SSKLNKIQSRISSLEKK.....LKQDFGPEKEFYSFHGRCFESKQGKYTYKVCAYKEA
G.arboreum 414 SAKLSKIQSRISSLTQK.....LKHDFGLEKEFYAFYDRCFEIKQNKYVYKVCPYKQA
G.soja 487 SAKLSKIQSRISSLKQK.....LKHDFGPAKEFYSFYDHCFEGKENKYTYKVCPYKQA
M.truncatula 488 SAKLSKIQSRISSLTQK.....QKLDFGPAKEFYSFYDRCFESKQNKYIYKVCPYKQA
H.saltator 402 EKAVSELQAEISQLEVK.....LRHNYGPDDEFASLYGECFEYTDMEYIYKLCLYDRA
L.hesperus 406 QEALQNIEREMQQIKEK.....LDMDYGPEDEYSPLDGQCFTFTDREYTYKLCPFNQV
A.aquasalis 425 ERHVREMDQEMRNIKEL.....LDKDYGREEEFAPLNGECFNFEDREYVYKLCPFDKA
C.quinquefasciatus 405 DRHVREIDQEMRNIEDA.....LNKDYGRDEEFAPLNGECINYEDREYVYKLCPFDKA
B.dorsalis 402 EQKVREIENEIRDIQDQ.....TSKDYGLNEEYAALDGECFTFEDREYLYTFCPFERA
M.domestica 405 ERTIREIDQEIRDIDDQ.....NSKDYGPHDEYAILEGECYKFEDREYVYTLCPFERA
A.suum 421 SNRIVELDSSIRDAESY.....LNGDFGVDSAWAPLKGKWLELDDSQYTYKLCLFERA
C.milii 413 EKSLKDLEDSIRSIEKE.....LTIDFGPYGEFAYLFSQCYEMNTNEYVYRLCPFNKV
M.musculus 374 ERSLKEMEESIRSLEQE.....ISFDFGPSGEFAYLYSQCYELTTNEYVYRLCPFKLV
B.taurus 386 ERSLKDMEESIRNLEQE.....ISFDFGPNGEFAYLYSQCYELTTNEYVYRLCPFKLV
H.sapiens 381 ERSLKDMEESIRNLEQE.....ISFDFGPNGEFAYLYSQCYELTTNEYVYRLCPFKLV
S.pombe 347 QRDLDAAEENEKSLEKEHTKLMHELEYHH.GWDLYRAIKGMETKREIGGYTYKVVFYENV
U.maydis 420 AAHLSRTTHELSSLKQK....LSEFSTRYGRSAEFKALENKCFSKDMGEYTYEYCFFGRA
A.otae 405 QSSLNSKKS...EITQL....KGDLEQDFGVDSVFRALKGACVSRDSGEYTYELCWMEQT
A.dermatitidis 408 ESSLNNIRS...QLDDH....KLDLDKDYGRDSVFRSMKGSCISKDSGEYTYELCWLEKT
T.marneffei 410 KYSLENTLR...SIKDH....QSDLEKYYGPDGIFRPLKDVCIQKDSGEYTYEHCFLAQT
P.digitatum 408 QESVTSLKN...KLRDQ....RADLEQDYGPSSIFRALKGVCITQDAGEYTYEHCFLDST
P.expansum 408 QDSVSSLKN...KLRDQ....RADLEQDYGPSSIFRALRGVCITQDAGEYTYEHCFLEST
B.spectabili 412 KSSLTGDQN...RLRDH....RQDLEKDYGPEDIFRPLKGRCISRDAGEYTYEHCFLEGT
A.clavatus 460 KTELAQSYT...ELKNH....QADLETDYGKAGVFRALKGVCISKDSGEYTYEHCFWDQT
A.oryzae 411 RKSIEDLKN...QVKDH....KEDLDTDYGVGSIFRALKGVCIQKDAGEYTYEHCFLDQT
E.dermatitidis 423 ERDVLNTEN...DLRRH....REDLEKDYGPDGIFRALKDTCISKDSGEYEYELCFMGQT
G.lozoyensis 401 SDDLNAKRV...TLGES....QTDLTKDYGPNDIFRALKNVCISKDSGEYEYELCWMENT
M.brunnea 414 SDEVSTKSS...KVGEL....RRDLEKDYGADDIFRALKDKCISKDSGEYEYELCWMGNT
S.brasiliensis 407 ENDLSSRQN...SLRED....EADLTKDYGADEIFRALKDKCTTTEVGEYDYEVCWLGQA
T.reesei 402 DQELGNKER...ELQQE....KTDLDKDYGPSGIFRSLKGKCAEIDAGEYTYELCWLDKT
M.album 404 ERELQDKIR...DRDSE....TEDLEKDYGPSDILRGIKGKCISIDAGEYEYELCYLDKT
F.oxysporum 402 KRDLSDKKS...TLDVE....QADLDFDYGPDDIFRALKDKCVTLEAGEYTYEQCWLGST
G.moniliformis 402 NRDLSDKKS...TLSVE....QADLDFDYGPDDIFRALKDKCVTLEAGEYTYEQCWLGST
C.higginsianum 404 NSDLLDKKR...QLEAE....EKDLEKDYGIDDIFRVLKNKCVSTELGEYEYELCWMDKT
N.crassa 406 RAEHTTKTS...QLADE....ERDLAKDYGPDDIFRALKGQCVSADVGEYEYELCWFDRT
N.tetrasperma 406 RAEHTTKTS...QLADE....ERDLAKDYGPDDIFRALKGQCVSADVGEYEYELCWFDRT


O.sativa 518 SQV..E.GHSTTSLGRWDKFEE.....SYRVMQFSNGDRCWNGPDRSL
T.cacao 461 IQE..E.GRSETRLGNWEKFGS.....SYRMMLFSNGDGCWNGPDRSL
A.thaliana 548 TQE..E.GYSKTRLGEWDKFEN.....SYQFMSYTNGEKCWNGPDRSL
G.arboreum 467 SQE..E.GHMTTRLGRWDKFED.....AYKMMVFSNGDKCWNGPDRSM
G.soja 540 SQE..E.GYSNTRLGSWDKFED.....SYRVMVFSNGDKCWNGPDRSL
M.truncatula 541 SQE..E.GHSTTRLGRWDKFED.....SYKVMVFSNGDHCWNGPDRSL
H.saltator 455 TQRSKSHGGSEVNLGQFNRFVGPT.....GNRFSSMEYDKGLTCWNGPPRST
L.hesperus 459 TQAPK.SGGSETRLGSWSSWSTAG.....D..GPAMIYDKGQSCWNGPQRST
A.aquasalis 478 VQQPR.SGGAETRLGTWEKWGHKG.....D..YSTMHYTNGAPCWNGPHRSA
C.quinquefasciatus 458 IQQPK.NGGSETRLGTWDRWEN.A.....D..YTAMRYSNGATCWNGPARSV
B.dorsalis 455 SQKQR.SSGHETNLGSYEQWIGEG.....DKKYQKQKYAHGTACWNGPQRLT
M.domestica 458 SQQGR.SGGAETTLGRWDQWLHDG.....EHKYAKQKYAHGASCWNGPQRSA
A.suum 474 VQKEK.NGHMETNLGYWREWSGGE.....NYKYKEQKYDKGQGCWNGPERST
C.milii 466 VQRPK.HGGTETSLGQWGLWSGPE.....GSKFSAMRYEHGTACWQGPSRST
M.musculus 427 SQKPK.HGGSPTSLGTWGSWAGPD.....HDKFSAMKYEQGTGCWQGPNRST
B.taurus 439 SQKPK.LGGSPTSLGTWGSWAGPD.....HDKFSAMKYEQGTGCWQGPNRST
H.sapiens 434 SQKPK.LGGSPTSLGTWGSWIGPD.....HDKFSAMKYEQGTGCWQGPNRST
S.pombe 406 FQDS.....ILLGNFA.....SQEGNVLKYENGQSCWNGPHRSA
U.maydis 476 TQIPN.NGGAQISLGTFTNFNPKHDKSAD.....EDAYWLQQIYARGQKCWNGPERSA
A.otae 458 KQKSK.KGRGDTTMGRFEKISSIVVDEVTPS..GQIVQKTKVTLLYTNGQTCWNGPSRST
A.dermatitidis 461 KQIPK.KGGSYTTMGTFSAFTTVTVDEQNAA..GKVVPQEKIALEYTNGQTCWNGPARST
T.marneffei 463 KQIPK.KGGATVTMGNFDAISSITVDDANTA..GEIRQIEKLALEYTSGQKCWNGPSRST
P.digitatum 461 KQNQR.KGGNSVSMGKFSHVGTTSVEEVNAA..GEVVNVEKMTIEYNRGQSCWNGPNRST
P.expansum 461 KQNQR.KGGSSVSMGKFSNVGTTSVEEVNAA..GEVVNVEKMTIEYNRGQSCWNGPNRST
B.spectabili 465 KQIPK.KGGSRVGMGNFVRFGSVTLDEINNY..GELIPVEMMTLEYANGQSCWNGPNRST
A.clavatus 513 KQIPK.KGGASVRMGQFVRIGSVTVDELNEA..GEMVPEERVSLVYANGQTCWNGPARST
A.oryzae 464 KQIPK.KGGSTARMGNFVRIGSVTIDQLNES..GEIVPEERISLEYAKGQTCWNGPARST
E.dermatitidis 476 KQKPK.KGGAHSNMGNFVGFDVEYVDEGVSLDGKGLGTGDRIVMKYENGQGCWNGPQRST
G.lozoyensis 454 KQKSK.KGGGSTGMGNFVRFDKAVYDEEVGADGKGVGKGERVTMSYENGQHCWNGPNRQT
M.brunnea 467 KQKSK.KGGSHTGMGNFVKFDKMVVNEEISADGKGLGRGERIVLSYENGQNCWNGPNRQT
S.brasiliensis 460 TQKSK.KGHGNTSLGRFERVDTEVADEEERTDGKSLGKGPRTVLRYENGLHCWNGPNRRT
T.reesei 455 MQKSK.KGHAATNMGNFERIEIAMADEDERVDGKSLGSGPRMVLRYENGQTCWNGPQRKT
M.album 457 MQKSK.KGHGHTNMGNFVRIDWQLADDEERLDGKSLGKGERMVLKYEDGQQCWNGPRRST
F.oxysporum 455 KQKSK.KGHGQSNMGNFKRIDREMADEEDRIDGKSLGRGERMVLRYEDGQQCWNGPQRRT
G.moniliformis 455 KQKSK.KGHGQSNMGNFKRIDREMADEEDRIDGKSLGRGERMVLRYEDGQQCWNGPQRRT
C.higginsianum 457 NQKSK.KGGGNTNMGNFVRIDKEMADDEERIDGKSLGQGLRMVLRYENGQGCWNGPQRRT
N.crassa 459 TQKSK.KGHGNTNMGNFERITTEIADEEDRVDGKGLGKGPRMVLRFENGQGCWNGPQRRT
N.tetrasperma 459 TQKSK.KGHGNTNMGNFERITTEIADEEDRVDGKGLGKGPRMVLRFENGQGCWNGPQRRT

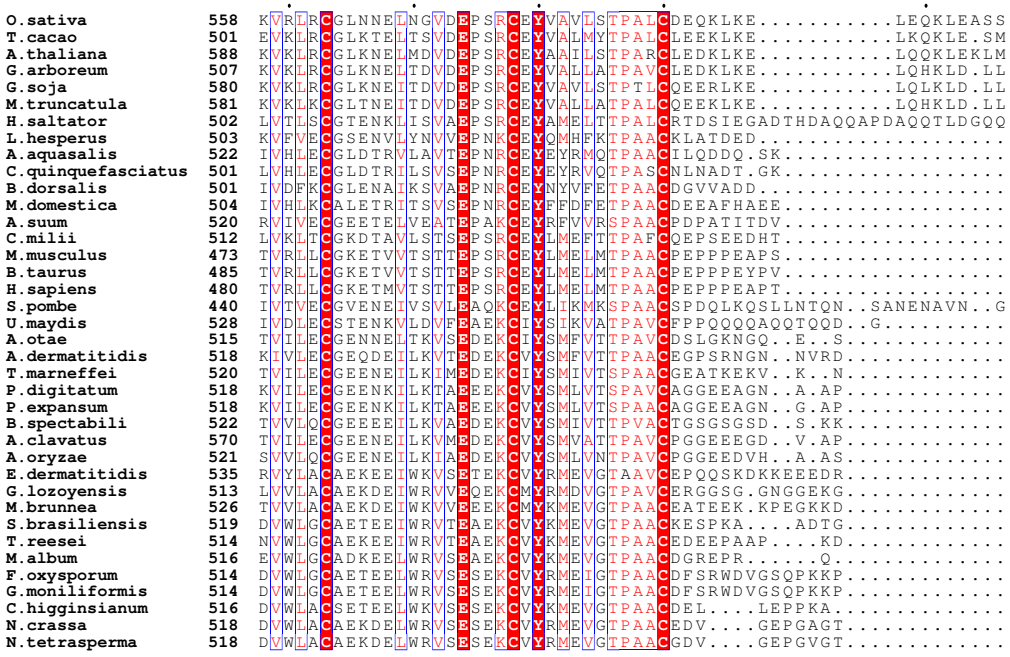

O.sativa 558 KVRLRCGLNNELNGVDEPSRCEYVAVLSTPALCDEQKLKE.....LEQKLEASS
T.cacao 501 EVKLRCGLKTELTSVDEPSRCEYVALMYTPALCLEEKLKE.....LKQKLE.SM
A.thaliana 588 KVKLRCGLKNELMDVDEPSRCEYAAILSTPARCLEDKLKE.....LQQKLEKLM
G.arboreum 507 KVKLRCGLKNELTDVDEPSRCEYVALLATPAVCLEDKLKE.....LQHKLD.LL
G.soja 580 KVKLRCGLKNEITDVDEPSRCEYVAVLSTPTLCQEERLKE.....LQLKLD.LL
M.truncatula 581 KVKLKCGLTNEITDVDEPSRCEYVALLATPALCQEEKLKE.....LQHKLD.LL
H.saltator 502 LVTLSCGTENKLISVAEPSRCEYAMELTTPALCRTDSIEGADTHDAQQAPDAQQTLDGQQ
L.hesperus 503 KVFVECGSENVLYNVVEPNKCEYQMHFKTPAACKLATDED.....
A.aquasalis 522 IVHLECGLDTRVLAVTEPNRCEYEYRMQTPAACILQDDQ.SK.....
C.quinquefasciatus 501 LVHLECGLDTRILSVSEPNRCEYEYRVQTPASCNLNADT.GK.....
B.dorsalis 501 IVDFKCGLENAIKSVAEPNRCEYNYVFETPAACDGVVADD.....
M.domestica 504 IVHLKCALETRITSVSEPNRCEYFFDFETPAACDEEAFHAEE.....
A.suum 520 RVIVECGEETELVEATEPAKCEYRFVVRSPAACPDPATITDV.....
C.milii 512 LVKLTCGKDTAVLSTSEPSRCEYLMEFTTPAFCQEPSEEDHT.....
M.musculus 473 TVRLLCGKETVVTSTTEPSRCEYLMELMTPAACPEPPPEAPS.....
B.taurus 485 TVRLLCGKETVVTSTTEPSRCEYLMELMTPAACPEPPPEYPV.....
H.sapiens 480 TVRLLCGKETMVTSTTEPSRCEYLMELMTPAACPEPPPEAPT.....
S.pombe 440 IVTVECGVENEIVSVLEAQKCEYLIKMKSPAACSPDQLKQSLLNTQN..SANENAVN..G
U.maydis 528 IVDLECSTENKVLDVFEAEKCIYSIKVATPAVCFPPQQQQAQQTQQD..G.....
A.otae 515 TVILECGENNELTKVSEDEKCIYSMFVTTPAVCDSLGKNGQ..E..S.....
A.dermatitidis 518 KIVLECGEQDEILKVTEDEKCVYSMFVTTPAACEGPSRNGN..NVRD.....
T.marneffei 520 TVILECGEENEILKIMEDEKCIYSMIVTSPAACGEATKEKV..K..N.....
P.digitatum 518 KVILECGEENKILKTAEEEKCVYSMLVTSPAVCAGGEEAGN..A.AP.....
P.expansum 518 KVILECGEENKILKTAEEEKCVYSMLVTSPAACAGGEEAGN..G.AP.....
B.spectabili 522 TVVLQCGEEEEILKVAEDEKCVYSMIVTTPVACTGSGSGSD..S.KK.....
A.clavatus 570 TVILECGEENEILKVMEDEKCVYSMVATTPAVCPGGEEEGD..V.AP.....
A.oryzae 521 SVVLQCGEENEILKIAEDEKCVYSMLVNTPAVCPGGEEDVH..A.AS.....
E.dermatitidis 535 RVYYLACAEKEEIWKVSETEKCVYRMEVGTAAVCEPQQSKDKKEEEDR.....
G.lozoyensis 513 LVVLACAEKDEIWRVVEQEKCMYRMDVGTPAVCERGGSG.GNGGEKG.....
M.brunnea 526 TVVLACAEKDEIWKVVEEEKCMYKMEVGTPAACEATEEK.KPEGKKD.....
S.brasiliensis 519 DVWLGCAETEEIWRVTEAEKCVYKMEVGTPAACKESPKA....ADTG.....
T.reesei 514 NVWLGCAEKEEIWRVTEAEKCVYKMEVGTPAACEDEEPAAP....KD.....
M.album 516 EVWLGCADKEELWRVSEAEKCVYKMEVGTPAACDGREPR.....Q.....
F.oxysporum 514 DVWLGCAETEELWRVSESEKCVYRMEIGTPAACDFSRWDVGSQPKKP.....
G.moniliformis 514 DVWLGCAETEELWRVSESEKCVYRMEIGTPAACDFSRWDVGSQPKKP.....
C.higginsianum 516 DVWLACSETEELWKVSESEKCVYKMEVGTPAACDEL....LEPPKA.....
N.crassa 518 DVWLACAEKDELWRVSESEKCVYRMEVGTPAACEDV....GEPGAGT.....
N.tetrasperma 518 DVWLACAEKDELWRVSESEKCVYRMEVGTPAACGDV....GEPGVGT.....

Figure C.3: Multiple sequence alignment of PRKCSH genes. *Appendix C*

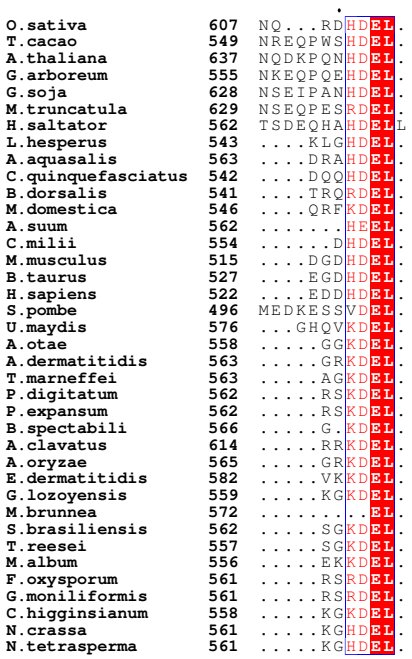


Table C.1: Key to species abbreviations on multiple sequence alignments.

Abbreviation	Species	UniProt Accession No.		
		GluI (MOGS)	GluIIα (GANAB)	GluIIβ (PRKCSH)
A.aquasalis	*Anopheles aquasalis*			T1DJN6
A.carolinensis	*Anolis carolinensis*	H9G3X3		
A.clavatus	*Aspergillus clavatus*			A1CD98
A.dermatitidis	*Ajellomyces dermatitidis*			F2T5U5
A.glabripennis	*Anoplophora glabripennis*		V5GYU2	
A.mellifera	*Apis mellifera*		A0A088A7S9	
A.oryzae	*Aspergillus oryzae*		Q2UA37	I8TZE5
A.otae	*Arthroderma otae*			C5FC10
A.suum	*Ascaris suum*			U1MAR3
A.thaliana	*Arabidopsis thaliana*	F4HTM3	Q9FN05	Q9FM96
B.cucurbitae	*Bactrocera cucurbitae*		A0A0A1WIR	
B.dorsalis	*Bactrocera dorsalis*			A0A034WM66
B.spectabili	*Byssochlamys spectabili*			V5FB10
B.taurus	*Bos taurus*		A6QNJ8	Q28034
C.capitata	*Ceratitis capitata*		W8C608	
C.elegans	*Caenorhabditis elegans*	Q19426		
C.familiaris	*Canis familiaris*	E2R8C9	E2R729	
C.higginsianum	*Colletotrichum higginsianum*			H1UX51
C.jacchus	*Callithrix jacchus*	F7E1L8	F7I173	
C.milii	*Callorhynchus milii*			V9KPP4
C.orthopsilosis	*Candida orthopsilosis*	H8WZK5		
C. porcellus	*Cavia porcellus*	H0WBE6		
C.quinquefasciatus	*Culex quinquefasciatus*			B0WS45
D.discoideum	*Dictyostelium discoideum*		Q94502	
D.melanogaster	*Drosophila melanogaster*	Q9VZ04		
D.rerio	*Danio rerio*	A1L1S8	B3DIZ3	
E.caballus	*Equus caballus*		F7C4H8	
E.dermatitidis	*Exophiala dermatitidis*			H6C5T8
F.catus	*Felis catus*	M3XE28	M3WAD1	
F.oxysporum	*Fusarium oxysporum*			W9LWD1
G.arboreum	*Gossypium arboreum*			A0A0B0MLL6
G.lozoyensis	*Glarea lozoyensis*			H0EE25
G.moniliformis	*Gibberella moniliformis*			W7LG94
G.soja	*Glycine soja*		A0A0B2R9G7	A0A0B2QL92
H.saltator	*Harpegnathos saltator*			E2BWI6
H.sapiens	*Homo sapiens*	Q13724	Q14697	P14314
I.punctatus	*Ictalurus punctatus*	W5UIZ1	W5UAR6	
K.pastoris	*Komagataella pastoris*		F2QS03	
L.hesperus	*Lygus hesperus*		A0A0A9X6S5	A0A0A9YLH4
M.album	*Metarhizium album*			A0A0B2X7T2
M.brunnea	*Marssonina brunnea*			K1X5K7
M.domestica	*Musca domestica*			T1PEB3
M.domestica	*Monodelphis domestica*		F7EBB3	
M.fascicularis	*Macaca fascicularis*		Q4R4N7	
M.mulatta	*Macaca mulatta*	I0FIG4		
M.musculus	*Mus musculus*	Q80UM7	Q8BHN3	O08795
M.truncatula	*Medicago truncatula*			G7JHL0
N.crassa	*Neurospora crassa*			Q7S6V9
N.fumigata	*Neosartorya fumigata*		Q4WEM1	
N.tetrasperma	*Neurospora tetrasperma*			G4UZP1
O.aries	*Ovis aries*	W5PT15	W5Q305	
O.sativa	*Oryza sativa*			A2WNF5
P.abelii	*Pongo abelii*		H2ND53	
P.digitatum	*Penicillium digitatum*			K9FX08
P.expansum	*Penicillium expansum*			A0A0A2J619
P.troglodytes	*Pan troglodytes*	H2QI63		
R.norvegicus	*Rattus norvegicus*	O88941		
S.brasiliensis	*Sporothrix brasiliensis*			A0A0C2J676
S.cerevisiae	*Saccharomyces cerevisiae*	P53008	P38138	
S.pombe	*Schizosaccharomyces pombe*	O14255	Q9US55	Q9USH8
S.scrofa	*Sus scrofa*		P79403	
T.cacao	*Theobroma cacao*			A0A061F039
T.marneffei	*Talaromyces marneffei*			A0A093VHJ8
T.nigroviridis	*Tetraodon nigroviridis*		H3DHB5	
T.reesei	*Trichoderma reesei*			A0A024SB06
U.maydis	*Ustilago maydis*			P0CT24
X.laevis	*Xenopus laevis*	B1WBB6	A1L3J7	

Bibliography

[1] Harvey, D. J., Merry, A. H., Royle, L., P Campbell, M., Dwek, R. A., and Rudd, P. M. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics*, 9(15): 3796–3801, August 2009.

[2] Imperiali, B. and Rickert, K. W. Conformational implications of asparagine-linked glycosylation. *Proceedings of the National Academy of Sciences*, 92(1):97–101, January 1995.

[3] Shental-Bechor, D. and Levy, Y. Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proceedings of the National Academy of Sciences of the United States of America*, 105(24):8256–8261, June 2008.

[4] Skropeta, D. The effect of individual N-glycans on enzyme activity. *Bioorganic & medicinal chemistry*, 17(7):2645–2653, April 2009.

[5] Hammond, C., Braakman, I., and Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proceedings of the National Academy of Sciences of the United States of America*, 91(3):913–917, February 1994.

[6] Parodi, A. J., Behrens, N. H., Leloir, L. F., and Carminatti, H. The role of polyprenol-bound saccharides as intermediates in glycoprotein synthesis in liver. *Proceedings of the National Academy of Sciences of the United States of America*, 69(11):3268–3272, November 1972.

[7] Li, E., Tabas, I., and Kornfeld, S. The synthesis of complex-type oligosaccharides. I. Structure of the lipid-linked oligosaccharide precursor of the complex-type oligosaccharides of the vesicular stomatitis virus G protein. *The Journal of biological chemistry*, 253(21):7762–7770, November 1978.

[8] Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., and Aebi, M.

N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science*, 298(5599):1790–1793, November 2002.

[9] Young, N. M., Brisson, J.-R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St Michael, F., Aberg, E., and Szymanski, C. M. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *The Journal of biological chemistry*, 277(45):42530–42539, November 2002.

[10] Helenius, J., Ng, D. T. W., Marolda, C. L., Walter, P., Valvano, M. A., and Aebi, M. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature*, 415(6870):447–450, January 2002.

[11] Frank, C. G., Sanyal, S., Rush, J. S., Waechter, C. J., and Menon, A. K. Does Rft1 flip an N-glycan lipid precursor? *Nature*, 454(7204):E3–4– discussion E4–5, July 2008.

[12] Rush, J. S., Gao, N., Lehrman, M. A., Matveev, S., and Waechter, C. J. Suppression of Rft1 Expression Does Not Impair the Transbilayer Movement of Man5GlcNAc2-P-P-Dolichol in Sealed Microsomes from Yeast. *The Journal of biological chemistry*, 284(30):19835–19842, July 2009.

[13] Breitling, J. and Aebi, M. N-Linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, 5(8): a013359–a013359, August 2013.

[14] Haeuptle, M. A. and Hennet, T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Human Mutation*, 30(12):1628–1641, December 2009.

[15] Parodi, A. J. and Cazzulo, J. J. Protein glycosylation in Trypanosoma cruzi. II. Partial characterization of protein-bound oligosaccharides labeled "in vivo". *The Journal of biological chemistry*, 257(13):7641–7645, July 1982.

[16] Marshall, R. D. Glycoproteins. *Annual review of biochemistry*, 41:673–702, 1972.

[17] Titani, K., Kumar, S., Takio, K., Ericsson, L. H., Wade, R. D., Ashida, K., Walsh, K. A., Chopek, M. W., Sadler, J. E., and Fujikawa, K. Amino acid sequence of human von Willebrand factor. *Biochemistry*, 25(11):3171–3184, June 1986.

[18] Miletich, J. P. and Broze, G. J. Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites. *The Journal of biological chemistry*, 265(19): 11397–11404, July 1990.

[19] Zielinska, D. F., Gnad, F., Wiśniewski, J. R., and Mann, M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell*, 141(5):897–907, May 2010.

[20] Matsumoto, S., Shimada, A., Nyirenda, J., Igura, M., Kawano, Y., and Kohda, D. Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(44): 17868–17873, October 2013.

[21] Pfeffer, S., Dudek, J., Gogala, M., Schorr, S., Linxweiler, J., Lang, S., Becker, T., Beckmann, R., Zimmermann, R., and Förster, F. Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nature communications*, 5:3072, 2014.

[22] Hebert, D. N., Foellmer, B., and Helenius, A. Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, 81:425–433, 1995.

[23] Nauseef, W. M., McCormick, S. J., and Clark, R. A. Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase. *The Journal of biological chemistry*, 270(9):4741–4747, March 1995.

[24] Tannous, A., Pisoni, G. B., Hebert, D. N., and Molinari, M. N-linked sugar-regulated protein folding and quality control in the ER. *Seminars in Cell & Developmental Biology*, 41:79–89, December 2014.

[25] Gonzalez, D. S., Karaveg, K., and Vandersall-Nairn, A. S. Identification, expression, and characterization of a cDNA encoding human endoplasmic reticulum mannosidase I, the enzyme that catalyzes the first mannose trimming step in mammalian Asn-linked oligosaccharide biosynthesis . *Journal of Biological Chemistry*, 274(30):21375–21386, 1999.

[26] Hosokawa, N., Wada, I., Hasegawa, K., Yorihuzi, T., Tremblay, L. O., Herscovics, A., and Nagata, K. A novel ER α-mannosidase-like protein accelerates ER-associated degradation. *EMBO Reports*, 2(5):415–422, May 2001.

[27] Vembar, S. S. and Brodsky, J. L. One step at a time: endoplasmic reticulum-associated degradation. *Nature reviews. Molecular cell biology*, 9 (12):944–957, December 2008.

[28] Stanley, P. Golgi Glycosylation. *Cold Spring Harbor Perspectives in Biology*, 3 (4):a005199–a005199, April 2011.

[29] Danilczyk, U. G. and Williams, D. B. The lectin chaperone calnexin utilizes polypeptide-based interactions to associate with many of its substrates in vivo. *The Journal of biological chemistry*, 276(27):25532–25540, July 2001.

[30] Frickel, E.-M., Riek, R., Jelesarov, I., Helenius, A., Wüthrich, K., and Ellgaard, L. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proceedings of the National Academy of Sciences*, 99(4):1954–1959, February 2002.

[31] Kozlov, G., Bastos-Aristizabal, S., Määttänen, P., Rosenauer, A., Zheng, F., Killikelly, A., Trempe, J.-F., Thomas, D. Y., and Gehring, K. Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. *Journal of Biological Chemistry*, 285(46):35551–35557, November 2010.

[32] Kozlov, G., Pocanschi, C. L., Rosenauer, A., Bastos-Aristizabal, S., Gorelik, A., Williams, D. B., and Gehring, K. Structural Basis of Carbohydrate Recognition by Calreticulin. *Journal of Biological Chemistry*, 285(49): 38612–38620, November 2010.

[33] Parodi, A. J., Mendelzon, D. H., and Lederkremer, G. Z. Transient glucosylation of protein-bound Man9GlcNAc2, Man8GlcNAc2, and Man7GlcNAc2 in calf thyroid cells. A possible recognition signal in the processing of glycoproteins. *The Journal of biological chemistry*, 258(13): 8260–8265, July 1983.

[34] Trombetta, S. E., Bosch, M., and Parodi, A. J. Glucosylation of glycoproteins by mammalian, plant, fungal, and trypanosomatid protozoa microsomal membranes. *Biochemistry*, 28(20):8108–8116, October 1989.

[35] Sousa, M. and Parodi, A. J. The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *The EMBO journal*, 14(17):4196–4203, September 1995.

[36] Takeda, Y., Seko, A., Hachisu, M., Daikoku, S., Izumi, M., Koizumi, A., Fujikawa, K., Kajihara, Y., and Ito, Y. Both isoforms of human UDP-glucose:glycoprotein glucosyltransferase are enzymatically active. *Glycobiology*, 24(4):344–350, April 2014.

[37] 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. 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(8):3404–3414, August 2008.

[38] Galli, C., Bernasconi, R., Soldà, T., Calanca, V., and Molinari, M. Malectin Participates in a Backup Glycoprotein Quality Control Pathway in the Mammalian ER. *PloS one*, 6(1):e16304, January 2011.

[39] Qin, S.-Y., Hu, D., Matsumoto, K., Takeda, K., Matsumoto, N., Yamaguchi, Y., and Yamamoto, K. Malectin Forms a Complex with Ribophorin I for Enhanced Association with Misfolded Glycoproteins. *The Journal of biological chemistry*, 287(45):38080–38089, November 2012.

[40] Bole, D. G., Hendershot, L. M., and Kearney, J. F. Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas. *The Journal of Cell Biology*, 102(5):1558–1566, May 1986.

[41] Braakman, I. and Hebert, D. N. Protein Folding in the Endoplasmic Reticulum. *Cold Spring Harbor Perspectives in Biology*, 5(5): a013201–a013201, May 2013.

[42] Rutkowski, D. T. and Kaufman, R. J. A trip to the ER: coping with stress. *Trends in cell biology*, 14(1):20–27, 2004.

[43] Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal*, 280 (Pt 2):309–316, December 1991.

[44] Henrissat, B. and Bairoch, A. *New families in the classification of glycosyl hydrolases based on amino acid sequence similarities.*, volume 293. Biochem. J., 1993.

[45] Davies, G. and Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure (London, England : 1993)*, 3(9):853–859, September 1995.

[46] Henrissat, B. and Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. *The Biochemical journal*, 316 (Pt 2):695–696, June 1996.

[47] Henrissat, B. and Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology*, 7(5):637–644, October 1997.

[48] Kornfeld, R. and Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annual review of biochemistry*, 54:631–664, 1985.

[49] Palcic, M. M., Scaman, C. H., Otter, A., Szpacenko, A., Romaniouk, A., Li, Y. X., and Vijay, I. K. Processing alpha-glucosidase I is an inverting glycosidase. *Glycoconjugate Journal*, 16(7):351–355, July 1999.

[50] Koshland, D. E. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews*, 28(4):416–436, November 1953.

[51] Sinnott, M. L. Catalytic mechanism of enzymic glycosyl transfer. *Chemical Reviews*, 90:1171–1202, 1990.

[52] Hettkamp, H., Legler, G., and Bause, E. 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(1):85–90, July 1984.

[53] Shailubhai, K., Pukazhenthi, B. S., Saxena, E. S., Varma, G. M., and Vijay, I. K. Glucosidase I, a transmembrane endoplasmic reticular glycoprotein with a luminal catalytic domain. *The Journal of biological chemistry*, 266(25): 16587–16593, September 1991.

[54] Romero, P. A., Dijkgraaf, G. J., Shahinian, S., Herscovics, A., and Bussey, H. The yeast CWH41 gene encodes glucosidase I. *Glycobiology*, 7(7):997–1004, October 1997.

[55] Dhanawansa, R., Faridmoayer, A., van der Merwe, G., Li, Y. X., and Scaman, C. H. Overexpression, purification, and partial characterization of Saccharomyces cerevisiae processing alpha glucosidase I. *Glycobiology*, 12(3): 229–234, March 2002.

[56] Faridmoayer, A. and Scaman, C. H. Binding residues and catalytic domain of soluble Saccharomyces cerevisiae processing alpha-glucosidase I. *Glycobiology*, 15(12):1341–1348, December 2005.

[57] Faridmoayer, A. and Scaman, C. H. Truncations and functional carboxylic acid residues of yeast processing α-glucosidase I. *Glycoconjugate Journal*, 24 (8):429–437, April 2007.

[58] 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. A Novel Disorder Caused by Defective Biosynthesis of N-Linked Oligosaccharides Due to Glucosidase I Deficiency. *The American Journal of Human Genetics*, 66(6):1744–1756, June 2000.

[59] 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. Glycosylation, Hypogammaglobulinemia, and Resistance to Viral Infections. *New England Journal of Medicine*, 370(17):1615–1625, April 2014.

[60] Kurakata, Y., Uechi, A., Yoshida, H., Kamitori, S., Sakano, Y., Nishikawa, A., and Tonozuka, T. 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(1):116–128, August 2008.

[61] Barker, M. K. and Rose, D. R. Specificity of Processing α-glucosidase I is guided by the substrate conformation: crystallographic and in silico studies. *Journal of Biological Chemistry*, 288(19):13563–13574, May 2013.

[62] Petrescu, A. J., butters, T. D., reinkensmeier, G., Petrescu, S., Platt, F. M., Dwek, R. A., and Wormwald, M. R. The solution NMR structure of glucosylated N-glycans involved in the early stages of glycoprotein biosynthesis and folding. *The EMBO journal*, 16(14):4302–4310, July 1997.

[63] Grinna, L. S. and Robbins, P. W. Glycoprotein biosynthesis. Rat liver microsomal glucosidases which process oligosaccharides. *The Journal of biological chemistry*, 254(18):8814–8818, September 1979.

[64] Kaushal, G. P., Pastuszak, I., Hatanaka, K., and Elbein, A. D. Purification to homogeneity and properties of glucosidase II from mung bean seedlings and suspension-cultured soybean cells. *The Journal of biological chemistry*, 265 (27):16271–16279, September 1990.

[65] Totani, K., Ihara, Y., Matsuo, I., and Ito, Y. Substrate specificity analysis of endoplasmic reticulum glucosidase II using synthetic high mannose-type glycans. *The Journal of biological chemistry*, 281(42):31502–31508, October 2006.

[66] Takeda, Y., Totani, K., Matsuo, I., and Ito, Y. Bioorganic & Medicinal Chemistry Letters. *Bioorganic & Medicinal Chemistry Letters*, 20(17): 5357–5359, September 2010.

[67] Trombetta, E. S., Simons, J. F., and Helenius, A. Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit. *Journal of Biological Chemistry*, 271(44):27509–27516, November 1996.

[68] Flura, T., Brada, D., Ziak, M., and Roth, J. Expression of a cDNA encoding the glucose trimming enzyme glucosidase II in CHO cells and molecular characterization of the enzyme deficiency in a mutant mouse lymphoma cell line. *Glycobiology*, 7(5):617–624, July 1997.

[69] Lovering, A. L., Lee, S. S., Kim, Y.-W., Withers, S. G., and Strynadka, N. C. J. Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *The Journal of biological chemistry*, 280(3): 2105–2115, January 2005.

[70] Frandsen, T. P. and Svensson, B. Plant alpha-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. *Plant molecular biology*, 37(1):1–13, May 1998.

[71] Lee, S. S., He, S., and Withers, S. G. Identification of the catalytic nucleophile of the Family 31 alpha-glucosidase from Aspergillus niger via trapping of a 5-fluoroglycosyl-enzyme intermediate. *The Biochemical journal*, 359(Pt 2):381–386, October 2001.

[72] Feng, J., Romaniouk, A. V., Samal, S. K., and Vijay, I. K. Processing enzyme glucosidase II: proposed catalytic residues and developmental regulation during the ontogeny of the mouse mammary gland. *Glycobiology*, 14(10): 909–921, October 2004.

[73] Ernst, H. A., Lo Leggio, L., Willemoës, M., Leonard, G., Blum, P., and Larsen, S. Structure of the Sulfolobus solfataricus alpha-glucosidase: implications for domain conservation and substrate recognition in GH31. *Journal of Molecular Biology*, 358(4):1106–1124, May 2006.

[74] Larsbrink, J., Izumi, A., Hemsworth, G. R., Davies, G. J., and Brumer, H. Structural enzymology of Cellvibrio japonicus Agd31B protein reveals α-transglucosylase activity in glycoside hydrolase family 31. *Journal of Biological Chemistry*, 287(52):43288–43299, December 2012.

[75] Sakai, K., Hirai, M., Minoshima, S., Kudoh, J., Fukuyama, R., and Shimizu, N. Isolation of cDNAs encoding a substrate for protein kinase C: nucleotide sequence and chromosomal mapping of the gene for a human 80K protein. *Genomics*, 5(2):309–315, August 1989.

[76] Reynolds, D. M., Falk, C. T., Li, A., King, B. F., Kamath, P. S., Huston, J., Shub, C., Iglesias, D. M., Martin, R. S., Pirson, Y., Torres, V. E., and Somlo, S. Identification of a locus for autosomal dominant polycystic liver disease, on chromosome 19p13.2-13.1. *American journal of human genetics*, 67(6): 1598–1604, December 2000.

[77] Li, A., Davila, S., Furu, L., Qian, Q., Tian, X., Kamath, P. S., King, B. F., Torres, V. E., and Somlo, S. Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease. *American journal of human genetics*, 72(3): 691–703, March 2003.

[78] Drenth, J. P. H., te Morsche, R. H. M., Smink, R., Bonifacino, J. S., and Jansen, J. B. M. J. Germline mutations in PRKCSH are associated with autosomal dominant polycystic liver disease. *Nature Genetics*, 33(3):345–347, February 2003.

[79] Waanders, E., te Morsche, R. H. M., de Man, R. A., Jansen, J. B. M. J., and Drenth, J. P. H. Extensive mutational analysis of PRKCSH and SEC63 broadens the spectrum of polycystic liver disease. *Human Mutation*, 27(8): 830–830, 2006.

[80] Hofherr, A., Wagner, C., Fedeles, S., Somlo, S., and Köttgen, M. N-glycosylation determines the abundance of the transient receptor potential

channel TRPP2. *Journal of Biological Chemistry*, 289(21):14854–14867, May 2014.

[81] Munro, S. The MRH domain suggests a shared ancestry for the mannose 6-phosphate receptors and other N-glycan-recognising proteins. *Current Biology*, 11(13):R499–R501, July 2001.

[82] D'Alessio, C., Fernández, F., Trombetta, E. S., and Parodi, A. J. Genetic evidence for the heterodimeric structure of glucosidase II: the effect of disrupting the subunit-encoding genes on glycoprotein folding. *The Journal of biological chemistry*, 274(36):25899–25905, 1999.

[83] Pelletier, M. F., Marcil, A., Sevigny, G., Jakob, C. A., Tessier, D. C., Chevet, E., Menard, R., Bergeron, J. J., and Thomas, D. Y. The heterodimeric structure of glucosidase II is required for its activity, solubility, and localization in vivo. *Glycobiology*, 10(8):815–827, August 2000.

[84] Stigliano, I. D., Caramelo, J. J., Labriola, C. A., Parodi, A. J., and D'Alessio, C. Glucosidase II β Subunit Modulates N-Glycan Trimming in Fission Yeasts and Mammals. *Molecular biology of the cell*, 20:3974–3984, 2009.

[85] Hu, D., Kamiya, Y., Totani, K., Kamiya, D., Kawasaki, N., Yamaguchi, D., Matsuo, I., Matsumoto, N., Ito, Y., Kato, K., and Yamamoto, K. Sugar-binding activity of the MRH domain in the ER alpha-glucosidase II beta subunit is important for efficient glucose trimming. *Glycobiology*, 19(10): 1127–1135, October 2009.

[86] Arendt, C. W. and Ostergaard, H. L. Two distinct domains of the β-subunit of glucosidase II interact with the catalytic α-subunit. *Glycobiology*, 10(5): 487–492, 2000.

[87] Trombetta, E. S., Fleming, K. G., and Helenius, A. Quaternary and domain structure of glycoprotein processing glucosidase II. *Biochemistry*, 40(35): 10717–10722, September 2001.

[88] Deprez, P., Gautschi, M., and Helenius, A. More Than One Glycan Is Needed for ER Glucosidase II to Allow Entry of Glycoproteins into the Calnexin/Calreticulin Cycle. *Molecular cell*, 19(2):183–195, July 2005.

[89] Rozeboom, H. J., Yu, S., Madrid, S., Kalk, K. H., Zhang, R., and Dijkstra, B. W. Crystal structure of α-1,4-glucan lyase, a unique glycoside hydrolase family member with a novel catalytic mechanism. *Journal of Biological Chemistry*, 288(37):26764–26774, September 2013.

[90] Sim, L., Willemsma, C., Mohan, S., Naim, H. Y., Pinto, B. M., and Rose, D. R. Structural basis for substrate selectivity in human maltase-glucoamylase and sucrase-isomaltase N-terminal domains. *Journal of Biological Chemistry*, 285(23):17763–17770, June 2010.

[91] Sim, L., Quezada-Calvillo, R., Sterchi, E. E., Nichols, B. L., and Rose, D. R. Human intestinal maltase-glucoamylase: crystal structure of the N-terminal catalytic subunit and basis of inhibition and substrate specificity. *Journal of Molecular Biology*, 375(3):782–792, January 2008.

[92] Ren, L., Qin, X., Cao, X., Wang, L., Bai, F., Bai, G., and Shen, Y. Structural insight into substrate specificity of human intestinal maltase-glucoamylase. *Protein & Cell*, 2(10):827–836, November 2011.

[93] Olson, L. J., Orsi, R., Alculumbre, S. G., Peterson, F. C., Stigliano, I. D., Parodi, A. J., D'Alessio, C., and Dahms, N. M. Structure of the Lectin Mannose 6-Phosphate Receptor Homology (MRH) Domain of Glucosidase II, an Enzyme That Regulates Glycoprotein Folding Quality Control in the Endoplasmic Reticulum. *Journal of Biological Chemistry*, 288(23): 16460–16475, April 2013.

[94] Olson, L. J., Orsi, R., Peterson, F. C., Parodi, A. J., Kim, J.-J. P., D'Alessio, C., and Dahms, N. M. Crystal Structure and Functional Analyses of the Lectin Domain of Glucosidase II: Insights into Oligomannose Recognition. *Biochemistry*, 54(26):4097–4111, July 2015.

[95] Inouye, S., Tsuruoka, T., and Nida, T. The structure of nojirimycin, a piperidinose sugar antibiotic. *The Journal of antibiotics*, 19(6):288–292, November 1966.

[96] Yagi, M., Kouno, T., Aoyagi, Y., and Murai, H. The structure of moranoline, a piperidine alkaloid from Morus species. *Journal of the Agricultural Chemical Society of Japan*, 11:571–572, 1976.

[97] Paulsen, H. Carbohydrates Containing Nitrogen or Sulfur in the "Hemiacetal" Ring. *Angewandte Chemie International Edition in English*, 5(5):495–510, May 1966.

[98] Hohenschutz, L. D., Bell, E. A., Jewess, P. J., and Leworthy, D. P. Castanospermine, a 1, 6, 7, 8-tetrahydroxyoctahydroindolizine alkaloid, from seeds of Castanospermum australe. *Phytochemistry*, 20(4):811–814, 1981.

[99] Saul, R., Ghidoni, J. J., Molyneux, R. J., and Elbein, A. D. Castanospermine inhibits alpha-glucosidase activities and alters glycogen distribution in animals. *Proceedings of the National Academy of Sciences of the United States of America*, 82(1):93–97, January 1985.

[100] Mellor, H., Neville, D., Harvey, D., Platt, F., and Dwek, R. Cellular effects of deoxynojirimycin analogues: uptake, retention and inhibition of glycosphingolipid biosynthesis. *Biochemistry Journal*, 381(Pt 3):861, August 2004.

[101] Elbein, A. D., Tropea, J. E., Mitchell, M., and Kaushal, G. P. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *The Journal of biological chemistry*, 265(26):15599–15605, September 1990.

[102] Elbein, A. D. Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annual review of biochemistry*, 56(1):497–534, 1987.

[103] Wong, C. H., Dumas, D. P., and Ichikawa, Y. Specificity, inhibition, and synthetic utility of a recombinant human. alpha.-1, 3-fucosyltransferase. *J Am Chem Soc*, 114:7321–7322, 1992.

[104] Horenstein, B. A., Zabinski, R. F., and Schramm, V. L. A new class of C-nucleoside analogues. 1-(S)-aryl-1,4-dideoxy-1,4-imino-D-ribitols, transition state analogue inhibitors of nucleoside hydrolase. *Tetrahedron Letters*, 34(45): 7213–7216, November 1993.

[105] Bols, M., Hazell, R. G., and Thomsen, I. B. 1-Azafagomine: A Hydroxyhexahydropyridazine That Potently Inhibits Enzymatic Glycoside Cleavage. *Chemistry (Weinheim an der Bergstrasse, Germany)*, 3(6):940–947, June 1997.

[106] Moriyama, H., Tsukida, T., Inoue, Y., Yokota, K., Yoshino, K., Kondo, H., Miura, N., and Nishimura, S.-I. Azasugar-based MMP/ADAM inhibitors as antipsoriatic agents. *Journal of medicinal chemistry*, 47(8):1930–1938, April 2004.

[107] Compain, P. and Martin, O. R., editors. *Iminosugars*. John Wiley & Sons, Ltd, Chichester, UK, October 2007.

[108] Asano, N. Glycosidase inhibitors: update and perspectives on practical use. *Glycobiology*, 13(10):93R–104R, October 2003.

[109] Bischoff, H. The mechanism of alpha-glucosidase inhibition in the management of diabetes. *Clinical and investigative medicine. Médecine clinique et experimentale*, 18(4):303–311, August 1995.

[110] Standl, E., Schernthaner, G., Rybka, J., Hanefeld, M., Raptis, S. A., and Naditch, L. Improved glycaemic control with miglitol in inadequately-controlled type 2 diabetics. *Diabetes research and clinical practice*, 51(3):205–213, March 2001.

[111] Rao Vunnam, R. and Radin, N. S. Analogs of ceramide that inhibit glucocerebroside synthetase in mouse brain. *Chemistry and Physics of Lipids*, 26(3):265–278, April 1980.

[112] Radin, N. S. Treatment of Gaucher disease with an enzyme inhibitor. *Glycoconjugate Journal*, 13(2):153–157, April 1996.

[113] Platt, F. M., Reinkensmeier, G., Dwek, R. A., and Butters, T. D. Extensive Glycosphingolipid Depletion in the Liver and Lymphoid Organs of Mice Treated with N-Butyldeoxynojirimycin. *The Journal of biological chemistry*, 272(31):19365–19372, August 1997.

[114] Fan, J. Q., Ishii, S., Asano, N., and Suzuki, Y. Accelerated transport and maturation of lysosomal alpha-galactosidase A in Fabry lymphoblasts by an enzyme inhibitor. *Nat Med*, 5(1):112–115, January 1999.

[115] Sawkar, A. R., Cheng, W.-C., Beutler, E., Wong, C. H., Balch, W. E., and Kelly, J. W. Chemical chaperones increase the cellular activity of N370S beta -glucosidase: a therapeutic strategy for Gaucher disease. *Proceedings of the National Academy of Sciences*, 99(24):15428–15433, November 2002.

[116] Hettkamp, H., Bause, E., and Legler, G. Inhibition by nojirimycin and 1-deoxynojirimycin of microsomal glucosidases from calf liver acting on the glycoprotein oligosaccharides Glc1-3Man9GlcNAc2. *Bioscience reports*, 2(11): 899–906, November 1982.

[117] Pan, Y. T., Hori, H., Saul, R., Sanford, B. A., Molyneux, R. J., and Elbein, A. D. Castanospermine inhibits the processing of the oligosaccharide portion of the influenza viral hemagglutinin. *Biochemistry*, 22(16):3975–3984, August 1983.

[118] Hentges, A. and Bause, E. Affinity purification and characterization of glucosidase II from pig liver. *Biological chemistry*, 378(9):1031–1038, September 1997.

[119] Alonso, J. M., Santa-Cecilia, A., and Calvo, P. Glucosidase II from rat liver microsomes. Kinetic model for binding and hydrolysis. *The Biochemical journal*, 278 (Pt 3):721–727, September 1991.

[120] Varrot, A., Tarling, C. A., Macdonald, J. M., Stick, R. V., Zechel, D. L., Withers, S. G., and Davies, G. J. Direct Observation of the Protonation State of an Imino Sugar Glycosidase Inhibitor upon Binding. *J Am Chem Soc*, 125 (25):7496–7497, June 2003.

[121] Gloster, T. M., Meloncelli, P., Stick, R. V., Zechel, D., Vasella, A., and Davies, G. J. Glycosidase Inhibition: An Assessment of the Binding of 18 Putative Transition-State Mimics. *J Am Chem Soc*, 129(8):2345–2354, February 2007.

[122] López, Ó., Qing, F.-L., Pedersen, C. M., and Bols, M. Enzyme inhibition by iminosugars: analysis and insight into the glycosidase-iminosugar dependency of pH. *Bioorganic & medicinal chemistry*, 21(16):4755–4761, August 2013.

[123] Schlesinger, S., Malfer, C., and Schlesinger, M. J. The formation of vesicular stomatitis virus (San Juan strain) becomes temperature-sensitive when glucose residues are retained on the oligosaccharides of the glycoprotein. *The Journal of biological chemistry*, 259(12):7597–7601, June 1984.

[124] Schlesinger, S., Koyama, A. H., Malfer, C., Gee, S. L., and Schlesinger, M. J. The effects of inhibitors of glucosidase I on the formation of Sindbis virus. *Virus Research*, 2(2):139–149, March 1985.

[125] Datema, R., Olofsson, S., and Romero, P. A. Inhibitors of protein glycosylation and glycoprotein processing in viral systems. *Pharmacology & therapeutics*, 33(2-3):221–286, 1987.

[126] 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. Loss of cytomegalovirus infectivity after treatment with castanospermine or related plant alkaloids correlates with aberrant glycoprotein synthesis. *Antiviral Res*, 10(1-3):11–26, November 1988.

[127] Block, T. M., Lu, X., Platt, F. M., Foster, G. R., Gerlich, W. H., Blumberg, B. S., and Dwek, R. A. Secretion of human hepatitis B virus is inhibited by the imino sugar N-butyldeoxynojirimycin. *Proceedings of the National Academy of Sciences*, 91(6):2235–2239, March 1994.

[128] Karpas, A., Fleet, G. W., Dwek, R. A., Petursson, S., Namgoong, S. K., Ramsden, N. G., Jacob, G. S., and Rademacher, T. W. Aminosugar derivatives as potential anti-human immunodeficiency virus agents. *Proceedings of the National Academy of Sciences*, 85(23):9229–9233, December 1988.

[129] Courageot, M. P., Frenkiel, M. P., Dos Santos, C. D., Deubel, V., and Desprès, P. Alpha-glucosidase inhibitors reduce dengue virus production by affecting the initial steps of virion morphogenesis in the endoplasmic reticulum. *Journal of virology*, 74(1):564–572, January 2000.

[130] Schul, W., Liu, W., Xu, H.-Y., Flamand, M., and Vasudevan, S. G. A dengue fever viremia model in mice shows reduction in viral replication and suppression of the inflammatory response after treatment with antiviral drugs. *The Journal of infectious diseases*, 195(5):665–674, March 2007.

[131] Zitzmann, N., Mehta, A. S., Carrouée, S., Butters, T. D., Platt, F. M., McCauley, J., Blumberg, B. S., Dwek, R. A., and Block, T. M. Imino sugars inhibit the formation and secretion of bovine viral diarrhea virus, a pestivirus model of hepatitis C virus: Implications for the development of broad spectrum anti-hepatitis virus agents. *Proceedings of the National Academy of Sciences of the United States of America*, 96(21):11878–11882, October 1999.

[132] Durantel, D., Branza-Nichita, N., Carrouee-Durantel, S., butters, T. D., Dwek, R. A., and Zitzmann, N. Study of the mechanism of antiviral action of iminosugar derivatives against bovine viral diarrhea virus. *Journal of virology*, 75(19):8987–8998, October 2001.

[133] Pavlovic, D., Neville, D. C., Argaud, O., Blumberg, B., Dwek, R. A., Fischer, W. B., and Zitzmann, N. The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10):6104–6108, May 2003.

[134] Steinmann, E., Whitfield, T., Kallis, S., Dwek, R. A., Zitzmann, N., Pietschmann, T., and Bartenschlager, R. Antiviral effects of amantadine and iminosugar derivatives against hepatitis C virus. *Hepatology*, 46(2):330–338, August 2007.

[135] Chang, J., Block, T. M., and Guo, J.-T. Antiviral Research. *Antiviral Res*, 99 (3):251–260, September 2013.

[136] Chang, J., Wang, L., Ma, D., Qu, X., Guo, H., Xu, X., Mason, P. M., Bourne, N., Moriarty, R., Gu, B., Guo, J. T., and Block, T. M. Novel Imino Sugar Derivatives Demonstrate Potent Antiviral Activity against Flaviviruses. *Antimicrobial Agents and Chemotherapy*, 53(4):1501–1508, March 2009.

[137] Qu, X., Pan, X., Weidner, J., Yu, W., Alonzi, D., Xu, X., Butters, T., Block, T., Guo, J.-T., and Chang, J. Inhibitors of endoplasmic reticulum alpha-glucosidases potently suppress hepatitis C virus virion assembly and release. *Antimicrobial Agents and Chemotherapy*, 55(3):1036–1044, March 2011.

[138] Yu, W., Gill, T., Wang, L., Du, Y., Ye, H., Qu, X., Guo, J.-T., Cuconati, A., Zhao, K., Block, T. M., Xu, X., and Chang, J. Design, Synthesis, and Biological Evaluation of N-Alkylated Deoxynojirimycin (DNJ) Derivatives for the Treatment of Dengue Virus Infection. *Journal of medicinal chemistry*, 55 (13):6061–6075, June 2012.

[139] Perry, S. T., Buck, M. D., Plummer, E. M., Penmasta, R. A., Batra, H., Stavale, E. J., Warfield, K. L., Dwek, R. A., Butters, T. D., Alonzi, D. S., Lada, S. M., King, K., Klose, B., Ramstedt, U., and Shresta, S. An iminosugar with potent inhibition of dengue virus infection in vivo. *Antiviral Res*, 98(1):35–43, April 2013.

[140] Du, Y., Ye, H., Gill, T., Wang, L., Guo, F., Cuconati, A., Guo, J.-T., Block, T. M., Chang, J., and Xu, X. N-Alkyldeoxynojirimycin derivatives with novel terminal tertiary amide substitution for treatment of bovine viral diarrhea virus (BVDV), Dengue, and Tacaribe virus infections. *Bioorganic & Medicinal Chemistry Letters*, 23(7):2172–2176, April 2013.

[141] 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. Small molecule inhibitors of ER α-glucosidases are active against multiple hemorrhagic fever viruses. *Antiviral Res*, 98(3):432–440, June 2013.

[142] Aguilar, A. L., Escribano, J., Wentworth, P., and Butters, T. D. Synthetic 1-Deoxynojirimycin N-Substituted Peptides Offer Prolonged Disruption to N-Linked Glycan Processing. *ChemMedChem*, October 2014.

[143] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., ding, J. S. o., Thompson, J. D., and Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:1–6, October 2011.

[144] Robert, X. and Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research*, 42(W1):W320–W324, July 2014.

[145] Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, 21(16):3369–3376, August 2005.

[146] Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B., and Hood, L. E. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071):674–679, June 1986.

[147] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343–345, April 2009.

[148] Je, Y. H., Chang, J. H., Choi, J. Y., Roh, J. Y., Jin, B. R., O'Reilly, D. R., and Kang, S. K. A defective viral genome maintained in Escherichia coli for the generation of baculovirus expression vectors. *Biotechnology Letters*, 23(8): 575–582, 2001.

[149] Zhao, Y., Chapman, D. A. G., and Jones, I. M. Improving baculovirus recombination. *Nucleic acids research*, 31(2):E6–6, January 2003.

[150] Hope, H. Cryocrystallography of biological macromolecules: a generally applicable method. *Acta Cryst. B*, 44:22–26, February 1988.

[151] Delageniere, S., Brenchereau, P., Launer, L., Ashton, A. W., Leal, R., Veyrier, S., Gabadinho, J., Gordon, E. J., Jones, S. D., Levik, K. E., McSweeney, S. M., Monaco, S., Nanao, M., Spruce, D., Svensson, O., Walsh, M. A., and Leonard, G. A. ISPyB: an information management system for synchrotron macromolecular crystallography. *Bioinformatics*, 27(22):3186–3192, November 2011.

[152] Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of Synchrotron Radiation*, 16(6):872–879, October 2009.

[153] Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography*, 67:293–302, March 2011.

[154] Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography*, 66(2):125–132, January 2010.

[155] 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. W., 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. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(Pt 4):235–242, April 2011.

[156] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography*, 40:658–674, July 2007.

[157] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66:486–501, March 2010.

[158] Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., and Womak, T. O. BUSTER.

[159] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 1):12–21, January 2010.

[160] Schrodinger, LLC. The PyMOL Molecular Graphics System .

[161] Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci*, 4(11): 2411–2423, November 1995.

[162] Savitzky, A. and Golay, M. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.

[163] Whitmore, L. and Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research*, 32(Web Server issue):W668–73, July 2004.

[164] Compton, L. A. and Johnson, W. C. Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical biochemistry*, 155(1):155–167, May 1986.

[165] Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., and Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(Pt 2):342–350, April 2012.

[166] Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5):1277–1282, October 2003.

[167] Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25:495–503, 1992.

[168] Grinna, L. S. and Robbins, P. W. Substrate specificities of rat liver microsomal glucosidases which process glycoproteins. *The Journal of biological chemistry*, 255(6):2255–2258, March 1980.

[169] Schweden, J., Borgmann, C., Legler, G., and Bause, E. Characterization of calf liver glucosidase I and its inhibition by basic sugar analogs. *Archives of biochemistry and biophysics*, 248(1):335–340, July 1986.

[170] Shailubhai, K., Pratta, M. A., and Vijay, I. K. Purification and characterization of glucosidase I involved in N-linked glycoprotein processing in bovine mammary gland. *The Biochemical journal*, 247(3):555–562, November 1987.

[171] Bause, E. and Schweden, G. A. O. B., Jurgen. Purification and characterization of trimming glucosidase I from pig liver. *European journal of biochemistry / FEBS*, 183(3):661–669, August 1989.

[172] Romaniouk, A. and Vijay, I. K. Structure-function relationships in glucosidase I: amino acids involved in binding the substrate to the enzyme. *Glycobiology*, 7(3):399–404, 1997.

[173] Palma, A. S., Liu, Y., Muhle-Goll, C., Butters, T. D., Zhang, Y., Childs, R., Chai, W., and Feizi, T. *Multifaceted Approaches Including Neoglycolipid Oligosaccharide Microarrays to Ligand Discovery for Malectin*, volume 478. Elsevier Inc, 1 edition, 2010.

[174] Torres-Rodríguez, B. I., Flores-Berrout, K., Villagómez-Castro, J. C., and López-Romero, E. Purification and partial biochemical characterization of a membrane-bound type II-like α-glucosidase from the yeast morphotype of Sporothrix schenckii. *Antonie van Leeuwenhoek*, 101(2):313–322, September 2011.

[175] Faridmoayer, A. and Scaman, C. H. An improved purification procedure for soluble processing alpha-glucosidase I from Saccharomyces cerevisiae

overexpressing CWH41. *Protein Expression and Purification*, 33(1):11–18, January 2004.

[176] Barker, M. K., Wilkinson, B. L., Faridmoayer, A., Scaman, C. H., Fairbanks, A. J., and Rose, D. R. Production and crystallization of processing alpha-glucosidase I: Pichia pastoris expression and a two-step purification toward structural determination. *Protein Expression and Purification*, 79(1): 96–101, September 2011.

[177] Kabsch, W. and Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22 (12):2577–2637, December 1983.

[178] Niesen, F. H., Berglund, H., and Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*, 2(9):2212–2221, September 2007.

[179] Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical biochemistry*, 332(1):153–159, September 2004.

[180] Hayashi, Y., Matsui, H., and Takagi, T. Membrane protein molecular weight determined by low-angle laser light-scattering photometry coupled with high-performance gel chromatography. *Methods in enzymology*, 172:514–528, 1989.

[181] Bigge, J. C., Patel, T. P., Bruce, J. A., Goulding, P. N., Charles, S. M., and Parekh, R. B. Nonselective and Efficient Fluorescent Labeling of Glycans Using 2-Amino Benzamide and Anthranilic Acid. *Analytical biochemistry*, 230 (2):229–238, September 1995.

[182] Suzuki, N. and Lee, Y. C. Site-specific N-glycosylation of chicken serum IgG. *Glycobiology*, 14(3):275–292, March 2004.

[183] Takeda, Y., Totani, K., Matsuo, I., and Ito, Y. Chemical approaches toward understanding glycan-mediated protein quality control. *Current Opinion in Chemical Biology*, 13(5-6):582–591, December 2009.

[184] Anumula, K. R. and Spiro, R. G. Release of glucose-containing polymannose oligosaccharides during glycoprotein biosynthesis. Studies with thyroid microsomal enzymes and slices. *The Journal of biological chemistry*, 258(24): 15274–15282, December 1983.

[185] Moore, S. E. and Spiro, R. G. Intracellular compartmentalization and degradation of free polymannose oligosaccharides released during glycoprotein biosynthesis. *The Journal of biological chemistry*, 269(17):12715–12721, April 1994.

[186] Grimm, C., Chari, A., Reuter, K., and Fischer, U. A crystallization screen based on alternative polymeric precipitants. *Acta Crystallographica Section D: Biological Crystallography*, 66:658–697, May 2010.

[187] Karplus, P. A. and Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Current opinion in structural biology*, 34: 60–68, July 2015.

[188] Eswar, N., Webb, B., Marti Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. y., Pieper, U., and Sali, A. *Comparative Protein Structure Modeling Using Modeller*. John Wiley & Sons, Inc., 2006.

[189] Sali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779–815, December 1993.

[190] Brandts, J. F. and Lin, L. N. Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry*, 29(29):6927–6940, 1990.

[191] Waldron, T. T. and Murphy, K. P. Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry*, 42:5058–5064, 2003.

[192] Matulis, D., Kranz, J. K., Salemme, F. R., and Todd, M. J. Thermodynamic Stability of Carbonic Anhydrase: Measurements of Binding Affinity and Stoichiometry Using ThermoFluor. *Biochemistry*, March 2005.

[193] Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N., and Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry*, 357(2):289–298, October 2006.

[194] Alonzi, D. S., Neville, D. C. A., Lachmann, R. H., Dwek, R. A., and Butters, T. D. Glucosylated free oligosaccharides are biomarkers of endoplasmic-reticulum α-glucosidase inhibition. *The Biochemical journal*, 409(2):571, January 2008.

[195] Walter, T. S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., and Grimes, J. M. Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure (London, England : 1993)*, 14(11):1617–1622, November 2006.

[196] Derewenda, Z. S. Rational protein crystallization by mutational surface engineering. *Structure (London, England : 1993)*, 12(4):529–535, April 2004.

[197] Brada, D. and Dubach, U. C. Isolation of a homogeneous glucosidase II from pig kidney microsomes. *European journal of biochemistry / FEBS*, 141(1): 149–156, May 1984.

[198] Sikora, J., Uřinovská, J., Majer, F., Poupětová, H., Hlavatá, J., Kostrouchová, M., Ledvinová, J., and Hřebíček, M. Bioinformatic and biochemical studies point to AAGR-1 as the ortholog of human acid α-glucosidase in Caenorhabditis elegans. *Molecular and Cellular Biochemistry*, 341(1-2):51–63, March 2010.

[199] Vieira, J. and Messing, J. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, 19(3):259–268, October 1982.

[200] Sreerama, N. and Woody, R. W. Computation and analysis of protein circular dichroism spectra. *Methods in enzymology*, 383:318–351, 2004.

[201] Holzwarth, G. and Doty, P. The Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc*, 87(2):218–228, January 1965.

[202] Greenfield, N. J. and Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, 8(10):4108–4116, October 1969.

[203] Saxena, S., Shailubhai, K., Dong-Yu, B., and Vijay, I. K. Purification and characterization of glucosidase II involved in N-linked glycoprotein processing in bovine mammary gland. *The Biochemical journal*, 247(3):563–570, November 1987.

[204] Alonso, J. M., Santa-Cecilia, A., and Calvo, P. Effect of bromoconduritol on glucosidase II from rat liver. A new kinetic model for the binding and hydrolysis of the substrate. *European journal of biochemistry / FEBS*, 215(1): 37–42, July 1993.

[205] Wilkinson, B. M., Purswani, J., and Stirling, C. J. Yeast GTB1 encodes a subunit of glucosidase II required for glycoprotein processing in the endoplasmic reticulum. *Journal of Biological Chemistry*, 281(10):6325–6333, March 2006.

[206] Quinn, R. P., Mahoney, S. J., Wilkinson, B. M., Thornton, D. J., and Stirling, C. J. A novel role for Gtb1p in glucose trimming of N-linked glycans. *Glycobiology*, 19(12):1408–1416, December 2009.

[207] Totani, K., Ihara, Y., Matsuo, I., and Ito, Y. Effects of Macromolecular Crowding on Glycoprotein Processing Enzymes. *J Am Chem Soc*, 130(6): 2101–2107, February 2008.

[208] Stigliano, I. D., Alculumbre, S. G., Labriola, C. A., Parodi, A. J., and D'Alessio, C. Glucosidase II and N-glycan mannose content regulate the half-lives of monoglucosylated species in vivo. *Molecular biology of the cell*, 22 (11):1810–1823, June 2011.

[209] Watanabe, T., Totani, K., Matsuo, I., Maruyama, J. i., Kitamoto, K., and Ito, Y. Genetic analysis of glucosidase II -subunit in trimming of high-mannose-type glycans. *Glycobiology*, 19(8):834–840, July 2009.

[210] Kang, M.-S., Okuyama, M., Mori, H., and Kimura, A. The first α-1,3-glucosidase from bacterial origin belonging to glycoside hydrolase family 31. *Biochimie*, 91(11-12):1434–1442, November 2009.

[211] Song, K.-M., Okuyama, M., Nishimura, M., Tagami, T., Mori, H., and Kimura, A. Aromatic residue on $\beta\alpha$ loop 1 in the catalytic domain is important to the transglycosylation specificity of glycoside hydrolase family 31 α-glucosidase. *Bioscience, biotechnology, and biochemistry*, 77(8):1759–1765, 2013.

[212] Saburi, W., Okuyama, M., Kumagai, Y., Kimura, A., and Mori, H. Biochemical properties and substrate recognition mechanism of GH31 a-glucosidase from Bacillus sp. AHU 2001 with broad substrate specificity . *Biochimie*, 108(C):140–148, January 2015.

[213] Tan, A., Van den Broek, L., Van Boeckel, S., Ploegh, H., and Bolscher, J. Chemical modification of the glucosidase inhibitor 1-deoxynojirimycin. Structure-activity relationships. *Journal of Biological Chemistry*, 266, number = 22, pages = 14504–14510,, 1991.

[214] Rawlings, A. J., Lomas, H., Pilling, A. W., Lee, M. J. R., Alonzi, D. S., Rountree, J. S. S., Jenkinson, S. F., Fleet, G. W. J., Dwek, R. A., Jones, J. H., and Butters, T. D. Synthesis and Biological Characterisation of Novel N-Alkyl-Deoxynojirimycin α-Glucosidase Inhibitors. *ChemBioChem*, 10(6): 1101–1105, April 2009.

[215] Yung-Chi, C. and Prusoff, W. H. Relationship between the inhibition constant (K I) and the concentration of inhibitor which causes 50 per cent inhibition (I 50) of an enzymatic reaction. *Biochemical Pharmacology*, 22(23):3099–3108, 1973.

[216] Lohse, A., Hardlei, T., Jensen, A., Plesner, I. W., and Bols, M. Investigation of the slow inhibition of almond beta-glucosidase and yeast isomaltase by 1-azasugar inhibitors: evidence for the 'direct binding' model. *The Biochemical journal*, 349(Pt 1):211–215, July 2000.

[217] Zechel, D. L., Boraston, A. B., Gloster, T., Boraston, C. M., Macdonald, J. M., Tilbrook, D. M. G., Stick, R. V., and Davies, G. J. Iminosugar Glycosidase Inhibitors: Structural and Thermodynamic Dissection of the Binding of Isofagomine and 1-Deoxynojirimycin to β-Glucosidases. *J Am Chem Soc*, 125(47):14313–14323, November 2003.

[218] Gorrec, F. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography*, 42(6):1035–1042, November 2009.

[219] Diederichs, K. and Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature structural biology*, 4(4): 269–275, 1997.

[220] van Driel, I. R., Goldstein, J. L., Südhof, T. C., and Brown, M. S. First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca2+ and monoclonal antibodies, but not lipoproteins. *The Journal of biological chemistry*, 262(36):17443–17449, December 1987.

[221] Fass, D., Blacklow, S., Kim, P. S., and Berger, J. M. Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature*, 388(6643):691–693, August 1997.

[222] Müller, P., Köpke, S., and Sheldrick, G. M. Is the bond-valence method able to identify metal atoms in protein structures? *Acta Crystallographica Section D: Biological Crystallography*, 59(1):32–37, January 2003.

[223] Brese, N. E. and O'keeffe, M. Bond-valence parameters for solids. *Acta Crystallographica Section B: Structural Science*, 47(2):192–197, April 1991.

[224] Brown, I. D. and Altermatt, D. Bond-valence parameters obtained from a systematic analysis of the Inorganic Crystal Structure Database. *Acta Crystallographica Section B: Structural Science*, 41(4):244–247, August 1985.

[225] Krissinel, E. and Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3):774–797, September 2007.

[226] Duarte, J. M., Srebniak, A., Schärer, M. A., and Capitani, G. Protein interface classification by evolutionary analysis. *BMC Bioinformatics*, 13:334, 2012.

[227] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic acids research*, 32(Web Server issue): W665–7, July 2004.

[228] Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T., and Ben-Tal, N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic acids research*, 33(Web Server): W299–W302, July 2005.

[229] Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic acids research*, 38(Web Server):W529–W533, June 2010.

[230] Evans, P. R. and Murshudov, G. N. research papers. *Acta Crystallographica Section D: Biological Crystallography*, 69:1204–1214, June 2013.

[231] Karplus, P. A. and Diederichs, K. Linking crystallographic model and data quality. *Science*, 336:1030–1033, 2012.

[232] Esnouf, R. M., Ren, J., Garman, E. F., Somers, D. O., Ross, C. K., Jones, E. Y., Stammers, D. K., and Stuart, D. I. Continuous and Discontinuous Changes in the Unit Cell of HIV-1 Reverse Transcriptase Crystals on Dehydration. *Acta Crystallographica Section D: Biological Crystallography*, 54 (5):938–953, September 1998.

[233] Newman, J. A review of techniques for maximizing diffraction from a protein crystal in stilla. *Acta Crystallographica Section D: Biological Crystallography*, 62(Pt 1):27–31, January 2006.

[234] Heras, B. and Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D: Biological Crystallography*, 61(Pt 9):1173–1180, September 2005.

[235] Tagami, T., Yamashita, K., Okuyama, M., Mori, H., Yao, M., and Kimura, A. Molecular basis for the recognition of long-chain substrates by plant α-glucosidases. *Journal of Biological Chemistry*, 288(26):19296–19303, June 2013.

[236] Sim, L., Quezada-Calvillo, R., and Sterchi, E. E. Human Intestinal Maltase–Glucoamylase: Crystal Structure of the N-Terminal Catalytic Subunit and Basis of Inhibition and Substrate Specificity. *Journal of molecular biology*, 375:782–792, 2008.

[237] Knott, T. J., Rall, S. C., Innerarity, T. L., Jacobson, S. F., Urdea, M. S., Levy-Wilson, B., Powell, L. M., Pease, R. J., Eddy, R., and Nakai, H. Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression, and chromosomal localization. *Science*, 230(4721):37–43, October 1985.

[238] Daly, N. L., Scanlon, M. J., Djordjevic, J. T., Kroon, P. A., and Smith, R. Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 92(14):6334–6338, July 1995.

[239] Zaiou, M., Arnold, K. S., Newhouse, Y. M., Innerarity, T. L., Weisgraber, K. H., Segall, M. L., Phillips, M. C., and Lund-Katz, S. Apolipoprotein E–low density lipoprotein receptor interaction: influences of basic residue and amphipathic α-helix organization in the ligand. *Journal of Lipid Research*, 41 (7):1087–1095, July 2000.

[240] Soussillane, P., D'Alessio, C., Paccalet, T., Fitchette, A.-C., Parodi, A. J., Williamson, R., Plasson, C., Faye, L., and Gomord, V. N-glycan trimming by glucosidase II is essential for Arabidopsis development. *Glycoconjugate Journal*, 26(5):597–607, January 2009.

[241] Lu, X., Tintor, N., Mentzel, T., Kombrink, E., Boller, T., Robatzek, S., Schulze-Lefert, P., and Saijo, Y. Uncoupling of sustained MAMP receptor signaling from early outputs in an Arabidopsis endoplasmic reticulum glucosidase II allele. *Proceedings of the National Academy of Sciences of the United States of America*, 106(52):22522–22527, December 2009.

[242] Rempel, B. P. and Withers, S. G. Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology*, 18(8):570–586, May 2008.

[243] Wu, S. F., Lee, C. J., Liao, C. L., Dwek, R. A., Zitzmann, N., and Lin, Y. L. Antiviral Effects of an Iminosugar Derivative on Flavivirus Infections. *Journal of virology*, 76(8):3596–3604, April 2002.